CHAPTER I

The Allure of the Ideal
Orienting the Quest for Justice

> A map of the world that does not include Utopia is not even worth glancing at.
>
> —Oscar Wilde

1 Orienting to Utopia

1.1 Beyond the Contemporary Debate and Its Categories

THERE ARE NUMEROUS UNDERSTANDINGS OF SO-CALLED IDEAL POLITI-cal theory—so many that the literature has now reached the stage in which taxonomies of the ideal/nonideal distinction are being presented. Laura Valentini identifies three different ways in which the contrast is employed—"(*i*) full compliance *vs.* partial compliance theory; (*ii*) utopian *vs.* realistic theory; (*iii*) end-state *vs.* transitional theory"[1]—while Alan Hamlin and Zofia Stemplowska identify other "dimensions": (*i*) full *v.* partial compliance; (*ii*) idealization *v.* abstraction; (*iii*) fact sensitivity *v.* insensitivity; and (*iv*), perfect justice *v.* local improvements.[2] Although such "conceptual cartography"[3] is helpful in organizing the now-large literature, it has important limitations. If we become too focused on classifications and distinctions, we are apt to miss how these different dimensions can be integrated (in various ways) into an overall, coherent, and compelling articulation of an ideal political philosophy. To be sure, when an idea is "messy"[4] because of its many dimensions the resulting debate may be confused;

---

[1] Valentini, "Ideal vs. Non-ideal Theory," p. 654.
[2] Hamlin and Stemplowska, "Theory, Ideal Theory and the Theory of Ideals," pp. 48–49.
[3] Valentini, "Ideal vs. Non-ideal Theory," p. 655.
[4] Ibid.

philosophers are apt to talk past each other. Here drawing sharp distinctions between different questions will be valuable. However, often philosophy is messy because the elements of the mess are intertwined in complex ways in a coherent view of the problem. We will see that almost all these different dimensions will come up in this book, as I explore a compelling, but somewhat complex, view of what a theory of an ideal may be, and when and why it is attractive.

Moreover, if we focus too much on the current debate, its categories and concerns, we are apt to fall into the all-too-common error of supposing that somehow these issues "have all originated in response to the methodological paradigm set by John Rawls."[5] No doubt the current round of literature has been spurred by themes in Rawls's work, but many of these issues have arisen, and been investigated, throughout the history of political thought, both recent and distant. In 1982 Barbara Goodwin and Keith Taylor presented a sophisticated answer to whether "utopian" thought must be realistic,[6] employing possible world analysis, inquiring whether an ideal world must be realistically achievable in one move, or could be reached in several moves as one navigates through intermediate possible worlds (an idea that I shall develop in detail in chapter II).[7] And other scholars have shown that the two millennia of utopian thought was often concerned with articulating *ideals* that provided the goals of progressive thought and practice.[8] Indeed, Karl Kautsky famously praised More's *Utopia* as articulating a socialist ideal, which satisfied important "realization" constraints. "More conceived of the realization of his ideals: he was the father of Utopian Socialism, which was rightly named after his *Utopia*. The latter is Utopian less on account of the inadequacy of its aims than on account of the inadequacy of the means at its disposal for their achievement."[9] I certainly do not wish to deny that the recent debate has stressed some new and important issues and has achieved new

[5] Ibid.

[6] Thus "utopian vs. realistic theory" is an overly simple rendering of the complexity within utopian theory; the current literature often wrongly supposes that utopian theories were resolutely antirealistic. See, e.g., Jubb, "Tragedies of Non-ideal Theory," p. 230.

[7] Goodwin and Taylor, *The Politics of Utopia*, pp. 210–14.

[8] Kenyon, "Utopia in Reality."

[9] Kautsky, *Thomas More and His Utopia*, p. 249.

insights, but contemporary philosophers too often see their concerns as new and unique when, in fact, they are echoes, as well as developments, of a long line of political thinking.

We thus need to be cognizant of the current debates, while stepping back and keeping in mind that we are exploring a larger and more enduring theme in political thinking. My aim in this chapter, then, is not to analyze or enter into the current debate, though I too shall engage in a bit of classification and line drawing (I am, after all, a philosopher). I shall identify several different enduring models of utopian-ideal thought, arguing that one stands out as meriting closer investigation. I argue that this is an attractive understanding of utopian-ideal theory, that it makes sense of the theory's appeal, and why those such as Oscar Wilde (in our epigraph) thought ideals are a necessary part of any "map" of political reform. I believe this understanding is broad enough to include a wide range of traditional utopian theory, as well as many current ideal theories. It also makes sense—if I may say so, much better sense—of many of the current facets of the ideal theory debate among contemporary philosophers, such as that between Amartya Sen and Rawls on the importance of ideals in pursuing justice. After I articulate this theory and its appeal in this chapter, the next two chapters analyze it in considerable depth. I certainly do not claim that all who would deem themselves "ideal theorists" or "utopians" are involved in this long-standing project, though I do think many more are committed to it than they realize. And it is a project that demands the attention of those of us who are skeptical that our diverse societies should be arranged around any conception of utopia.

1.2 Of Paradise

Right from the beginning political philosophy has sought to describe the ideal state, which, even if not fully achievable, gives us guidance in constructing a more just social world. As Plato, the first of the ideal theorists, acknowledged, it is in "the nature of things that action should come less close to truth than thought," and so our ideal constructions will not be "reproduced in fact down to the least detail."[10]

---

[10] Plato, *The Republic*, p. 178 [v. 473]. For discussions of Plato as an instance of uto-

On this view, as Ingrid Robeyns has put it, the ideal functions as "a mythical Paradise Island" that tells us where "the *endpoint* of our journey lies." Although the ideal does not "necessarily tell us anything about the route to take to Paradise Island," it orients our journey.[11] Only after identifying the ideal can we take up the task of figuring out how to get there (or, if we cannot quite get to the ideal, to come as close to it as possible). As Rawls says, "By showing how the social world may realize the features of a realistic Utopia, political philosophy provides a long-term goal of political endeavor, and in working toward it gives meaning to what we can do today."[12] To this he adds, "[the] idea of realistic Utopia is importantly institutional."[13] We wish to identify the institutional structures and patterns of interaction of an achievable ideally just social world, for it is this that ultimately provides the guidance we need to reform our own social world's institutions. Of course, we may never arrive at the ideal social world, but with an ideal guiding us the hope is that we can rest assured that our efforts to secure justice have at least moved us in the right direction.

If the goal of the ideal is to orient our navigation through less-than-ideal social worlds, we need to understand where we are now in relation to it. The ideal can orient us only if we have some idea of where it is, and where we presently are. Consequently, this *orienting function* of ideal political philosophy seeks not only to inform us about the long-term goal of creating a perfectly just society, but also to ground at least some significant class of judgments as to whether a move from, say, our present social world to a near social state moves us closer to, or further from, the ideal.[14] Rawls believed that ideal justice provided

---

pian thought, see Berneri, *Journey through Utopia*, chap. 1; Kumar, *Utopia and Anti-Utopia in Modern Times*, chap. 1; Kenyon, "Utopia in Reality."

[11] Robeyns, "Ideal Theory in Theory and Practice," p. 345. Emphasis was in original. The motif of the voyage is a feature of classical utopian writing. In Bacon's *New Atlantis* a ship arrives at a marvelous island. In *Utopia* More reports the life in a far-off island, that has distinct similarities to the features of England—a far off version (reached by a long journey) of the current world. See also Goodwin and Taylor, *The Politics of Utopia*, chap. 2.

[12] Rawls, *The Law of Peoples*, p. 138.

[13] Ibid., p. 16. This is also stressed by Hamlin, "Feasibility Four Ways." Traditional utopian writing concurred. See Kenyon, "Utopia in Reality," p. 154.

[14] See Robeyns, "Ideal Theory in Theory and Practice," pp. 346–47; Stemplowska and Swift, "Ideal and Nonideal Theory," p. 379. By "significant class of judgments" I mean

guidance for thinking about justice in our nonideal societies, assisting to "clarify difficult cases of how to deal with existing injustices" and to orient the "goal of reform," helping us to see "which wrongs are more grievous and hence more urgent to correct."[15] Existing institutions are thus to be judged in light of ideal justice, and ideal theory thus provides a goal for societies that pursue justice.[16] Famously, the back cover of *Justice as Fairness* (2001) informs us that "Rawls is well aware that since the publication of *A Theory of Justice* in 1971 American society has moved *further away* from the idea of justice as fairness."[17] Rawlsians thus not only seek to depict a perfectly just society but can employ this knowledge to orient their comparative judgments about, say, the justice of American society in 1971 and 2001.

1.3 Climbing

We shall discover in chapter II that this orienting function of the ideal turns out to be surprisingly complicated. It seeks to combine two tasks—(*i*) identification of the ideally, optimally, or perfectly just society, and (*ii*) comparative justice judgments of less-than-ideal societies. Our all-things-considered judgments about what changes are recommended by justice (§I.1.5) critically depend on judgments about where the ideal is, and how far from it (in a sense that needs to be explained) we are. As Amartya Sen observes, this implies that to make an all-things-considered judgment as to whether justice recommends a move from our current world to nonideal world a, or nonideal world b, we must know which is closer to the ideal, utopian point u.[18] We may have

---

to imply that the ordering of less-than-ideal social worlds need not be complete, but neither is it so incomplete that the theory approaches what I shall call "dreaming" (§I.1.4).

[15] Rawls, *Justice as Fairness*, p. 13.

[16] Rawls, *A Theory of Justice*, p. 216. (Unless explicitly noted, all references are to the 1999 revised edition.)

[17] Rawls, *Justice as Fairness*, back cover. Emphasis added.

