

The Tyranny of a Metaphor

David Wiens

Methodological debates among political philosophers have become preoccupied with the relevance of ideal theories for guiding action in the real world. Following Sen (2006, 2009), these controversies have been deeply shaped by the metaphor of navigating a mountainous terrain. The rough idea is this. Ideally just societies are analogous to high points in a landscape of social possibilities, with lower altitude points being analogous to less just social worlds. Questions about the relevance of ideal theories for guiding action are then transposed to questions about the relevance of information about these high points for navigating the terrain in search of higher ground (i.e., greater justice).

Metaphors are useful tools for framing and guiding theoretical inquiry. (So are fables, thought experiments, and the like — generally, *models*.) They help by making the subject of inquiry both more abstract and more concrete. Metaphors make a subject more abstract by isolating central components of a subject, bracketing complicating factors that can obscure our view of the relationships between these central components. Metaphors render the abstraction tractable by invoking a particular concrete object to represent — to stand in place of — the more complex subject we wish to study. We then study aspects of our subject by exploring the workings of its concrete representation. Thomas Schelling (2006) has given us vivid examples: we can learn about the complex dynamical processes underlying certain macroeconomic regularities by studying the game of musical chairs; we can learn about the processes underlying the cyclical rise and fall of disease rates by studying the operation of a heating and cooling system regulated by a thermostat.

Metaphors are useful aids because they impose limitations on inquiry — they obscure certain aspects of a subject so as to highlight and focus our attention on others. But these benefits bring a risk: that our use of metaphors not only masks certain aspects of a subject temporarily, but renders us indefinitely blind with respect to them. This can occur when a particular metaphor comes to *define* the contours of a subject instead of being used as a tool for exploring a particular aspect of a subject. For example, the cyclical feedback loop represented by a thermostat might be useful for studying the feedback relationship between aggregate disease rates and aggregate vaccination rates. But if we

Author's note. Thanks to Jerry Gaus, Brian Kogelmann, and Ryan Muldoon for helpful comments on earlier versions.

take that metaphor to define public health problems, we will neglect other important aspects, such as the effect of budget constraints or local culture on vaccination choices.

Gerald Gaus's *The Tyranny of the Ideal* offers a case study of the ways in which metaphors can illuminate but also blind. In it, Gaus develops a model that presents the most thorough articulation of the navigation metaphor to date. His detailed exploration yields new insight on central issues in existing debates regarding the relevance of ideal theories to practical action. This model also provides a fruitful medium for Gaus to explore important limitations on our ability to map the space of social possibilities, thereby deepening existing skeptical arguments against the relevance of ideal theories for guiding action in the real world.

I worry, however, that Gaus's heavy reliance on mountains and maps leads him to neglect important questions about the relationship between ideal theories and nonideal theories.¹ In particular, by presupposing a standard for measuring the justice of particular social possibilities, this metaphor brackets the logic by which we are justified in identifying certain possibilities as high points in the landscape. Put differently, by emphasizing the logic of navigation—the logic guiding our efforts to navigate our way to ideally just societies—this metaphor obscures questions about the logic by which ideal theories are justified. As a result, Gaus fails to notice the ways in which his theory of the Open Society resembles the ideal theories he aims to dismiss. Ironically, Gaus winds up neglecting the ways in which the Open Society might tyrannize our efforts to realize greater justice.

My remarks should be taken as cautionary rather than critical. Exploring the implications of a single metaphor deeply, even to the exclusion of other aspects of a subject, can be worthwhile. In doing this, Gaus has exposed many of the limitations that are internal to existing debates about the practical relevance of ideal theory. This is an important

¹ Gaus (personal communication) wants to say that, while his formal model starts from the navigation metaphor, its development ultimately enables us to drop the metaphor (with its inherent limitations) and carry on our inquiry using only the model. I think this is right to some extent, and suggest as much toward the end of section 2. This said, I will set aside a more nuanced treatment of the relationship between the navigation metaphor and Gaus's model for the remainder of the paper. But, then, am I not being unfair in charging Gaus with "heavy reliance" on the metaphor? Two brief replies. First, everything I say about the ways in which metaphors create blindspots applies to formal models too. I focus on the limitations of the navigation metaphor more generally (rather than the limitations that are specific to Gaus's model) because I am ultimately interested in pointing out the ways in which this metaphor, which has in many ways come to define the terms for an entire debate, threatens to create *collective* theoretical blindspots. Second, Gaus develops his model to fit the navigation metaphor (see Gaus, 2016, 48). This is understandable, as it is rhetorically prudent for Gaus to work from a familiar starting point. Additionally, our collective interpretation of the model will be guided by the metaphor, at least until there is a sufficiently mature understanding of the model that is independent of the metaphor. But these points mean that Gaus's model inherits many of the limitations inherent to the metaphor.

The Tyranny of a Metaphor

contribution. My aim is to show how reflecting on *The Tyranny of the Ideal* can also expose some collective blindspots of existing debates, especially the limitations imposed by a near-exclusive focus on the issue of “practical relevance”.² I hope that, by taking note of these potential blindspots, we become alert to new avenues for exploring the uses and abuses of ideal theory in normative reasoning about politics.

1. OF MOUNTAINS AND MAPS: INTRODUCING A METAPHOR

How is ideal theory relevant to our normative reasoning about politics? The answer depends, of course, on how we understand "ideal theory", "relevance", and "normative reasoning about politics". Given their relative immaturity, it is unsurprising that debates about ideal theory have rarely been explicit about these matters.³ One way to narrow the question is by locating a common reference point. Notably, much of the recent literature operates in the shadow of Rawls's remark that ideal theory is necessary to guide nonideal theory (e.g., Rawls, 1999, 8, 216). This narrows "relevance" to some form of "guidance", and "normative reasoning about politics" to something called "nonideal theory". But "guidance" and "nonideal theory" are still vague, and we have yet to define "ideal theory".

Enter Amartya Sen. Sen (2006, 2009) refines the general question thus: how can an analysis of a perfectly just society help us figure out what we ought to do to address current injustices and realize greater justice in the real world? On this formulation, an "ideal theory" is an analysis of a perfectly just society, while a "nonideal theory" offers normative prescriptions for morally progressive action in our unjust world. But how are we to understand "guidance"? "Help us figure out what we ought to do", how? Sen introduces a metaphor to help us articulate the question.

Following our refinement of "nonideal theory", suppose the point of a theory of justice is to help us move from our unjust status quo to a more just state of affairs. If we think of justice as analogous to altitude, then we can think of this movement as a change in altitude, from a low point to a higher point in a landscape. Continuing with the metaphor, we can draw an analogy between the highest point in the landscape and a perfectly just society. The question of the relevance of ideal theory is transposed thus: do we need to know anything about the highest point in a landscape if we are to navigate our way from our current position to higher ground?

² David Estlund's work is an important exception (e.g., Estlund, 2011). Even so, Estlund's work often becomes entangled with issues of practical relevance, obscuring (I think) alternative avenues of inquiry. I gesture toward one of these avenues at the end of this paper.

³ Although several surveys have mapped some of the theoretical terrain; see Hamlin and Stemplowska (2012); Stemplowska and Swift (2012); Valentini (2012, 2017).

As is well known, Sen answers his question in the negative — ideal theory is neither necessary nor sufficient to guide nonideal theory. Ideal theory is not necessary to guide nonideal theory because we do not need to know anything about the highest point in a landscape to sort out the difference in altitude between any two points (Sen, 2009, 101f). We need a way to measure altitude, of course. But, by analogy, this means our efforts to advance justice only require a theory of comparative justice — a theory that aggregates the relevant criteria for comparing social possibilities — not a theory of perfect justice. Ideal theory is not sufficient to guide nonideal theory because identifying the highest point in a landscape and analyzing its features does not determine the measures we should use to map the surrounding landscape. More specifically, points in a landscape can deviate from the highest point along numerous dimensions: latitude, longitude, altitude. Perhaps it is relevant for our purposes to include the length of the available paths to the highest point, or the difficulty of traveling the available paths (some paths might be direct but steep, while others might be circuitous but on a gentler slope). Similarly, possible states of affairs can deviate from perfect justice along numerous dimensions: liberty, equality, welfare, security, and so on. A theory of perfect justice doesn't tell us how to measure these dimensions or how to weigh deviations along these dimensions relative to each other (Sen, 2009, 98ff). What we need is a theory of comparative justice, one that pays explicit attention to the aggregation of several criteria for comparing feasible options for advancing justice.

