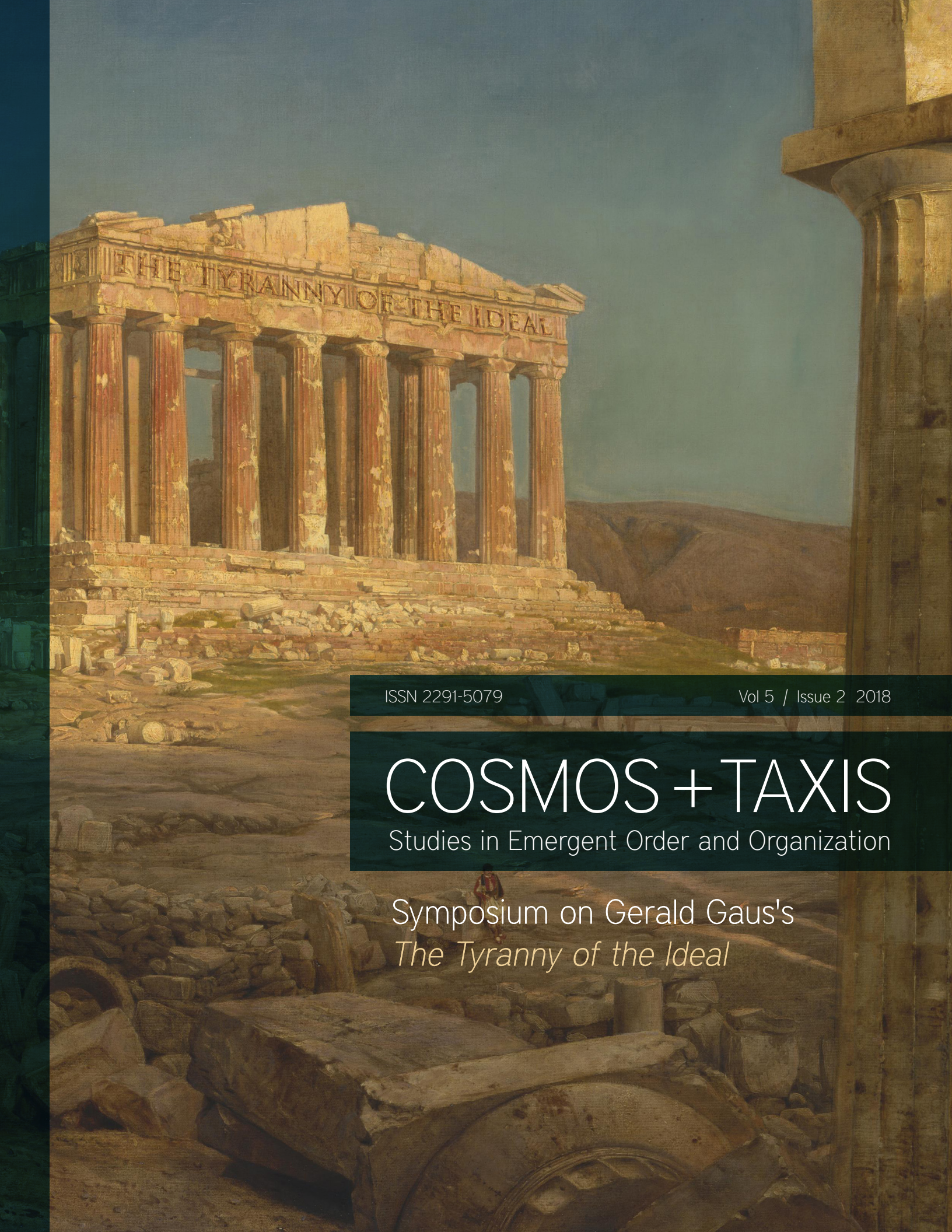

ISSN 2291-5079 Vol 5 / Issue 2 2018

COSMOS + TAXIS

Studies in Emergent Order and Organization

Symposium on Gerald Gaus's *The Tyranny of the Ideal*

COSMOS + TAXIS

Studies in Emergent Order and Organization

VOL 5 / ISSUE 2 2018

COVER ART (WITH APOLOGIES):
Frederic Edwin Church
Parthenon, 1871
Metropolitan Museum of Art, New York

IN THIS ISSUE

EDITORIAL BOARDS

Introduction to Symposium on Gerald Gaus' *The Tyranny of the Ideal*

RYAN MULDOON

Department of Philosophy
University at Buffalo
135 Park Hall
Buffalo, NY 14260-4150

Email: rmuldoon@buffalo.edu
Web: http://www.buffalo.edu/~rmuldoon

The Tyranny of the Ideal is a powerful challenge to the dominant approach to political philosophy. Gaus argues that the pursuit of a political ideal is fundamentally problematic. Indeed, once we understand the structure of the problem that ideal theorizing in political philosophy has, we find that our ability to identify a political ideal is deeply constrained. Homogenous societies have an extremely constrained "vision" of alternative social worlds. Diverse societies may be able to see more, but at the cost of disagreement over what it is we should be looking for. Indeed, even if we can find an ideal we can agree on, we now face what Gaus calls The Choice: "…we must choose between relatively certain (perhaps large) local improvements in justice and pursuit of a considerably less certain ideal." (p. 142) It is in The Choice where Gaus identifies the *tyranny* of ideal theory. Commitment to an (theoretical, possibly mistaken) ideal can cause us to make our world *less just* while we chase our vision of justice.

By combining careful use of formal models with a rich philosophical framework, Gaus demonstrates how ideal theorists have glossed over the core challenges posed by social complexity on the one hand and human diversity on the other. By starting from the understanding that social systems are complex systems, Gaus offers a reorientation of political philosophy. Rather than focus on identifying the ideal, with a notion of a well-ordered society in which all share a conception of justice, Gaus presents us with a framework for better debating different visions of the ideal. This reorientation is necessary: in diverse, complex systems, there isn't going to be a realizable ideal. Instead of ignoring these fundamental constraints, Gaus encourages us to choose to leverage diversity to help us navigate social complexity.

In this version of the Open Society, the aim is to allow for as wide a scope of different perspectives to hold each other accountable to a public morality while allowing for different republican communities to develop their visions of the ideal. As Gaus makes clear, this is not a vision of a society in which everyone shares common values and lives in perfect harmony with each other. Instead, it is a potentially frustrating world of disagreement, debate and discovery. A commitment to the Open Society is a commitment to not always getting your way, just like everyone else.

This issue is devoted to exploring and challenging the ideas contained within *The Tyranny of the Ideal*. The lead essay, by Scott Page, engages with the formal challenge of identifying and attaining just social worlds in a complex environment, and the ways in which diversity can contribute to this endeavor. Next, David Wiens challenges Gaus' use of the landscape metaphor, and suggests that this metaphor has led to a larger issue in understanding the relationship between ideal and non-ideal theories. The third essay, by Fred D'Agostino situates Gaus' work in a broader discussion of how we should go about engaging in political philosophy. This leads us to our final contributors. Kevin Vallier presents a puzzle with how we might reconcile *Tyrrany of the Ideal* with Gaus' previous book, *The Order of Public Reason*, given that his earlier book relies on public justification as an equilibrium concept, whereas his latest book argues that we must be able to break free of equilibria and move to new ones. Blain Neufeld and Lori Watson offer a robust defense of Rawlsian approaches, and particularly the conception of a well-ordered society, against Gaus' arguments. The issue concludes with a reply from Gaus.

The Imperative of Complexity

SCOTT E. PAGE

Leonid Hurwicz Collegiate Professor Complex Systems, Political Science, and Economics
University of Michigan
Ann Arbor, MI 48109

External Faculty, The Santa Fe Institute

Email: spage@umich.edu
Web: https://sites.lsa.umich.edu/scottepage/

In *The Tyranny of the Ideal,* Gerald Gaus constructs a deep, wide ranging argument in favor of a diverse, open, adaptive society and against the necessity of a social ideal as a guidepost. He hangs much of his argument on mathematical models, constructing formal assumptions and working through a logical progression that improves upon the blurry conceptualizations and logical leaps of others. The book provokes, challenges, and inspires. I am privileged to have the opportunity to offer commentary.

I restrict my thoughts to the part of his analysis that frames the processes of identifying and attaining just social worlds as problem solving. I devote particular attention to his analyses of the contributions and limits of diversity in complex domains. This confined engagement stems in part from the centrality of problem solving to his argument, and in part from my lack of depth in philosophy. Gaus would surely applaud my self-imposed confinement as it respects his claim of conjectural and evaluative myopia, i.e. the idea that we can only imagine and understand social arrangements near to our own.

Gaus's core argument relies on the following sequence of logical claims. First, there exists an enormous, incomprehensible number of possible arrangements of our world. By this he means feasible collections of laws, organizational structures, and institutions. Second, the features of that world interact in symbiotic and conflicting ways to produce a *rugged landscape* with many local optima. The ruggedness of the landscape creates mirroring paradoxes: locally improving the social world could move society further from the ideal arrangement, and moving toward the ideal arrangement could come at an immediate cost to our well-being.

Third, we have at best local knowledge of the landscape. That limitation combined with the landscape's ruggedness can undermine the possibility of knowing the ideal.

Fourth, diverse perspectives appear to offer a fix by smoothing the landscape. However, diversity only takes us so far, because as we approach what we thought we wanted, we learn about ourselves and our world and alter our conception of the ideal. Thus, he warns against positioning our rudder and setting sail.

In sum, Gaus argues that to conceive of the ideal and apply it as a polestar doubly tyrannizes. We may move away from a certain improvement. And we may head toward an incompletely considered ideal.

THE HONG-PAGE FRAMEWORK

To build his argument, Gaus constructs a formal model that elaborates on the Hong-Page framework for problem solving (Hong and Page 2001). The Hong-Page framework assumes a value function defined over a set of possibilities along with problem solvers who possess *perspectives*: representations of the set of possibilities, and *heuristics:* algorithms or methods for maneuvering within their representations.

The framework formalizes computational optimization in which a problem domain is encoded in a formal language and algorithms find improvements. The traveling salesperson of finding the shortest route between a collection of cities is a canonical example of such a problem. To find a solution, a problem solver must first develop a representation for all possible routes. The natural encoding (a Hong-Page perspective) would be an ordered list of cities. Heuristics then manipulate that ordering to find shorter routes.

Common heuristics include switching adjacent pairs of cities and switching cites separated by a city.

The parsimony of the Hong-Page framework allows it to be applied across any number of domains: finding a molecule that cures hepatitis C, designing a car engine, selecting a location for a movie shoot, or writing a national health care plan. Within each, we can define a value function and a domain of alternatives. We can also imagine multiple representations of the possible solutions (*perspectives*) and various ways to search among them (*heuristics*).

Applying a single stark framework across such vast disciplinary boundaries allows for transdisciplinary. As an analogy, applying a basic network model of contagion to study the spread of disease, information, fads, and technologies allows one to derive general insights for how the degree distribution of the network, the contact rate, and the probability of transmission combine to determine the likelihood of contagion.

At the same time, any one size fits all approach glosses over details. Some assumptions do not apply; square pegs get forced into round holes. To return to the contagion example, as we take deeper dives into the various domains, we discover the limits of lumping the measles, hybrid corn, Lady Gaga, and fax machines in the same conceptual bucket. We find that the differences in network structures across the domains matter. A person can give the measles to at most a few people at a time. By tweeting, President Trump can reach tens of millions.

Further, we see that while disease transmission occurs over an undirected network—I can give my wife the flu, and she can spread the flu to me—social influence occurs through a directed network, as does information. I may wear the same shoes as LeBron, but so far as I know, Lebron cares not a whit about my choice of kicks.

A FRAMEWORK FOR EVALUATING THE IDEAL

When Gaus applies the Hong-Page framework to the problem of creating an ideal society, he lumps. In doing so, he can invoke a general insight—that diversity leads to better solutions. The unavoidable misfits oblige him to refine the model to better fit his domain of interest. That refinement elucidates gaps and opens new possibilities for analysis.

To prove the general point, he must reinterpret the Hong-Page assumptions. Gaus first defines a set of possible social worlds {X}, an analog to the set of possible solutions in the Hong-Page framework. We can think of these social worlds as possible arrangements of society. As an analog of the outcome function, he assumes a *social realizations condition,* T, that assigns either a value or ranking to each alternative in {X}. An alternative's ranking corresponds to its level of justice.

Gaus also defines a social ideal, denoted by u. If we assume that {X} is finite and that T admits no ties, a unique ideal necessarily exists. To this, Gaus adds an *orientation condition* that captures the proximity between the ideal, u, and some alternative arrangement, x. This added feature of the model plays a key role. In the cases where T scores justice as a real value, {X} acts as the domain of the function T, and the social justice levels can be thought of as the range of T.

Proximity operates as a distance measure applied to the domain of T. Similarly, justice levels can be used to determine distances in the range of T. Conflict between these two distance measures causes the first problem with the ideal. We can move closer to the ideal according to one measure and further from the ideal in another.

THE {MARKET, BUREAUCRACY, DEMOCRACY} MODEL

To unpack Gaus's argument, I introduce a stark model that emphasizes the role of institutions in creating the ideal (Rawls 1999). Within the model, I can define proximity and show how interdependencies produce rugged landscapes.

In the model, a society must allocate resources and opportunities across a set of domains. Within each domain, the society chooses among three pure institutional types: a market (M), a bureaucratic organization (B), or a democratic mechanism (D).

For example, to select a construction firm to build roads, a society could hold an auction among qualified firms (M), it could construct a bureaucracy that develops criteria for selecting a firm (B), or it could hold a vote among elected representatives for the winner (D).

If there exist ten such domains, then the set {X} consists of all vectors of length ten whose entries belong to the set {M,B,D}. Though a simplified characterization of the world, the model allows for a combinatorial explosion of social arrangements—59,049 distinct possibilities to be precise.

Assume that the ideal arrangement from this set, MMMBBBDDDD, is known. As a benchmark, set its social justice value equal to 100. Consider the following two alternative arrangements with their respective social justice values:

Alternative x: MMMBBBDDMB Social Justice Value 80
Alternative y: MMDMMMBBBB Social Justice Value 90

Let the proximity between two alternatives equal the number of institutions on which they disagree. Given this convention, a lower proximity score corresponds to a nearer alternative. The proximity between x and the ideal equals 8, and the proximity between y and the ideal equals 2. Alternative x is more proximate than y to the ideal, but it produces lower social justice than y.

When this occurs, when the nearby alternative that most increases social justice has lower proximity to the ideal, society faces what Gaus calls *The Choice*. Does society increase justice or does it move in the direction of the ideal? In more mathematical terms, do we move vertically toward the ideal (by pursuing better outcomes) or horizontally (by following our moral compass)?

The first question we must address is how *The Choice* could arise. It could be that more proximate social worlds(i.e. movements in the direction of the ideal) always increase social justice. In that case, the social justice function creates a single peaked, Mt Fuji landscape. As we move toward the ideal, social justice values increase. And as we move away, values decline. On a Mt Fuji landscape, society avoids having to make *The Choice*. It can reach the ideal by climbing the social justice gradient.

The Choice only arises when less proximate social worlds have higher values. This property holds whenever there exist non-ideal alternatives with the property that all nearby alternatives have lower values. Metaphorically, these local optima can be thought of as peaks on a rugged landscape. Thus, rugged landscapes imply *The Choice*. This is an important insight.

The Choice imposes only temporary losses if we know the position and value of the social ideal. We must endure the path through the Sierras to reach the gold of California.

Here is where Gaus's assumption of myopia enters. In the Hong-Page model, the landscape values are knowable, by that I mean that given an element of the domain, we know or can determine its value. For instance, if we plug in a route for the traveling salesperson to follow, we can compute its distance using algebra. Gaus finds this assumption problematic in his context. We cannot toss out a novel reconfiguration of society and know its value. I agree.

To make the myopia assumption formal, Gaus elaborates on the Hong-Page framework as follows: He adds a set of *evaluative standards* (ES), a set of relevant *world features* (WF), a *mapping* that applies the evaluative standards to the world features (MP), and a *similarity ordering* (SO) along with a corresponding *distance measure* (DM) over the possible social worlds.

This more detailed construction enables him to introduce the concept of a *neighboring social world*. In the toy, institutional model introduced earlier, the *neighbors* of a collection of institutional choices might consist of all choices that differ by a single institution.

Figure 1 shows all twenty-seven possible institutional arrangements in a three-domain world.

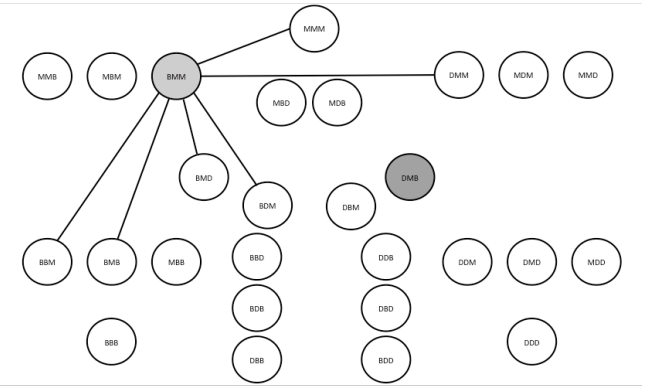

Figure 1: A Three Domain World

In the figure, the ideal arrangement, DMB, is shaded dark grey. The status quo arrangement, BMM, is shaded lighter grey. The lines extended from BMM connect to its six neighbors none of which is the ideal arrangement.