[18] Sen, *The Idea of Justice*, pp. 98–99. I consider Sen's own proposal in §IV.1.2. In Valentini's terms, the ideal is functioning here as an "end-state." "Ideal vs. Non-ideal Theory," pp. 660–62. Wiens refers to this as a "target" function of the ideal, an idea that he challenges in "Prescribing Institutions without Ideal Theory." Hamlin and Stemplowska are also critical: "Theory, Ideal Theory and the Theory of Ideals," pp. 51ff.

firm grounds for concluding that *a* is more just than *b*, but unless we also know where *a* and *b* are in relation to *u*, we do not know whether moving to *a* or *b* would be recommended by justice. Given these complexities, Sen argues that we should simply focus on what we are really concerned about, the relative justice of *a* and *b*, and forget about comparing them to *u*, which is not only difficult but, happily, unnecessary. He writes:

> The possibility of having an identifiably perfect alternative does not indicate that it is necessary, or indeed useful, to refer to it in judging the relative merits of two alternatives; for example, we may be willing to accept, with great certainty, that Mount Everest is the tallest mountain in the world, completely unbeatable in terms of stature by any other peak, but that understanding is neither needed, nor particularly helpful, in comparing the peak heights of, say, Mount Kilimanjaro and Mount McKinley. There would be something off in the general belief that a comparison of any two alternatives cannot be sensibly made without a prior identification of a supreme alternative.[19]

This passage is crucial for understanding the contrast between ideal theories and Sen's comparative approach. An ideal theory begins with identifying an ideal within a set of possible worlds—or, using a somewhat more formal language, a "global optimum" in the domain {*X*}—and evaluates all options in relation to it, whereas Sen's analysis does not concern itself with an ideal but only "whether a particular social change would enhance justice."[20] The latter, Sen frequently tells us, is a comparative, essentially pairwise, exercise. What we need to know is whether *a* is more just than *b* (whether Mount McKinley is higher than Mount Kilimanjaro); we need not know anything about *u* (or Mount Everest) to make this decision. We seek a theory that allows us to make comparisons about "the advancement or retreat of justice."[21]

A. John Simmons offers a counteranalysis:

[19] Sen, *The Idea of Justice*, p. 102.
[20] Ibid., p. ix.
[21] Ibid., p. 8.

While Sen's point about Everest and determining the heights of smaller mountains is obviously true, its use in Sen's analogy is, I think, potentially misleading.... Which of two smaller "peaks" of justice is the higher (or more just) is a judgment that matters conclusively only if they are both on equally feasible paths to the highest peak of perfect justice. And in order to endorse a route to that highest peak, we certainly do need to know which one that highest peak is. Perhaps for a while we can just aim ourselves in the general direction of the Himalayas, adjusting our paths more finely—between Everest and K2, say—only when we arrive in India. But we need to know a great deal about where to find the serious candidates for the highest peak before we can endorse any path to them from here.[22]

If we focus on the metaphor of a mountain range, our aim may be to reach the highest peak; if that is the goal of our journey it will not help much to know which of two local peaks is higher (*a* or *b*). Even if *a* is higher than *b*, we want to know whether climbing *a* takes *us closer or further from highest peak in the range*. We do not only want to know whether the elevation *a* is greater than *b* (whether Mount McKinley is higher than Mount Kilimanjaro); we want to know something about the longitude and latitude—whether moving our society to *a* or *b* moves us closer to *u*.[23]

The dispute between Sen and Simmons turns on the relevant dimensions involved in deciding whether justice recommends a move to *a* or *b*. Here we must introduce more rigor to get beyond instructive, but loose, metaphors and really grasp what the debate is about. As is well known, Sen has made fundamental contributions to axiomatic social theory, which concerns the properties of preference orderings and methods of aggregating two or more such orderings into a social

[22] Simmons, "Ideal and Nonideal Theory," p. 35. Geographic metaphors are commonplace in the literature; see Stemplowska, "What's Ideal about Ideal Theory?," p. 336; Robeyns, "Ideal Theory in Theory and Practice," p. 345; Schmidtz, "Nonideal Theory."

[23] This debate is not only between Simmons and Sen. For an analysis supporting the necessity of ideals, see Valentini, "On the Apparent Paradox of Ideal Theory." For a criticism, see Lawford-Smith, "Ideal Theory—a Reply to Valentini."

preference ordering. Now individual i's complete transitive preference ordering O_i is an ordering of options of some option set, such that for any two elements $\{a, b\}$ in that set, either a is preferred to b, b is preferred to a, or a and b are indifferent.[24] What is important about any single O_i is that it is, *in the end*, unidimensional, in the sense that no matter how many dimensions of evaluation may have been involved in i's deliberations in coming to her ranking, once she has a well-formed ordering, all the options can be arrayed along a single dimension from best to worst.[25] Given this social choice presupposition, Sen is absolutely correct to suppose that the height of the mountains (a metaphor for how highly ranked in terms of justice an option is) exhausts the relevant dimensions, for there *is* only one dimension in an individual's ordering and, as we shall see, for Sen, judgments of justice are about the orderings of impartial spectators over social worlds (§IV.1.2).[26]

We can give two interpretations—one substantive and one formal—of Sen's claim that only pairwise judgments are necessary; that is, that in determining the justice of $\{a, b\}$, we need not worry about how a and b compare to the best, or optimal, social world, u. On what we might call the *inherent binariness of judgments of justice* interpretation, Sen is advancing a substantive thesis about justice, according to which a judgment of the relative justice of any two social worlds $\{a, b\}$ never depends on information about a third alternative, the optimum social world, u. (Or, more strongly, the $\{a, b\}$ judgment never depends on any third option c, whether or not $c = u$).[27] This is a substantive thesis as it

[24] The assumptions of completeness and transitivity can be weakened; investigating incomplete preferences over rankings of states of affairs in terms of justice is a central concern of Sen's in *The Idea of Justice*. See my "Social Contract and Social Choice."

[25] Trivially, the person's ordering will be single-peaked with the best at the top, declining down to the least preferred, showing that it is unidimensional. See my *On Philosophy, Politics, and Economics*, pp. 162–64.

[26] If we aggregate different individual orderings into an overall social ordering, this aggregation may reveal multiple dimensions. In his work on justice, Sen is careful to employ only the most modest forms of aggregation, in which the pathologies associated with such aggregative multiple dimensionality (e.g., cycles) do not appear (§IV.1.2).

[27] On binariness and its relation to such independence, see Sen, *Collective Choice and Social Welfare*, pp. 7ff., 16ff. For a deeper analysis, see his "Choice Functions and Revealed Preference."

implies that, in evaluating the justice of social worlds a and b, information about the proximity of a and b to u (or any c) is irrelevant. The justice of a and the justice of b depend solely on the features of the states of affairs that constitute them, and this does not include their relation (say in terms of proximity) to other alternative, c. This is perhaps the most natural interpretation of Sen's advocacy of his pairwise approach, and of many of his criticisms of Rawls's insistence on the importance of knowing the ideal when making judgments about less-than-ideal social worlds. However, his pairwise approach by no means necessarily assumes it. Another, more formal, interpretation might be called the all-things-considered *ordering view*. Suppose that person i's preference ordering of the relative justice of a set of options $\{a, b, c\}$ depends not only on how well a and b do when evaluated in terms of the relevant standards of justice, but also on the relative distance of, respectively, a and b, to a third, superior option c (which may or may not be the ideal state, u). That is, in thinking through the issue of whether a was more or less just than b, a person considers both the inherent justice of worlds a and b, as well as how "far away" (again, leaving aside for now what that might mean) a and b are from c. If our person i nevertheless arrived at a complete transitive ordering of $\{a, b, c\}$ in terms of each element's justice on both these dimensions, the "distance from c" information would already be accounted for in i's preference ordering, and the place of a and b in it. Given this, it would still be strictly true that *given* this ranking, to say whether a or b is more just, does not require knowing what social world is at the top of the ordering; knowing the top of the ordering is not relevant to a pairwise choice. Yet, unlike on the *inherent binariness* interpretation, to arrive at the ordering, distance information would be required, and so knowledge of the ideal would be presupposed in developing the ranking.

Given that Sen is deeply skeptical that knowledge of the ideal is really of much use at all, we should definitely reject this second interpretation of his social choice approach. However, neither should we saddle him with a pure version of the *inherent binariness* interpretation. Sen need not be absolutely committed to ignoring all "distance" or "directional" information (again, for now let us use these vague ideas drawn from Simmons). For example, suppose that though a's features

were more just than *b*'s, small modifications of *b* would lead to a far more just world, *c*, while this is not the case with *a*. Nothing Sen says indicates that he could not take this into account, and so judge that all things considered *b* is more just than *a*, because of its "proximity" to *c*. Nevertheless, it seems safe to read Sen as saying that such judgments are not fundamental to a theory of justice, and for the most part we should focus on the fundamental inherent features of a social world when determining its justice. Thus, basically, in determining the justice ranking of {*a, b, c*} we should focus on the justice of the inherent (not relational vis-à-vis other worlds) features of each world and how people fare in it. Consequently, knowledge of the ideal is of little or no use when making pairwise comparisons about nonideal worlds. In seeking more justice, we basically wish to climb up the ordering of more inherently just social worlds. To do that we do not need to see the top of the ordering, or make many "distance" judgments. It is essential to realize that Sen's climbing-up-the-order approach in no way implies that improvements must be incremental or conservative; if we are now at the social world judged fortieth and our option set includes the thirty-ninth and tenth options in our ordering, Sen certainly does not hold that we first move to the thirty-ninth. It is fundamental to distinguish Sen's claim that judgments of justice are basically binary, and so do not rely on their relation to an ideal point, from the common misunderstanding that they are therefore incremental.[28]

Simmons's counterpoint to Sen is best interpreted as insisting that the social choice ordering model is inappropriate because we should keep distinct at least two dimensions of evaluation in comparing the justice of *a* and *b*. Fundamental to thinking about justice is not simply the inherent justice of *a*'s and *b*'s social structures, but "how close" in terms of, say, similarity or feasibility,[29] *a* and *b* are to the ideal point (or global optimum) *u*. Thus because there are now two dimensions of evaluation being kept separate,[30] it is perfectly possible for *a* to be

[28] Cf. Gilabert, "Comparative Assessments of Justice, Political Feasibility, and Ideal Theory."

[29] I shall distinguish these in some detail in chapter II.

[30] Recall that one possibility for a social choice analysis was the all-things-considered *ordering view*, which aggregates both dimensions into an overall social ordering; thus the importance of keeping them distinct.

closer to u on one dimension (say, the justice of its institutional structure), but further away on the distance dimension (in some proximity
sense it is further from u than is b).