Sen's argument has garnered numerous critical replies. Some of these have noted the limitations of Sen's metaphor — pointing out, for example, that ideal theory is not concerned with analyzing perfectly just institutional arrangements but with specifying general normative principles (Gilabert, 2012; Valentini, 2011); or, relatedly, that reasoning about perfect justice can help construct the metrics required for comparison by exposing considerations that are appropriate for evaluating feasible options (Boot, 2012; Swift, 2008). Perhaps the most acclaimed reply shows that Sen's understanding of his own metaphor is incomplete. Following Rawls (1999), Simmons (2010) argues that the point of nonideal theory is not simply to guide us to higher ground (more just social arrangements), but to identify transitional paths to the highest ground (a perfectly just social arrangement). To avoid wandering aimlessly or getting stuck at a lesser peak, we need to chart the open paths to the highest ground; for that, we certainly need to locate the highest peak. Hence, we need a theory of ideal justice, not simply a theory of comparative justice, to guide nonideal theory.

2. EXTENDING THE METAPHOR: ENTER GAUS

We could dispute Simmons's view (among others) of nonideal theory as a transitional guide to perfect justice, arguing instead that the point of nonideal theory is to help us avoid low points in the terrain (Schmidtz 2011, 774; cf. Wiens 2012). Gaus takes a different tack. Assume Simmons is right—the point of nonideal theory is to lead us to perfect justice. There remains a question: under what conditions do we require a theory of ideal justice to develop a nonideal theory that can lead us to perfect justice? Why isn't a theory of comparative justice enough?

Notice how the metaphor of navigating a mountainous terrain can be extremely helpful for exploring this question. Transitioning from the status quo to more just social arrangements and, perhaps eventually, to perfectly just social arrangements is an incredibly complex endeavor, raising myriad interconnected questions about the potential of various institutional schemes to facilitate information transmission and coordinate social activity, the sequencing of policy reforms given their likely consequences over the medium- and long-term, and so on (cf. Heath, 2017, 6–8). While these details are obviously relevant for transitioning toward perfect justice, they present distracting complications here. Our question is not about which particular transitional reforms to implement or how to implement them, but about how far ahead we must forecast if our reform efforts are to help us realize perfect justice and whether our efforts must hit particular intermediate targets along the way. Indeed, the metaphor helps articulate our question: we are not asking about how to take particular steps along the path to the highest peak, but about whether our efforts to reach the highest peak require a map that locates the highest peak.

Put metaphorically, then, Gaus starts from the following question: what must we assume about the topography of a terrain if, as Simmons claims, our attempts to navigate our way to the highest peak require that we orient ourselves to that peak? Clearly, if the terrain of social possibilities has only a single peak—if it is akin to a “Mount Fuji” landscape—then Sen is right: we have no need to map the terrain, much less orient the map around the highest peak. Simply climbing to increasingly higher ground is sufficient to take us to the highest peak. A theory of comparative justice is enough (Gaus, 2016, 62, 73).⁴ If Simmons is correct—if ideal theory is a necessary guide to nonideal theory—then the terrain must be “rugged”, that is, it must at least have several peaks (67).⁵ But how rugged must it be?

4 Hereafter, bare page references in the text refer to Gaus (2016).

5 Gaus presents some reasons for thinking that the terrain of social possibilities is not a Mt Fuji landscape (pp. 63–7). We can set this aside here.

To address this question, we need to make sense of the “ruggedness” of the terrain of social possibilities. We can easily see how a physical terrain can be more or less rugged. But this is because physical terrains have two horizontal dimensions (latitude and longitude) in addition to the vertical dimension of altitude. Thus far, our terrain of social possibilities — “social worlds”, in Gaus’s terminology — only has a “vertical” dimension, justice. To translate our intuitive sense of ruggedness from the metaphor, we need to devise a conceptual apparatus to help us make sense of how social worlds can be arranged not only along a vertical dimension, but a horizontal dimension too.

Gaus formalizes the notion of an *evaluative perspective* to help us translate our intuitions about physical terrains to insights about the need for ideal theory. An evaluative perspective includes seven elements (43–4, 53–56)⁶:

- (1) a set of *evaluative standards* for assessing social worlds with respect to justice;
- (2) a specification of how to aggregate multiple evaluative standards into an *overall justice evaluation* (i.e., a specification of the importance of each standard vis-à-vis the others);
- (3) a specification of *justice-relevant world features*, that is, the features of a social world to which a perspective is sensitive when evaluating social worlds;
- (4) a *set of models* that estimates how particular justice-relevant features are likely to interact to engender social outcomes;
- (5) a *justice score* assigned to each social world;
- (6) a *similarity ordering* of worlds that encodes descriptive (rather than evaluative) similarity among social worlds in terms of their justice-relevant features;
- (7) a *distance metric* that quantifies world similarity.

The justice score constitutes the vertical dimension and is analogous to altitude in the metaphor; the distance metric constitutes the horizontal dimension and can be seen as analogous to latitude in the metaphor. (There is no equivalent to the metaphor’s second horizontal dimension, longitude.)

We should pause to nurture some of the intuition behind Gaus’s model; figure 1 gives a schematic representation. Let’s start with justice score assignments, which determine

⁶ By Gaus’s count, an evaluative perspective has five elements. Our counts differ because Gaus groups items (2), (4), and (5) as a single element, called a “mapping function”; he calls (4) the “modeling task” of the mapping function and (2) the “overall evaluation task” of the mapping function.

The Tyranny of a Metaphor

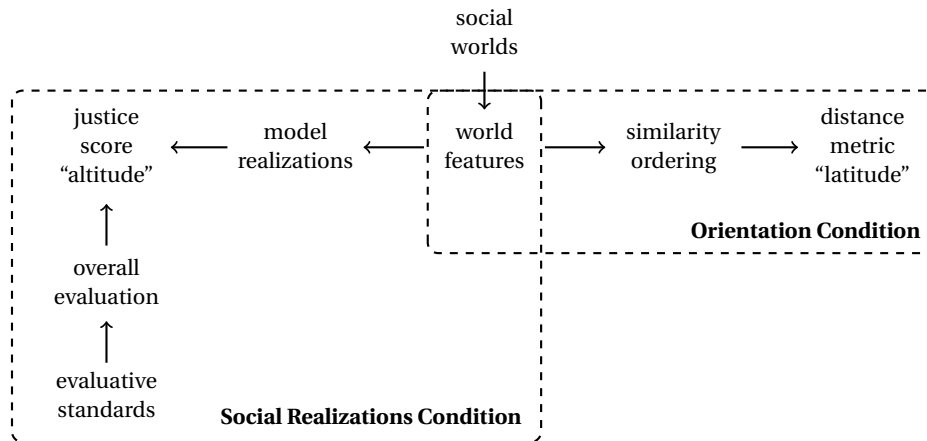


Figure 1. A schematic representation of Gaus's model

a world's place along the vertical dimension. We start with a specification of justice-relevant world features (item 3). Just what is supposed to count as a justice-relevant world feature isn't entirely clear, but Gaus's example of a "bleeding-heart libertarian" perspective offers some guidance. Here, we find social worlds described as "prohibiting (permitting) government budget deficits", "prohibiting (permitting) tax increases", or "prohibiting (permitting) spending cuts to vital services" (63–4). Following this example, we can take justice-relevant features to be any descriptive feature of a world that is relevant to our normative assessment of that world: a specification of the mechanisms by which resources are distributed among individuals; a schedule of who enjoys which rights; a specification of the social arrangements by which rights are enforced; so on. Once we have settled on a world's justice-relevant description, we must construct a model of a world that fits the description to predict "how such a world works" (item 4). For example, suppose a particular world is described by the following features: an equal income distribution and occupational assignments enforced by state intimidation and coercive force. We must now construct a model of a world with the indicated features in an effort to estimate the broader consequences of these features. What are the welfare consequences of such a world? How much coercion is required to enforce occupational assignments? (Do people adhere to their assignments willingly?) This modeling task completes our description of the world: its justice-relevant features and the broader consequences of their interaction. We then normatively evaluate the world as depicted by the model in accordance with our comprehensive evaluative standard (item 2), which represents an aggregation of

multiple evaluative standards (item 1).⁷ This normative evaluation yields a cardinal justice score for each world, which indicates a world's "inherent justice", that is, a world's justice independent of its location relative to the ideally just world (see 40f).⁸ Although Gaus does no such thing, we can use these justice scores to define a notion of "evaluative similarity": two worlds x and y are evaluatively more similar than x and z if and only if $|j(x) - j(y)| < |j(x) - j(z)|$, where $|j(x) - j(y)|$ denotes the absolute difference in the justice scores of x and y .