Three possibilities exist. First, there could be a neighboring alternative, say DMM, that improves on BMM. If so, given that DMB is a neighbor of DMM, climbing the gradient would lead to the ideal. *The Choice* would not occur. Moving up would also move toward. Second, it might be that a more distant alternative BBM is the best choice among the neighbors. Moving to BBM would improve social justice but be less proximate to the ideal. That would imply a rugged landscape and *The Choice*.

Finally, the status quo BMM could be a *local optimum*. This would involve a weak form of *The Choice*. Society could move toward the ideal and receive lower social value or stay put.

The key insight, which bears repeating, concerns conflicts between the two measures of distance: an arrangement can be closer to the social ideal but result in less justice, or be further from the social ideal yet increase justice. Both correspond to ruggedness on the landscape.

Gaus does not delve into the micro-foundations of ruggedness. I do so here given the logical necessity of landscape ruggedness for his argument. Establishing micro-foundations requires greater realism. Any number of examples could be constructed. Here, I consider here the {M,B,D} construction applied to the following three university resource allocation problems: assigning professors to classes, allocating revenue among professors for salaries, and registering students to the class offerings. Thus, the arrangement DMB corresponds to a world in which professors vote on teaching assignments, salaries are determined by the market, and students choose classes by fixed bureaucratic rules.

If making the institutional choice associated with the ideal improved any arrangement, then the landscape would be equivalent to a Mt Fuji. Landscape ruggedness requires that in some cases making the "ideal" choice on one of the domains lowers the value of the outcome

Here I describe how information synergies could cause ruggedness. By that I mean, information produced by one institution could benefit another institution. Suppose, for example, that DMB is the ideal social arrangement, and that we start from the arrangement BMM. In that arrangement, teaching assignments are made by fixed bureaucratic rules, salaries are determined by an external market, and students bid for classes using fixed budgets of university created currency.

In BMM, the market for classes among students produces a market clearing price for each faculty class pairing. These prices reveal information about student satisfaction with professorial assignments

DMM assigns teachers to classes democratically, endowing each faculty member with equal power and input. Democratic institutions do not always make optimal choices. They can produce voting blocs and vote trading.

Given that in the current arrangement, BMM, the bureaucracy can exploit the information revealed by the prices created by students in the classroom market, a bureaucracy might make more efficient assignments than a democracy. In other words, given a market for classes, a "non-ideal" choice of institution for allocating teachers may be more just. The informational synergy means that BMM, though less proximate to the ideal than DMM, could produce higher social value.

If instead, the students do not compete in a market for classes, then they no longer produce the information the bureaucracy needs to make efficient allocations. Absent that information, the bureaucracy likely makes poor decisions, and democracy, warts and all, becomes the better alternative. Thus, the information created by one institution creates a synergy with another, producing ruggedness.

Information is not the only possible cause of ruggedness. Behavioral synergies across institutional pairings can also produce local peaks and create conditions for *The Choice*. If individuals interact primarily in market mechanisms, they may develop repertoires of behaviors well suited to markets in other domains (Bednar et al 2007, 2015, 2018).

Those behaviors may both give some market institutions a leg up and hinder the performance of other institutions. For instance, self-interested behaviors that lead to good outcomes in a market may perform poorly in a bureaucracy.

To take this logic to its extreme, self-reinforcing behavior could make all three homogenous institutional arrangements DDD, MMM, and BBB local optima (Greif and Laitin 2004). That would mean that landscape would have at least three peaks, creating multiple instantiations of *The Choice*.

Third, collections of institutions that function as ensembles can create ruggedness. This is true within a federal system, where institutions have overlapping jurisdictions. In robust federations, institutions complement one another by providing safeguards against transgressions and overreach (Bednar 2009). Thus, the optimal design of any one institution depends on the designs of others. Those interdependencies would also result in a rugged landscape. Of course, capturing this type of ruggedness would require a more elaborate model than the stark {M,B,D} framework.

KAUFFMAN'S NK MODEL

The institutional model that I have just described bears a strong resemblance to Kauffman's NK rugged landscape model based on binary strings (Kauffman and Weinberger 1989). In the NK model, the parameter N corresponds to the number of binary attributes, and the K corresponds to the number of interactions between each attribute and others. The value of a binary string equals the sum of the contributions of the attributes. The contribution of an attribute, in turn, depends on the state of the attribute (either on or off) as well as the states of K other randomly chosen attributes.

We can use the NK model to link interactions to ruggedness. If K=0, each attribute contributes independently. The landscape forms an N dimensional Mt Fuji. Climbing locates the ideal. As K increases, it makes the landscape rugged. In the extreme case where K=N-1, on average one out of every N alternatives will be a local optimum. We can therefore think of the parameter K as a ruggedness dial. The

analog of K within the institutional model would be the number of institutional interactions resulting from informational spillovers, behavioral spillovers, or institutional complementarities.

We can now connect two dots. Kauffman showed that increasing these interactions adds ruggedness. Gaus shows that ruggedness creates difficulties for the pursuit of the ideal. Interactions therefore cause the tyranny of the ideal.

We might then think that we should eradicate these interactions as they make optimization difficult for Kauffman and result in *The Choice* for Gaus. That would be a mistake. Interactions are a desirable feature of our social world. Without interactions, the whole cannot exceed the parts. In Kauffman's model the value of the ideal, the peak on the landscape, occurs at an intermediate value of K. In Bednar's model, interactions produce robustness. As a rule, interactions create the potential for great things. Of course, we may not be able to find that best among them. On a rugged landscape, we could get stuck on a local peak.

THE DIVERSITY SOLUTION

Hong and Page propose a solution. They show how a collection of problem solvers with different perspectives and heuristics can smooth a rugged landscape by reducing the set of local optima. Landemore (2013) interprets and extends their model within the context of political philosophy.

To see how this can happen we need to return to figure 1. This perspective encodes the alternatives as triples of institutional choices. The proximity between two alternatives corresponds to the number of different institutional choices. The lines from BMM define the set of neighbors of proximity one. We can equivalently think of these as the alternatives located by the heuristic: *search all neighbors within proximity one*.

Figure 2 represents the alternatives using the same features—triples of institutional choices—but arranges them based on the number of bureaucratic institutions. This is a different perspective. The alternatives now belong to different neighborhoods. In this perspective, the neighborhood of BMM includes the social ideal BDM because each alternative includes a single bureaucratic institution.

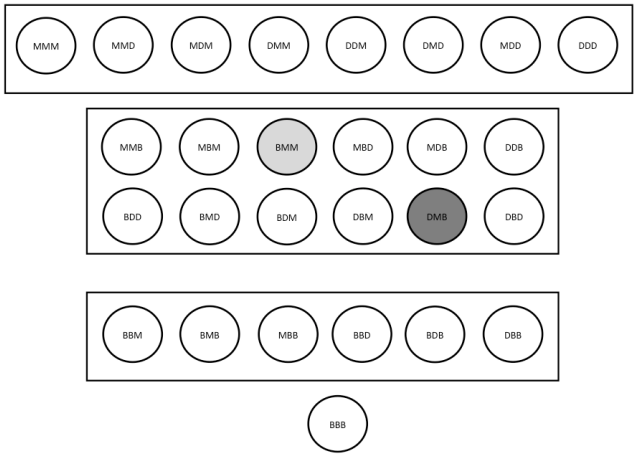

Figure 2. An Alternative Perspective on the Three Domain World

Different perspectives (and heuristics) produce different neighborhoods and therefore need not agree on their local optima. For instance, though the alternative BMM could be a local optimum in the first perspective, it cannot be in the second because the social ideal lies in BMM's neighborhood.

Given a group of people, their set of local optima equals the intersection of the individuals' sets of local optima. Herein lies the power of diversity. Hong and Page derive the following conclusion: *A sufficiently diverse group of people only sees the social ideal as a local optimum.* Any other alternative will not be a local optimum for some person. The logic is straightforward, for a solution to be a group local optimum, it must be a local optimum for every person.

We can show this logic using the two perspectives developed so far. Figure 3 shows sets of local optima for each perspective with grey shading. The intersection of the local optima for both perspectives are drawn thicker boundaries.

As we add new and diverse perspectives, we will eventually add one in which MDM lies in the neighborhood of BMD, the social ideal. Expanding on that insight, it follows that with sufficient diversity, at some point only the social ideal will be a local optimum for everyone. Through diversity, the landscape is made smooth. Society lands on the social ideal.

Gaus challenges that panglossian view by taking the position that if a proposed alternative lies a great distance from the status quo within someone's perspective, then that alternative will be difficult for the person to evaluate.

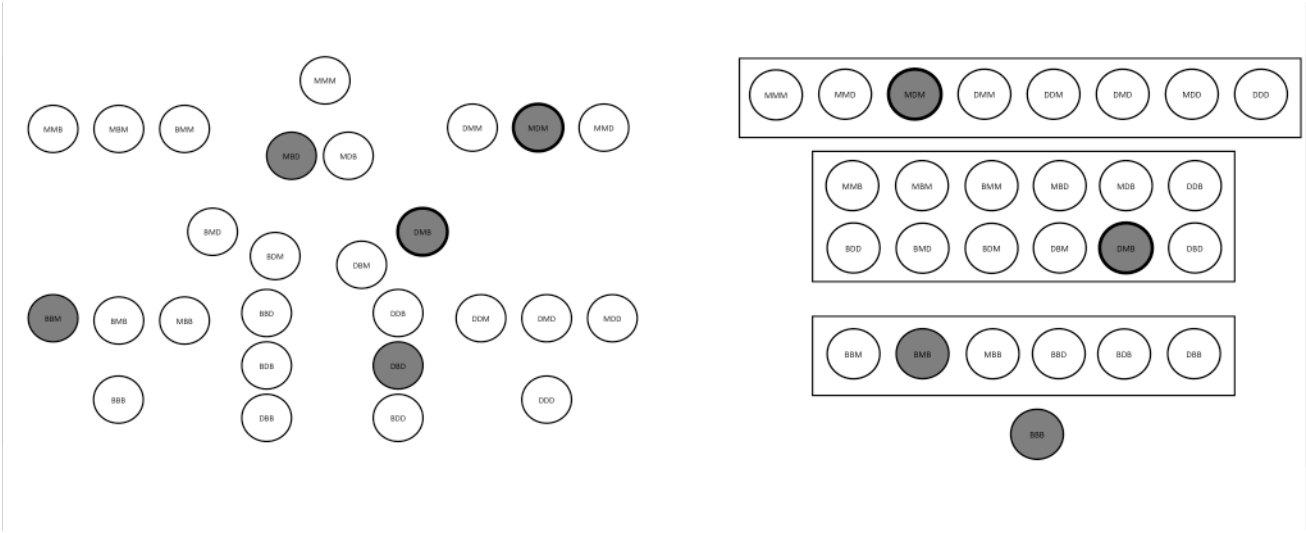

Figure 3.

Gaus formalizes this idea as the *Neighborhood Diversity Dilemma*: as the diversity of perspectives increases so too does the number of neighboring alternatives. An alternative viewed through any one perspective has a small number of neighbors. Thus, the large collective neighborhood necessarily includes difficult to evaluate, non proximate alternatives. We might rephrase the dilemma as follows: my friend's neighbor (in her perspective) need not be my neighbor (in my perspective).

WHAT DIVERSITY GIVETH IT TAKETH AWAY (PARTIALLY)

To appreciate Gaus's argument, we must first reinterpret the core insight of the Hong-Page model as follows: diverse perspectives allow a short step by one person to be a giant leap for another. In the perspective shown in Figure 2, DBD, and MMB belong to the same neighborhood: each arrangement includes a single bureaucracy. The movement between those two alternatives is a short step in that perspective. In the first perspective shown in figure 1, DBD and MMB differ on all three choices. They are not proximate.

Gaus argues that long distances preclude accurate evaluations of arrangements. In the example of the university, DBD corresponds to a world with democratically assigned classes, salaries determined by a bureaucracy, and students who vote among themselves to determine who can take which classes. If DBD represents the status quo, we can imagine that people within that university community would develop beliefs, behaviors, norms, and social networks suited to those institutional choices. For people operating in that environment, BMM, a world consisting of a bureaucratic process that allocates students to classes and market mechanisms that assign professors to classes and determine salaries would be difficult to imagine, let alone evaluate.

Hong and Page are not unaware of this potential problem. At least four workarounds have been proposed (Page 2007). As we work through them, we see that none overcome Gaus's critique.

First, there could exist an oracle that determines the value of any proposed alternative. The alternative, regardless of how far away it is, can be presented to the oracle, and the oracle will calculate its value. For a team writing computer code, designing an airfoil, or performing chemical reactions, oracles, or at least near oracles exist. For a society choosing among a collection of institutions, it surely does not.

Second, even without an oracle, if individuals can make accurate and independent pairwise comparisons, then by the logic of the Condorcet Jury Theorem, with a sufficiently large group, uphill can be distinguished from downhill. Given a choice between DBD and MBD, a diverse group would land on the correct answer most of the time.

Third, if each member of the group makes a crude evaluation of the social justice value of a proposed alternative, so long as the ways by which people make evaluations differ, i.e. if people rely on diverse predictive models, then by the *Diversity Prediction Theorem*, the crowd will be accurate.

In fact, even granting Gaus's claim that those people for whom the alternative lies a long way away will make inaccurate assessments, collective accuracy could be achieved by placing assigning weights on predictions that negatively correlate with distance from the status quo. That is, the prediction of someone for whom a proposed alternative is close will receive more weight than the prediction of someone for whom the alternative lies a long way away.

Fourth, no correlation between perspective proximity and ability to estimate value need hold. Imagine a Hollywood production team selecting a city in which to film a new sitcom. The original script penciled in Boston as a placeholder. The team now must make a choice.

One natural perspective for the set of US cities would be to represent each by its latitude and longitude. Given Boston, a person using this perspective might toss out Providence or Portland, Maine as alternatives. A second, equally natural perspective would be to arrange the cities in alphabetical order. A portion using alphabetic ordering might propose Boulder, Colorado or Berkeley, California. From the geographic perspective, these alternatives lie way outside the box. Though not proximate, neither would be difficult to evaluate.

This last example reveals a key distinction. To evaluate an alternative a person needs to know the mapping from attributes, what Gaus calls features, to the value. Boulder, Colorado could be an unexpected, out of the box proposal to someone, but it could still be evaluated accurately. Its lack of proximity causes no evaluative difficulties.

In contrast, consider the mapping from DNA to phenotype for some species. If we were to manipulate one or two genes with well-known functions, we might be able to predict the effects. If we were to do wholesale manipulations of large sections of the genome, we may have no idea what to expect.

Gaus's claim is that the problem of finding the social ideal is more like the mapping from DNA to phenotype. He has a strong case. We have little to no idea of how large scale reorganizations of society would play out. Given that, he is also correct on his second point. Unless we reside in a place darn close to the best of all possible worlds, we should be reluctant to move toward that ideal.

DIVERSE MODELS AND STEP SIZE

Gaus's offers a second best solution: an open, adaptive, diverse world that explores locally and abandons the ideal. Though pragmatic, his solution begs the question of how we achieve the appropriate diversity: How do we look far enough to avoid getting stuck but not so far as to head to a false ideal?

To clarify the problem, I again rely on the DMB institutional model. Here, I expand the number of domains to twenty creating roughly 3.5 billion institutional arrangements. Each of these arrangements has some social justice value determined by a function V, that is unknown.

Any function V can be assigned a *step size* that corresponds to the number of choices that must be coordinated during search to locate the social ideal.[1] If V has a step size of one, then it is a Mt Fuji problem. If the step size equals two, then the landscape has more than one peak, but any non-ideal peak can be escaped by changing two institutional choices simultaneously.

A value function's step size differs from its K value in Kauffman's NK model. Kauffman's K measures the size of interactions built into the function V. Step size measures the number of coordinated actions that must be taken to reach the ideal. A function could have a small K yet have a large step size. Alternatively, a function could have a large K, but if all interactions were positive, the landscape would be a Mt Fuji.

To see why step size matters, consider two scenarios. In scenario one, V has a step size of two. That means that to find the ideal, society must have the capability of accurately assessing alternatives of proximity two. With twenty domains, an alternative has twenty choose two, or 190 neighbors of proximity two. For society to ensure the ideal, it needs the capacity to think through that relatively large set of neighbors. That requires diversity of thought—people must be able to imagine all 190 alternatives. And it requires the capability to make accurate assessments of each.

In scenario two, V has a step size of four. That means there exists some institutional arrangement that requires four changes to escape. The number of possible changes of size four equals twenty choose four, or 4845. To ensure escaping local optima in that world requires greater diversity and more evaluative expertise.

Evolutionary systems solve this problem by relying on populations to search the space of possibilities. Ideally, exploration and exploitation find a balance appropriate to the problem's step size. On a Mt Fuji landscape, uphill changes in DNA quickly spread. On a more rugged landscape, a successful mutation for one string of DNA may not be for another. Thus, diversity will be maintained.

This tuning of search breadth to problem difficulty also occurs within simulated annealing algorithms which adjust

the neighborhood size to ruggedness. Optimal annealing algorithms search broad neighborhoods on rugged landscapes and narrow neighborhoods on smoother landscape.

Gaus's open society offers no such guarantee. What clues or signals might society get to adjust the breadth of exploration? Even more troubling, on more rugged landscapes, society needs to evaluate longer leaps. In practice, the opposite should be true: the smoother the landscape the easier to extrapolate multiple changes.