On this multidimensional analysis, then, it matters a great deal
where the ideal is, because some metric of distance to the ideal is a
fundamental element of the overall comparative judgments of whether
justice recommends a move to a or b. And the multidimensional analysis allows us to see why the ideal would be necessary in orienting our
judgments about justice; we wish to make not only evaluations of the
binary inherent justice of two states, but also determinations as to
whether one or the other brings us closer in the proximal sense to the
ideally just condition. We can now see a way to more rigorously thinking about how the ideal can orient our judgments of justice, a task to
which I turn in the next two chapters. In this book, then, I shall explore
multidimensional ways of thinking about justice, for they provide the
most compelling response to Sen's elegant unidimensional analysis—
an analysis that makes the ideal otiose.

1.4 Dreaming

As I have stressed—and as we will see in upcoming chapters—this multidimensional, orienting-through-the-ideal approach to political philosophy is complicated. We need to know the justice of various social
states as well as their proximity to the ideal (where these are different,
not well correlated, dimensions of justice). One way to avoid the inevitable complications is Sen's: unleash comparative judgments from reliance on an ideal, and so render the ideal unnecessary.[31] Then we need
to know only the relative justice of social states. Something like the
opposite simplifying strategy is to depict only the ideally just state or
condition, and make no claims about the justice of less-than-ideal situations. There is nothing incoherent about this; it could be that all less-
than-ideal states are incommensurable or equal (everyone but the
winner is equally a loser), and so cannot be ranked in terms of strictly
better or worse than one another. It is too strong to say that such a

---

[31] Or, alternatively, develop a single ordering where "more just" always implies
"closer to the ideal" as in what I called the all-things-considered interpretation of the
social choice approach (§I.1.3).

theory is entirely useless,[32] but a political philosophy that is unable to describe any but the top, ideal social state, is of little use in helping us sort through the options for justice that confront us.[33] It is as if we have developed a clear conception of the ideal square, but are unable to say which of three drawings, a square, a rectangle, or a circle, is closest to it.

One might contend that a political philosophy that presents a vision only of the fully just society can have value. This contention would seem supported by the historical importance of Martin Luther King's "I Have a Dream" speech. Recall, however, that his litany of dreams of justice and harmony in America was offered as a source of hope and faith to help overcome the alienation and bitterness of his audience, arising from the fact that their actual, here-and-now demands for justice were met with contempt and hatred. The speech commenced not with mere dreams, but an assertion that the American founders issued "a promissory note to which every American was to fall heir. This note was a promise that all men would be guaranteed the inalienable rights of life, liberty, and the pursuit of happiness." So, King proclaims to the marchers, "we have come to cash this check—a check that will give us upon demand the riches of freedom and the security of justice."[34] A world where the requirements of complete justice—or perhaps even a world where the commitments of the American founders—are fully met is perhaps an unattainable ideal. I do not claim that a useful ideal theory must posit attainable ideals. But King nevertheless insisted that the ideal orients current political demands. Even if the check that King demanded be cashed was one that fell far short of the promised value, the critical point is that the ideal orients political demands in our manifestly nonideal world. King's "dream" is not one from which we awake—a dream of a perfect world—that is unrelated to the pursuit of justice in nonideal social worlds. By "dreaming" I mean to isolate a different view of ideal justice that not only proposes an unattainable

---

[32] See Stemplowska and Swift, "Ideal and Nonideal Theory," pp. 377–80.

[33] Swift argues in favor of this ordering function of philosophical analysis. "The Value of Philosophy in Nonideal Circumstances."

[34] King, "I Have a Dream...."

ideal, but is willing to admit that this ideal may be, and often is, entirely irrelevant to improvements in justice in nonideal conditions.

Divorced from judgments of justice about the social world in which we find ourselves, and those to which we might move, mere dreams of ideal justice may inspire or give hope, though they may also lead to hopelessness, despair, and cynicism. To wake from a dream of a world of perfect justice and confront the realities of our social world, without any way to connect the dream to the problems and questions arising in our social world, is all too likely to disorient thinking about justice.[35] We might think of it as surrealist justice, seeking to depict in our world a vision of a dream world that does not really translate.[36] What one "knows" about justice is discontinuous with the questions of justice arising from life in one's actual and near social worlds. David Miller is quite right: on such a view "there is nothing left for political philosophy but lamentation over the size of the gap that unavoidably exists between the ideals it defends and the actual conditions of human life."[37]

David Estlund upholds such a "hopeless" view of justice:

Consider a theory that held individuals and institutions to standards that it is within their ability to meet, but which there is no reason to believe they will ever meet…. It would be morally utopian if the standards were impossible to meet, but, again, by hypothesis, they are not. Many possible things will never happen. The imagined theory simply constructs a vision of how things should and could be, even while acknowledging that they won't be…. So far, there is no discernable defect in the theory, I believe. For all we have said, the standards to which it holds people and institutions

---

[35] In Bacon's *New Atlantis* (p. 23), a very few of those sailors who have happened on utopia have elected to return to the less-than-ideal world. "What those few that returned may have reported [about our utopia] abroad I know not. But you must thinke, Whatsoever they have said, could bee taken where they came, but for a Dreame."

[36] On the other hand, writing of the Bolshevik avant-garde, Buck-Morss maintains that "By not closing the gap between dream and reality, the artwork of the avant-garde left both dream and reality free to criticize each other." *Dreamworld and Catastrophe*, p. 65.

[37] Miller, *Justice for Earthlings: Essays in Political Philosophy*, p. 231.

may be sound and true. The fact that people will not live up to them even though they could is a defect of people, not of the theory. For lack of a better term, let us call this kind of theory a version of *hopeless realism*.[38]

I have claimed that such a theory is of little use. Estlund admits it may well be: "The value of unrealistic theory is not my point.... The point is that a hopeless normative theory might be the true theory. Admittedly, not all truths are of great value. The telephone book contains many relatively unimportant truths. We are talking about the truth about justice, however, and I am inclined to think that there is more importance here, but perhaps this is only because we would have thought that it would be of practical value."[39] There may well be no realistic prospect that we will do as justice "requires," and thus it may also be that given this, such justice "might have no practical value at all" once we ask what, given our imperfections, we ought now to do.[40] Estlund seems to rather relish this possibility. He makes a great deal of cases where "it ought to be the case that Alf *Xs*" even though it would be deeply inadvisable for Alf to seek to *X*, because he is likely to fail, and in trying to *X* he is apt to bring about worse results than if he had pursued a more "realistic" alternative.[41] But Estlund's thesis is even more radical: it stresses that knowing that *X* ought to be the case might not be of relevance to practice in the sense of giving some reasons to look for ways to secure *X*, if not now, then in the future. Thus knowledge that *X* ought to be the case may have "no practical value at all."

Like others, Estlund draws on Lipsey and Lancaster's "theory of the second best," according to which if *u* is the Pareto-efficient state, it does not follow that a state of affairs that is almost identical to *u* will be almost Pareto efficient.[42] Thus, at first blush, it would seem that the

[38] Estlund, *Democratic Authority*, p. 264. For extended analyses, see Estlund's "Utopophobia" and "Human Nature and the Limits (If Any) of Political Philosophy."

[39] Estlund, "Utopophobia," p. 133.

[40] Ibid., pp. 132–33.

[41] See Estlund, "Prime Justice." This sort of concern has led political philosophers to elevate the role of deontic logic, and analyses of "ought implies can." "Can it be the case that 'you ought to *X*' if you are unlikely to *X* even if you 'try'?" (Or sometimes even if you "keep trying"?) See Estlund, "Prime Justice," and "Human Nature and the Limits (If Any) of Political Philosophy"; Lawford-Smith, "Non-ideal Accessibility."

[42] Lipsey and Lancaster, "The General Theory of Second Best." See Estlund, "Utopo-

goal of approximating the ideal might be ill founded: if we cannot reach it, getting close may be of no use at all in securing justice.[43] Knowing ideal justice thus could be more like knowing rather abstruse theorems in mathematics than a guide to action.[44]

Estlund's point is, broadly speaking, a conceptual one: 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.[45] It is not, however, my concern to adjudicate this dispute; for the most part, I set it aside. My chief concern is analyzing a class of ideal theories according to which the specification of an ideal *is* of critical importance in making relative judgments of justice and, in some way, of orienting practice. Still, while not seeking to jump into this particular fray, I am supposing that throughout its long history, political thinkers so often have been drawn to utopian ideals because they have been convinced that they *do* help orient our less-than-ideal judgments, and *do* provide recommendations about how to think about social and political change, even when this change seems infeasible now, or even in the near future. This is not a conceptual truth, but a truth about the utopian tradition in political philosophy. As I have stressed, it is simply a misreading of utopian thought to equate it with hopeless, unrealizable ideals: it sim-

––––––––––––

phobia," p. 121; Wiens, "Against Ideal Guidance," p. 441; Wiens, "Political Ideals and the Feasibility Frontier," p. 22.

[43] Appeals to Lipsey and Lancaster's work on the second best in economics is suggestive, but political theorists have been rather too quick to assume that analyses of efficiency carry over into theories of the ideal. Chapter II of this volume presents a model that explains when, in a theory of the ideal, approximations to the structure of the ideal state also approximate its justice, and when they do not. Illuminating this, we shall see, is a distinct advantage of this model.

[44] Others have likened moral inquiry to physics. See Enoch, "The Disorder of Public Reason." Yet another model of inquiry into justice has been "epistemology"—the theory of knowledge (see Swift, "The Value of Philosophy in Nonideal Circumstances," pp. 366ff.). Even this manifestly more appropriate (because it is also a normative inquiry) comparison does not seem quite apt. The epistemologist might give us an account of the conditions for justified belief, but while a person such as Betty, who continues to use a Ouija board, violates the norms of rationality, she does not fail to give anyone else what is owed them, nor do others have a right to demand that she cease and desist using her Ouija board. Martin Luther King was neither merely dreaming of perfect justice, nor was he merely imparting to white Americans important theoretical knowledge about the norms of justified racial relations.

[45] See Farrelly, "Justice in Ideal Theory: A Refutation."

ply does not follow that one who resists a "hopeless utopia" is a "utopophobe." It is hard not to be worried about a trend in academic philosophy that searches for "true justice" and "utopia" by turning its back on so much of this tradition, resolutely defending the possibility that its subject is utterly useless.