We now turn to the distance metric, which determines a world's place along the horizontal dimension. We start again with justice-relevant world features, making similarity comparisons between worlds based on their variation with respect to justice-relevant features. To illustrate, consider three worlds: x distributes material resources via unregulated markets in labor, goods, and services; y distributes resources using a combination of markets and redistributive taxation; z distributes resources using centralized government planning. Intuitively, x and y are more *descriptively similar* to each other than x and z ; in Gaus's notation, $[(x \sim y) > (x \sim z)]$. We arrive at a complete similarity ordering of social worlds (item 6) by making increasingly fine-grained similarity judgements on triples of worlds.⁹ From this similarity ordering of worlds, we construct a distance metric (item 7), which provides a cardinal measure of world proximity in terms of descriptive (rather than evaluative) similarity.¹⁰

Gaus's model presents a conceptual apparatus that can extend the theoretical reach of our navigation metaphor, providing normative theoretic analogues for two dimensions,

7 Gaus doesn't say much about how this aggregation is supposed to happen, other than to indicate that the comprehensive standard represents a specification of the relative importance of the different component standards. Note that this aggregation must assume some degree of comparability across these component standards if Gaus intends for an overall evaluation to be a genuine aggregation of multiple component standards (and not just a replication of a single standard), while also admitting a potentially diverse array of standards. The need for this assumption follows from Arrow's (1963) theorem and the fact that cardinal justice scores entail a complete and transitive ordering of social worlds (cf. Sen, 1970, chaps. 7 and 7*).

8 Note that the cardinal justice score is primitive in Gaus's model. The justice function on worlds is not a representation of a more primitive set of pairwise comparative evaluations of worlds, in the way that (cardinal) utility functions represent preference relations that satisfy certain structural conditions. Rather, the cardinal justice score precedes and, thus, "guarantees a comparable (noncyclical) ranking" of worlds (44f). In this way, a world's cardinal justice score is akin to the measurable psychological correlates of pleasure and pain that underwrite utility functions in classical theories of utilitarianism. This is a strong assumption about the measurability of our various evaluative standards.

9 Gaus is a bit loose when he says that "pairwise similarity is the basic relation" (53). In fact, the basic relation involves triples of worlds, a point to which he alludes on p. 52, footnote 24. Notice, also, that we could easily generate a multidimensional similarity ordering, creating several "horizontal" dimensions. Gaus takes a single descriptive dimension to be sufficient for his purposes.

10 Gaus's Appendix A gives details on the construction of a distance metric.

The Tyranny of a Metaphor

altitude and latitude. Yet his effort to “make sense of the metaphor of a mountain range” (48) is not *ad hoc*; it models a principled view of ideal theory. As he argues, if a theory of ideal justice is to be required to help us navigate the terrain of social worlds, then it must satisfy two conditions (40f):

- An ideal theory satisfies the *Social Realizations Condition* if it provides the evaluative resources required to rank social worlds with respect to justice (or some other normative standard).
- An ideal theory satisfies the *Orientation Condition* if its overall assessment of social worlds must refer to worlds’ proximity to the ideal world, where proximity is understood in terms of descriptive rather than evaluative similarity.

A theory that comprises the elements by which we determine a world’s justice score or “altitude” (items 1–5) satisfies the Social Realizations Condition; a theory that comprises the elements by which we determine a world’s distance from the ideal or “latitude” (items 3, 6, and 7) satisfies the Orientation Condition (56).

We can now make sense of how the space of social worlds can be more or less “rugged”: ruggedness is a function of the correlation between evaluative and descriptive similarity. At one extreme — a smooth, single-peaked landscape — worlds that are descriptively similar are also evaluatively similar; thus, movement from the status quo to a descriptively similar social world leads to a relatively small change in overall justice score. At the other extreme — a “high-dimensional” or “maximally rugged” landscape (68–9) — two worlds that are “neighbors”, descriptively speaking, can have wildly divergent justice scores; thus, movement within a neighborhood of descriptively similar worlds can lead to wild fluctuations in justice. A terrain is more or less rugged depending on where it fits between these two extremes.

As already indicated, at the simple extreme, ideal theory is not necessary to help us navigate to ideal justice. Wherever we are in the landscape, we need only use our theory of comparative justice to help us sort out which neighboring social worlds represent a justice improvement and implement the necessary reforms. Incremental “gradient climbing” is sufficient to reach the highest peak (62, 72–3). At the other, maximally rugged extreme, ideal theory is useful for navigation only if it can provide maximally precise and accurate comparative judgments of justice (70). In a maximally rugged landscape, small steps to descriptively nearby social worlds will take us to worlds that are evaluatively divergent from the status quo. If our theory of justice can only make rough and ready comparative judgments, we cannot be confident about which small steps will lead to justice improvements or to significant setbacks.

So Gaus's enhancement of our navigation metaphor yields an underappreciated insight: a theory of ideal justice is necessary to guide nonideal theory—that is, Simmons is right—only if we are navigating a moderately rugged landscape (73). In a moderately rugged landscape, we can be confident about which steps will lead to higher ground and which ones will lead to lower ground. But we cannot expect to reach the highest peak by simply moving toward higher ground; climbing to higher ground in our local neighborhood might lead us away from the globally highest peak.

If ideal theory is needed to guide nonideal theory, then we are likely to be confronted with tough choices: make local improvements in justice that lead us away, descriptively speaking, from ideal justice; or pursue reforms that take us closer, descriptively speaking, to ideal justice at the cost of justice setbacks over the short- to medium-term. We need a map that includes the ideal to tell us when we face such choices and to chart the paths to the ideal to clarify the tradeoffs involved in such choices.¹¹ But a map helps us make these choices only insofar as we can be confident in its accuracy. If our map is fuzzy in places—if, for certain points, our map can only indicate a range of possible altitudes or latitudes—then we cannot chart paths through the terrain with much confidence. If we conjecture that the highest peak—the ideally just world—is located in a fuzzy area of our map, then we face The Choice “between relatively certain (and perhaps large) local improvements in justice and pursuit of a considerably less certain ideal, which would yield optimal justice” (82). If The Choice characterizes our epistemic situation, then the injunction to pursue ideal justice faces a considerable burden of proof, for it enjoins us to forego sure justice improvements to undertake what might very well be a wild goose chase. If this is the best a theory of ideal justice can offer, why bother with ideal theory at all?

Having characterized the topographical conditions under which we require ideal theory to help us navigate to the highest peak, we must now ask whether we are able to draw a map that can accurately depict the location of the highest peak. It is a virtue of Gaus's model that it invites us to investigate questions about our capacity to draw accurate maps, questions that have been largely neglected by defenders of ideal theory. What's more, by providing an analogue for latitude, Gaus's model extends the navigation metaphor in a way that focuses such investigations by indicating where to look. Looking to the metaphor, we notice that it is relatively straightforward to draw an accurate map of

¹¹ Gaus argues that ideal theorists must always insist on pursuing ideal justice, otherwise there is no point to doing ideal theory (142). This seems hasty to me. One point to doing ideal theory could be to help us get clear on the tradeoffs involved in choosing one path over another; once these tradeoffs are in view, an ideal theorist might argue that it is best to pursue the local peak. Of course, getting a clear picture of these tradeoffs assumes we can construct a reasonably clear picture of the ideal, which is something Gaus disputes.

The Tyranny of a Metaphor

the area surrounding our current position — we can easily judge the altitude and relative distance of nearby points. The trouble arises in trying to estimate the altitude and relative distance of far-off points without leaving our current position.