Gaus offers no escape from the ruggedness dilemma—the fact that more rugged problems require the ability to understand a larger neighborhood (Gaus's *Diversity Dilemma)*, which, given the increased ruggedness, will be more difficult to do.

Though I have only described the ruggedness dilemma within the institutional model, it also arises when constructing a social contract. Iterative improvements in the social contract demand new, diverse ways of thinking.

If we imagine those iterations as requiring a bargain *and* if we allow for a diversity of categories, then Muldoon (2017) offers a possible solution. He assumes people see projections of a higher dimensional world. The result of the bargain would be the union of their proposed emendations. The union of three steps of size two that intersect on a single change creates a step of size four.[2] Once again, diversity, along with a large dose of tolerance, comes to the rescue.

THE IMPERATIVE OF COMPLEXITY

My analysis so far has considered a fixed value function. That assumes that the world in which we must construct a social contract or develop an ensemble of ideal institutions remains unchanged. Advances in science and technology along with population growth, the emergence of new nation states, and climatic changes deny that assumption.

The landscape does not remain fixed. It dances. It shifts under our feet. Yesterday's ideal may well lie in tomorrow's trough. If the dancing landscape maintains the same ruggedness, that is if we imagine that the step size stays the same, then we can adjust once and for all our levels of diversity and openness and hope to keep climbing even as the sand shifts under our feet.

All evidence suggests that our world is becoming more complex. I mean that not in some metaphorical way but measurably. Our world is more diverse, more connected, more adaptive, and more interdependent (Page 2015). The landscape is dancing and adding new peaks with each twirl.

Gaus implicitly promotes diversity and sophistication as the imperatives of complexity, that we must become more open, more expansive in what we think possible and we must develop the capacity evaluate novel proposals. Muldoon would argue that we must be tolerant, more willing to bargain, and to cede changes that others demand that we fail to understand. Neither can derive much value from the ideal.

Their critiques of the ideal assume the landscape to be fixed. When the landscape dances, the ideal becomes even less relevant. I am inclined to agree. Rather than pursue an ever moving target, we should devote our efforts to developing diverse ways of discovering near improvements.

NOTES

1 What I call step size here, I call *cover size in P*age (1994)

2 One person wishes to change A and B, a second to change A and C, and a third to change A and D. The bargain results in A,B,C, and D all being changed. Given their diverse categorizations, each person only sees their own changes.

REFERENCES

Bednar, J. (2009). *The Robust Federation*. Cambridge: Cambridge University Press.

Bednar, J. and Page, S. E. (2007). Can Game(s) Theory Explain Culture? The Emergence of Cultural Behavior Within Multiple Games. *Rationality and Society* 19(1):65–97.

Bednar, J., Jones-Rooy, A., and Page, S. E. (2015). Choosing a Future based on the Past: Institutions, Behavior, and Path Dependence. *European Journal of Political Economy* forthcoming.

Bednar, Jenna and Scott E. Page 2018). "When Order Affects Performance: Culture, Behavioral Spillovers, and Institutional Path Dependence" *American Political Science Review.* 112(1)

Greif, A. and Laitin, D. D. (2004). A Theory of Endogenous Institutional Change. *American Political Science Review* 98(4):633–652.

Hong, L. and Page, S. E. (2001). Problem Solving by Heterogeneous Agents. *Journal of Economic Theory* 97:123-163.

Kauffman, S. A. and Weinberger, E. (1989). The Nk Model of Rugged Fitness Landscapes and Its Application to the Maturation of the Immune Response. *Journal of Theoretical Biology* 41, 211-245

Landemore, H. (2013). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many.* Princeton: Princeton University Press.

Landemore, H. and Page, S. E. (2015). Deliberation and Disagreement: Problem Solving, Prediction, and Positive Dissensus. *Philosophy, Politics, and Economics* 14(3), 2015: 229-254.

Muldoon, R. (2016). Social Contract Theory for a Diverse World: Beyond Tolerance. London and New York: Routledge.
Page, S. E. (1996). Two Measures of Difficulty. *Economic Theory* 8: 321-346.
—(2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.
— (2015). What Sociologists Should Know About Complexity. *Annual Review of Sociology,* 41:21-41.
Rawls, J. (1999). *The Law of Peoples,* Cambridge MA: Harvard University Press.

The Tyranny of a Metaphor

DAVID WIENS

Department of Political Science
University of California, San Diego
9500 Gilman Drive, Social Sciences Building 323, #0521
La Jolla, CA 92093-0521

Email: dwiens@ucsd.edu
Web: http://dwiens.ucsd.edu

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 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 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.

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, pp. 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?

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 incom-

plete. 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.

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, p. 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, pp. 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, pp. 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?

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 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.

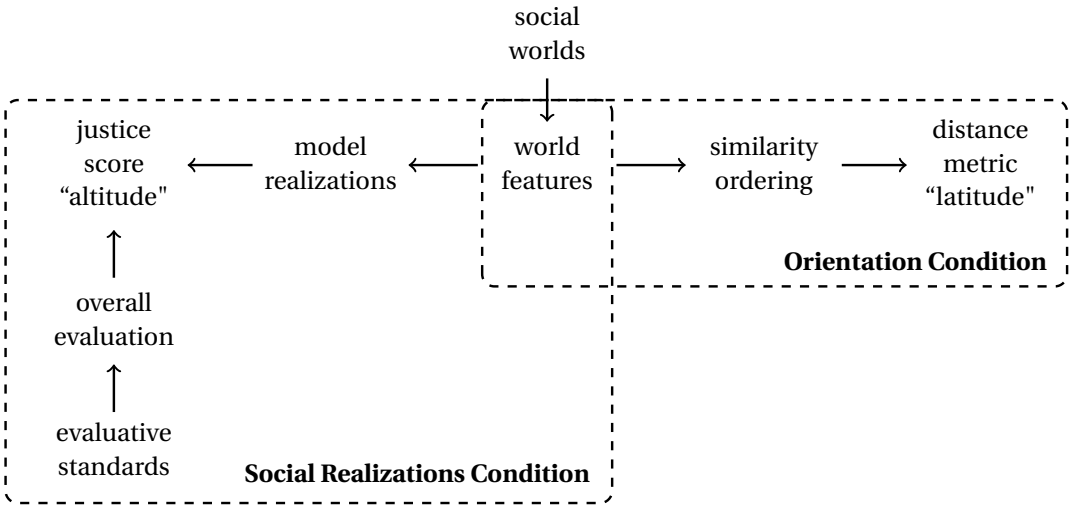

Figure 1. A schematic representation of Gaus's model

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 *descrip-*

tively 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, 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 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.

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 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 $X = \{x, y, z,...\}$ 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 $f : X \to \mathbb{R}$ if and only if R satisfies certain structural properties (e.g., if R is complete, transitive, and continuous).
- Let $K = \{k_1,..., k_n\}$ 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 $C(R, S) = \{x \in S|xRy$ 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 2015a, 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 op-

timal 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.

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 2015a). 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 2015a, 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 2015a).

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]

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 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.

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.

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". 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 a feature 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, people are assumed to have limited ability to predict the consequences of implementing social reforms, particularly 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 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 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 characteris-

tic 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 2015a, pp. 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 2015a,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 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.

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 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 forthcom-

ing). 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.[25]

NOTES

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.

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).

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.

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.

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' 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 comparative (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.

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.

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

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.) I 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 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.

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.

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.

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.

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.

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).

25 Thanks to Jerry Gaus, Brian Kogelmann, and Ryan Muldoon for helpful comments on earlier versions.

REFERENCES

Arrow, K. J. (1963). *Social Choice and Individual Values*. 2nd ed. New Haven: Yale University Press.

Barry, C. and Wiens, D. (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, M. (2012). The Aim of a Theory of Justice. *Ethical Theory and Moral Practice* 15:7–21.

Estlund, D. (2011). What Good Is It? Unrealistic Political Theory and the Value of Intellectual Work. *Analyse & Kritik* 2: 395–416.

— (2014). Utopophobia. *Philosophy & Public Affairs* 42(2):113–134.

— (2017). Prime Justice. In: *Political Utopias: Contemporary Debates*. ed. Kevin Vallier and Michael Weber. New York: Oxford University Press.

Farrelly, C. (2007). Justice in Ideal Theory: A Refutation. *Political Studies* 55(4):844–864.

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

Gilabert, P. (2012). Comparative Assessments of Justice, Political Feasibility, and Ideal Theory. *Ethical Theory & Moral Practice* 15(1):39–56.

Hamlin, A. and Stemplowska, Z. (2012). Theory, Ideal Theory and the Theory of Ideals. *Political Studies Review* 10:48–62.

Hausman, D. (1992). *The Inexact and Separate Science of Economics*. New York: Cambridge University Press.

Heath, J. (2017). On the Scalability of Cooperative Structures: Remarks on G. A. Cohen, *Why Not Socialism?* University of Toronto, unpublished manuscript.

Johnson, J. (2014). Models Among the Political Theorists. *American Journal of Political Science* 58(3):547–560.

Mills, C. W. (2005). 'Ideal Theory' as Ideology. *Hypatia* 20(3):165–184.

Muldoon, R. (2016). *Social Contract Theory for a Diverse World: Beyond Tolerance*. New York: Routledge.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Rawls, J. (1999). *A Theory of Justice*. 2nd ed. Cambridge, MA: Harvard University Press.

Schelling, T. C. (2006). *Micromotives and Macrobehavior*. London and New York: W.W. Norton & Co.

Schmidtz, D. (2011). Nonideal Theory: What It Is and What It Needs to Be. *Ethics* 121(4):772–796.

Sen, A. (1970). *Collective Choice and Social Welfare*. San Francisco: Holden-Day.

— (2006). What Do We Want From A Theory of Justice? *Journal of Philosophy* 103(5):215–238.

— (2009). *The Idea of Justice*. Cambridge, MA: Harvard University Press.

Simmons, A. J. (2010). Ideal and Nonideal Theory. *Philosophy & Public Affairs* 38(1):5– 36.

Stemplowska, Z. and Swift, A. (2012). Ideal and Nonideal Theory. In: *The Oxford Handbook of Political Philosophy*, ed. David Estlund. New York: Oxford University Press.

Swift, A (2008). The Value of Philosophy in Nonideal Circumstances. *Social Theory and Practice* 34(3):363–387.

Valentini, L (2011). A Paradigm Shift in Theorizing About Justice? A Critique of Sen. *Economics and Philosophy* 27:297–315.

— (2012). Ideal vs. Non-ideal Theory: A Conceptual Map. *Philosophy Compass* 7(9):654–664.

— (2017). On the Messy 'Utopophobia vs. Factophobia' Controversy: A Systematization and Assessment. In: Vallier and Weber (2017) pp. 11–35.

Vallier, K and Weber, M. (eds). (2017). *Political Utopias: Contemporary Debates*. New York: Oxford University Press.

Wiens, D. (2012). Prescribing Institutions Without Ideal Theory. *The Journal of Political Philosophy* 20(1):45–70.

— (2013). Demands of Justice, Feasible Alternatives, and the Need for Causal Analysis. *Ethical Theory & Moral Practice* 16(2):325–338

— (2015a). Against Ideal Guidance. *Journal of Politics* 77(2):433–446.

— (2015b). Political Ideals and the Feasibility Frontier. *Economics and Philosophy* 31(3):447–477.

— (2017). Will the Real Principles of Justice Please Stand Up? In: Vallier and Weber (2017).

How Can We do Political Philosophy?

FRED D'AGOSTINO

President of the Academic Board
Office of the President of the Academic Board
The University of Queeensland
Brisbane St Lucia, QLD 4072
Australia

Email: f.dagostino@uq.edu.au
Web: http://researchers.uq.edu.au/researcher/1345

The Tyranny of the Ideal, by Gerald Gaus (hereafter *TI*), is an examination, in two parts, of the prospects for a type of political philosophy that provides direction for reform while remaining in contact with the contingencies, diversities, and infirmities of the human condition.[1]

In the first, longer part, which is critical in intention, Gaus considers the viability of an approach, labelled *ideal theory,* that has been influential for millennia and concludes, to put it crudely, that this approach is misbegotten, in particular, because of the complexity of the relations among the congeries of factors that are relevant to defining a political ideal, including, in particular, the diversity of individuals' perspectives on their good and the good for their society. My comments on this part of the book will largely take the form of a reconstruction, via a different mode of presentation, of some of Gaus's main points. The key idea here will be that the problem of political philosophy is cast within a framework of *layered complexity.*

In the second part, which is constructive in intention, Gaus, first, offers an alternative approach, couched in terms of Karl Popper's idea of "the open society" as a privileged configuration of social arrangements that is designedly friendly to precisely the forms of diversity that, on Gaus's account, undermine the ideal theory approach, and, second, essays the prospects that such a configuration might be a point of convergence for individuals with (as it turns out, not quite) the full range of perspectives. One of my critical points will be about the prospects for "the open society" in current socio-political circumstances, characterized, as they seem increasingly to be, by various forms of polarization, fundamentalism, denialism, extremism, irrationalism, and rejectionism.[2]

These matters have an importance beyond the immediate argumentative dialectic, or so I believe. In particular, Gaus's approaches, both critical and constructive, point towards a new way of doing political philosophy, one that has been emerging in his own work and, increasingly, in the work of others, such as Ryan Muldoon's (2016) recent *Social Contract Theory for a Diverse World: Beyond Tolerance* and Julian Müller's forthcoming *Capitalizing on Political Disagreement: the Case for Polycentric Democracy.* So, as a way of framing *TI,* I begin by saying something about this emergent new program in political philosophy.

THE NEW PROGRAM IN POLITICAL PHILOSOPHY

To speak of a "new" program in political philosophy is to imply the existence of an existing or established or simply "old" program and that, of course, is the Rawlsian one, initiated, for all intents and purposes, by the 1971 publication of the book *A Theory of Justice,* which, along with its various successors, has been a point of focus for Anglosphere or more broadly "analytic" political philosophy ever since.[3] Indeed, both Gaus and I have contributed to that very program of research. But less reluctantly than others, or maybe it's just more recklessly, we have concluded that the Rawlsian project has failed, to adopt the terminology of footnote 1, to effectively manage "the essential tension" between the diversity of individuals' interests, values, "comprehensive conceptions of the good" and so on (all acknowledged by Rawls) and the necessity, as Rawls saw it at least (though we do not), that a public conception of justice provide a basis, grounded in a deep moral consensus, for the coordination of social action for mutual benefit. As Gaus puts it (*TI,* 153-4):

> Rawls insisted that a theory of justice was characterized by choice from a certain normalized perspective, but his later view allows multiple partially normalized perspectives that yield different conceptions of justice. However, if one acknowledges that there are other reasonable normalizations that yield inconsistent conceptions, in what sense can one plausibly claim that one has identified *the* principles of justice for the definitive ordering of social claims in a well-ordered society…?

Indeed, what seems to have happened is that Rawls abandoned, during the course of his long career, both the goal he set for political theory and the fundamental modelling device that he adopted as a basis for pursuing that goal. In particular, Rawls in effect abandoned the idea that political theory ought to and could successfully aim at the identification of a public conception of justice fit to order competing social claims. (This is the idea of the "well-ordered society".) And he also abandoned what had been a key, and highly arresting, heuristic for pursuing that goal, namely, the normalization of stakeholders, or, in particular, the abstraction, during theorizing, from their various differences, empirically—in their values priorities, in their social roles, in their "identities" (e.g. race, gender) etc. (This is the device of "original position" argumentation, where the original position is a unique privileged instance of a collection of initial situations of choice.) What we see here, in the reversal or abandonment of key features of Rawls's theoretical approach, is what one of my teachers, Imre Lakatos (1970), called a "degenerating research programme". So, as Gaus now puts it (*TI,* 153), "[o]ne has to be an especially devout disciple of Rawls not to conclude that by the close of his political liberalism project the theory of justice was in disarray."

Of course, Thomas Kuhn (1970, p. 151) already argued, indeed around the same time as *Theory* was published, that a well-entrenched approach to knowledge-making or theory-building would persist, even in the face of internal and external difficulties, so long as there were no reasonably well-delineated alternative to or replacement for it. Perhaps that explains the persistence of work on the Rawlsian project. But such an alternative is, arguably, now coming to hand, in the form of a new approach that Gaus has played a conspicuous role in developing.

One feature of *the new program* is its treatment of the empirical diversity of ethico-political stakeholders, which it simply accepts as a given and works with without any very elaborate abstraction or idealization in the interests of homogenization. Indeed, Gaus and others (including Muldoon, Müller, and me) all see precisely this diversity as a resource for the organization of social life, providing, as it does, the basis for a division of labor in the economy and for the better understanding of partially shared values and cognitive perspectives (through their encounters with different variants on common themes). So, whereas Rawls starts with *stakeholders are different so let's normalize their differences in the interests of constructing a consensus on the principles of social coordination*, the new program starts with *stakeholders are different so finding principles of social coordination will depend on finding points of convergence from an accepted diversity of unnormalized perspectives.*

Indeed, this way of putting the matter suggests a second conspicuous feature of the new program, namely its preference for convergence rather than consensus approaches to social coordination. This represents a real sticking-point for many participants in the Rawlsian project, who worry that the lack of a shared substantive basis for coordination is ipso facto lacking in normative authority or stability, and, in fairness, this issue does constitute a challenge for the new program that ought to be, indeed already to some extent is, on its to-do list (or, as Lakatos would put it, is part of its "positive heuristic"). See, for example, Kevin Vallier's (2014) recent book *Liberal Politics and Public Faith.* Because of the emphasis on convergence rather than consensus, we can expect, and indeed find, that sponsors of the new program put emphasis on argumentative or more broadly justificatory mechanisms that are convergence-compatible, such as

- bargaining, which, as Muldoon, following Hayek, points out, "has the advantage of requiring very little agreement between the parties involved" (Op. cit., p. 69);
- path-dependent social-evolutionary processes, as sketched by Gaus, for example, which can discover and stabilize a coordinating equilibrium even in the presence of considerable divergence of underlying motivations among participants;[4]
- reconciliation by separation or the recognition of separate spheres, as for example in systems of so-called jurisdictional rights, as analysed by Gaus (*TI*, IV.2.4);
- the appeal to the gains to trade, broadly understood, that are available via a division of labor built on diverse perspectives, as for instance, in an economistic sense in Muldoon's analysis (Op. cit., sec. 5.3) and more broadly by Gaus, especially in the form of his "Fundamental Diversity Insight" (*TI*, 133), and including in particular,

those gains resulting from Millian "experiments in living",[5] a concept which has become much more common in the political-philosophical literature of recent times.