1.5 Recommending—Rescuing Justice from Uselessness

In the remainder of this book, then, I set aside ideal theory as mere dreaming; whether or not achieving the ideal is itself *hopeless*, my concern is the type of political theorizing in which the ideal is not *useless*—in which the ideal serves as a criterion that assists in regulating, directing, or facilitating less-than-ideal judgments of comparative justice. On these views knowledge of the ideal is informative in making less-than-ideal judgments. But even such a theory of the ideal, which is not mere dreaming—in which knowledge of the ideal informs evaluative judgments about the relative superiority of nonideal conditions—could still be a sort of purely academic enterprise, conveying theoretical knowledge of comparative evaluations that never provides anyone with guidance. From the point of view of working toward justice, such a theory would be useless all the way down: the whole enterprise would result in, perhaps, a beautiful theory of justice that yields evaluations of a wide variety of imperfect social states oriented by the ideal, but never manages to yield recommendations about what actual agents ought to do or strive for. The theory would be thoroughly useless.

I am not so naive as to suppose that this is not an attractive goal for many philosophers, but it does make rather puzzling the paraphernalia of the practice of justice: demanding, complaining, blaming, claiming, agitating, "the justice system," punishments, editorials, social movements, consciousness-raising, and revolutions. Estlund's theory, while allowing that "prime justice" may be hopeless, admits the category of "concessive" justice—the justice that applies to Alf when he knows that he will not conform to prime or ideal justice. A theory that is hopeless all the way down is truly a truncated theory of justice.[46] Such a theory fails to make sense of the way justice enters into our form of life.

[46] Estlund, "Prime Justice."

Again, this is not to say that ideal justice itself must be implementable; even in a nondreaming account, ideal justice still can be a hopeless (but not a useless) vision that orients, inspires, and guides us. Nor is it to say that for every well-formed judgment that one social state is more just than another, a theory of justice must yield a practical recommendation to move to the more just condition. Note the range of practical activities listed in the previous paragraph; claims and demands can be quite specific, but social movements and consciousness-raising guide action in much vaguer ways. The point is merely that, as I shall consider it here, ideal theory is not simply about theoretical evaluative judgments of social states. The orientation of *judgments* of justice cannot be the entire story; there must arrive a point at which a political philosophy concerned with justice yields recommendations about moving to more just social states. An ideal might be simply aspirational, but not the entire field of justice.

When a political philosophy focused on justice moves from judgments of justice to recommendations, it must move away from the sphere of pure moral philosophy, understood simply as the justification of normative judgments.[47] At this point, political philosophy must draw on social and economic theory, political science, and psychology in seeking to formulate policies, recommendations, and demands.[48] Depending on the context, this will require employing different social science toolkits and predictive models. A political theorist advocating a constitutional structure, one advocating a tax policy, and one advocating a change in particular laws (say, about school segregation), will obviously draw on different data, toolkits, and models, but it will be rare indeed that no significant social science will be drawn on. Even at the judicial stage, where it may seem that "recommendations" are simply derivations from legal requirements, courts regularly invoke social theory and science. Thus, for example, in the landmark *Brown v. Board of Education of Topeka*, which was without doubt about basic claims of justice, the Supreme Court noted the psychological harms of school segregation.[49] Because the recommendation task of a theory of justice

[47] This is not to say that it ever must have occupied such a sphere.

[48] See Wiens, "Demands of Justice, Feasible Alternatives, and the Need for Causal Analysis."

[49] *Brown v. the Board of Education of Topeka*, 347 U.S. 483; 74 S. Ct. 686; 98 L. Ed. 873

or political philosophy requires a set of tools, data, and models, analyzing this aspect of a theory of justice requires understanding the ways that political philosophy draws on social science.

2 Social Realizations and the Ideal

2.1 Perfect Principle Conformity and Ideal Societies

I have thus far dealt with "the ideal" as an unspecified variable for optimal or perfect justice. Although I wish the analysis to remain reasonably general—it certainly should not turn on a precise understanding of the ideal—it is important to identify the constraints on characterizing an ideal as I will examine it. My aim is a sufficiently broad characterization such that it captures the essentials of an interestingly large group of theories while, at the same time, lending itself to a coherent and fruitful analysis.

G. A. Cohen's account of justice is often depicted as a sort of "ideal theory," indeed often as an instance of the most extreme type.[50] Cohen—and in this respect he shares much with Estlund[51]—is adamant that a claim that some fundamental principle of justice J is the correct principle cannot be defeated by a demonstration that J is outside the bounds of what humans plausibly can be expected to do. That is, on his view J's claim to be a fundamental principle cannot be defeated by the finding that, given psychology and other social sciences, it is highly improbable that people will conform to it. The fundamental truths of justice, it is said, do not depend on such practicality considerations, nor do they derive from any set of facts about a group's current beliefs and convictions.[52] This suggests a characterization of the ideal, the

> *Principle-Defined Ideal*: Supposing that we have identified principles of justice J that do not depend on feasibility constraints or sociological facts (such as current beliefs about justice), for any

---

(1954). The court specifically cited a social science study on doll choice by children: white dolls were chosen consistently over black dolls.

[50] Farrelly, "Justice in Ideal Theory: A Refutation," pp. 846–47.

[51] For their differences, see Estlund, "Utopophobia," pp. 129–30.

[52] Cohen, *Rescuing Justice and Equality*, pp. 20–21, chaps. 6 and 7.

social state, S, it is an ideal social state if and only if in S everyone affirms J and perfectly conforms to J. All other social states are nonideal.

Cohen's text actually does not support this interpretation of the ideal. "Ideal theory," he says, is "what transpires in the ideal society, and nonideal theory . . . applies to settings in which, *among other things*, citizens do not affirm and act upon the correct principles of justice."[53] The implication of the italicized phrase is that nonideal states are not merely defined by lack of endorsement and conformity to the relevant principles. In contrasting the ideal liberal and socialist societies, Cohen tells us that "in the ideal socialist society, equal respect and concern are not projected out of society and restricted to the ambit of an alien superstructural power, the state. If the right principles are, as Marx thought, the ones that are right for everyday, material life, and if they are practices in everyday life, as the socialist ideal utopianly envisages they will be, then the state can wither away."[54]

We see here that the ideal is characterized by not only conformity to the relevant principles, but a specification of the aspects of the social world to which the principles apply ("the practices"); the socialist and liberal dispute about the ideal does not involve a dispute simply about principles, but about the parts of the world that are to be regulated by the principles. Whereas, Cohen suggests, the liberal sees ideal justice as regulating the state (and, we might add, more generally, the basic structure of social cooperation), the socialist holds that the principles apply to various practices of everyday life. This, though, means that the identification of the ideal society must involve specification of various types of institutions, practices, or spheres of life;[55] if there is no state or no basic structure, for example, Cohen indicates that simply *cannot* be liberal justice. But then a liberal ideal must provide

---

[53] Ibid., p. 221. Emphasis added.

[54] Ibid., p. 1. Note Cohen's rather interesting reading of Marx as a utopian socialist. Certainly Cohen seems to qualify as one. Recall Engels's characterization of the utopian socialists: "To all these socialism is the expression of absolute truth, reason, and justice. . . . And . . . absolute truth is independent of time, space, and of the historical development of man." Engels was not a great fan of such views. "Socialism: Utopian and Scientific," p. 693.

[55] See here Habermas, "Reconciliation through Public Reason," p. 129.

an account of what constitutes state structures, or what institutions are part of the basic structure. In Rawls's liberal ideal, for example, the family sits (somewhat uncomfortably) in the basic structure.[56] Rawls's liberal ideal thus differs from that of a liberal who would exclude the family, yet both may endorse Rawls's two principles of justice.[57] And conceptually, the liberal and socialist might agree on the same principles, yet their visions of the "ideal society" could still radically differ with regard to the parts of social life to which they apply.

Although, then, we might start off thinking of "ideal justice" as simply universal endorsement of, and action in accordance with, the correct, true, or best principles of justice, to make sense of ideal justice we need to suppose a set of institutions, practices, or spheres, and the agents that act in them—and all this requires some account of how institutions or agents (could possibly, are apt to, can without excessive strain) operate, and how various spheres are demarcated.[58] As Henry Sidgwick observed, "even an extreme Intuitionist would have to admit that the details of Justice and other duties will vary with social institutions."[59] Even the idea of agency itself needs specification. For example, it needs to be determined whether groups are agents subject to principles of justice and, if so, what *are* group agents, and for what can they possibly be held accountable.[60] At the very minimum, our vision of Paradise must specify features of the social world to which the

---

[56] Rawls, "The Idea of Public Reason Revisited," pp. 595ff.

[57] "(I) Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system for all. (II) Social and economic inequalities are to be arranged so that they are both: (a) to the greatest benefit of the least advantaged, consistent with a just savings principle [the difference principle], and (b) attached to offices and positions open to all under conditions of fair equality of opportunity." Rawls, *A Theory of Justice*, p. 266.

[58] See §V.3 on the charge that this confuses justice with "rules of regulation." The point here is that even if one seeks to sharply distinguish principles of justice from some of their institutional embodiments, an ideal cannot be simply equated with maximal fulfillment of principles of justice.

[59] Sidgwick, *The Methods of Ethics*, p. 20.

[60] See, for example, List and Pettit, *Group Agency*. For worries that there may be much more diversity within group agency than List and Pettit's model accommodates, especially concerning moral responsibility, see my "Constructivist and Ecological Modeling of Group Rationality."

principles apply, and some suppositions about the workings of the agents regulated.[61]

2.2 Justice and Its Social Realization

I have been stressing that even theories that, prima facie, identify the ideal in terms of perfect conformity to principles also include reference to social institutions and the way they function. More fundamentally, though, any plausible notion of the ideally just society must take seriously the way such a society would actually work out. Recall Sen's charge that Rawls succumbs to "institutional fundamentalism." Although he focuses on institutionalism and rules, the general category Sen is attacking is rather broader, encompassing much of what is called "deontology," whether that be focused on institutions, rules, or fidelity to principles.[62] Sen rejects such "fundamentalism" because "justice is not just a matter of judging institutions and rules, but of judging societies themselves."[63] I believe that Sen is correct that there is something extraordinarily severe and unattractive about a conception of justice that is entirely blind to the social realization of conformity to its principles, as if the only thing relevant to justice is that a certain set of principles be obeyed, come what may. Sen is right that such a view would be implausible. What is wrong, however, is his charge that Rawls or other plausible versions of principle-based justice are blind in this way.[64] It is certainly true that Rawls sought to include "an *important element* of pure proceduralism" in the design of institutions: to a significant extent he aims to discover a set of institutions whose outcomes, whatever they might be, would be just.[65] But these

---

[61] Estlund, for example, writes that "requirements of social justice morally require things of societies as such." As he recognizes, this means that collective responsibility is a matter for ideal justice. "Human Nature and the Limits (If Any) of Political Philosophy," p. 210.