Theories of ideal justice typically depict social worlds that are quite distant from the actual world; moreover, they aspire to estimate the justice and relative similarity of these unfamiliar social worlds. Given the Social Realizations Condition, if we are to have any confidence in our evaluative judgments about distant social worlds, we must be confident in our ability to predict how those distant worlds are likely to operate. Thus, theories of ideal justice can accurately locate the ideal for the purposes of navigating a moderately rugged terrain only insofar as we can accurately model the operation of unfamiliar social worlds (“the modeling task”, in Gaus’s terms).

Gaus’s argument against ideal theory enumerates the serious epistemic barriers to confidently modeling the operation of unfamiliar social worlds and the prospects for overcoming these barriers (76–149). The main conclusions are as follows.

- The Diversity Prediction Theorem: our ability to accurately map the terrain of social worlds depends on having diverse perspectives and, hence, diverse models of social worlds (95).
- The Neighborhood Diversity Dilemma: increasing the number and diversity of perspectives leads to disagreement about the features of nearby worlds and about our relative distance to the ideal (116).
- The Fundamental Diversity Dilemma: increasing diversity leads to sharp disagreements about which social worlds are ideal (131).
- Pluralistic liberal societies will never agree on the ideal; that is, they will never agree on which social world to pursue (145).
- Our best hope for effectively realizing justice improvements is to establish an Open Society, a society which accommodates a plurality of diverse perspectives and permits free moral exploration (xix, 148, 174–6).
- Ironically, the aim of ideal theory is best served by abandoning the pursuit of a single unified ideal (246).

The details of this argument are beyond my concern here (although, for what it’s worth, I’m sympathetic to the epistemic concerns motivating the argument; see Wiens 2015*b*). Gaus’s arguments on these points do not rely much on the metaphor of navigating a rugged terrain, and my purpose is to enumerate the ways in which reliance on this

metaphor shapes Gaus's investigation of ideal theory. My aim in reconstructing Gaus's model at length is to illustrate its value for extending our use of the navigation metaphor to generate and investigate intuitions about the practical relevance of ideal theory. The metaphor and the formal model work in tandem: the metaphor supplies theoretical intuitions for exploration and alerts us to questions that need answering; the formal model gives us a means to translate our intuitions from the metaphor to the subject of theoretical inquiry, while also enabling us to articulate questions precisely and explore potential answers rigorously. More specifically, Gaus's formalization of the metaphor makes two important contributions to debates on ideal theory. First, it makes transparent that ideal theory is needed to guide nonideal theory only if the terrain of social worlds is moderately rugged; this, in turn, implies that ideal theory is needed only if we are likely to face tradeoffs between making local justice improvements and pursuing ideal justice. The second contribution, which follows on the heels of the first, is to flag the serious possibility that we are unable to map a moderately rugged terrain of social worlds with much accuracy, leaving us with precious little information for making informed choices about which transitional paths to pursue. While Gaus has argued for pessimism on this point, I take his real contribution to be indicating how debates about the practical relevance of ideal theory might turn on epistemic issues that have been largely neglected to this point.

3. NAVIGATION VERSUS JUSTIFICATION: REMODELING IDEAL THEORY

While Gaus's model—and the metaphor of navigating a rugged terrain more generally—provides important insight regarding the practical relevance of ideal theory, heavy reliance on the metaphor also risks important theoretical blindspots. In the remainder of this essay, I will limit myself to highlighting one blindspot in particular. A handy way to put my point might be this: Gaus's model—and the navigation metaphor more generally—brings into focus the logic of navigation but neglects the logic of justification. More precisely, the model highlights key aspects of the tradeoffs we face while navigating practical political choices, drawing our attention in particular to the kinds of information we need to estimate and appreciate these tradeoffs and the limits on our ability to obtain the required information. But, by presupposing a measure for altitude—a feature which it shares with the more general metaphor—the model brackets questions about the reasoning by which we justify claims identifying particular social worlds as ideally just. In Gaus's words, his model “formalize[s] *the pursuit* of the ideally just society” (73, my emphasis). Regarding normative justification, however, his model obscures the reasoning by which we identify particular worlds as ideally just by simply assuming a cardinal justice

The Tyranny of a Metaphor

function. Once we attend to this black box, we will see that Gaus's Open Society has more in common with ideal theory than he appreciates.

I use a contrastive method to make my point. I start by presenting a model that enables us to systematically explore the logic by which ideal theories are justified. I then contrast this model with Gaus's own, thereby revealing his comparative silence on matters of justification.

As I will use the term, an *optimization problem* has four components: a domain of options that represent possible solutions; an evaluative standard by which these options are comparatively ranked; a set of constraints to restrict attention to a subset of options considered to be "available" or "feasible" in some sense; derived from these last two items, a characterization of the important features of the best available options, that is, the available options that are most highly ranked according to the specified evaluative standard. Optimization problems can be represented by a purely formal, uninterpreted mathematical object, like so:

- Let $X = \{x, y, z, \dots\}$ be a set of possibilities.
- Let R be a subset of the product $X \times X$. R is a set of ordered pairs, $R = \{(x, y) | x \in X \text{ and } y \in X\}$. We call R a binary relation on X and say that x is at least as highly ranked as y if and only if $(x, y) \in R$, which we also write as xRy . We can represent R with a real-valued function $f : X \rightarrow \mathbb{R}$ if and only if R satisfies certain structural properties (e.g., if R is complete, transitive, and continuous).
- Let $K = \{k_1, \dots, k_n\}$ be a set of constraints on X , which specifies properties that may or may not be instantiated by possibilities in X . Let $S(K) \subseteq X$ be the subset of possibilities that instantiate the properties in K .
- For all $S \subseteq X$, we say that a possibility $x \in S$ is optimal if and only if xRy for all $y \in S$. Let $C(R, S) = \{x \in S | xRy \text{ for all } y \in S\}$ denote the set of optimal possibilities.

Clearly, such an object is too abstract to offer much insight for particular theoretical purposes. To be theoretically useful, the mathematical object must be given a specific interpretation. A familiar example of a theoretically specific interpretation is consumer choice theory: X is the set of possible bundles of consumer goods; R encodes the individual's preferences over bundles and is represented by a utility function, $u(x)$; K specifies a budget constraint, $\sum_i p_i x_i \leq Y$, where p_i is the price of good i , x_i is the quantity of good i , and Y is the individual's income; $S(K)$ is the set of consumption bundles that satisfy the budget constraint. $C(R, S(K))$ is the set of optimal feasible consumption bundles. From the conjunction of a utility function and a budget constraint, we can characterize

key features of the optimal consumption bundles. Other familiar interpretations can be found in statistics, computer science, choice theory, and population biology.

In previous work, I have argued that the reasoning by which ideal theorists justify claims about the ideal conforms to the logic of an optimization problem (Wiens, 2015*a*, 2017). The key to substantiating this claim is to offer a theoretically compelling interpretation of the structural components of a general optimization problem using concepts that are central to ideal theorists' reasoning about ideal worlds. Such an interpretation starts with a recognition that a theory of ideal justice conjectures a normatively optimal solution to a well-specified problem. For example, on my view, Rawls's (1999) theory of justice conjectures a normatively optimal solution to the problem of specifying terms of social cooperation that are rationally acceptable to free and equal individuals who are committed to fairness and mutual reciprocity. Similarly, Nozick's (1974) theory of the libertarian minimal state conjectures a normatively optimal solution to the problem of biased and unreliable enforcement of people's natural rights given a Lockean state of nature. I then illustrate by example how Rawls's and Nozick's efforts to justify their theories of ideal justice exhibit the following structural features.¹²

- A set of possible social worlds; we can let X denote this set.
- A set of evaluative principles, which specify the basic evaluative considerations by which they comparatively rank social worlds. This can be modeled as a binary relation R on X .
- A set of constraints given by their broad assumptions about how the social world works, which we can denote K . These constraints fix some of the properties that are borne by the worlds under consideration. We can let $S(K)$ denote the set of worlds that satisfy the properties specified by K .
- A set of principles that characterizes certain general features of the set of ideally just worlds, that is, principles that characterize certain general features of the set $C(R, S(K))$. We call this set of principles a "theory of ideal justice".