A third important feature of the new program is its general understanding of the political-philosophical enterprise. In particular, the new program adopts what I call an "engineering" approach to political philosophy, seeing the goal of the enterprise as the design of devices or technologies (see *TI*, 183) that are fit for the purposes of ethico-political stakeholders in their pursuit of workable social arrangements. There is convergence from seemingly unrelated spheres of philosophical enquiry on this understanding, including, for instance, and indeed as an especially vivid example, the recent work of Elijah Millgram, who characterizes his own discipline, metaphysics, as "a design science (like architecture, or computer science, or mechanical engineering)" and who explains, at greater length, why we might better understand certain philosophical enterprises "as intellectual ergonomics". Millgram (2015, p. 15) says:

> Throughout its history… metaphysics has been the answer to the question: how do we have to understand the world for reasoning about it to be possible? There is a *practical* spin to put on the question: how can we make reasoning and inference tractable and effective? One approach to the question so understood will focus on the design—and on the redesign—of intellectual devices that make it feasible to think about the world through which we must navigate ourselves. If that is what we are interested in, when we are doing metaphysics, then we should be attending not in the first place to what expressions *mean,* but to what the devices *do.*

(Others, including Muldoon and Gaus's student Chad van Schoelandt (2015) have all explicitly endorsed this approach.)

Perhaps most notably, though this point remains to be fully explicated, the new program more or less abandons the project of providing an end-state description of a justified social order, preferring, instead, and entirely in keeping with the engineering conception, to see its job as that of identifying a *method* of (i.e. tools for) thinking about the problems of political and social order,[6] rather than of deriving or otherwise justifying descriptive statements (of normative intent) about that order. While it cannot be precluded (though it also can't be guaranteed) that the new approach, carefully applied, will provide some guidance to stakeholders, what it will not provide is a substantive account of what their social arrangements ought to be. That account will emerge from *their* activities, perhaps using the tools identified by the philosopher. Political philosophy, on this account, lets go of what the sociologist Zygmunt Bauman (1987) calls the "legislative" aspiration and cedes the legitimation of their social arrangements to the stakeholders who make those arrangements. And, although this is radically overdetermined, this is so, at least in part, because, unlike the situation in other forms of knowledge-making such as science, there is, despite a charming faith to the contrary widespread in the philosophical community, no external source or standard against which to measure the success of a process of reasoning or evolution in delivering a legitimate system of social arrangements. All that matters, really, is how that arrangement came into being… and that is what a theory of method in political philosophy can give us.[7]

So much for the circumstances in which *TI* might be understood. How about its main arguments? Let me begin with the critical or negative part of Gaus's development of ideas.

THE ENIGMA OF IDEAL THEORY

I have already mentioned the idea that political philosophy might aspire, indeed has typically been seen to aspire, to the derivation or justification of a description of an ideal social state. For political philosophy to be practical, even in a restricted sense that does not entail an engineering approach, the notion of an ideal social state has to include, at some level of granularity, a description of the specific institutional and cultural forms that sustain the realisation of, if you will, the social ideal. And the social ideal will be given, in turn, by some account of the ways in which, at least in typical situations and perhaps at a high level of abstraction, the various social goods are balanced optimally against one another (so that, crudely, there is no other balancing which would deliver net-gains in some appropriate way).

This is *ideal theory,* then, and, in the first 150 or so pages of *TI,* Gaus undertakes to expose its pretensions. I should say, at the outset, that I found Gaus's critical points entirely persuasive and have, as far as I can see, no corrections or amendments to suggest. What I would like to do, however, is present an account from a different perspective that will track Gaus's at least in its more important aspects. Perhaps this account will open up some points for my readers, as it

did for me. Let me begin by describing a model of *layered complexity* in relation to social order. (I note that this model in my view survives, in its main features, any critique of the pretensions of ideal theory and hence remains available for appropriation by advocates of the new program.)

At the base, we have a set **A** of arrays of social institutions, $\{A_1, A_2,\ldots, A_n\}$, where each array A_i consists of a system of specific institutional and cultural elements $<I^i_1, I^i_2,\ldots I^i_n>$, where the relations among the elements within a given array vary from loosely to tightly coupled.[8] Relations between different arrays are relations of similarity and difference, so that, for example, one array might differ in one particular institutional or cultural element from another and hence could be considered quite similar and indeed perhaps as quite accessible if we were thinking of social transformations. (It will be part of Gaus's argumentative strategy to assume, *arguendo* and as ideal theorists probably must assume, that we can define a similarity metric Σ over the set **A**. It will, of course, be a partial ordering rather than a complete one.)

One of the things we know about such arrays, when they are even moderately "realistic" in relation to our own social circumstances, is that it can be very difficult to understand all the relations between or among elements. Given coupling, how does changing one element affect others in the array? How, because of the complexity of the interrelations among elements, does changing one element affect the configuration of the array as a whole? We know the difficulty of these questions, purely empirical though they be, through the concept of "unintended consequences", already familiar to Adam Smith, but put into play in our own time by the sociologist Robert Merton (1936). Because or insofar as the relations among elements of an array are obscure or complex, changing an element can have consequences that we cannot predict and hence will not anticipate.

I have, on other occasions (D'Agostino 2010, esp. Ch. 7), written about complexity and so won't repeat the exposition, but the key point is that any given element in an array might be coupled with others so that changing it changes them (or vice versa) and not always in smoothly linear ways. For example, changing some element might have different kinds of effects on different elements with which it is coupled. Perhaps increasing the stringency of some social rule reduces a certain form of associated activity while also increasing another, also associated activity. All this is purely empirical, how changes to one element affect others… it is a matter for observationally informed social science, difficult, probably, to model in a purely theoretical sense, except, perhaps, by simulations (perhaps involving agent-based modelling).[9]

The base level or first layer consists, then, of a set of arrays, each of which is a complexly organized collection of social institutions and cultural elements.

At the "top" of the layered complexity model is an evaluative instrument which identifies a number of metrics, μ_1, $\mu_2,\ldots, \mu_n$, against which the various arrays can be assessed.[10] Again, these metrics will be, variously, tightly and loosely coupled with one another, so if we are at all "realistic" relative to what we know of our own systems of evaluation, it will be characteristic that the various metrics will interact complexly with one another, so that changes (in institutional arrangements, of course) which increase value against one of these measures may well decrease value against others of them. (A change to a legal rule may increase liberty but decrease equality because of the way the effects of that change propagate through an existing system of social institutions.)[11] We may also, of course, have what I will call a reconciliation $\prod$ of these various metrics, trading them off against one another in some appropriate way that yields an *overall* assessment of the given arrays. And, given the complex coupling of metrics with one another, we can expect overall assessments to exhibit the usual nonlinearities.

For example, starting with the array A_p, perhaps we believe, even rightly believe, that changing the element I^p_j will increase the value of the resulting array against the measure μ_i. Given the way the metrics are coupled, this may not result, however, in an overall increase in value. Perhaps increases against μ_i are coupled with decreases against μ_j such that, because of μ_i/μ_j trade-offs, the $\prod$ value of the new array is lower than of the starting-point array. That is one complexity. But there is a second and it's internal to the empirical layer in our scheme. Remember that changing I^p_j may change other elements in the given array, perhaps, in this case, I^p_k. But now, even without the complexities at the evaluative layer, evaluating the effects of the I^p_j change are more obscure because any evaluation has to identify the I^p_k changes as well and consider how those changes affect the μ_i value of the array, and, of course, it cannot be guaranteed, in the face of complexity, that there is overall increase in μ_i value, given that the positive I^p_j changes might well be outweighed by negative I^p_k changes.

Returning to the idea of an ideal theory, we can see now, I think, that it will, just as Gaus proposes using other terminology, have two key elements… though notice how these must be linked. One element is the description, in terms of institutional elements, of the ideal social state, defined, of

course, as that member of the array **A**, A* for which the ∏ value, π* is maximized. To make this theory practical, we rely on the similarity measure Σ, so that we can determine, for arbitrary A_i, how distant it is from A* or, more crudely, how many changes to institutional elements, are required to transform A_i into A*.

Such a theory would tell us, then, what the ideal social state is like and would, at least in principle, imply a pathway between our existing social situation and that ideal social state, for example, by a series of small changes to the institutional elements, one by one, until all the differences between A_i and A* has been eliminated.[12] (Crucially, because of the nonlinearities, there may be no clear and consistent pathway from A_i to A*; the first step in a given direction may be value-increasing and distance-decreasing, whereas the next step, in the same direction, might be, again, distance-decreasing but value-decreasing.)

In calling this section of my paper "The *Enigma* of Ideal Theory", I meant to refer, in fact, to the notorious German war-time Enigma coding machine, the one whose workings were, according to the mythology, unpacked by Alan Turing. And I chose this locution because, in fact, the relation between a change in some institutional element and the overall evaluation of the resulting array is as obscure, in most reasonably "realistic" cases, as the relation, in the Enigma machine, between the letters typed in and the letters that were then lit up on the board, having passed through various enciphering devices within the machine. And that, in essence, is Gaus's argument about ideal theory. Because of the two complexities identified above (among others), there is no realistic prospect that we will be able to use the similarity measure as a guide to realising the ideal social state… the array whose overall value is at a maximum, the array that best realises the social ideal. Using the idea of "high-dimensional landscapes" to refer to what I've called the complexity of couplings, Gaus puts it (*TI,* 68-9) like this:

> In terms of our ideal theory model, in a maximally high-dimensional landscape there is no systematic relation between the justice of social world *i* [my A_i] and the justice of the worlds that are adjacent to it. Note that in such a landscape there is no point in getting *close* to the ideal point, *u* [my A*], but not achieving it: its near neighbors may not be at all just.… The crux of maximally high-dimensional landscapes is that the justice of any one rule or institution is a function of all others, producing what Kauffman called "a complexity catastrophe."… [O]ur concern here is a political theory that seeks to judge the justice of various social worlds, and recommends moves based on its evaluations of these worlds. In this context, the idea of a complexity catastrophe is entirely apropos, for the system will be too complex—really chaotic—for the theory to generate helpful judgments and recommendations.

While there are other, indeed fascinating, bells and whistles, this is the crux of Gaus's critical analysis of the situation of political philosophy. If political philosophy is defined by an ambition to develop an ideal theory that meets certain reasonable criteria, then it is bound to fail and, accordingly (*TI,* 245-6), there is need for "a break with contemporary social and political thought", because the idea on which it is based, that of ideal theory, "is ultimately a mirage, yet one that tyrannizes over our thinking and encourages us to turn our backs on pressing problems of justice in our own neighbourhood."

Of course, it is one thing to undermine the legitimacy of the project of ideal theory and quite another to reject the model of layered complexity that in fact enables that demonstration to be undertaken in a rigorous way. While it may not be possible, any longer, to take seriously the idea that we can use the similarity measure Σ to set an unwavering course towards the ideal social state, A*, that we can take with confidence,[13] it is still possible, indeed highly desirable, to use the apparatus of empirical and evaluative layers as an aid to theoretical enquiry about the project of political philosophy. And, in particular, one of the points that emerges quite clearly from the contemplation of this model is that it will be highly relevant to any respectable project of political philosophy to learn more about the variety of institutional arrays, the relations among institutional elements that are characteristic of the more common arrays, the issues around transformation of one array into another, and so on. All these are empirical questions, not to be answered simply by conceptual analysis, although they might well be modelled, for instance through simulation exercises. Using Lakatos's terminology, the injunction to study these arrays, for instance through institutionalist enquiries in political science,[14] is a reasonable element of the "positive heuristic" for the new program in political philosophy. Insofar as the PPE project—the project of recognizing the complementarities among the three key disciplines of politics, philosophy and economics—is now reasonably well consolidated,[15] it perhaps could be extended by the more systematic inclusion

of institutionalist analyses, coupled, perhaps, with studies in political sociology and indeed political social psychology.[16]

I said, earlier, that there was a positive, as well as the reported negative part to Gaus's analysis. Let me turn, briefly, to that account, framed, as it is, largely in terms of Popper's idea of "the open society", another one of Gaus's targets for intellectual rehabilitation.

THE OPEN SOCIETY AND ITS EXTREMITIES

If the critical argument of *TI* establishes that ideal theory is not viable as a paradigm for political-philosophical enquiry, then the positive argument is meant to provide an alternative approach, one that, as indicated earlier, starts from the inescapable and indeed potentially beneficial fact of diversity and seeks, via an engineering approach, to establish the possibility of, and conditions for, a convergence on non-trivial principles of institutional design. This takes the form of what Gaus, following Popper, calls "the open society", or, in particular, a social state that is characterized by certain key features that enables its stakeholders to, as Gaus puts it (*TI,* 176), "live under a shared moral framework that is diversity accommodative because it is accommodative to diversity per se".

The open society, on Gaus's account, is not primarily characterized by what we might call fine-grained, descriptively rich, action-guiding normative injunctions. Because there is no ideal social state (the critical part of the argument), there can be no legitimate array of injunctions (and other cultural elements) of this particular kind. (Recall that the so-called new program abjures such a "legislative approach"; the failure of ideal theory is a reason.) What there might be, however, is a collection of (meta-level) principles that will provide design criteria for any array of concrete institutions that can claim to exemplify the open society. The new program, then, consistently with its proceduralism, and with its orientation to methodology rather than theory, delineates a set of abstract principles that will have to be instantiated (though in any of a variety of ways) in order to be "diversity accommodative" in a society which cannot agree about the social ideal, but which still seeks to obtain the benefits of mutuality. This is, I believe, what Gaus calls (*TI,* 149) "a moral, liberal framework… which abjures the pursuit of the ideal while providing a framework for diverse individual perspectives on justice."

And the key meta-level design principles for such an open society include, according to Gaus, at least the following:

- That there be opportunities for an individual to interact (a) with others who share their particular interpretation of the social ideal within a "republican" sub-community (*TI,* 146-7) and (b) with those who have different interpretations of the social ideal, typically via their shared interest, despite these differences, in some common practical problem (*TI,* 185).[17]
- That the substantive normative injunctions, whatever they might turn out to be, be understood, by all relevant parties, via what Gaus calls (*TI,* 187) "the principle of natural liberty", or, in other words, the meta-level interpretive principle that "[w]hatever is not prohibited… is permitted", on the grounds, specifically (*TI,* 196) that "a natural liberty system encourages experimentation and discovery" and hence is diversity accommodative.
- That prohibitions rather than permissions be the default form for specific normative injunctions (*TI,* IV.2.3.3) on the grounds that, in conjunction with the principle of natural liberty as a "closure principle", this will, again, encourage discovery and innovation because (*TI,* 196) "[m]oral experimenters… need not first convince themselves that a new action falls under a previous permission".
- That the system of concrete normative injunctions makes appropriate use of so-called "jurisdictional rights", which (*TI,* 199, 200), for participants in an open society, "decouple the perspectives and so lessen the complexity of the system", thus "allowing high levels of change in some perspectives without affecting the shared public world".
- That market transactions and exchange more generally be a primary modality of social coordination, since, as Gaus puts it (*TI,* 203), they "provide bridges between different perspectives [because although] each sees the object in different ways… [they] typically share enough so that… they can agree on what is being traded, and that each is better off".
- That specifically legal rules be relied on (*TI,* 207) "to provide shared classifications of prohibited behaviour among those with, essentially, different perspectives", thus permitting coordination of expectations and behaviour among a diversity of individual stakeholders.

As you can see, these are, as foreshadowed, meta-level design principles, rather than specific and concrete social ideals or behavioural norms. If we undertake to deliver something positive by way of political philosophy, it will take this form. If we imagine ourselves considering how to create an open society, in the absence of, and indeed on ac-

count of the absence of an agreed social ideal (with its corresponding ideal social state), then we should, on Gaus's account, do so by seeking, in particular concrete circumstances, to identify or create an array of social institutions which meets these meta-level requirements. Indeed, reverting to our set of arrays **A,** we see, I think, that these design principles will impose a partition on that set between the arrays which do and those which do not honor them. While this does not tell us, concretely, what to do to improve our social situation, it does tell us how to begin… by moving, from wherever we are now, towards an array in which there are jurisdictional rights, exchange relations, and a principle of natural liberty that provides a closure rule over specific prohibitions. Of course, because of the layered complexities considered earlier, our pathway may not be a uniformly smooth one, and we will want to proceed, as Popper recommends under the heading of "piecemeal social engineering" (Popper 1962), by small and ideally reversible steps.