[62] Sen, *The Idea of Justice*, pp. 21, 210. Drawing on the philosophic tradition of India, Sen associates these ideas with *niti*—a severe, rule- or principle-based approach to justice. See §IV.1.2.3.

[63] Ibid., p. 20. Emphasis added.

[64] But see Sen, "A Reply," pp. 327–28.

[65] Rawls, *A Theory of Justice*, pp. 57–58. Emphasis added.

institutions are designed in light of their known and predicted effects: the aim is a "fair, efficient, and productive system of social cooperation" that is "maintained over time."[66] Recall Rawls's characterization of teleology and deontology: "Utilitarianism is a teleological theory whereas justice as fairness is not. By definition, then, the latter theory is deontological, one that does not specify the good independently from the right, or does not interpret the right as maximizing the good. (It should be noted that deontological theories are defined as nonteleological ones, not as views that characterize the rightness of institutions and acts independently from their consequences. All ethical doctrines worth our attention take consequences into account in judging rightness. One that did not would simply be irrational, crazy)."[67]

A sharp distinction between an ideal theory focused on principles and one focused on social structures is, ultimately, untenable. All principle-based theories "worth our attention" are sensitive to their social realizations.[68] And this includes Kantian principled theories as well. We can better appreciate how this is so by looking at Rawls's interpretation of Kant's universal law formulation of the categorical imperative, which he models in terms of a four-step "CI procedure." The first three steps of the CI procedure are fairly straightforward:

(1) I am to do X in circumstances C in order to bring about Y. (Here X is an action and Y a state of affairs).

The second step generalizes the maxim at the first to get:

(2) Everyone is to do X in circumstances C in order to bring about Y.

At a third step we are to transform the general precept at (2) into a law of nature to obtain:

(3) Everyone always does X in circumstances C in order to bring about Y (as if by a law of nature).[69]

In the fourth step we are to consider the "perturbed social world" that would result from the addition of this new law of nature; we seek to

---

[66] Rawls, *Justice as Fairness*, p. 50.
[67] Ibid., p. 26.
[68] Ibid.
[69] Rawls, "Themes in Kant's Moral Philosophy," p. 499.

understand the new "equilibrium state" on which this perturbed social world would settle. We are then to ask ourselves whether, when we regard ourselves as a member of this new social world, we can "will this perturbed social world itself and affirm it should we belong to it."[70]

Thus on Rawls's analysis of Kant's "deontology," our endorsement of a moral imperative takes place against the background of a model of the social world and the effects of introducing a maxim. To characterize an ideally just condition is not simply to posit that certain principles are in fact followed; we must consider the social realization of conformity and consider whether we can maintain our judgment that the realized world would be ideally just. Consider a case that worried Rawls: we come to the conclusion that a particular principle is the best principle of justice, and we reflect on the social world where everyone conforms to it. Assume additionally, however, that when we consider the social realization of this perfect compliance individuals become deeply alienated from their sense of justice.[71] Perhaps they curse their very sense of justice, and bemoan the day they discovered true justice, which sets them at odds with their deepest convictions about their place in the universe. Cynicism and alienation are rife. We might wonder: is this a just world, in which those devoted to justice so suffer for their devotion? If we think not, we have at least two options: either conclude that under these social realizations, perfect compliance with this principle is not ideally just, or posit a social world where the social realization of universal justice is not so dire and identify *that* as the ideal.[72] Either option acknowledges that the social realization of universal compliance with our proposed principle is relevant to a claim that it regulates the ideal social world.

In an important series of papers, David Wiens has also cast doubt on whether a plausible ideal theory can be cast entirely in terms of a set of substantive principles, with comparative justice determined simply by comparative satisfaction of the correct principles.[73] Wiens

[70] Ibid., p. 500.

[71] Rawls, *A Theory of Justice*, p. 295. See also Weithman, *Why Political Liberalism?*, p. 53, and my "The Turn to a Political Liberalism."

[72] This second route is taken by Barry and Valentini, "Egalitarian Challenges to Global Egalitarianism," p. 509.

[73] Wiens, "Against Ideal Guidance"; "Will The Real Principles of Justice Please Stand Up?"; "Prescribing Institutions without Ideal Theory."

distinguishes *evaluative criteria* from *directive principles*. On his interpretation, for example, "Rawls's principles are meant to characterize an ideal balance of certain basic values, that is, the balance of basic values realized by a fully just institutional scheme."[74] The original position is thus understood as a way in which a complex set of criteria—liberty, equality, reciprocity, fairness, and so on—can be employed to yield directive principles for a set of institutions. "These moral conditions modeled by the original position operationalize certain 'commonly shared' basic values and, thus, represent basic evaluative criteria (as I understand that concept)."[75] Wiens argues that there are social worlds where the set of values that lie behind Rawls's two principles of justice would not be best satisfied by recommending worlds whose institutions most closely satisfy the two principles. Consider, for example, social worlds *a* and *b*. In social world *a* the difference principle is best satisfied; in social world *b* it is not quite as well satisfied, but *b*'s institutions guarantee much more equality than in world *a*. Wiens posits the relations in figure 1-1.

Let us stipulate that there are no effects on the satisfaction of the first principle of justice or fair equality of opportunity: they are equally well met in *a* and *b*. Wiens's insight here is that the values that the original position seeks to model would plausibly favor world *b* over *a* even if, *ex hypothesi*, the two principles are better met in *a*. The two principles of justice, we might say, only adequately model Rawls's underlying evaluative standards in a (proper) subset of the social worlds to be evaluated. This reading is confirmed by Rawls, who argued that the "General Conception of Justice," not his famous two principles, is the best articulation of the underlying values modeled in the original position in some empirical circumstances (essentially those of pressing scarcity).[76] Thus Rawls explicitly acknowledged the inadequacy of evaluating the justice of all social worlds in terms of the two principles

[74] Wiens, "Against Ideal Guidance," p. 436.

[75] Ibid. The quoted phrase is from Rawls, *A Theory of Justice*, p. 16.

[76] "All social values—liberty and opportunity, income and wealth, and the social basis of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage." Rawls, *A Theory of Justice*, p. 54. The general conception has a diminished place in the revised (1999) edition of *Theory*; note that it appears in the 1971 edition under the final statement of his principles (1971, p. 302), but is omitted from the relevant passage (p. 267) in the 1999 edition.

	Principle	Least advantaged	Most advantaged
World a	Difference principle	x	$130x$
World b	Stricter equality	$9/10x$	$45x$

Figure 1-1. Distributive shares under two principles (modified from Wiens, "Against Ideal Guidance," p. 443)

(which he called the "Special Conception" of justice). It must follow, then, that if he advances a coherent theory, which sometimes employs the two principles to evaluate social institutions and in other cases the General Conception, there must be deeper underlying evaluative standards than the Special Conception.

One might dispute the importance of the case of Rawls. One might grant that both his "four step CI-procedure" and his recognition that the two principles are inappropriate in some circumstances do indeed indicate that Rawls tests principles in terms of the acceptability of their social realizations given fundamental values or evaluative standards. But one might argue that Rawls is driven to this because of his conviction that justice is the preeminent virtue of social institutions: if, as for Rawls, the overall acceptability of institutions essentially turns on their justice, then the requirements of justice must converge with our judgments of the overall acceptability of a social world and its institutions. Because Rawls holds justice to be such a central virtue of social life and institutions, a social world's overall evaluation must closely track its justice. Others, however, insist that justice is simply one of many virtues:[77] even from the perspective of political philosophy, they claim, our overall evaluation of a social world depends on many values, of which justice is but one. Think again of Wiens's worlds a and b, and suppose we agree with him that b is morally superior to a, even though a better satisfies Rawls's difference principle. But if justice is simply one of many virtues of social institutions, one might say that, while b is better, it is not "through-and-through just"[78]—it fails to meet the principles of justice as well as does a.

[77] Compare Rawls, *A Theory of Justice*, p. 3; Cohen, *Rescuing Justice and Equality*, pp. 302ff.

[78] Cohen, *Rescuing Justice and Equality*, p. 7; see also generally chap. 2.

This is a conceivable view. However, to the extent that justice is one of many virtues of the morally relevant social world, "truths" about justice do not tell us a great deal about the acceptability of our social world. Just as one might say, "Yes, yes, this social world is pretty inefficient, but inefficiency isn't that important" one could then say, "Yes, yes, this social world is pretty unjust, but justice isn't that important." Ideal justice is simply about everyone conforming to some principle come what may, but it may not be especially praiseworthy that such an ideal is achieved. In this case, while the ideal may still orient our thinking about justice, justice no longer orients our evaluation of social and political worlds. Justice is insulated from its social realizations by diminishing its significance.[79] Of course between these radical extremes—of justice being preeminent and it being just one of many virtues of social worlds—justice may be very important, but not quite the first virtue of social institutions. But the more important justice is to our overall evaluation, the more important becomes its social realization, for the closer we expect our evaluations of justice to track our overall evaluations.

2.3 How Well Justified Are Our Principles of Justice?

We might press a more fundamental worry: can we really be so confident of any derivation of principles of justice that we are warranted in insisting that a society that is miserable or deeply dysfunctional, but yet perfectly conforms to them, is the ideally just social world?[80] When we search for a plausible basis for such a radical insulation of the justification of our judgments of justice from their social realizations, we seem to be led to something like a strongly foundationalist account of justification. Roughly, on a strongly foundationalist account, some belief β_1 (*i*) is the basis of the justification of other beliefs $\{\beta_2 \dots \beta_n\}$ such as beliefs about just institutions; (*ii*) β_1 is not justified by any other

---

[79] We might say that although Cohen seeks to save justice from Rawls, he seeks to do so by diminishing it.