On my reconstruction of Rawls's and Nozick's reasoning, a theory of ideal justice is justified by showing that, among the worlds that are consistent with the theorist's assumptions about how the world works, those that are highest ranked with respect to the specified evaluative principles satisfy the proposed principles of justice.

¹² The formalization used here differs from — and is more general than — the formalization I used in Wiens 2015*a*.

The Tyranny of a Metaphor

I have surely said too little to vindicate the claim that ideal theorists' efforts to justify their claims about the ideal conform to the logic of an optimization problem. I won't rehearse the details here (see Wiens, 2015*a*). Instead, I wish to show how the model reveals two key implications of optimization reasoning for normative justification. First, the reasoning by which a theory of ideal justice is justified is fundamentally comparative. Ideal theorists select among candidate theories of justice by comparing (typically implicit) models of social worlds constructed to represent the realization of candidate theories of justice. These comparisons are made with respect to certain specified criteria — for example, by assessing the degree to which different model worlds respect individual autonomy, or realize social equality, or promote total welfare. A justified ideal theory is one that is modeled by the social world that is deemed best with respect to the specified evaluative criteria, all things considered. (This comparative mode of reasoning is explicit in Rawls's argument for justice as fairness; in an appendix to Wiens 2015*a*, I show that similar comparisons are implicit in Nozick's reasoning.)

Second, these comparative evaluations are always limited to a proper subset of all possible social worlds. This is basically to make the comparative exercise tractable. It is cognitively impossible for us to compare all social possibilities vis-à-vis each other, and it is cognitively impossible for our evaluative comparisons to attend to every detail of candidate social worlds. To focus their inquiry, ideal theorists make assumptions about certain relevant features of the worlds that will be subject to comparative assessment: the motivations of agents within society, or the technology available to facilitate communication and coordination, and so on. These assumptions restrict the set of social worlds under consideration to those that satisfy the specified assumptions. (For example, if we assume that individuals are hedonistic egoists while specifying an ideal theory of state legitimacy, our comparative assessment will be limited to social worlds at which individuals are hedonistic egoists.) My point is not about ideal theorists' intentions. Ideal theorists don't typically justify their assumptions by reference to their tractability consequences; I'm not even sure their choice of assumptions consciously attends to the ways in which these assumptions both facilitate and limit their inquiry. My point is about the consequences of these assumptions, in particular the implication that ideal theorists' comparative evaluations are always constrained by their assumptions about the basic features of the social worlds under consideration (cf. Wiens, 2017). This feature of their optimization reasoning has consequences for the practical relevance of ideal theories (Wiens, 2015*a*).

The preceding points illustrate a more general point: my model of an optimization problem concerns the *logic of the reasoning by which ideal theorists justify their claims*

about ideal justice. This is so because the model enumerates the core components of the reasoning by which ideal theorists justify their conclusions and specifies the logical relationships among these components. This might be a mistaken model of the logic of justification; it is not my aim here to defend this model against criticisms. Rather, my aim is to demonstrate what it means to model the logic of justification rather than the logic of navigation.

Throughout his book, Gaus characterizes our attempts to navigate a rugged landscape as a “complex optimization problem”. Indeed, one might take his model to formalize an optimization problem as I have just analyzed that notion. To wit, the terrain to be navigated is modeled as a set of social worlds, denoted $\{X\}$ by Gaus; the “altitude” of social worlds is modeled by an evaluative standard, namely, cardinal justice scores; we make assumptions about how particular social worlds work to construct models that enable us to predict the broader consequences of their justice-relevant features; from all this, we are able to identify the ideally just world and characterize its central features. Given this reconstruction, if ideal theorists’ arguments to justify their ideal theories conform to the logic of optimization, Gaus’s model must not obscure the logic of justification, as I charged above.

But this reconstruction of Gaus’s model is hasty. Gaus uses the term “optimization problem” to characterize the logic of *navigation*: we are *searching for* the best or “optimal” world. Such an exercise presupposes that we are able to distinguish optimal worlds from suboptimal ones. This is where the logic of justification is relevant. On Gaus’s model, the reasoning by which we identify the ideal world (i.e., by which we justify the claim that a particular world is ideal) is contained in the black box from which justice scores emerge. We know that justice scores are given by a function that aggregates evaluative standards, but we are told little more about the reasoning underpinning our assignment of justice scores. Gaus’s model offers no “reasoning template” for assigning justice scores to worlds; worlds are simply assumed to come with cardinal justice scores. Contrast this with my optimization model above, which constructs a “template” for justifying ideal theories by identifying the generic components of ideal theoretic reasoning and the logical relationships among these components.¹³

13 Wait — am I not guilty of black boxing how (e.g.) evaluative principles are constructed on my model? Have I thereby failed to adequately model the logic of justification? (Thanks to Brian Kogelmann for raising this point.) As it happens, I have work in progress that develops a model of evaluative principles, which largely follows Dietrich and List (2016). I also provide an example of how evaluative principles work in section 4 below. But it’s important to notice that, to adequately model the logic by which we identify particular worlds as ideal, I don’t need to say much about how evaluative principles are constructed; these details can be largely left to particular ideal theorists. All I must do is point out that a set of principles with a specific

The Tyranny of a Metaphor

Here's a different way to put my point. If asked, "How do we determine which social worlds are ideal?" Gaus's answer must be, "Look at their justice scores". My answer is similar: "Look at the choice set". But my model outlines the basic logic by which worlds are placed in the choice set. In contrast, Gaus's model is virtually silent on the reasoning that leads to the assignment of justice scores. It's the inner workings of that black box that constitutes a logic of justification.

We can get a feel for the extent of Gaus's silence on matters of justification by considering how things might work inside the black box. Start by noticing that Gaus presupposes a model of normative reasoning that is fundamentally non-comparative. While it's true that the justice function ensures a comparative ordering of worlds, this justice function is logically primitive in Gaus's model and does not represent a more primitive set of pairwise comparisons (as I noted in footnote 8). Consider an example to illustrate the point. The Human Development Index (HDI) is an effort to rank countries with respect to human development, basically, expected quality of life. The HDI is a composite index, constructed from three component indices measuring life expectancy at birth, average educational attainment, and income per capita. We can set aside the technical details of how this index is constructed.¹⁴ The key point here is that a country's HDI score is determined by collecting and analyzing data from each country in isolation from the others. To calculate Angola's HDI score, for instance, we analyze data on life span, educational attainment, and household income that is collected *from Angola*. We can, of course, use the HDI to rank countries with respect to human development. But the construction of the HDI is fundamentally non-comparative. Compare: "Norway is the highest-ranked country with respect to human development because it bests all other countries in a series of head-to-head comparisons", versus "Norway is the highest-ranked country with respect to human development because it has, statistically speaking, the best overall balance of life expectancy, average educational attainment, and income per capita". The first comparative claim essentially depends on a set of pairwise comparisons among countries; the basis for the second comparative claim is countries' non-comparatively determined HDI scores.¹⁵

The assignment of justice scores in Gaus's model is essentially non-comparative in just this way (46–8): we determine a world's justice score by measuring its "inherent justice". In terms of the rugged terrain metaphor, we can think of an accurate map being

function (viz., comparatively ranking worlds) is required and demonstrate the role they play in justification. I do this above.

¹⁴ For details, see <http://hdr.undp.org/en/content/human-development-index-hdi>.

¹⁵ Gaus's example using figures 3-1, 3-2, and 3-3 (108–11) confirms that my HDI illustration captures the essentials of how normative reasoning works on Gaus's model.

drawn by collecting and analyzing data (so to speak) from each world in isolation from the others. This implies that the reasoning by which we identify the ideal is essentially non-comparative. A particular world is ideal not because it bests all other worlds in a series of head-to-head comparisons (as on my model), but because it has the highest inherent justice score. To reason successfully in this non-comparative mode, we must have (at a minimum) a set of cardinally measurable evaluative standards, which we can use to determine each world's inherent justice.¹⁶ Since Gaus brackets the construction of these evaluative standards and their use in our normative reasoning, the logic by which we identify a particular world as ideal is left obscure.¹⁷

I have tried in this section to provide an example of a model that explicitly investigates the logic of normative justification as a means to make clear how Gaus's formalization of the navigation metaphor leads him to comparative silence on such matters. The model I have presented constructs a "reasoning template" by outlining the ways in which different kinds of considerations fit together to justify a theory of ideal justice. In contrast, Gaus's model simply presupposes a procedure by which we assign justice scores to worlds — and, thus, the reasoning by which we identify the ideal world — but sets aside many of the controversial issues raised by the assumed procedure. This might still seem like a puzzling charge. Gaus obviously explores questions about the use of his assumed justice function for mapping the space of social worlds — in particular, whether, given our epistemic limitations, we can obtain the information required to assign justice scores to distant worlds. But notice that this is a question about our ability to "collect quality data" to feed into the assumed justice function; it says nothing about how the justice function is constructed. Put another way, then, my point is that Gaus's model leads him to neglect questions about the *logical structure* of the reasoning process by which we assign justice scores and, in addition, whether the assumed reasoning process is something that captures the logic by which ideal theorists justify their claims about the ideal.