Of course, in speaking of these meta-level principles as *design* principles, and, indeed, in speaking, earlier, of the *engineering* approach, I seem to commit myself to a position which, actually, I do not accept and which Gaus doesn't either. So, in particular, and to avert to the not entirely coincidental title of the journal in which this article appears, I seem to be presupposing that, in our political philosophizing, we are thinking about the characteristics of a made order, a taxis in Hayek's terminology, rather than a cosmos, or grown order.[18] I am not. Sometimes, of course, engineers look to the products of untheorized practice to identify principles that can be articulated, systematised, and then rigorously tested. These design principles might well be, on this account, the deliverances of a reflective examination, by political philosophy, of the various grown orders revealed by history and political sociology. To develop some more sophisticated account of the relation between so-called design principles and grown orders is, I believe, a project for the new program, and one worthy of endorsement by the positive heuristic of that program.

Let me conclude by way of Gaus's forthright approach to what he calls (*TI,* 208) "the critical question of justification", one which is especially acute because, absent the impossible dream of an ideal social state, there are a diversity of perspectives on the social and because justification consists, in the broad framework in which Gaus works, in finding some surface of convergence (on the open society) from all (or most) of these various perspectives.

In fact, as Gaus is quick to recognize (*TI,* 215), whether there *is* a surface of convergence depends more or less crucially on stakeholders "not insist[ing] on… the 'optimizing stance'—that the only rule that is acceptable is one's top-ranked rule." It is on this point that I want to focus, on account, specifically, of the prospect that more people are more inclined now to insist of precisely that, though not, perhaps, in exactly those terms (which are, after all, terms of art).

In particular, I believe that it is an open, sociological question—a question about the middle level in our model of layered complexity, the layer of social agents—whether individuals who differ in their interpretations of the social ideal, who would, upon reflection, offer different, especially opposed, accounts of the ideal social state are now able, in sufficient numbers, to recognize each other, nevertheless, as potential partners in a social contract to form an open society. This, I think, is the lesson of many recent political events: as evidenced especially in the social media (which have a known tendency to polarize judgment),[19] people are increasingly likely to treat those who adopt different social ideals as pariahs, as unworthy of moral regard. To put the matter crudely and in terms of the contemporary political situation in the United States, it is not merely that (cartoon) Red Staters have a different way of looking at the social world than (cartoon) Blue Staters. It is, rather, that at least some members of each of these two cohorts thinks of many members of the other cohort that they are not on an equal moral footing and are not worthy of moral regard and, indeed, are not the sorts of beings that you can bargain or deliberate with about jurisdictional rights or the benefits to each of the diversity of perspectives made available by the presence of the other.

And, indeed, Gaus is fully aware of this possibility and its implications, though he doesn't phrase the matter in the same way. He says (*TI,* 222):

> [I]n the end, some perspectives will conclude that even the most fundamental elements of the Open Society are worse that no moral constitution at all. Some perspectives are, in the end, unable to share a framework of moral accountability with diverse others. Even the Open Society must be prepared to normalize to some, hopefully to a very small extent. Such "Excluded Perspectives," which cannot find sufficient space in the Open Society, will almost surely be those that are committed to the optimizing stance, or some near approximation to it.… Such perspectives may live along with, but are not part of, the Open Society, treating

its rules as at best mere descriptive norms rather than moral injunctions.

The key question about this characterization is whether the *In*cluded Perspectives, and their adherents, are sufficiently numerous for it to be possible to form an Open Society even at the expense of excluding some other, "marginal" perspectives. It is not clear to me, as a purely empirical matter, what the relative dimensions now are in many of the so-called North Atlantic democracies of those individuals able and those unable to recognize that their ideals can't be insisted upon as a non-negotiable basis for the more broadly social ideals in their community. These, then, are the "extremities" of the playful heading of this section.

Of course, as Gaus reasonably remarks (*TI,* 222), "[t]here is no reason why we should leave ourselves at the mercy of those who refuse to live on terms that others can endorse." As a key task, the new program has, I believe, to consider (let this be an injunction of its positive heuristic) first of all how we can assess the relative balance of power, in our societies, between diversity-respecting and diversity-abhorring cohorts and, secondly, should the diversity-respecting be "outnumbered" by the diversity-abhorring, what we are, practically speaking, to do about that. There's no reason to suppose that these questions are going to be easy to answer. There is less reason to believe that even an answer will delineate a feasible program of social reform that will, if they are currently lacking, restore the preconditions for the open society. But the questions and issues are primarily empirical ones, both about the sociology and about the social psychology of diverse societies and this suggests a final positive heuristic injunction: We need to continue and extend the work of Cass Sunstein and others on the conditions under which such extremism can flourish. It, rather than bare diversity, may well represent the biggest challenge to the new program for political philosophy, and taking up this challenge will require expanding the domain of political philosophy to include other disciplines.

I called this paper "How can we do political philosophy?" With every new publication, Gaus is showing us how to do political philosophy in a way that is multi-disciplinary, diversity-respectful, and practically oriented. Most importantly, he has loosened the ties in his own work to established paradigms, partly by the rehabilitation of those left behind by the canonization of those paradigms (Popper on this occasion, P. F. Strawson in *The Order of Public Reason*), and partly by extending the tool-kit of concepts and modes of analysis that are drawn on in his work (as, for example, with his appropriation of game-theoretical modelling). In my view, this work nails the coffin shut on the project of ideal theory. And it delineates a rich field of enquiry with numerous open questions, characterized by me in terms of the injunctions they set up as part of the positive heuristic for the new program. In particular, the new program needs to consider extremism, polarisation, demonization of "the Other", and not just the "tamer" forms of diversity. Otherwise, it risks making the same misstep as, I think, Rawls made when he tried to control diversity, first by means of normalization, and then by limiting it to those differences that arise between "reasonable" conceptions. While some idealisation is unavoidable even if we are not practicing ideal theory, we can't abstract away from extremism if we hope for the new program to be morally relevant to our current situation. That is a challenge worthy of the times, and it is one, I am happy to report, that Gaus is fully alive to, as shown, for instance, in his recent article "The Open Society and Its Friends", online in *The Critique,*[20] which is especially valuable, in my view, because, unlike a lot of horrified commentary on recent political events (though Gaus too recognizes the dangers), Gaus's analysis foregoes the complacent comforts of marginalizing "the Other". That, surely, is the counsel of wisdom. If, as the engineering metaphor would suggest, we are to take seriously the problem of getting from where we are to someplace better, we'd better be reasonably "it is what it is" about where we are. Otherwise, we are engaged in fantasizing not just the end-point, but the starting-point on "The Road to Utopia". Gaus takes the path less (frequently) chosen, and that might yet make all the difference.

NOTES

1 Just as we speak, in scientific methodology, of the essential tension between the conditions that facilitate innovation and those that support tradition, so too might we speak, or so I have argued (D'Agostino 1996), of a tension, unavoidable if not "essential", between remaining in theoretical contact with empirical contingencies, on the one hand, and providing some critical/normative leverage over those contingencies, on the other hand. This tension is, I believe, intrinsic to any political philosophy, including, on one interpretation anyway, ideal theory.

2 See for example Sunstein 2009.

3 According to Google Scholar, at 18 January 2017, this book has been cited nearly 64,000 times.
4 See, for instance, Gaus, 2011, III.7.3.
5 Muldoon, op. cit., secs. 2.7-8. See also Gaus, *TI,* II.4.1.
6 Muldoon is particularly good on this point. See op. cit., p. 118.
7 There are faint echoes here of Rawls's notion of pure proceduralism (Rawls 1971, p. 85).
8 See Weick 1976.
9 See for instance Epstein 2006.
10 An important intermediate layer is that of individual social agents, who will themselves have their own modes of evaluation and will experience in their own persons the effects of changes to social institutional arrangements.
11 On these matters generally, see D'Agostino 2003.
12 As Gaus documents (*TI,* 87), Karl Popper foresaw the extravagance of the conditions required if this notion of ideal theory were not to be empty. "This line of analysis led Popper to conclude that 'the Utopian approach can be saved only by the Platonic belief in one absolute and unchanging ideal, together with two further assumptions, namely (*a*) that there are rational methods to determine once and for all what the ideal is and (*b*) what the best means of its realization are. Only such far-reaching assumptions could prevent us from declaring the Utopian methodology to be utterly futile.'" And utterly futile it surely is, and is surely shown to be by Gaus's careful analysis.
13 The matter of confidence is meant to signal Gaus's excellent discussion of "The dilemmas of diversity" at *TI,* III.2.
14 See for instance Peters 2012.
15 See for instance Brennan 2010.
16 See for instance Jost and Major 2001.
17 In case (a) we can expect that individuals will mutually enhance each others' always probably only partially shared understanding of the social ideal. In case (b), we can hope for the development, among individuals who might otherwise be at odds with one another, of some fellow-feeling that may, in turn, support their mutual recognition.
18 See in particular Hayek, 1973, ch. 2.
19 See Sunstein 2006.
20 http://www.thecritique.com/articles/open-society-and-its-friends/, accessed 27 January 2017.

REFERENCES

Bauman, Z. (1987). *Legislators and Interpreters*. Cambridge: Polity Press.
Brennan, G. (2010). PPE: An Institutional View. *Politics, Philosophy & Economics* vol. 9(4).
D'Agostino, F. (1996). *Free Public Reason*. New York: Oxford University Press.
— (2003). *Incommensurability and Commensuration*. Aldershot: Ashgate.
—(2010). *Naturizing Epistemology*. Houndsmill: Palgrave Macmillan.
Epstein, J. (2006). *Generative Social Science*. Princeton: Princeton University Press,
Gaus, G. (2011). *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.
—(2016). *The Tyranny of the Ideal*. Princeton & Oxford: Princeton University Press.
Hayek, F. (1973). *Law, Legislation and Liberty*. Vol. 1. London: Routledge & Kegan Paul.
Jost, J. and Major, B. (eds.) (2001). *The Psychology of Legitimacy*. Cambridge: Cambridge University Press.
Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
Lakatos, I. (1970). Falsification and the Methodology of Scientific Research Programmes. In: *Criticism and the Growth of Knowledge*, eds. Imre Lakatos & Alan Musgrave. Cambridge: Cambridge University Press.
Merton, R.(1936). The Unanticipated Consequences of Purposive Social Action. *American Sociological Review* vol. 1.6.
Millgram, E. (2015). *The Great Endarkenment*. Oxford & New York: Oxford University Press.
Muldoon, R. (2016). *Social Contract Theory for a Diverse World: Beyond Tolerance*. London & New York: Routledge.
Müller, J. (forthcoming). *Capitalizing on Political Disagreement: the Case for Polycentric Democracy*. London & New York: Routledge.
Peters, G. B. (2012). *Institutional Theory in Political Science*. 3rd edn., New York & London: Continuum.
Popper, K. (1962). *The Open Society and Its Enemies*. Vol. 1. London: Routledge & Kegan Paul.
Rawls, J. (1971). *A Theory of Justice*. Oxford: Oxford University Press.
Schoelandt, C. van (2015). Rawlsian Functionalism and the Problem of Coordination. Paper delivered to the Pacific Division of the American Philosophical Association.
Sunstein, C. (2006). *Infotopia*. Oxford & New York: Oxford University Press.
— (2009). *Going to Extremes*. Oxford: Oxford University Press, 2009.
Vallier, K. (2014). *Liberal Politics and Public Faith*. London & New York: Routledge.
Weick, K. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly*, vol. 21.1

Public Reason in the Open Society

KEVIN VALLIER

Department of Philosophy
Bowling Green State University
305 Shatzel Hall
Bowling Green, OH 43403

Email: kevinvallier@gmail.com
Web: http://www.kevinvallier.com

A TENSION IN THE IDEA OF A PUBLICLY JUSTIFIED DISCOVERY SYSTEM

This paper attempts to bridge some insights in Jerry Gaus's two most recent books, *The Order of Public Reason* and *The Tyranny of the Ideal* (Gaus 2011, 2016: hereafter *OPR* and *TI*). In *OPR*, Gaus argues that the "social-moral" rules that comprise our shared moral order must be *publicly justified* to each person in order to sustain the moral practices that bind society together and make social life possible and beneficial (2011, p. 2). A rule is publicly justified when each moral person has sufficient reason to internalize the rule as requiring her to engage in certain lines of conduct in the relevant circumstances. If we ensure that the social, moral rules to which we are subject are publicly justified, Gaus argues that we can sustain a shared social life and enjoy its benefits (2011, p. 263) because publicly justified rules sustain our moral practice of *holding others responsible* for wrongdoing. When we hold others to rules that are publicly justified for them, they acknowledge that they are culpable and blameworthy for rule violations, which motivates conformity to the rule. Publicly justified rules render our moral practice consistent with our jointly recognized freedom and equality (2011, p. 14) and help us to avoid relating to one another in an authoritarian fashion (2011, pp. 32-4).

A critical feature of publicly justified social-moral rules is that they are in some sense *self-stabilizing*. These rules do not persist as social practices merely because they are coercively enforced; instead, persons generally comply with publicly justified social-moral rules because they each see sufficient reason of their own to comply with the rule so long as others do likewise. The flip side of this is that no one can do better from her own perspective and simultaneously sustain cooperation with others by unilaterally deviating from such a rule; accordingly, her deviation from the rule will bring more costs than benefits.[1] Under these conditions, we can therefore say that a publicly justified social-moral rule is a social *equilibrium* (2011, p. 390). As such, the idea of public justification is a kind of *equilibrium concept* because it elaborates the conditions under which a social-moral rule can remain in equilibrium under moral conditions.

In *TI*, Gaus is focused on showing how a diverse "open society" can generate important social benefits (2016, pp. 133-8), in particular the benefits of social discovery (2016, p. 96). But capturing these benefits requires having a system of rules that can *escape* equilibrium. As Gaus argues, we want a system of rules that will tend to return to *some* equilibrium, but not necessarily the *same* equilibrium when disrupted; this is the difference between a system being *robust* and a system being *stable* (2016, p. 231). While we still want a degree of stability for our shared social rules in an open society, we also want to allow for discovery through social change, that is, through a change in the rules that govern our common lives. For there may exist rules that would be better than the rules in equilibrium as judged from each person's evaluative perspective. So while we want our shared moral rules to be self-stabilizing, we also want to be able to move to other rules. This means that social-moral rules in equilibrium have a greater social cost in *TI* than in *OPR*. They can be too hard to change in an open society of citizens who recognize the possibility of improving their shared social rules.

We now confront an apparent tension between *TI* and *OPR*, since *OPR* stresses public justification as an equilibrium concept, and *TI* stresses the costs of keeping social rules in equilibrium. I argue that the apparent tension can be resolved. Gaus stresses that an open society is composed of local "republican" communities (2016, p. 145), within which experiments can take place. The idea, I take it, is that

social-moral rules should be able to vary *within* these communities, whereas the social-moral rules that govern all communities should be harder to change. This means that we may want some social-moral rules to be *stable* for the open society as a whole, but local republican communities should focus more on robustness. The difficulty with this solution, however, is that Gaus argues that all moral violations are "everyone's business" (2011, p. 188), such that moral innovators within local republican communities may be subject to ostracism and punishment from members of the open society that are not members of her local republican community. The threat of such interference may discourage social discovery within local republican communities, which will push those communities away from robustness and back to an excessively conservative notion of stability.

Even with the threat of interference, a great deal of moral innovation can take place. In *TI*, Gaus allows moral innovation when a moral innovator wishes to move her society to another rule within its "optimal eligible set" (2016, p. 214), or among rules that no member of the open society has sufficient reason to reject in comparison with no rule. Moreover, if social-moral rules do not prohibit an activity, then by a principle of natural liberty, persons may experiment with a new "act-type" (2016, p. 187).

However, there is an important type of moral experimentation that *OPR*'s model of public justification appears to forbid. In some cases of moral innovation, we *don't know* whether a rule favored by the moral innovator can be publicly justified, in part because many members of local republican communities cannot evaluate the rule until they can see how the rule works in practice. It appears that *OPR*'s model of public justification thereby renders the moral innovator's action impermissible, since it is not obvious whether the rule can sustain our shared moral life. So by requiring that a society stick with an obviously justified rule when a new rule *might* be better, as judged from the perspective of all, we are forbidden from finding out whether that rule is better in practice. In this way, *OPR*'s model of public justification may forbid innovation that an open society could otherwise accommodate.

I believe that *OPR*'s model of public justification can accommodate this form of moral innovation. To demonstrate, I develop the idea of a *jurisdictional rule* based on Gaus's idea of a jurisdictional right (2011, p. 370); a jurisdictional rule establishes *who is permitted to enforce* publicly justified social-moral rules. If jurisdictional rules are publicly justified for all members of an open society, they may protect moral innovators within local republican communities from ostracism and punishment by non-members, and so no one outside the moral innovator's community will punish or ostracize her for experimentation. With publicly justified jurisdictional rules, therefore, moral innovators can attempt to move their local republican community to a new rule without fear of reprisal or ostracism by the open society as a whole. She need only face resistance from members of her local republican community.

To illustrate, consider the case of Mormon polygamy. Joseph Smith, the founder of Mormonism, believed that his new religious community had been granted divine authority to engage in "plural marriage" where a man may take multiple wives. This is, in many ways, the quintessential moral experiment. But given how long polygamy had been morally and legally prohibited in Western civilization, the polygamy rule may well have been unjustifiable for most members of American society. It was certainly treated as such. In 19th century American society, plural marriage was considered a moral abomination. Even with Mormon polygamy confined to the Utah Territory, other Americans saw it as their business to stop Mormon polygamy even through the use of military power.