[80] Cf. Gilabert and Lawford-Smith: "It is ways of implementing principles in the world that are feasible or infeasible, not principles themselves." "Political Feasibility," p. 811. See also Barry and Valentini, "Egalitarian Challenges to Global Egalitarianism," pp. 507–11.

beliefs; and (*iii*) given the degree of self-evident justification of β_1, there is no other belief β^* that could possibly defeat (i.e., undermine) the justification of β_1. In terms more familiar to moral and political philosophy, β_1 may be among our "deep intuitions"[81] that simply cannot be defeated by pointing out other conflicting values, including those associated with "facts" about the social realizations of the principles expressing β_1.[82] So on this view, if people who conform perfectly to the correct principle of justice, which expresses β_1 in social world *S*, are deeply alienated and cynical, or lead impoverished lives, this might lead us to think that people in *S* simply are not up to justice, but it couldn't lead us to seriously question whether we are correct about the principles of justice. The correct principle is so near self-evident that no mere facts about the disaster of implementing it, or that reflective and goodwilled people do not will such a world, could defeat its justification.[83]

Strongly foundationalist accounts of justified belief have a number of intractable problems, the most important of which is their central claim that "deep intuitions" have very high (approaching unity) levels of credence; it is this that insulates them from defeat by other values, including values that are realized in facts concerning social realizations. Although on some accounts perceptual beliefs may approach this (*some* are tempted to say: "If I believe I perceive red, I cannot be wrong, and this could not be defeated by other beliefs"),[84] it is very

[81] Cohen, *Rescuing Justice and Equality*, p. 22. See also §V.3.

[82] I consider the idea of defeat at some length in *Justificatory Liberalism*, chaps. 5–7. On foundationalism, see pp. 91ff. What I say in the following paragraphs is not directly applicable to (nor is it entirely irrelevant to) Cohen's "fact insensitivity of principles" thesis, which involves a number of claims about the structure of justification and the role of "principles" per se. A detailed discussion of this much-discussed thesis would take us too far afield here; my concern is simply whether (*i*) a claim that some principles are the correct account of justice can be defeated by facts about their social realizations, not whether (*ii*) the justification of those principles depends on social facts. The relation of claims (*i*) and (*ii*) is too complex to explore here.

[83] Cf. Estlund, who seems to grant that only facts about impossibility could defeat a claim of justice. "Human Nature and the Limits (If Any) of Political Philosophy," e.g., p. 210. It is not clear, however, whether Estlund is arguing that a principle of justice immune to facts about dire social realizations is plausible, or simply not incoherent.

[84] And even this is implausible: a person's percepts may be undefeatable, but not that person's perceptual beliefs. That the stick *looks* crooked to me when half im-

hard to accept that intuitions about social justice could possibly have this status. On any plausible view, the level of self-justification of beliefs about justice should be set low enough to admit defeat. Note that this does not mean that basic beliefs about justice cannot be self-justified and foundational (and so in *this sense* they still can have "fact independent" justification), but simply that whatever its degree of self-justification, it is sufficiently less than unity such that its (self-) justification can be defeated by other considerations.[85] Once we allow that regardless of how deep it is, an intuition about justice can be defeated, we can press that the dastardly character of the social world that best realizes our intuitive principles could, at least in principle, defeat the claim that a world that meets these principles is just (in Kantian terms, we could not will that these principles be a law of nature in that social world).

I am not claiming that our moral judgments must be "coherentist" in any formal sense, nor must we embrace "reflective equilibrium." I do, however, believe that we have very good grounds for (*i*) denying that any intuition that a substantive principle of justice is correct has such a high credence that it could not be defeated by other considerations; (*ii*) that when specifying a just social world, the realizations of universal or widespread acting on a principle are relevant defeaters for the claim that it is the most just principle in this social world. And (*iii*) so too for the ideally just society: to say that universal conformity to a principle is an aspect of the ideally just society can be defeated by the social realizations of acting on it in the ideal social world. None of this is to say that any and all adverse social realizations defeat the claim that a principle correctly identifies justice (if, say, we have a social world of evil agents, it may still be the correct principle), but only that social realizations are always of potential relevance, and so cannot be simply "bracketed" on the grounds that since "we know the truth about justice, no mere facts could ever change this."

I believe that this is an important point about the justification of principles of justice and their social realizations. It is the deep reason

---

mersed in water seems privileged, and I cannot be wrong; that it is a crooked stick is often erroneous.

[85] I go into these matters in some depth in *Justificatory Liberalism*, chap. 7.

why we should reject what Sen might call "principle fundamentalism": that is to say, so long as the principles are satisfied, justice doesn't concern itself with the way people's lives go. However, the core analysis of this and the following chapters does not depend on accepting this point. All that is strictly required is that, for whatever reason, the ideal theorist acknowledges that determining the comparative justice of social worlds partly depends on the worlds' institutional structures and their dynamics, given the population operating, and subject to, them.

3 MODELING THE IDEAL (AND NONIDEAL)

3.1 Setting the Constraints Regulating Coherent Social Worlds (One Sense of Feasibility)

We now can see that our initial distinction between the orienting role of ideal theory and political philosophy as recommending was too stark (§I.1.5). While we might have initially viewed the mere orientation function of ideal theory as one of "pure moral philosophy," which is unconcerned about recommending moves to new social structures, and thus also unconcerned with predictive models of social worlds, we now see that even the most ideal of theories of justice must engage in modeling the ideal social world, as we need to make some effort to inquire about the social realizations of our ideal principles. Plato's *Republic* is a good example of ideal theorizing, for it seeks to model the ideal state based on true justice: if we accept Plato's account, we have some idea as to how ideal justice would pan out, at least under ideal circumstances. Notice that Plato does not say: ideally, acting on the principles of ideal justice must work out well—after all, they are ideal! Rather, he specifies some conditions (the nature of economic life, human psychology, international relations, and so on), and then engages in a long modeling exercise to show that under these conditions the social realizations of his understanding of justice are acceptable.

This is of even more importance for Rawls, who stresses not only that the ideal of justice must suppose a certain background social world, but that the postulated basic features of the social world must not stretch credulity:

> Some philosophers have thought that ethical first principles should be independent of all contingent assumptions, that they should take for granted no truths except those of logic and others that follow from these by an analysis of concepts. Moral conceptions should hold for all possible worlds. Now this view makes moral philosophy the study of the ethics of creation: an examination of the reflections an omnipotent deity might entertain in determining which is the best of all worlds. Even the general facts of nature are to be chosen. Certainly we have a natural religious interest in the ethics of creation. But it would appear to outrun human comprehension.[86]

Rawls abjures the ethics of creation, where just about all parameters of possible social worlds, including the laws of nature, are open to specification; instead, he seeks to develop "the conception of a perfectly just basic structure … under the fixed conditions of human life," assuming that given these fixed conditions the social world is characterized by "reasonably ideal" circumstances.[87]

The last decade has witnessed intense debate among political philosophers whether a theory of "ideal distributive justice" should, as does Rawls's, postulate social facts as, for instance, that without incentives people will not voluntarily work at the most socially useful job, or that they will not voluntarily pay taxes without any coercive threats.[88] This is often understood as a debate about feasibility, in one of the many senses of this complex concept: what constitutes a "feasible ideal" in the sense of one that identifies an ideal social world that is fixed by credible parameters of the socially, economically, and politically possible.[89] For our present purposes, the important point is not

---

[86] Rawls, *A Theory of Justice*, pp. 137–38.

[87] Ibid., pp. 216, 211.

[88] "Thus even under reasonably ideal conditions, it is hard to imagine, for example, a successful income tax scheme on a voluntary basis." Ibid., p. 211. Cf. Cohen, *Rescuing Justice and Equality*, chap. 1.

[89] We can distinguish numerous applications of the notion of feasibility as applied to worlds u, the ideal, and some other world, a (such as our own), including: (*i*) in u, are the "oughts" generated by its governing principles feasible for humans (can/will they comply?); (*ii*) in our present world a, are the oughts that characterize u still true/valid though not feasible?; (*iii*) is it feasible to move from world a to u?; (*iv*) is the way that an ideal theory depicts the functioning of world u (and, we can add, a) feasible? The recent literature has tended to focus on versions of (*i*) and (*ii*). Importantly (*i*) is

whether Rawls has correctly identified the parameters, but that any
account of the ideal must settle on some parameters or constraints in
modeling the social realization of the principles of justice if its ideal is
not to "outrun human comprehension." What Rawls calls the "ethics of
creation" allows, as it were, social philosophers to treat almost every
parameter as open to setting at any value they see fit: manna falls at
will from heaven so all resources are available in infinite quantities,
genders can change at will, one can be a critical critic *and* fisherman
at the same moment since time-space variables allow simultaneous
activity by the same person in any number of time-space coordinates,
one lives forever, we have worlds where actions have no opportunity
costs, or perhaps one is reincarnated in various lives such that by the
end of time all bad luck will be equalized, or whatever. If the principles
of justice hold in all possible worlds, they hold in all these creations,
for the idea of "possibility" is inherently one that is set against a set of
constraints. There is no such thing as "absolutely impossible": every
claim of possibility is of the form "*X* is possible (is not precluded by)
the set of constraints (including laws) *L*."[90]

Of course no theory of ideal justice allows *that* many variables to
remain open. A plausible current proposal is to distinguish "hard" con-
straints, such as the laws of nature and the natural sciences, and "soft"
constraints, such as those concerning present human motivations.[91]
The idea is that the hard constraints entirely rule out some sorts of

---

not equivalent to (*iv*): everyone may obey the "ought" in world *u* (and so it is feasible
to expect them to), yet the upshot may not be as the ideal theorist postulates—the
social realizations might not be what the theorist anticipates (my point in §I.2.2). My
present concern here is (*iv*); I consider (*iii*) in chapter II, §1.3. I have very little to add
to the extensive debate about (*i*) or (*ii*), which, to my mind, seems more about moral
than political philosophy. Feasibility (*iv*) is much more important than (*i*); it largely
encompasses the interests of political philosophy in (*i*). Even within, say, (*iii*) or (*iv*)
there are numerous dimensions of feasibility. Majone characterizes feasibility as a
"cluster concept" in "On the Notion of Political Feasibility," p. 261. See Hamlin's excel-
lent analysis in "Feasibility Four Ways."