16 Additionally, we must be able to collect the relevant data from each social world. I take it that this is the point of Gaus's "modeling task". We can thus read Gaus's later skepticism about modeling distant worlds as expressing skepticism about our ability to collect the data required to measure the justice of distant worlds.

17 Gaus eschews a "strictly comparative" approach because he worries that pairwise comparisons among worlds might be intransitive, in which case, we are not guaranteed to have an ideal world (47). Gaus goes with primitive justice scores to ensure a transitive ordering and, hence, an ideal world. But this doesn't address the worry so much as it skirts it by fiat. Since taking cardinal justice scores as primitive imposes transitivity on our comparisons among worlds anyway, why not just stipulate that "proper" pairwise comparative assessments of worlds must be transitive (or, more weakly, acyclic, which is sufficient to guarantee an ideal world)? At least the latter route can avoid the strong assumption of cardinal measurability required by Gaus's approach. Not only is this cardinal measurability left unexplained, thus obscuring the reasoning by which we identify the ideal; it is surely more controversial than assuming transitive pairwise comparisons.

The Tyranny of a Metaphor

To be clear, none of this is meant as a criticism of Gaus's model. Recall one of my starting points: metaphors usefully limit our attention as a means to facilitate theoretical inquiry on a complex subject. I've pointed out the ways in which Gaus's model fruitfully extends what has become a central metaphor in debates on the practical relevance of ideal theory. Gaus's explicit purpose in constructing his model is to investigate the relevance of ideal theory as a guide to navigating a rugged terrain of social possibilities — to explore the logic of navigation, as I called it earlier. Inquiries into the logic of navigation concern the information we need to reach the highest peak and the possibilities for acquiring that information. This presupposes that we have a procedure that, given the required data, enables us to identify the highest peak. To pause and wring our hands about how we do this is beside the point of the inquiry. So Gaus isn't to be faulted for constructing a model that ignores the logic of justification.

Gaus's neglect of the logic of justification becomes problematic, though, because it blinds him to the ways in which his argument for the Open Society manifests the optimization reasoning used by ideal theorists like Rawls and Nozick. In the next section, I show how the Open Society is closer to an ideal than Gaus appreciates.

4. GAUS'S IDEAL THEORY OF THE OPEN SOCIETY

In arguing for the Open Society, Gaus explicitly distances himself from ideal theorists. The Open Society is dubbed "The Nonideal", for example. Gaus uses the concept of "normalization" to achieve further distance. Traditional theories of ideal justice presuppose a fully normalized evaluative perspective: all individuals are assumed to share the same set of basic evaluative considerations when reasoning about the optimal solution to the problem of social cooperation (150–3). More recently, political liberals like Rawls and social choice theorists like Sen have reasoned about justice from a partially normalized perspective: individuals are assumed to disagree about fundamental evaluative matters, but everyone is assumed to share the same description of the social worlds under consideration (153–8). In contrast, Gaus, following Muldoon (2016), seeks a perspective on justice that abandons normalization altogether: our reasoning about justice assumes radical diversity with respect to both matters of normative evaluation and the description of the social worlds under consideration (165–9). To the extent that normalization is an element of ideal theories (142, 144, 150), Gaus's rejection of normalization places his theory of the Open Society among nonideal theories.

It's true that Gaus's reasoning to justify the Open Society assumes that people do not share evaluative or descriptive perspectives. But, as I will now argue, his argument for the Open Society manifests the kind of optimization reasoning that I argue is at least

implicit, if not explicit, in conventional theories of ideal justice. This effectively assumes a normalized evaluative perspective from which to compare social worlds and a normalized description of the social worlds under consideration.

To see this, we must recall the problem that emerges from Gaus's argument against the practical relevance of ideal theory: our best prospects for mapping the terrain of social possibilities requires fostering diverse evaluative perspectives, but deep diversity hinders communication among divergent perspectives. So pluralistic liberal societies will not converge on a unified ideal of justice. This threatens our ability to reap the cognitive benefits of diversity and, in turn, our prospects for making moral progress. The central problem, then, is to construct a social framework that facilitates cooperation and communication among people who have divergent perspectives on evaluative and descriptive matters. In Gaus's words, the last half of the book is an "inquiry into a justified liberal framework: under what conditions can we live under a shared moral framework that is diversity accommodative because it is accommodative to diversity per se, and so is an open society?" (176) Importantly, if this liberal framework is to harness the potential of diversity to bring moral progress, it must be one that diverse moral communities can endorse "as a bona fide just way to relate" (174). As a solution to this problem, Gaus conjectures that "such a framework of liberal diversity seems most likely when our public social world is shaped by a set of characteristic features of the Open Society" (176). How does Gaus substantiate this conjecture? By arguing that, among the options for organizing a diverse liberal society so as to effectively realize moral improvements, the normatively optimal options with respect to certain evaluative criteria are characterized by certain core features of an Open Society. Put simply, Gaus aims to justify the Open Society by showing that it is a solution to a normative optimization problem.

Note that, on my model of optimization reasoning, this kind of argument requires some normalization: we must settle on the set of worlds under consideration, as well as the criteria to be used to comparatively evaluate these worlds. We need not normalize the perspectives (evaluative or descriptive) of *the people in our model worlds*; that would bracket the problem in this case. Rather, for the argument to be minimally cogent, it must assume a normalized perspective from which to consider the problem of constructing a social framework for a pluralistic liberal society.

Let's start by briefly describing Gaus's constraints on the set of worlds under consideration. Obviously, he restricts his attention to worlds characterized by deep perspectival diversity; without this restriction, Gaus's central problem does not clearly emerge. Additionally, this perspectival diversity is assumed to hinder communication among individuals, preventing convergence on shared ideals of justice (130f, 145). Relatedly,

The Tyranny of a Metaphor

people are assumed to have limited ability to predict the consequences of implementing social reforms, particular when the reforms would lead to unfamiliar social terrain (102–3). It follows that people also have limited ability to determine the inherent justice of unfamiliar social worlds. Importantly, Gaus assumes several constraints on the conditions under which social cooperation can occur. For example, coordination requires promoting shared behavioral expectations among members of a group; thus, to sustain social cooperation, a group must establish a social practice of holding each other accountable to common moral rules (180–2).

These and other assumptions about how the world works are important for defining the problem for Gaus. Accordingly, his attention is restricted to the set of worlds that satisfy these constraints; to consider worlds that violate these constraints is to consider worlds where his central problem may not even arise. By calling these constraints “assumptions”, I do not mean to say that Gaus is ultimately unjustified in adopting them. I simply mean to indicate that he does not spend much effort connecting these conditions to the actual world—showing, for example, that his model tracks the depth and extent of perspectival diversity we encounter in actual societies, or the empirical conditions under which deep perspectival diversity inhibits the formation of shared behavioral expectations within a group. I don’t think this is a problem *per se*; Gaus is free to define the central problem as he likes. My only point is that the way in which he defines the problem restricts the set of social possibilities under consideration and, in turn, which possibilities are candidates for the optimal world.

We now turn to Gaus’s efforts to substantiate his claim that certain core features of the Open Society characterize the normatively optimal worlds within his constrained set of options. But, first, what are the core features in question? The normatively optimal worlds, on Gaus’s view, are characterized by the Principle of Natural Liberty (187–9), a system of jurisdictional rights (199–201), and markets for facilitating exchange (202–5).¹⁸ Put differently, worlds that are characterized by these three features are normatively superior to worlds that lack any one of these features.