In this case, Americans recognized no jurisdictional rule that protected Smith and his followers in experimenting with plural marriage. A publicly justified jurisdictional rule, however, would have protected Smith from punishment by American society broadly, even though members of his community would be free to push back. And had the experiment been allowed to proceed in the open, American society might have come to a better understanding of the justifiability of different marital norms, even if polygamy turned out to be unjustified. In this way, jurisdictional rules can allow *OPR*'s model to capture the benefits of this kind of moral experiment. That is the way in which the tension between *OPR* and *TI* can be resolved.

I proceed in four parts. In section II, I bring out the tension between *OPR* and *TI* as one concerned with the costs and benefits of understanding public justification as an equilibrium concept. Section III explores the resolution of the tension that I think Gaus has in mind in *TI*. Section IV introduces the idea of a jurisdictional rule, which is somewhat at variance with Gaus's conception of a moral rule. I conclude in section V by using the idea of a jurisdictional rule to show that *OPR*'s model of public justification can capture the benefits of moral innovation better than one might think. Note throughout that I argue that all the tools required to resolve the tension between *OPR* and *TI* can be found in different parts of Gaus's work.

PUBLIC REASON AS AN EQUILIBRIUM CONCEPT

On my reading of *OPR* and *TI*, Gaus's social theory is focused on how *moral relations* can be established and maintained between persons (2011, p. 13).[2] Moral relations can be understood as a series of relationships between people that are mediated by a *practice of moral responsibility* (2016, p. 182), understood as the practice of holding persons responsible for culpable errors in public behavior and judgment and providing the conditions under which the reactive attitudes of guilt, resentment, and indignation can be rationally sustained (2011, p. 205). The idea of public justification in Gaus's work is a specification of the conditions under which holding others responsible, blaming them, and holding the reactive attitudes against them is appropriate (2011, p. 254). We normally expect persons to follow a wide array of what Gaus calls social-moral rules (2011, p. 2), rules that are socially recognized, that are generally internalized as morally binding on community members, and that meet certain formal conditions for moral requirements like reversibility, generality, and a modest common good requirement (2011, pp. 172-3). Violations are normally met with the reactive attitudes and punishment.

But our practice of insisting that others comply with social-moral rules is determined in part by the conditions of culpability at work in a violation, that is, when we think that persons are morally responsible for violations of social-moral rules. In many cases, we excuse persons from violating rules because we think certain appropriateness conditions for holding others responsible have not been met. Excusing others is appropriate when we see that they could not have known better than to act as they did. To put it another way, persons are judged accountable or excused based on our model of their commitments and their cognitive capacities. If Reba breaks a promise to John, we hold her responsible because we think she knows that she made the promise, that she recognizes promise-keeping rules as applying to her, and that no exculpatory conditions have been met. Under these conditions, then, we can appropriately blame Reba for breaking her promise and be indignant with her as a result.

One of the exculpatory conditions that Gaus identifies is engaging in a "respectable amount" of reasoning (2011, p. 254), where the person violating the rule does so because she concludes, after careful consideration, that she lacks sufficient reason to comply with the rule. Our normal practice is to only hold persons to moral rules that we think that another person, after considering it, should have recognized herself as bound by. We cannot justifiably hold persons accountable to rules whose rationale is beyond the ordinary exercise of her cognitive capacities.[3] It is true that we often hold a person responsible for breaking moral rules in cases where she was unaware of the violation. However, we only do so when we think the person *should have known better*. Gaus interprets what the agent should have known as what an agent *would have seen* after a respectable amount of reasoning.

The idea of public justification can be understood as a specification of the conditions under which someone does or should be able to recognize a rule as in effect and as binding on her moral agency. We say that a moral rule is publicly justified for an agent when she has sufficient reason, after a respectable amount of reasoning, to internalize the rule as binding on her in the relevant circumstances.

A rule is publicly justified for all community members (which Gaus terms "members of the public") when each community member recognizes the rule as in force in her community and sees herself as having sufficient reason to abide by and internalize the moral rule as applying in some particular set of circumstances. This gives meaning to the idea of *public* justification, such that a moral rule is justified to a public on that public's own terms.

To remain in force, and to sustain the rational reactive attitudes, a social rule must also achieve a measure of stability among members of the public. This does not mean that there must be some enforcer agent who coerces others to follow a moral rule (though enforcement is sometimes required), but rather that the rule is sustained as a social norm by the actions of members of the public, by compliance and holding violators accountable.[4] This means that a publicly justified norm is stable based on the *moral* reasons of citizens. They do not stabilize the rule as in force merely out of fear of reprisal or violent threats; rather, they comply with the rule based on their own evaluative attitudes and psychological drives.

In *OPR*, then, the object of public justification is a social-moral rule that exists as a kind of *social equilibrium*: compliance with a publicly justified social-moral rule is each agent's best response to the actions of others (2011, p. 390), where the "best response" is understood as the balance each agent engages in between the satisfaction of her own evaluative standards and what she takes to be the social good of ensuring cooperative relations with others. That is, each person in a real sense consents to a publicly justified moral

rule because she can regard it as normatively compelling so long as others comply with it. For the rule both comports with her own moral point of view and sustains cooperative relationships with others (2011, pp. 398-9).

The advantage of treating publicly justified moral rules as equilibria is that the rules form an ongoing basis for social life on moral terms. Rules in equilibrium create moral relations between persons by sustaining our practice of moral responsibility and mediating our otherwise strategic relationships with each other. We should understand public reason as an equilibrium concept, therefore, because the problem we wish to solve is how to establish stable moral relations with each other despite our differences.

TI identifies a problem with conceiving of our practice of moral responsibility in terms of equilibrating on social rules. In many cases, our evaluative standards contain a certain kind of social *ideal*, a social configuration towards which we would like to push our social order (2016, pp. 39-40). Many people simply are not satisfied with a shared social morality that establishes moral relations between persons; they wish to pursue a more perfect union. Gaus thinks there are some ways of pursuing our ideal that are compatible with our practice of moral responsibility, but our practice depends upon *sustaining* rules in equilibrium, while our ideals lead us to *change* rules in equilibrium.

Importantly, however, some ways of pursuing one's ideal are problematic. First, if we impose rules on others that cannot sustain our practice of moral responsibility, we lose the great good of moral relations with them. There is also the problem that we might impose "The Choice" on other people (2016, pp. 140-2), where we decide to make others worse off in the short-term in the hopes of making them better off in the long-term. Imposing the Choice on others raises a number of moral problems, not least among them that it is in an important sense tyrannical.

But the central reason *TI* identifies for opposing the imposition of our ideals on the unwilling is that we are likely to lose out on the best social mechanism for arriving at our own ideal; for by making our society more uniform, we lose out on the prospect of using diverse ideals and agents to explore the social space required to discover how to realize *our own* ideal in practice or how to formulate our ideals in the first place (2016, p. 130). Thus, if we impose our present ideal on others, we are at risk of ending up in a worse place *even as judged by our own evaluative standards*. Gaus defends this point by arguing that we know quite a bit less than we think we do about how to institutionalize our social and political ideals. Consequently, we must discover how to understand and realize our ideals through social experimentation (2016, pp. 89, 133). If we want a more just society, then given how little we know, we should embrace a diversity of views and opinions in order to better map the territory of how societies might be better organized. This is a reason we wish to avoid social change, since we should not try to make our society more homogenous.

But just as there are some problems with social change, there are also critical benefits. Thus, we want *OPR*'s model of public justification to allow for persons to change the social-moral rules that apply to them if people discover better rules and superior forms of social organization. Now, allowing for change is by no means problematic in *OPR*. Social change is actually *required* when a social-moral rule is not publicly justified; a social-moral rule should be discarded, no longer regarded as normatively binding, or replaced, if some members of the public have sufficient reason to reject the rule. Further, there is no particular problem with moving from a mutually acceptable rule to a second rule that is broadly acceptable but is considered superior to the present rule by some members of the public.[5] In Gausian terms, there may be multiple social-moral rules in a society's "optimal eligible set" of moral rules that can govern some issue (2011, p. 323). Moves within the optimal eligible set are morally permitted, even if we move from one rule that some members of the public rank as best to another rule that those members (though not all members) rank as inferior.

OPR's model of public justification instead opposes two other types of social change. First, it condemns persons who try to push their society from a publicly justified social-moral rule to a defeated rule, one outside of the optimal eligible set. Holding others responsible for violating the new rule makes one a small-scale authoritarian (2011, p. xvi). I think that much is clear, and the model is right to condemn social change of this variety. But I also think *OPR*'s model is uncomfortable with a second kind of social change—when persons try to push their society from a publicly justified social-moral rule to a rule whose justificatory status is *unknown*. Moral innovators insist that the new rule will prove superior to the present, eligible rule. However, given that many members of the public are unaware of how the rule will function or have yet to acquire the ability to assess the rule, they might turn out to have defeaters for the rule once they've experienced the rule. Thus, members of the public are likely to want to stick to the extant rule. But in *TI*, we may want to allow for moral innovators to push for rules outside of the optimal eligible set in order to discover the effects of new rules. In comparison to *TI*'s model, then, *OPR*

seems wedded to a certain kind of conservatism that is in tension with *TI*'s stress on discovery.

Perhaps the main source of *OPR*'s conservatism is that there are shared returns to current moral requirements (2011, p. 398). Rules that already exist have a normative advantage over rules that do not: extant rules coordinate our interactions. As a result, we have additional reason to sign on to extant rules than we do for rules that are at present mere proposals, for they already establish moral relations between persons. We should also stick to extant eligible rules because they have already reduced uncertainty about how to live together (2016, p. 171).

So the reformer must give up on moving her society to a rule whose justificatory status is unknown. This is because *OPR* understands public justification as an equilibrium concept, leaving us with a kind of conservatism that resists the experimental orientation of *TI*.[6]

A GAUSIAN RESOLUTION

I believe Gaus is aware of the problem I raised in the previous section. In this section, I characterize what I take to be Gaus's solution and explain why I think it needs further development.

In *TI*, Gaus argues that an open society with a diversity-accommodating moral constitution can be liberal overall but nonetheless "contain numerous republican communities" that can "reap the benefits of diverse (but not too diverse) searches" (2016, p. 146). So we can have a moral constitution of social-moral rules that bind all members of a diverse society together but that expressly assign local communities the liberty to carry out their own experiments. As Gaus notes,

> … often the same society will be characterized by a variety of sets of rules, regulating different areas of social life, different types of problems, over different areas. And often the same society will be characterized by competing sets of rules, followed by different parts of the population (2016, p. 184).

So we can allow local variation in our social-moral rules in order to capture the benefits of diversity and experimentation. Now, to do this, the system needs "*relatively* stable social and moral rules" (2016, p. 171). We need these stable rules because we cannot otherwise reduce uncertainty about how to interact and so how to live together on moral terms. But we can have relatively more stable moral rules at the highest levels of social organization and allow for relatively more social change in moral rules at the local level, along with a variety of rules across different local areas.

Gaus also defends a more experimental model of social life by establishing a morality of "natural liberty" where persons confront new "act-types as permissible" such that when someone proposes a new way of living together, she is "free to engage in a new type of action that is not covered by existing prohibitions" (2016, p. 195). Gaus thinks recent empirical work on moral reasoning suggests that people implicitly adopt a morality of natural liberty since moral learning proceeds by assuming permissions to act and then learning about prohibitions. We assume that liberty is the default; restrictions on liberty are learned and justified in order to override the appearance of a moral liberty. This accommodates the innovator insofar as she proposes a new rule that refers to new act-types. So if a moral innovator sees no analogy between the new act-type and prohibited act-types, "he will conclude that morality allows his innovative activity" (2016, p. 196). Thus,

> Moral experimenters—those who are exploring a new perspective on justice—need not first convince themselves that a new action type falls under a previous permission; they proceed as long as they do not conclude that the new type falls under a current prohibition (ibid).

So insofar as moral experimenters are experimenting with *new act types*, Gaus has an answer for the tension I've outlined.

By defending a moral constitution that allows for local variation and arguing that we implicitly endorse a principle of natural liberty in exploring new act-types, we may be able to resolve the tension between *OPR* and *TI*. We can argue that a moral innovator who wants to venture outside of her society's optimal eligible set should confine her experiments to her local moral communities.[7] She should not attempt to drag other communities along until they have more information about how the innovator's moral proposals work out in practice. So long as these moral communities adopt a principle of natural liberty, anyone pursuing new act types should be morally permitted to do so.

Neither of these solutions adequately addresses the case of moral innovation I discussed in the previous section, however, where a moral innovator tries to establish a new social-moral rule whose justificatory valence is unknown.[8] A social morality of natural liberty only permits the moral

innovator to engage in act-types not already regulated by rules, but the moral innovator often proposes a new form of social regulation for *recognized* act-types. The moral innovator, then, is vulnerable to ostracism and social punishment for trying to move her society to a rule that may, for all we know, be outside of the optimal eligible set, even if she sticks to moral innovation in her local moral community. This means that she will be discouraged from engaging in social innovation.

Perhaps the case of moral innovation I examine is rightly prohibited by the moral constitution of an open society. Gaus could argue that the moral innovator should restrict herself to pursuing new act-types and moving around within the optimal eligible set of an open society. She must do this because the social-moral rule that exists at present is publicly justified and establishes moral relations between persons; the new rule is much less certain and clear. By pushing for a new rule in this case, we undermine the great goods provided by the extant rule. Gaus can then argue that an open society will not lose out on valuable innovation if we confine the moral innovator in this way. She already has plenty of avenues for experimentation. After all, she can pursue experimentation within the optimal eligible set, she can pursue new act-types, and if she has defeater reasons for the social-moral rule in question, then she is morally free to disobey it. Perhaps these forms of moral experimentation are sufficient to capture the benefits of innovation.

Nonetheless, confining moral experimentation in this way prohibits a large class of experiments. Given Gaus's stress on the limits of our knowledge, it will often be difficult to determine which rules are publicly justified, especially new rules that have been proposed. Since such rules might be publicly justified, we may do better even from *everyone's* perspective by allowing this form of moral experimentation. Once we acknowledge our considerable fallibility about the justificatory status of rules that may be outside of the open society's optimal eligible set, we can see that barring action to move us to rules whose justificatory valence is unknown might be quite restrictive.

It is not enough to allow experiments with rules in the optimal eligible set of her local republican community that are not also in the optimal eligible set of the open society as a whole. This is because the level of interaction and cooperation between different republican moral communities in an open society is sufficiently rich that violations of social-moral rules are still the business of all community members, such that all members of the open society will hold local experimenters responsible for violations. As of yet, we have not grappled with the tension between Gaus's insistence that moral violations are everyone's business and the liberty of republican communities to experiment.

So my goal henceforth is to draw on other parts of Gaus's social theory in order to show that *OPR's* model of public justification can accommodate the social experimentation emphasized in *TI*.

JURISDICTIONAL RULES

Resolving the *OPR-TI* tension requires appealing to the idea of a *jurisdictional rule*. A jurisdictional rule is a social-moral rule that *constricts the community of those subject to a social-moral rule*. An ordinary social-moral rule, according to Gaus, is a social rule that both permits persons to authoritatively direct the actions of others in accord with an act-type and meets the minimal standards for a rule to count as a genuine moral requirement. A jurisdictional rule is a social-moral rule that specifies that *another* social-moral rule only applies to a local community within an open society. The jurisdictional rule prohibits members of the open society from demanding that members of the local republican community act in accord with the social-moral rule, since the rule is not the rule of the open society. Conversely, a jurisdictional rule can make it the case that a social-moral rule that would otherwise apply to all members of an open society does *not* apply to a local republican community, such that members of the republican community are exempt from appropriate moral punishment on the part of members of the open society as a whole for violating the rule. So jurisdictional rules are a kind of *second-order* moral rule that makes reference to another social-moral rule and fixes the scope of the group to whom the rule applies.

In both *OPR* and *TI*, Gaus stresses the importance of what he terms "jurisdictional rights" or rights that give individuals or groups the authority to control the expression of their evaluative standards within a restricted domain of social space (2011, pp. 370-4; 2016, p. 200). Within a moral jurisdiction, persons are permitted to act in accord with their evaluative standards without the interference or permission of others. Similarly, a jurisdictional rule requires that a social-moral rule only applies to a sub-group or that the sub-group is exempt from an open society's social-moral rules on some issues. So a jurisdictional rule creates a kind of jurisdictional freedom, though it differs somewhat from the sort of freedom secured by a jurisdictional right, since it is focused not so much on the expression of local evaluative standards, but on the scope of other social-moral rules.

To illustrate the idea of restricted scope, let us return to the case of Mormon polygamy. If a *marital* jurisdictional rule is publicly justified for an open society, this may allow that, within the Utah territory, polygamy is morally permitted and polygamous marriages impose moral and legal duties on spouses. Assuming the polygamy-permitting rules do not apply to American society broadly, the jurisdictional rule in this case prohibits the American public from punishing and ostracizing the Mormon community for engaging in polygamy. Now, in practice, such a jurisdictional rule may not have been publicly justified to American society. However, the 19th century emphasis on federalism, even following the Civil War and the 14th amendment, which limited the extent of federalism, suggests that federalism was enough of a part of their social morality that confining polygamy to the Utah territory may have been publicly justified. But *had* the jurisdictional rule been publicly justified, then the Mormon community would have had moral freedom from American society as a whole to experiment with polygamy.