[90] This is a reason why the current fascination with "ought implies can" seems un-
likely to yield much insight. What one "can do" does not crucially depend on whether
a theory defines "can" in terms of possibility or a conditional probability, but rather
on the constraints a theory has set. It is this that is critical.

[91] See Gilabert and Lawford-Smith, "Political Feasibility," pp. 812ff.; Carey, "Towards
a 'Non-ideal' Non-ideal Theory," pp. 149ff.; Räikkä, "The Feasibility Condition in Po-
litical Theory," p. 32; Wiens, "Political Ideals and the Feasibility Frontier," p. 7.

social realizations as infeasible, while the soft constraints can be seen
in more probabilistic terms, yielding judgments of more-or-less feasi-
ble. It is also supposed that hard constraints cannot be removed by
humans, while soft constraints describe "that some particular fact
could be changed, now or at some point in the future."[92] While no
doubt there is something to this, when we begin to model social worlds
and predict social realizations, the distinction pretty much evaporates
for at least three reasons.

(*i*) In determining whether a social world, *u*, is feasible, we need to
know not only that constraint *C* cannot be violated, but what social
states *C* rules out. Once we add technological innovations, for exam-
ple, whether *C* is or is not a constraint on whether social state *u* can be
realized can radically vary and, indeed, be manipulated, often more
easily than "soft" constraints, such as the degree to which humans are
public-spirited. Consider Thomas Malthus's two "hard" constraints in
his *Essay on Population*:

> First, That food is necessary to the existence of man.

> Secondly, That the passion between the sexes is necessary and
> will remain nearly in its present state.[93]

For Malthus these "laws, ever since we have had any knowledge of
mankind, appear to have been fixed laws of our nature; and, as we
have not hitherto seen any alteration in them, we have no right to
conclude that they will ever cease to be what they now are, without an
immediate act of power in that Being who first arranged the system of
the universe; and for the advantage of his creatures, still executes, ac-
cording to fixed laws, all its various operations."[94] In contrast, William
Godwin's utopian anarchism (to which Malthus was partly replying)
saw the second as a "soft" constraint, which could be overcome by
intellectual progress.[95] Godwin accepted that his anarchist-socialist
ideal, which resisted the struggle for existence, needed some way to

[92] Carey, "Towards a 'Non-ideal' Non-ideal Theory," p. 150.
[93] Malthus, *Malthus—Population*, p. 4.
[94] Ibid.
[95] Godwin, *An Enquiry concerning Political Justice*, vol. 2, pp. 850ff.

control population.[96] So what is hard and what is soft is by no means uncontroversial. Leaving aside this dispute, suppose one does accept that enjoyment of, and consequently the having of, sex is a hard constraint. To cope with this constraint a "socialist" utopia such as Thomas More's included institutions such as colonization to control population.[97] However, as John Stuart Mill recognized, one could accept Malthus's "hard constraint" (that people will continue to enjoy sex), but deny that it precluded a more socialist ideal by technological change— the diaphragm.[98] Thus under some technologies the hard constraint that people will continue to enjoy sex does not constrain utopian social worlds.

(*ii*) Because technological variables are difficult to determine, the ideal theorist seldom is in the position to assert with anything like a probability of 1 that some hard constraint definitely rules out desired social states. Moreover, it is often difficult to determine whether some hard constraint is "redundant" in the sense that, while it is a constraint on possible social realizations, it in fact does not constrain the set of relevant options in an ideal theory. Assuming that it is a hard constraint that we cannot go faster than the speed of light, this may well not impact any utopian scheme and so is to be ignored. But, perhaps, it *could* affect social realization, if instantaneous communications was part of paradise. As Giandomenico Majone concludes, because "it is often difficult to determine a priori which limitations are binding … feasibility statements must then be interpreted in probabilistic terms."[99] This, then, means that, as far as theoretical understanding of the ways a utopia might function, the distinction between probabilistic "soft constraints" that may be overcome, and "hard" constraints that definitely rule out possibilities, is not of great import.

[96] On his rejection of Malthus's "struggle," see Godwin, *Of Population*, pp. 619ff.

[97] More, *Utopia*, "Of Their Traffic." As late as the nineteenth century colonization was seen as a method of population control, which could assist the laboring classes. See Mill, *Principles of Political Economy*, pp. 194–95, 748–49, 958ff.

[98] Mill was arrested as a young man for handing out instructions for constructing such devices. See Schwartz, *The New Political Economy of J. S. Mill*, appendix 2, "The Diabolical Handbills."

[99] Majone, "On the Notion of Political Feasibility," p. 261. See also Hamlin, "Feasibility Four Ways."

(*iii*) Lastly, because social realizations depend on the interaction of a number of variables, what appears as a number of soft constraints can set into a much harder one. For example, suppose that a theory posits five soft constraints each of which yields a 20 percent chance that the utopia will arise, and all come to bear. Given the joint probabilities, there is only about a .00032 chance utopia will come about.

In modeling whether some social world is feasible given a set of constraints (C) that govern its behavior, then, pretty much all constraints should be understood probabilistically. Given this, for any set C of constraints there is a probability distribution of possible social worlds that might emerge on a given model. Let us call this probability distribution *the ideal theory's predictive modeling of the social realizations of meeting C*. It is fundamental that we should not fall into the trap of thinking that for every set of constraints C an ideal theory could predict a unique social world arising out of it. Every ideal involves specifying constraints or parameters (I shall not distinguish these here), taking some as very likely to affect social outcomes, and others as less apt to.

3.2 The Aim of Ideal Theory

On a naive view—which, alas, seems widely embraced in political philosophy—the aim of an ideal political theory is uncovering "the truth," as if, in making decisions between rival theories, we decide on the basis of a clear and unidimensional goal—truth. Not even natural science is well understood as having a monistic goal of "the truth." As Thomas Kuhn has shown, a true scientific theory aims at maximizing a set of values: accuracy of fit with the data (and of course that requires a decision about what the relevant data are), simplicity (this can mean different things, such as ease of computation or axiomatic parsimony), consistency (not simply internal, but also with related theories), scope (is the theory a comprehensive explanation of a range of phenomena?), and fecundity (does the theory open up fruitful lines of research?).[100] And in some fields, perhaps social utility is a desidera-

---

[100] Kuhn, *The Essential Tension*, p. 322.

tum.[101] Some think accuracy is all that is necessary, but as Kuhn stresses, on accuracy grounds alone, Copernicus's system was not, until revised by Kepler, more accurate than Ptolemy's, yet Kepler had already made a theory choice in favor of Copernicus's system before he began to work on revising it.[102] In some cases a theory may rank higher on all these criteria, and so dominate competitors, but often one theory will rank high on some, another theory on others. And as Kuhn effectively argued, there simply is no algorithm for combining these values into a single value to be maximized.[103] Different scientists will trade off these values in different ways, sometimes leading to different theory choice, even among those who share the same values and agree on the data.

Theory choice in political philosophy too has multiple desiderata. According to Rawls, a political philosophy should aim (*i*) "to focus on deeply disputed questions and to see whether, despite appearances, some underlying basis of philosophical and moral agreement can be found";[104] (*ii*) "to orient us in the (conceptual) space, say, of all possible ends, individual and associational, political and social";[105] (*iii*) at "reconciliation… to calm our frustration and rage against society and history," leading us "to accept and affirm our social world positively, not merely be resigned to it";[106] and (*iv*), as a variation of reconciliation, to propose a realistic utopia, "probing the limits of practical possibility.[107] Our hope for the future of our society rests on the belief that the social world allows at least a decent political order, so that a reasonably just, though not perfect, democratic regime is possible."[108] Without claim-

---

[101] Kuhn notes that this has entered into theory choice in chemistry. Ibid., p. 335. Of course it is obviously relevant to the social sciences.

[102] Ibid., p. 323.

[103] Ibid., pp. 328ff. See also D'Agostino, *Naturalizing Epistemology*, esp. chaps. 3, 5, and 6.

[104] Rawls, *Justice as Fairness*, p. 2.

[105] Ibid., p. 3.

[106] Ibid.

[107] This is especially clear in Rawls, *The Law of Peoples*, §1.

[108] Rawls, *Justice as Fairness*, p. 4. Rawls, I think, does not mean mere possibility qua not impossible, but a realistic possibility. On the importance of Rawls's quest to identify a reasonably just social world that is within our grasp, see Weithman, *Why Political Liberalism?*, pp. 8ff.

ing that these are canonical desiderata, we can readily see that, depending on how they are weighted, different theorists will endorse different theories, and so model the ideal in different ways. Importantly, a political philosopher who gives significant weight to the third and fourth aims is unlikely to decide that a political philosophy of hopeless realism is the way to go (§I.1.4). Estlund's theory of the hopeless ideal is hardly apt to "calm our frustration and rage against society and history," by inducing us "to accept and affirm our social world positively, not merely be resigned to it." Given Rawls's understanding of the aims of political philosophy, he is committed to an ideal that fixes parameters within, but at the limits of, some plausible range. In contrast, a theory that stresses orienting, or one that focuses on the importance of conformity with strong intuitions about justice, is apt to fix less of the parameters in the modeled social worlds, allowing for possible but hopeless ideals.[109] What must be stressed is that to invoke "possibility" as a constraint is itself to fix *some* parameters; it is the area (conceptual, empirical, etc.) defined by the fixed parameters that define the space of possibility or, as I shall say, the option space of an ideal theory. So it is not as if Rawls fixed parameters in his ideal social world whereas Estlund's utopia does not: the difference is in the parameters that are identified, and the range of variance that is allowed in others, and that in turn will depend on the criteria employed in theory choice.

3.3 Abstraction and Idealization

Onora O'Neill has insisted on the "fundamental, and frequently overlooked" distinction between abstraction and idealization, the latter of which many take as the heart of ideal theory.[110] In the strict sense, she argues, abstraction is a form of "*bracketing*, but not of *denying*, predi-

---

[109] Estlund makes much of the possibility constraint in "Human Nature and the Limits (If Any) of Political Philosophy."