“Normatively superior” in what ways? What are the criteria Gaus uses to substantiate this comparative claim? The main criterion concerns the extent to which worlds can accommodate perspectival diversity. Recall Gaus’s arguments earlier in the book about the ways in which perspectival diversity promotes moral progress. If these arguments are correct, then we have strong reasons to rank worlds at which diversity is accommodated

¹⁸ I say “optimal worlds”, plural, because these three features can be realized by numerous worlds that are otherwise quite divergent. Indeed, this is, for Gaus, part of the Open Society’s appeal: it gives societies wide latitude on many details.

higher than worlds at which diversity is suppressed (174–6).¹⁹ Since deep diversity can limit communication and, in turn, moral learning, we have strong reasons to rank worlds at which diversity is productively and cooperatively harnessed higher than worlds at which diversity leads to conflict and confusion (183f). For Gaus, this effectively means that we have strong reasons to rank worlds that realize “a cooperative social life supported by a practice of accountability” higher than worlds that fail to realize this (220; cf. 180–3).

Let’s state these points more explicitly in the form of comparative criteria. For the subset of worlds $S(K) \subseteq X$ that satisfy Gaus’s constraints (denoted K), and all worlds $x, y \in S(K)$,

- x is normatively superior to y if x more effectively accommodates perspectival diversity than y , other things equal;
- x is normatively superior to y if x realizes a cooperative social life supported by a practice of accountability to a greater extent than y , other things equal.²⁰

Gaus does not state his evaluative criteria so explicitly; indeed, he does not even articulate these points *as evaluative criteria*. But the fact that they play this role in his reasoning comes out when we reconstruct his arguments in favor of the Principle of Natural Liberty and a system of jurisdictional rights.

The Principle of Natural Liberty (PNL) says “[w]hatever is not prohibited. . . is permitted” (187).²¹ Gaus compares this principle with two alternatives. The All Liberal Liberties Are Specifically Justified Principle (SJP) says, in effect, “[w]hatever is not permitted is prohibited” (192, 191); the Proceed with Justification Principle (PJP) says, in effect, “whatever is not currently permitted is permissible if and only if it can be justified to others” (cf. 196).

A society characterized by the PNL is superior to one characterized by the SJP because the former more effectively accommodates diversity. As emerging perspectives catalyze innovative activity in a society, the innovations will be permitted by the PNL by default, but prohibited by the SJP (194–5). Consequently, a society characterized by the PNL more

19 This assumes, of course, that we have reasons to prefer worlds that facilitate moral progress. We might say, then, that the most fundamental evaluative criterion concerns the extent to which worlds facilitate moral progress. I take the diversity accommodation criterion in the text to be a more precise specification of this moral progress criterion in light of Gaus’s assumptions about human cognitive limitations.

20 I’ve articulated these criteria to permit continuous variation along both dimensions. But I don’t mean to suggest that these dimensions in fact vary continuously; I wish to remain neutral on the matter. My articulations do not rule out the possibility that these are binary (“on/off”) variables.

21 This is stated more precisely for an interpersonal context on p. 189, but the details don’t matter here.

The Tyranny of a Metaphor

effectively “encourages experimentation and discovery” than a society characterized by the SJP (196). A society characterized by the PNL is superior to one characterized by the PJP for two reasons. First, the PNL gives more decisive guidance than the PJP and, thus, more effectively coordinates social activity. So the PNL is more effective for realizing a cooperative social life supported by a practice of accountability (190, 197). Second, the PJP stifles experimentation and innovation by placing a heavy burden of justification on innovators; thus, the PNL better accommodates diversity (197–8). In sum, worlds characterized by the PNL are normatively optimal among the worlds considered with respect to the specified evaluative criteria.

A system of jurisdictional rights creates “autonomous zones” for individuals in society, spheres in which individuals are free from the interference of others to believe and act as they see fit (200). Examples include freedom of conscience, freedom of association, private property rights, and so on. The purpose of jurisdictional rights is to reduce the complexity of interactions within a diverse society, to ensure that perspectival changes in one part of society do not reverberate throughout society (199). This point is most easily seen by contrasting a system of jurisdictional rights with a “corporatist” society, in which shared moral rules are secured by a complex network of social bargains that enumerates a finely balanced schedule of specific claims that different groups have against each other. In a corporatist society, the introduction of, say, a new religious perspective brings new demands, creating the need for new social bargains with the other groups to settle specific religious accommodations, specific claims on public funds, specific claims to representation in policy forums, and so on. These new claims must be fit with the existing system of specific social bargains, which is already rife with interdependencies. Insofar as the new claims can only be accommodated by circumscribing existing claims, the required bargains threaten to unravel the entire network of social bargains. In contrast, a system of jurisdictional rights insulates society from the effects of perspectival changes in one part of society; a new religious perspective enjoys the same zone of autonomy as the existing perspectives and is only asked to respect others’ jurisdictional rights (200).

The contrast between a system of jurisdictional rights and corporatist social bargains provides for an implicit comparative evaluation: by reducing the complexity of social interactions, a society characterized by a system of jurisdictional rights more effectively accommodates diversity than a corporatist system of social bargains (200). Gaus more explicitly compares a system of jurisdictional rights with socialism (of both the state-led and democratic varieties). Private property rights are said to be “quintessentially jurisdictional” (201), while socialist societies put property under collective control. Socialism is inferior for the same reasons as corporatism: its system of collective decision-making has

extreme difficulty making tradeoffs among numerous and diverse demands. When implemented, socialism typically suppresses diversity, stifling innovation and experimentation. In contrast, a system of jurisdictional rights nurtures diversity, thereby encouraging innovation and experimentation. In sum, a society characterized by a system of jurisdictional rights is normatively optimal among the worlds considered with respect to the specified evaluative criteria.

Given this reconstruction, Gaus justifies the Open Society by arguing that it solves a normative optimization problem: among the worlds that are consistent with the specified constraints, worlds that satisfy three characteristic features of the Open Society are argued to represent the normatively optimal solution to the problem of organizing social activity in diverse societies. Thus, Gaus's mode of reasoning is structurally identical to that of ideal theorists like Rawls and Nozick (cf. Wiens, 2015*a*, 435–7). Like Rawls, Nozick, and other ideal theorists, Gaus's argument for the Open Society can at best indicate the highest peak in a particular subregion of the terrain of social possibilities.

What, then, are we to take away from Gaus's argument for the Open Society? More specifically, is Gaus's theory of the Open Society practically relevant? Does his argument justify a set of practically relevant normative prescriptions? Does his argument give us sufficient reason to think that the Open Society characterizes a set of social arrangements that we should aim to realize? In previous work, I have argued that the constraint-relative property of ideal theories precludes them from being useful for justifying practically relevant normative prescriptions (Wiens, 2015*a,b*, 2017). With respect to the Open Society in particular, and speaking metaphorically, Gaus has shown us the highest peak in one subregion of the territory of social worlds. Numerous questions remain before we are justified in setting a course for the Open Society. How does this peak compare to the high points in other subregions?²² Given these other high points, why should we aim for the Open Society rather than other high points?²³ Why should we set a course for the subregion picked out by Gaus's specified constraints? Has Gaus been exploring the

²² Given the way ideal theorists proceed, we can't even answer this question until we do a comparative analysis of these high points across subregions. Given that these subregions are defined by distinct sets of constraints, this cross-sectional comparative analysis requires increasingly general assumptions about how the world works, which may subvert the tractability of such analyses.

²³ To this question, Gaus might answer that his is a "meta-peak", that the Open Society is a superior kind of ideal because it permits exploration of all peaks. If he wishes to go this way, then I'd point out that this makes the Open Society a particular peak in a "meta-terrain", namely, the terrain of strategies for achieving moral progress within society. The Open Society picks out the optimal cluster of strategies for realizing moral progress given certain constraints. But this doesn't show that the Open Society picks out the globally optimal cluster of strategies; nor does it show that we should be selecting strategies from the specified constrained set.

The Tyranny of a Metaphor

subregion in which the status quo is located? If so, have we any reason to believe that we can attain the highest peak? If not, why should we travel to the subregion Gaus has been exploring rather than another? (What if the path from the status quo to the Open Society is more costly to traverse than the paths to lesser peaks in the same region or to peaks in a different subregion?) Until we answer these and related questions, we simply can't say anything about the practical relevance of the Open Society.