Importantly, Gaus's work suggests that the idea of a jurisdictional moral rule is incoherent because violations of all social-moral rules are necessarily the business of all community members. For Gaus, our moral practice assumes that we "hold ourselves to have standing to insist on actions on [another person's] part" (Gaus 2011, pp. 190-191). Following Kurt Baier, Gaus argues that moral violations where reactive attitudes are relevant are ones where we think the violation is "[our] business" because we "have standing to insist on performance and standing to hold the violator responsible for what she has done" (Gaus 2011, p. 224). Baier argued that moral violations cannot be *entirely* the business of the person who engaged in the moral violation: "whether a person conforms to the mores and laws of the group is not entirely his own business." (Baier 1958, pp. xviii-xix). But both Gaus and Baier have cause for concern about the suggestion that moral violations can be *entirely* the private business of some individual or group. The concern is that social-moral rules are public entities that are created, enforced, and maintained by the community, such that moral violations license indignation among those who observe an infraction of t he rule, and license resentment by those who were harmed or insulted by the infraction. So jurisdictional rules, by establishing that violations of some social-moral rules are *not* everyone's business, contradict the Gausian understanding of one of the central features of social-moral rules.

And yet, it seems obvious that there are many jurisdictional rules in effect. In the Catholic Church, for instance, there are social-moral rules that require Catholics to confess their sins or obey the directives of the church hierarchy. If an atheist insists that her Catholic friend go to confession, even when the Catholic friend herself acknowledges that she should go to confession, the atheist lacks standing to insist on compliance with the Catholic moral rule of confessing sin. In this case, the Catholic friend is liable to think that *the atheist* has violated a social-moral rule of minding her own business because she is not a member of the group to whom the social-moral rule applies.

To show that the idea of a jurisdictional rule is coherent, we need not entirely reject Gaus and Baier's claim that moral violations are everyone's business. Instead, we can begin with the default assumption that moral violations are everyone's business, but that social-moral rules can apply to sub-communities within an open society so long as the jurisdictional rule is publicly justified. Otherwise, members of the open society may think that a community's having a unique social-moral rule or rejecting a broadly accepted social-moral rule is cause for indignation, ostracism, and punishment. We therefore arrive at the possibility of a social-moral rule whose violation *is not* everyone's business because the scope of those whose business it is to care about the violations of that rule is restricted by a social-moral rule whose violation *is* everyone's business.

One might wonder why Gaus can't simply acknowledge that some social-moral rules apply to the open society whereas other social-moral rules only apply to local republican communities. Can't the fact that some rules just happen to be part of a local community be enough to ensure that non-members have no standing to enforce the rule since the rule does not apply to them? If so, perhaps we can do without the idea of a jurisdictional rule.

In reply, I argue that we need the idea of a jurisdictional rule because the degree of social unity and social cooperation within an open society suggests that all social-moral rules should apply to all persons within that system of social cooperation. In *OPR*, Gaus discusses the incompleteness of moral relations between citizens of different nation-states (2011, p. 474), and so he allows that social-moral rules can vary across nation-states such that moral relations are incomplete between members of these states. He also notes, however, that moral relations can extend across the boundaries of states in virtue of the market interactions between persons across these states (2011, pp. 471-74). This condition suggests that social morality extends across groups of

people who regularly interact with one another, just as we can expect members of different local republican communities to interact with one another. So the presumption is that persons who interact on a regular basis share a social morality, and we need jurisdictional rules to explain how their justified social moralities can differ.

A GAUSIAN RESOLUTION VIA JURISDICTIONAL RULES

The tension between *OPR* and *TI* is that *OPR* treats public reason as an equilibrium concept in a way that seems to prohibit some kinds of valuable moral experimentation, in particular moral experimentation where a moral innovator wishes to push her republican community to a rule whose justificatory valence is unclear. I believe the model of public justification in *OPR* can allow for this sort of moral experimentation.

Jurisdictional rights play a prominent role in both *OPR* and *TI*. In solving the problem of diverse evaluative standards, Gaus argues, "jurisdictional rights reduce complexity by decoupling the public moral constitution from changes in perspectives, allowing high levels of change in some perspectives without affecting the shared public world" (2016, p. 200). So jurisdictional rights, like the right to privacy and freedom of association, are ways of permitting persons to engage in moral experiments. The idea of a jurisdictional rule plays a similar role, since it allows persons to attempt to move their local community or association to a new rule without fear of reprisal by the larger community of the open society. This is because the jurisdictional rules specify that not all social-moral rules apply equally to all members of the open society. Some rules apply only to local communities, and some local communities are morally exempt from rules that apply to everyone else. In this way, a publicly justified jurisdictional rule can allow for moral innovators to try to move the open society outside of its eligible set by first engaging in a moral experiment in her local republican community that is protected from moral control by members of the open society as a whole.

One problem remains, however, and this is the question of whether the moral innovator is permitted to try to move her local republican community outside of *its* optimal eligible set. Returning to the case of Mormon polygamy, we might imagine that prior to the establishment of polygamy, polygamy was not clearly in the Mormon community's optimal eligible set, but it would in fact turn out to be. Thus, in moving the Mormons to polygamy, Joseph Smith was moving his society to a new marital rule outside of its optimal eligible set. So what is the moral status of Smith's innovation? Remember that jurisdictional rules play no role in this case, since we are only focused on the optimal eligible set of the moral community protected by a jurisdictional rule. This suggests that *OPR*'s model of public justification may prohibit Smith's experimentation within his own community, since the new rule undermines moral relations between Mormons.

In this case, I submit that protection from the open society's indignation and punishment should be enough to capture the benefits of experimentation. When a moral innovator tries to disrupt extant publicly justified moral rules in her community in favor of a rule whose justificatory valence is unknown, she is unjustifiably subversive. Fortunately, however, members of local republican communities can have social-moral rules that are at great variance with the rest of the open society, such that moral innovators in different communities will be able to engage in quite different forms of life. Moreover, the discovery of new act-types is also permitted within these communities by the morality of natural liberty. So *OPR*'s model of public justification is still somewhat conservative when applied within local republican communities, but it is more liberal and open when applied to the open society as a whole. That seems to me to strike the right balance. And all we need to achieve this balance is to introduce the idea of a jurisdictional rule.

NOTES

1 Where "costs" and "benefits" are understood expansively to account for whatever her evaluative perspective regards as a cost or benefit.

2 Gaus discusses the idea of moral relations many times in *OPR*. See Gaus 2011, pp. 8, 13, 174, 183-4, 193, 199-200, 282, 426, 431, 463, 475.

3 Though we can justifiably hold her accountable to rules whose validity she may not see given her present reasoning, but would see if she reasoned a fair or respectable amount. This is where *idealization* enters into the idea of public justification.

4 And, in some cases, holding persons who refuse to enforce the rule accountable as well.

5 Though if the current rule is in equilibrium, then the process of moving to the new rule will come with costs. Deviators may still be blamed and resented on the ba-

sis of the existing rule. So even moves within the optimal eligible set can be resisted in *OPR*. I thank Paul Billingham for this point. Since movement within the optimal eligible set can often be justified, I set aside this case of experimentation as compatible with the experimental emphasis of *TI*.

6 And, again, *OPR*'s model gives us reason to prefer extant rules in the optimal eligible set to *other* rules in the optimal eligible set, so it may be even more conservative than I have argued.

7 Here I am focused on the open society's optimal eligible set, not the optimal eligible set of the local community, which might be different.

8 The justificatory valence for the open society as a whole, that is.

REFERENCES

Baier, K. (1958). *The Moral Point of View: A Rational Basis of Ethics*. Ithaca: Cornell University Press.

Gaus, G. (2011). *The Order of Public Reason*. New York: Cambridge University Press.

— (2016). *The Tyranny of the Ideal*. Princeton: Princeton University Press.

The Tyranny–or the Democracy–of the Ideal?

BLAIN NEUFELD AND LORI WATSON

INTRODUCTION

Gerald Gaus's *The Tyranny of the Ideal* is an ambitious book that covers an impressive range of topics in political philosophy and the social sciences. The book launches a systematic critique of 'ideal theorizing' about political and social justice and aims to defend a vision of an 'Open Society' that "forsakes a collective ideal of justice" (Gaus 2016, p. xvi).[1] Gaus charges that the dominant philosophers of justice in recent decades, at least within the Ango-American tradition, have been seduced by the allure of identifying 'the' ideal conception of justice. Part of this approach to theorizing about justice involves positing "well-ordered societies, where we [the citizens] all agree on the correct principles of justice, our institutions conform to them, and we are all committed to them" (Gaus 2016, p. xix). Ideal theorizing, so construed, is understood by many political philosophers to provide a useful guide for reforming and reshaping our present non-ideal and unjust societies. Gaus argues, though, that this kind of ideal theorizing about justice "tyrannizes over our thinking, preventing us from discovering more just social conditions" (ibid.). The book aspires, then, to show political philosophers that many of them have been labouring under a yoke that they have failed to recognize. More than this, the book also aims to articulate an alternative approach to ideal theorizing, one that frees theorists from this tyranny.[2]

Unsurprisingly, John Rawls's political philosophy is a central focus of Gaus's 'liberation project.' Rawls's work, including its form of ideal theorizing,[3] has significantly shaped the field of political philosophy since the publication of *A Theory of Justice* in 1971. Gaus argues, though, that Rawls's later attempt to accommodate the fact of reasonable pluralism in *Political Liberalism* (2005) leaves his overall theory of justice "in disarray" (Gaus 2016, p. 153). Further, Gaus contends that once we fully confront the depths of diverse points of view concerning justice, not only must Rawlsians abandon the ideal of the well-ordered society, they must reorient their thinking about ideal theory altogether.

In contrast to Rawls, Gaus holds that a morally heterogeneous society—a society characterized by deep differences concerning justice and not merely 'conceptions of the good' and 'comprehensive doctrines'[4]—is necessary for us to advance our knowledge of the requirements of justice. Abandoning the 'myth' of the well-ordered society and replacing it with the idea of the Open Society, including especially its embrace of diverse and evolving perspectives on justice, is much more likely to "make the world better for all, and allows us to better understand our different moral truths" (Gaus 2016, p. xx). Thus instead of trying to cultivate the one 'perfect rose,' so to speak, Gaus recommends that political philosophers encourage 'a thousand flowers to bloom.'

In this article we defend the Rawlsian view against Gaus's criticisms.[5] Specifically, we dispute Gaus's claim that Rawls's idea of the well-ordered society cannot survive the move to political liberalism. We formulate a 'political liberal' version of the well-ordered society, and show that Gaus's Open Society, rather than a radical alternative to the political liberal well-ordered society, in fact closely resembles it. We also challenge Gaus's claim that Rawlsians committed to the principles of justice as fairness are confronted with 'The Choice.' According to The Choice, roughly, ideal theorists must either: (1) pursue 'nearby' relatively certain 'local' gains in justice for their society, or (2) forgo these local gains in order to pursue the more ambitious but far less certain goal of 'ideal' justice. (The goal of ideal justice is 'less certain' both in terms of its likely achievement as well as the likelihood that it is in fact *the* ideal.) We challenge Gaus's claim regarding The Choice, at least as applied to the Rawlsian view, by explaining how addressing local injustices naturally can *lead* some citizens to develop conceptions of full justice, including 'realistically utopian' versions of their societies. The kinds of political proposals that plausibly follow from this account of public reasoning indicate that Rawlsians in fact do not confront The Choice (or, at the very least, that some additional argument is needed to show that they do). Thus, despite the many interesting points that Gaus raises in his book, we conclude that his arguments do

not ultimately threaten the Rawlsian approach to thinking about political justice.

I. TWO PUBLIC REASON LIBERALISMS

Before we get to our replies to Gaus's criticisms of Rawlsian ideal theorizing (in Sec II-III), we think that it would be helpful to make some general points about Gaus's and Rawls's respective projects in political philosophy. As is well known, Rawls resurrected philosophical interest in political contractualism with the publication of *A Theory of Justice*. In that work, Rawls employs the device of the 'original position' (See Rawls 1999, Sec 4 and Ch. III)—a device that, in Gausian terminology, 'normalizes' the perspectives of diverse citizens (Gaus 2016, pp. 42-51, 105-114) by (inter alia) placing the parties who represent those citizens behind a 'veil of ignorance,' thereby depriving them of any particular knowledge concerning the identities of the citizens whom they represent. The parties within the original position consequently all reason in the same way and have access to the same (general) information. Rawls proposes in *Theory* that the parties would select the two principles of 'justice as fairness,'[6] and that a society 'well ordered' by those principles would be stable over time. Such a society would be stable because all of the citizens who belong to it would endorse and support over time—via their developed and rationally maintained 'sense of justice' (Rawls 2001, pp. 18-19)—the institutions of their just 'basic structure.'[7]

Eventually Rawls came to have doubts about *Theory*'s account of the stability of the well-ordered society.[8] Simplifying greatly, this is because part of that account rested on the acceptance by all citizens of a broadly Kantian ideal of autonomy. Rawls came to think that not *all* citizens within the well-ordered society could endorse this ideal. Instead, the citizens of any society that conformed in its basic structure to the principles of justice as fairness, through the free exercise of their reason, invariably would come to endorse a plurality of reasonable 'comprehensive doctrines' (religious, moral, and philosophical views). This 'fact of reasonable pluralism' (Rawls 2005, p. 441) motivated Rawls to develop his theory of political liberalism, of which the idea of 'public reason' is a central component.

In his writings on political liberalism Rawls holds that decisions concerning fundamental political questions—those having to do with "constitutional essentials" and "matters of basic justice" (Rawls 2005, pp. 214-15, 227-30, 235)—should be made by means of *shareable* public reasons (reasons that all reasonable citizens find acceptable, despite their adherence to different comprehensive doctrines). The idea of public reason, Rawls proposes, should be understood as "part of the idea of democracy itself" (Rawls 2005, p. 441). This is because by deciding fundamental political questions via shareable public reasons, or by ensuring that their political representatives do so, citizens can relate to one another as equal co-*sovereigns*. Citizens also are the *subjects* of political decisions. Political power is ultimately coercive in nature,[9] but the exercise of such power over citizens can be normatively legitimate—it can satisfy what Rawls calls the 'liberal principle of legitimacy'—if it is authorized by a constitutional structure that is justified in terms that all citizens find acceptable (Rawls 2005, pp. 37, 446-447).

Public reasons are directed at a moderately idealized justificatory constituency: citizens whom Rawls labels 'reasonable persons.' Reasonable persons acknowledge the fact of reasonable pluralism and are committed to what Rawls call the "criterion of reciprocity" (Rawls 2005, pp. 48-58). According to this criterion, decisions regarding constitutional essentials and matters of basic justice must be *acceptable* to those citizens subject to them (even if those decisions are not the most preferred ones of all citizens). Satisfying the criterion of reciprocity, then, involves citizens providing mutually acceptable justifications for their shared exercise of political power.[10]

Among Rawlsian public reasons are ecumenical democratic ideals and civic virtues, like transparency and toleration, as well as general rules of inquiry, such as those concerning evidence, logic, and so forth. Public reasons also can be drawn from 'reasonable political conceptions of justice.' Conceptions of justice are 'reasonable' *if* they satisfy the criterion of reciprocity.[11] Such conceptions, in order to satisfy this criterion, must secure a set of specially ranked 'basic liberties' for all citizens (including liberty of conscience, freedom of association, and the political liberties of democratic citizenship), as well as adequate resources (such as education and wealth) for all citizens to exercise effectively those liberties over the course of their lives (Rawls 2005, p. 450). A reasonable conception of justice is 'political' if it is compatible with the various comprehensive doctrines endorsed by reasonable citizens, that is, if it is 'freestanding' in nature. A political conception of justice also is limited in its scope to the basic structure of society;[12] hence it does not apply to all domains of social life. ('Comprehensive' conceptions of justice, in contrast, presuppose the truth of particular comprehensive doctrines, such as utilitarianism, and/or apply directly to domains of social life beyond the basic structure.)

A reasonable political conception of justice—because it satisfies the criterion of reciprocity, is compatible with citizens' different comprehensive doctrines (is freestanding), and is limited in its scope to the basic structure—can be the focus of what Rawls calls an "overlapping consensus" (Rawls 2005, Lecture IV). In such a consensus, roughly, citizens who endorse different comprehensive doctrines *also* can support the reasonable political conception of justice. They may do so by integrating the political conception of justice into their broader sets of beliefs and values.[13] A society in which there is an overlapping consensus on a reasonable political conception of justice consequently can be *stable* over time through the free allegiance of its members. Hence Rawls claims that a well-ordered society characterized by the fact of reasonable pluralism is possible. But, as Gaus stresses in *Tyranny*, in his later writings Rawls *also* acknowledges that there is a *family* of reasonable political conceptions of justice. It is not clear whether the idea of a well-ordered society based on an overlapping consensus coheres with Rawls's acknowledgement of a plurality of reasonable political conceptions of justice. (We try to explain why these ideas are compatible, albeit with some modifications to the idea of a well-ordered society, in Sec II.)

In recent years Gaus has developed an alternative 'convergence' account of public reason justification[14] to Rawls's 'consensus' account. With respect to legislation in general, and not simply constitutional essentials and matters of basic justice (the domain, recall, to which Rawlsian public reasons primarily apply),[15] Gaus applies what he calls the "Public Justification Principle." This principle states: "*L* is a justified coercive law only if each and every member of the public *P* has conclusive reason(s) *R* to accept *L* as binding on all" (Gaus 2010, p. 244). Hence Gaus holds that a political decision can be legitimate if all relevant citizens—the moderately idealized justificatory constituency of the "members of the public" (Gaus 2011, p. 26)—have (at least) *a* sufficient reason to support it. Different members of the public, however, can rely upon different, even incompatible, reasons to support laws. For instance, some citizens might use reasons drawn from their respective religious doctrines while others might appeal to philosophical views like utilitarianism. It is for this reason that Gaus's account of public justification often is referred to as a 'convergence' account: diverse justifications can 'converge' in supporting a law, and thereby secure the legitimacy of that law, even if there is no 'consensus' amongst all members of the public on any of those justifications.