[110] See O'Neill, *Towards Justice and Virtue*, p. 40. See also O'Neill, "Abstraction, Idealization and Ideology in Ethics"; and O'Neill, "Justice, Gender and International Boundaries." For discussions focusing on ideal theory, see Valentini, "On the Apparent Paradox of Ideal Theory"; Lawford-Smith, "Ideal Theory—a Reply to Valentini"; Jubb, "Tragedies of Non-ideal Theory"; Estlund, "Utopophobia."

cates that are true of the matter under discussion."[111] Reasoning from
an abstraction, O'Neill maintains, claims neither that the predicates
hold *nor* do not hold, and so reasoning from an abstract premise does
not lead validly from a truth to a falsehood. "Idealization is another
matter: it can easily lead to falsehood. An assumption, and derivatively
a theory, idealizes while it ascribes predicates—often seen as enhanced
'ideal' predicates—that are false of the case in hand, and so denies
predicates that are true of that case."[112] In her view Rawls's account of
justice is based on (suspicious) idealization assumptions. "For exam-
ple, in *A Theory of Justice* the veil of ignorance, that constitutes the
original position and forms part of the procedure for identifying jus-
tice, is defined by reference to an *ideal* that requires mutual indepen-
dence between the preferences of distinct agents, and assumes that
there is a restricted set of primary goods of which agents always pre-
fer more to less."[113]

Whatever the merits of this distinction in some contexts, when we
apply to it modeling choice situations and social worlds, it is extraor-
dinarily difficult to grasp.[114] In modeling terms, we can define "abstrac-
tion" as leaving some variable out of our model (we bracket it and put
it aside) whereas "idealization" employs a variable but assigns it a
value that departs from the best estimation, thus "denying… predi-
cates that are true of the matter under discussion." Suppose, then, we
are modeling the realization of a principle of justice in a social world,
and in one case (*i*) we simply do not include some variable, such as the
fact that people are gendered, and in another (*ii*) we suppose falsely
that all persons remain a single gender for their entire lives. Now it
seems clear that the abstraction in (*i*) may lead us much further away
from a sound understanding than the idealization in (*ii*), in trying to
grasp the social realizations of our principles. Abstracting away from
some factors can lead to fundamental errors, while positing false val-
ues for some variables may actually improve the model. For example,
to suppose that everyone prefers more primary goods over less from a

---

[111] O'Neill, *Towards Justice and Virtue*, p. 40.
[112] Ibid., p. 41.
[113] Ibid., p. 45.
[114] See here Hamlin and Stemplowska, "Theory, Ideal Theory and the Theory of Ide-
als," pp. 50–51.

restricted list is strictly false, but trying to model an entire population's distribution of preferences over all possible primary goods could lead us into such complexities that we end up throwing up our hands in despair because it's all so complicated. This is not, of course, to say that all abstractions are objectionable while all idealizations are helpful; it is to say, however, that all theoretical understanding supposes that we set some issues aside, and that all models require that we posit values that we know, say, are overly uniform, too optimistic, or too simple, but they can be part of the best understanding of a social world, either ideal or real.[115] It thus seems misguided to identify "ideal" theory with one that employs "idealizations."[116]

In economics "ideal" sometimes refers to a model that makes radical idealizing assumptions, such as perfect information or zero transaction costs. Such an "ideal model" creates a radically simplified and quite impossible world (there are no worlds of zero transaction costs) as a way to understand important dynamics in real-world economies. Here it is not simply the case that the model makes false (idealized) assumptions, but that the world it describes is, strictly speaking, impossible. We should not confuse this sense of an "idealized" world with an ideally just social world that we seek to analyze with a model that makes counterfactual, simplifying assumptions. In modeling the ideal world we seek to capture how it will function as well as we can, and to do this requires some abstraction and idealization. But the world so described is our best estimate of what a real, ideal world would actually look like, not an impossible world that we use to judge our own. Very few think that the ideally just social world is strictly impossible, even though many seem to allow that it might be, for now, deeply infeasible. As almost all insist, "ought implies can," so if the ideally just social world is to ground any sort of "ought judgment," then the "ought" must not be strictly impossible.[117]

---

[115] The classic defense of "idealization" in economics is, of course, Milton Friedman's "The Methodology of Positive Economics."

[116] "In spite of its popularity the notion of ideal theory tends to be employed somewhat loosely, to indicate any theory constructed under false, that is, idealized assumptions, which make social reality appear significantly 'simpler and better' than it actually is." Valentini, "On the Apparent Paradox of Ideal Theory," p. 332.

[117] See Estlund, "Human Nature and the Limits (If Any) of Political Philosophy."

4 Two Conditions for Ideal Theory

In this chapter I have set out what I believe to be the most powerful
case for a political philosophy that seeks to identify the most just,
or—more broadly—the best social world from the perspective of the
political.[118] Such a political theory would orient our thinking about the
justice in our own society; the ideal might be beyond our ability to
fully implement, but it would still serve as our "mythical Paradise Is-
land," which provides the goal of our quest for greater justice in the
worlds that we can bring about. So, as I have said, in a useful political
theory, the ideal cannot be mere *dreaming* of a utopia that is discon-
nected from our evaluations of less exalted social worlds (§I.I.4). I am
not claiming that such dreaming has no value, but it has precious little
value for answering the question that confronts the political theorist:
how are we to rank the alternatives that, at least potentially, are open
to us, and what moves (reforms) does our political philosophy recom-
mend? Moreover, I have argued that if orientation via the ideal is truly
necessary, then the task confronting political theory must be more
complex than Sen's proposal of simply climbing up the ranking of
social worlds (§I.1.3). What I shall call the "optimization" problem
must include more than simple pairwise judgments (is social world *a*
more just than *b*?); it must be important to also ask "is *a* or *b* closer to
the ideal?" Unless this second question is distinguished from the first,
and is important in its own right, Sen would be quite correct that
"there would be something off in the general belief that a comparison
of any two alternatives cannot be sensibly made without a prior iden-
tification of a supreme alternative" (§I.1.3). Thus if a theory of the ideal
is not to collapse into a version of Sen's climbing model, it must iden-
tify, in addition to a dimension comparing the justice of different social
states in the theory's option space, a dimension that, in some way,
relates the present world to the ideal in a different way. Ideal theory, if
it is to be a distinct enterprise, must have at least two dimensions of
evaluation that must be kept distinct. Under this condition the judg-

---

[118] To simplify, I shall simply speak of the "most just" social world, but the analysis
of this book generalizes to most understandings of "the politically best," where this
includes the realization of values that go beyond justice.

ments of ideal theory cannot be reduced to Sen's unidimensional model.

I have also stressed that pursuit of the ideal cannot be insensitive to the social realizations of our principles of justice (or, more broadly, our standards of evaluation). It is, as Rawls said, "by showing how the social world may realize the features of a realistic Utopia... [that] political philosophy provides a long-term goal of political endeavor, and in working toward it gives meaning to what we can do today" (§I.1.2).[119] But for this hope of greater meaning to be achieved, we must understand the social world we have designated as the realistic utopia, and we must understand how, in our long-term endeavor, the less-than-ideal social worlds we create along the way approximate it. Understood thus, ideal theory is a complicated endeavor: we need to be clear about the option space—the parameters that are assumed in our specifications of social worlds (aspects of human nature, the laws of economics, ecology, the laws of biology, the nature of agency), and how they can vary. Having identified the ideal social world, we must be able to compare it to our own, and understand its relation to intermediate social worlds. If we can do all that, we can orient our quest for greater justice, knowing where we are, where we would like to go, and the social worlds that lie in the direction of the best.

We can formalize these requirements in terms of two conditions that any acceptable theory of the ideal (T) must meet:

> *Social Realizations Condition*: T must evaluate a set (or domain) of social worlds $\{X\}$. For all social worlds i, which are members of $\{X\}$, T evaluates i in terms of its realization of justice (or, more broadly, relevant evaluative standards). This must yield a consistent comparative ranking of the members of $\{X\}$, which must include the present social world and the ideal, in terms of their overall justice.

> *Orientation Condition*: T's overall evaluation of nonideal members of $\{X\}$ must necessarily refer to their "proximity" to the ideal social world, u, which is a member of $\{X\}$. This proximity measure cannot be simply reduced to an ordering of the members of

---

[119] Rawls, *The Law of Peoples*, p. 128.

{X} in terms of their inherent justice as in the Social Realizations
Condition.

I have stressed that ideal theory does not require that all the worlds in
{X} need be achievable (for example, u, the ideal, might not be on some
views). We could also accept that whatever the criteria of "realistic" or
"feasible" are (see §II.1.3) T need not hold that movements to all social
worlds in {X} (given where we are) must be feasible for us. But unless
T holds that many of the social worlds in {X} are more than mere
hopeless possibilities, but in some sense real options for us, T will not
be able to serve the function of orienting our quest for justice. Unless
some of the worlds in {X} are realistic alternatives, T will not be able
to issue any recommendations for reform, which, again, would show
that T fails to orient our quest for justice. I assume that these adden-
dums are part of the Orientation Condition.

The Social Realizations Condition expresses the requirement that
the theory of the ideal aims to evaluate the justice of various social
worlds, which realize to different extents, and in different ways, the
relevant principles or standards of justice. This is an implication of our
social realization thesis (§I.2), and our antidreaming conclusion (§I.1.4).
The Orientation Condition expresses the idea that, while a simple
ranking (à la Sen) of the inherent justice of each social world can sat-
isfy the Social Realizations Condition, an overall judgment of the rela-
tive attraction of any two members of {X}, a and b, must necessarily
depend (though not decisively) on a's and b's relative distance and
direction in relation to u. This, of course, means that T must provide
some coherent concept of social world proximity (or, more generally,
location) that does not collapse into an ordering in terms of inherent
justice (recall again Sen's argument and Simmons's reply, §I.1.3). The
addendum that we must have a realistic chance of achieving some
worlds in {X} interprets the thought that a theory of the ideal seeks to
orient our quest for, not simply our judgments about, justice.

The question for ideal theory, then, is whether the Social Realiza-
tions and Orientation Conditions both can be met or, more usefully:
under what assumptions can they be met, and are these assumptions
plausible? I now turn to these matters.