To be clear, I don't take these to be criticisms of Gaus's argument for the Open Society; rather, I only mean to caution against taking it as justifying practically relevant normative prescriptions. To put my points in the previous paragraph less metaphorically: before we can justify taking the Open Society as providing practical direction, Gaus has much work to do to connect the Open Society to the status quo. In particular, Gaus has yet to diagnose the mechanisms underlying the social problems we face and show how the Open Society might be an effective solution to these problems (see Wiens, 2012); additionally, Gaus has yet to explore the causal mechanisms by which we could bring about the Open Society given the mechanisms that are operative at the status quo (see Wiens, 2013). Perhaps ironically, these are exactly the kinds of gaps that stand in the way of taking traditional ideal theories as justifying practically relevant normative prescriptions.

5. TWO LAST WORDS

It is not a failing of *The Tyranny of the Ideal* that Gaus's argument for the Open Society is insufficient to justify practically relevant normative prescriptions. Perhaps Gaus doesn't intend his theory of the Open Society to be practically relevant; he never expresses that aspiration. Moreover, failing to be practically relevant need not prompt criticism. Contrary to much that has been written in debates about ideal theory, an ideal theory need not be practically relevant to be useful. Indeed, I think that Gaus's model of the Open Society can do useful *conceptual* work, even if it is not practically relevant (cf. Johnson, 2014).²⁴ In this case, Gaus's Open Society provides a model for exploring how we might reconcile liberty, diversity, and morally progressive social cooperation. By enabling us to explore potential relationships among these concepts in a concrete environment—in particular, by exploring potential tradeoffs involved in instantiating these concepts under

²⁴ Perhaps this is Estlund's point against utopophobia, although I'm not certain. Estlund often casts his arguments for idealistic political philosophy in terms of "discovering the truth about justice" (cf. Estlund, 2011, 2017), although he sometimes suggests that he is after truths about deontic matters, our "true" obligations of justice or somesuch (cf. Estlund, 2014). While I am sympathetic to the thought that idealistic political philosophy can do useful conceptual work, I'm not at all sympathetic with the claim that it can help discover our "true" obligations (see Wiens, 2017).

various conditions — such models help to enrich our understanding of these concepts (cf. Hausman, 1992, chs. 1–5). Although I do not have the space to defend the claim, this is useful and important work; indeed, it is useful and important *for the purposes of crafting practical normative theories*. In contrast with recent calls to abandon ideal theory (e.g., Farrelly, 2007; Mills, 2005; Sen, 2009), the need for rich conceptual exploration gives political philosophers reasons to continue doing ideal theory.

A final caveat brings me full circle. While I think exploring idealistic models of society can enrich our understanding of key normative concepts like liberty, equality, social cooperation, and so on, we should be wary of treating any single model or small class of models as defining the subject of inquiry. This risks unnecessary and undesirable theoretical blindness. Political philosophers should proliferate idealized models, exploring the ways in which (e.g.) liberty can be reconciled with authority or equality reconciled with personal responsibility under multifarious social conditions. And we shouldn't limit ourselves to exploring idealized models. Enriching our understanding of normative concepts requires attending to potential tradeoffs that arise in nonideal conditions too (see Barry and Wiens, forthcoming). Yes, our models should simplify, idealize, and abstract from potentially distracting complications. But we must also seek ways to integrate conceptual insights across different classes of models with the aspiration of gaining a more comprehensive view.

REFERENCES

- Arrow, Kenneth J. 1963. *Social Choice and Individual Values*. 2nd ed. New Haven: Yale University Press.
- Barry, Christian and David Wiens. Forthcoming. What Second Best Scenarios Reveal About Ideals of Global Justice. In *Oxford Handbook of Global Justice*, ed. Thom Brooks. New York: Oxford University Press.
- Boot, Martijn. 2012. "The Aim of a Theory of Justice." *Ethical Theory and Moral Practice* 15:7–21.
- Dietrich, Franz and Christian List. 2016. "What Matters and How it Matters: A Choice-Theoretic Representation of Moral Theories." London School of Economics, unpublished manuscript.
- Estlund, David. 2011. "What Good Is It? Unrealistic Political Theory and the Value of Intellectual Work." *Analyse & Kritik* 2:395–416.

The Tyranny of a Metaphor

- Estlund, David. 2014. "Utopophobia." *Philosophy & Public Affairs* 42(2):113–134.
- Estlund, David. 2017. Prime Justice. In *Political Utopias: Contemporary Debates*, ed. Kevin Vallier and Michael Weber. New York: Oxford University Press.
- Farrelly, Colin. 2007. "Justice in Ideal Theory: A Refutation." *Political Studies* 55(4):844–864.
- Gaus, Gerald. 2016. *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton and Oxford: Princeton University Press.
- Gilabert, Pablo. 2012. "Comparative Assessments of Justice, Political Feasibility, and Ideal Theory." *Ethical Theory & Moral Practice* 15(1):39–56.
- Hamlin, Alan and Zofia Stemplowska. 2012. "Theory, Ideal Theory and the Theory of Ideals." *Political Studies Review* 10:48–62.
- Hausman, Daniel. 1992. *The Inexact and Separate Science of Economics*. New York: Cambridge University Press.
- Heath, Joseph. 2017. "On the Scalability of Cooperative Structures: Remarks on G. A. Cohen, *Why Not Socialism?*" University of Toronto, unpublished manuscript.
- Johnson, James. 2014. "Models Among the Political Theorists." *American Journal of Political Science* 58(3):547–560.
- Mills, Charles W. 2005. "'Ideal Theory' as Ideology." *Hypatia* 20(3):165–184.
- Muldoon, Ryan. 2016. *Social Contract Theory for a Diverse World: Beyond Tolerance*. New York: Routledge.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Rawls, John. 1999. *A Theory of Justice*. 2nd ed. Cambridge, MA: Harvard University Press.
- Schelling, Thomas C. 2006. *Micromotives and Macrobehavior*. London and New York: W. W. Norton & Co.
- Schmidtz, David. 2011. "Nonideal Theory: What It Is and What It Needs to Be." *Ethics* 121(4):772–796.
- Sen, Amartya. 1970. *Collective Choice and Social Welfare*. San Francisco: Holden-Day.
- Sen, Amartya. 2006. "What Do We Want From A Theory of Justice?" *Journal of Philosophy* 103(5):215–238.

- Sen, Amartya. 2009. *The Idea of Justice*. Cambridge, MA: Harvard University Press.
- Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy & Public Affairs* 38(1):5–36.
- Stemplowska, Zofia and Adam Swift. 2012. Ideal and Nonideal Theory. In *The Oxford Handbook of Political Philosophy*, ed. David Estlund. New York: Oxford University Press.
- Swift, Adam. 2008. "The Value of Philosophy in Nonideal Circumstances." *Social Theory and Practice* 34(3):363–387.
- Valentini, Laura. 2011. "A Paradigm Shift in Theorizing About Justice? A Critique of Sen." *Economics and Philosophy* 27:297–315.
- Valentini, Laura. 2012. "Ideal vs. Non-ideal Theory: A Conceptual Map." *Philosophy Compass* 7(9):654–664.
- Valentini, Laura. 2017. On the Messy 'Utopophobia vs. Factophobia' Controversy: A Systematization and Assessment. in Vallier and Weber (2017) pp. 11–35.
- Vallier, Kevin and Michael Weber, eds. 2017. *Political Utopias: Contemporary Debates*. New York: Oxford University Press.
- Wiens, David. 2012. "Prescribing Institutions Without Ideal Theory." *The Journal of Political Philosophy* 20(1):45–70.
- Wiens, David. 2013. "Demands of Justice, Feasible Alternatives, and the Need for Causal Analysis." *Ethical Theory & Moral Practice* 16(2):325–338.
- Wiens, David. 2015a. "Against Ideal Guidance." *Journal of Politics* 77(2):433–446.
- Wiens, David. 2015b. "Political Ideals and the Feasibility Frontier." *Economics and Philosophy* 31(3):447–477.
- Wiens, David. 2017. Will the Real Principles of Justice Please Stand Up? in Vallier and Weber (2017).