Gaus thus denies a central claim that Rawls advances in *Political Liberalism*: namely, that a political conception of justice requires a *freestanding* justification, and that that freestanding justification must be 'political' in nature (that is, draw upon shareable, ecumenical political ideas, like the ideals of citizens as free and equal, and society as a fair system of social cooperation) rather than 'comprehensive.'[16] While Gaus does not explicitly refer to the Public Justification Principle in *Tyranny*, he does make use of the same justificatory structure with respect to the relation between citizens' diverse conceptions of, or perspectives on, ideal justice, and the set of social rules (including, but not limited to, coercively-enforced laws) that apply to all, and to which all consequently are subject and mutually-accountable.

Gaus's 'Open Society' is regulated by what he calls a 'public moral constitution.' Simplifying somewhat, this public moral constitution consists of a set of rules (that evolve over time) for social cooperation, and with respect to which persons hold one another accountable (Gaus 2016, pp. 177-240). Thus we can say that there is a kind of 'overlapping consensus' within the Open Society; however, the public moral constitution need not have its own *independent* justification (a justification that draws on shared normative ideas) that serves as the basis for public reasoning (Gaus 2016, pp. 177-78). Rather, the Open Society arises *out* of its members' practice of public reasoning together (Gaus 2016, p. 179), in which they draw upon their respective non-shared values, beliefs, and perspectives on ideal justice in order to arrive at mutually acceptable rules and practices.

In developing his account of the Open Society, Gaus recognizes that some 'partial normalization' of perspectives is required—namely, some normative assumptions about our co-operators' attitudes, beliefs, and dispositions. These features include the fact that the members of the Open Society wish to engage in a cooperative scheme, endorse norms of mutual accountability, see certain reactive attitudes as justified, and—most importantly for our discussion—do not take up what Gaus calls an "optimizing stance" (Gaus 2016, p. 215), that is, they do not insist that their ideal "be instituted in the face of disagreement by other perspectives" (Gaus 2016, p. 218) no matter what. One might restate this as: the co-operators in the Open Society are not dominators. Or, to put the point in Rawlsian terms, they desire a social world in which they can cooperate with others as free and equal on terms acceptable to them all.

So not all persons or all perspectives on ideal justice can belong to the justificatory constituency for the public moral constitution. Gaus (2016, p. 222) writes:

> Some perspectives are, in the end, unable to share a framework of moral accountability with diverse others. Even the Open Society must be prepared to normalize to some, hopefully very small extent. Such 'Excluded Perspectives,' which cannot find sufficient space in the Open Society, will almost surely be those that are committed to the optimizing stance […]. The Excluded perspectives can live only by those [rules] that they think best, and so cannot endorse the characteristic institutions of the open society.

So like Rawls's reasonable political conception of justice, Gaus's public moral constitution cannot be supported by *all* of the members of society. Despite their differences, then, both Rawls and Gaus appeal to *restricted* justificatory constituencies: reasonable persons in the case of Rawls, perspectives on justice not committed to the optimizing stance in the case of Gaus. Simply put, both approaches restrict their justifications to constituencies committed to a notion of reciprocity.

II. THE WELL-ORDERED SOCIETY AND THE OPEN SOCIETY

After laying out the arguments against Rawlsian ideal theory and presenting his own positive view of the role of ideal theorizing within the Open Society, at the end of the book Gaus concludes that we must bid "adieu" to the idea of the well-ordered society as a guiding ideal. This ideal is not a feasible or attractive goal; it can give us no useful practical guidance about what to do here and now, nor can it "help reconcile us to this conflict-ridden and often manifestly unjust social world" (Gaus 2016, p. 245). While different communities of citizens or research groups within the Open Society should be allowed or even encouraged to develop their respective conceptions of ideal justice—and to engage in constructive criticism and debate with each other over those conceptions—the hope that citizens might someday come to all endorse the *same* conception of ideal justice must be abandoned.

In the concluding chapter of *Tyranny*, Gaus summarizes two main charges against the ideal of the well-ordered society:

1. It is a mirage.
2. As such, it "tyrannizes over our thinking and encourages us to turn our backs on pressing problems of justice in our own neighbourhood" (Gaus 2016, p. 246).

With respect to the first charge, the ideal of a society in which all (reasonable) citizens accept the same conception of justice and know that all others accept it is a mirage, argues Gaus, because even *if* we could arrive at complete agreement about the correct principles of justice, we would not agree about which "social states best satisfied them" (Gaus 2016, p. 246). Moreover, it is hopelessly unrealistic to even think that we *could* settle upon one ideal of justice. One reason for this is that, as we get 'closer' to an ideal on which we may have agreed upon earlier, our knowledge concerning that ideal expands such that a *new* and improved ideal almost certainly will emerge.

With respect to the second charge, Gaus holds that the ideal of the well-ordered society can act as a will-o'-the-wisp, leading theorists to futilely pursue it and thereby ignore the immediate injustices of their society. This is because political philosophers' theorizing about ideal justice inevitably is subject to what Gaus calls the "Neighbourhood Constraint": "we have far better information about the realization of justice in our own neighbourhood than in far-flung social worlds" (Gaus 2016, p. 102).[17] That is, we can apprehend with relatively high confidence what a 'nearby' social world produced by modest institutional changes would look like; in contrast, we cannot be especially confident in our judgements about more 'distant' social worlds, ones produced by more radical or extensive institutional changes. And given the limitations of our knowledge, powers of prediction, accurate modeling, and so on—limitations reflected in the Neighbourhood Constraint—ideal theorists consequently face "The Choice": "In cases where there is a clear optimum within our neighbourhood that requires movement away from our understanding of the ideal, we often must choose between relatively certain (perhaps large) local improvements in justice and pursuit of a considerably less certain ideal, which would yield optimal justice" (Gaus 2016, pp. 82, 246). The ideal of the well-ordered society thus can tempt us to make 'the perfect the enemy of the good,' so to speak.

In order to mitigate (though likely not eliminate entirely) The Choice faced by the advocates of any particular ideal of justice, they must recognize that they will (almost certainly) improve *their* understanding of their own ideal if they have access to *other*, rival ideals of justice—and, moreover, can

expose their ideal of justice to criticisms, challenges, and recommendations for reform from others who do not share it. This epistemic gain from diversity, though, comes with a cost: a society that fosters a diversity of perspectives on ideal justice will (almost certainly) never converge on a *single* one of those perspectives. (Some conceptions of justice may come to be abandoned over time, of course, but others will be created, others will evolve, and so forth.) As Gaus puts this point: "Only those in a morally heterogeneous society would have a reasonable hope of actually understanding what an ideal society would be like, but in such a society we will never be collectively devoted to any single ideal" (Gaus 2016, p. xix). Thus philosophers cannot have their 'ideal justice cake' and eat it too. Society nonetheless can benefit from ideal theorizing about justice—but only once the aims of such theorizing are appropriately *chastened*. The activity of widespread 'non-tyrannical' ideal theorizing, by many different communities of citizens with different perspectives on ideal justice, can help society progress towards greater overall (non-ideal) justice. Hence ideal theorizing need not be pointless or counterproductive—even if *no* individual perspective will ever turn out to be endorsed by all citizens in any given society.

The main complaint that Gaus levels against Rawls's views as formulated in *Theory*, then, is that the argument for the two principles of justice as fairness rests on the 'normalization' of perspectives, that is, the elimination of diversity among deliberators (via the original position device). Such normalization is neither realistic nor will it produce an accurate ideal of justice. The "deep dilemma" that a normalized approach to ideal justice faces is this: "the very normalization that defines the 'correct' perspective on justice cannot effectively identify its own ideal" (Gaus 2016, p. 150). Rather, as explained above, Gaus holds that advancing our knowledge of justice requires a deep diversity of perspectives, because any given perspective (that is, any particular approach to thinking about justice, such as the Rawlsian one) aiming to identify its own ideal may do so, but probably only will have arrived at a 'local optimum' because it is blind to certain relevant considerations. This is because: "It is other, different but related perspectives, that are most likely to see overlooked superior alternatives—ones that the original perspective can appreciate it has overlooked" (Gaus 2016, p. 135). To use the metaphor of mountain climbing,[18] from any given perspective you may identify the highest peak in *your* range, but you may fail to identify *higher* peaks further afield, let alone the highest peak of all.

To illuminate this claim, Gaus reminds readers of the various insights feminist perspectives have brought to bear on our thinking about justice over the past forty years (Gaus 2016, p. 134). Indeed, early feminist criticism of *Theory* led to important revisions in Rawls's formulation of the original position.[19] Sex was added explicitly to the list of features of persons about which deliberators in the original position were deprived of knowledge behind the veil of ignorance. The specification of the parties in the original position as 'heads of households' also was dropped. This is but one example of how a diversity of perspectives can help improve any particular approach to thinking about justice.

Recall Rawls's original formulation of the well-ordered society: "it is a society in which (1) everyone accepts and knows that others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles" (Rawls 1999, p. 4). In a society that is so ordered, citizens "acknowledge a common point of view from which their claims may be adjudicated"; moreover, a shared public sense of justice "establishes the bonds of civic friendship" (Rawls 1999, p. 5). The idea of the well-ordered society is importantly tied to defending justice as fairness as capable of securing stability for the right reasons in Part III of *Theory*. The representatives in the original position assess conceptions of justice according to their potential for securing stability: more stable principles of justice are preferable to less stable schemes.[20] Rawls aims to show that justice as fairness will be more stable than alternative conceptions of justice (or at least stable enough):[21] it generates its own support and is more in line with principles of moral psychology than alternative conceptions (Rawls 1999, Part Three).

In the move from *Theory* to *Political Liberalism* Rawls makes clear, as we noted earlier, that justice as fairness is in fact but *one* member of a family of reasonable political conceptions of justice. This is a consequence of taking the fact of reasonable pluralism seriously—such reasonable pluralism does not simply include the range of reasonable comprehensive doctrines and conceptions of the good, but also pluralism with respect to justice itself. Thus, according to the final version of Rawlsian political liberalism, there is a family of reasonable political conceptions of justice from which reasonable citizens can draw when making public reason arguments in favour or against proposals concerning fundamental political questions. It is in light of this development that Gaus concludes: "One has to be an especially devout disciple of Rawls not to conclude that by the close of his political liberalism project the theory of justice

was in disarray" (Gaus 2016, p. 153). This is because Gaus holds that the idea of the well-ordered society cannot survive in light of the developments in *Political Liberalism*.

Gaus proposes that a liberal Open Society—and *not* a well-ordered society—provides the best framework for securing the necessary terms of social cooperation for citizens (including communities of citizens committed to different ideals of justice) to coordinate effectively their actions and hold each other accountable. Such an Open Society contains a wide range of perspectives on justice (as well as religious and philosophical views, conceptions of the good, and so forth). The Open Society nonetheless possesses an overarching "public moral constitution"—"a stable, shared, moral, and political framework for living together" (Gaus 2016, p. 178)—which (most of) the perspectives on ideal justice find acceptable, and thus with which they can *comply* over time.

We think that Gaus is right about the version of the idea of the well-ordered society formulated in *Theory*. However, it remains an open question whether the idea of the well-ordered society can be revised or reconstructed in light of political liberals' recognition that there is no single political conception of justice that we can realistically expect full agreement upon. In other words, once political liberals acknowledge that there is a range of reasonable political conceptions of justice, they must either revise the idea of a well-ordered society or abandon it. Here we aim to do the former: we provide a reconstruction of the idea of the well-ordered society as it functions in political liberalism. For the sake of convenience, we will refer to this as the 'PL WOS' (political liberal well-ordered society).

Our account of the PL WOS does not rely on a problematic complete normalization of citizens' perspectives (such as that criticized by Gaus with respect to *Theory*'s formulation of justice as fairness). Reasonable persons hold a diversity of political views. While they all are committed to the criterion of reciprocity—this commitment (along with their acceptance of the fact of reasonable pluralism) is what makes them 'reasonable'—they do not all hold that the original position is the best way to satisfy this criterion when thinking about political justice. As such, the members of the PL WOS cannot be expected to endorse a single political conception of justice (such as justice as fairness). Though, of course, citizens can draw upon their favoured conceptions of justice when making public reason arguments, doing so does not commit them to an 'optimizing perspective' whereby they *insist* that their conception of justice alone be the basis of their society's laws. The idea of public reason, based upon the criterion of reciprocity, thus creates the conditions for the exchange of public reasons, a process by means of which citizen can arrive at mutually acceptable justifications for the institutions and policies of their shared basic structure.

Philosophers cannot know in advance of the exercise of public reasoning what the outcomes of citizens' deliberations will be. However, as mentioned in Sec I, in order to satisfy the criterion of reciprocity, and thereby be 'reasonable,' a political conception of justice must include certain features. Specifically, Rawls holds that all members of the family of reasonable political conceptions of justice will contain the following three features:

> First, a list of certain rights, liberties, and opportunities (such as those familiar from constitutional regimes);
> Second, an assignment of special priority to those rights, liberties, and opportunities, especially with respect to the claims of the general good and perfectionist values; and
> Third, measures ensuring for all citizens adequate all-purpose means to make effective use of their freedoms (Rawls 2005, p. 450).

Political conceptions of justice that include some version of these three features embody respect for all citizens as free and equal persons.[22]

While reasonable citizens typically will find most plausible, and hence endorse, only *one* reasonable political conception of justice,[23] they will judge *all* conceptions that satisfy the criterion of reciprocity to be 'acceptable.' A citizen finds a political conception of justice 'acceptable' insofar as she can appreciate the justification(s) for that conception, and—because that conception satisfies the criterion of reciprocity, has a justification that is freestanding in nature, and contains principles that apply only to the basic structure—can willingly abide by its institutional requirements should it be implemented democratically in her society's basic structure. This is so even if that citizen would prefer a *different* conception of justice to be realized in her basic structure, that is, even if she regards an alternative political conception to be 'more reasonable.'[24] What is important for our purposes here is the idea that for a reasonable person to find a conception of justice acceptable is (ceteris paribus) sufficient reason for that person to *comply* with the various institutions and laws justified by that conception (if they are realized via legitimate political procedures).

With the idea of a family of reasonable political conceptions of justice in hand, we can formulate the PL WOS—a

society regulated by a public conception of justice (Rawls (2001, p. 8)—as a society with the following features:

1. All citizens (as reasonable persons) endorse *a* reasonable political conception of justice (one member of the family of reasonable political conceptions of justice).
2. The basic structure is organized in compliance with (at least) one member of the family of reasonable political conceptions of justice.[25]
3. All citizens (reasonable persons) know (1) and (2) ((that is, the 'publicity condition' (Rawls 2005, pp.66-71) is satisfied)).
4. A public political culture obtains as characterized by a reasonable overlapping consensus and a shared commitment (among reasonable citizens) to public reason.

Such a PL WOS can be 'stable for the right reasons,' namely, through the *free* support of the reasonable citizens who belong to it. And it is 'realistically utopian' in that citizens can try to more closely realize it through their political activity, and in particular by deciding fundamental political questions via public reasons, *without* expecting that everyone must endorse the same political conception of justice.

Thus, contrary to Gaus's claims, the idea of the well-ordered society, a pluralist society that is stable for the right reasons, can survive the move to political liberalism. The PL WOS does not require that all citizens share the same conception of justice; nor does it require the use of a single normalizing perspective (such as that of the original position). The PL WOS relies on the criterion of reciprocity, the family of reasonable political conceptions of justice, and the idea of public reason. The core of the ideal of the well-ordered society is that of a system of fair social cooperation (across generations) amongst free and equal citizens; the PL WOS realizes this ideal, despite pluralism with respect to justice.

But what is the fate of Rawls's specific conception of justice in the PL WOS? Gaus observes, "In 1958's 'Justice as Fairness' a 'family' of distributive views was also justified, and Rawls saw this as a core *weakness* in a theory of justice," because that family could not provide "a determinate ranking of claims" (Gaus 2016, p. 152, our emphasis).[26] Are we transported back to 1958, so to speak, in the PL WOS? Our answer is 'yes and no.' The 'yes' part of our answer involves recognizing that, for society as a whole, decisions regarding the 'ranking of claims' concerning justice are to be made by *citizens* as co-sovereigns and public reasoners. Philosophy cannot identify a single conception of justice on which all rational and reasonable citizens must converge. At the same time, though, for particular citizens our answer is 'no.' This is because the conception of justice as fairness will remain (at least for the foreseeable future) a member of the family of reasonable political conceptions of justice that citizens can endorse and draw upon when providing public reason justifications for political proposals.

Even in his final writings on political liberalism Rawls maintains that justice as fairness is the *most* reasonable conception of justice (See Rawls 2005, pp. xlvi-xlvii, 450-51). Holding this view is compatible with acknowledging that other conceptions are reasonable, and thus also can serve as the bases for legitimate laws. Political philosophers occupy no privileged position in the PL WOS. "[T]here are no philosophical experts," Rawls writes, "Heaven forbid!" (Rawls 2005, p. 427). But this is not to say, of course, that philosophers cannot *contribute* to the public political culture of their society. "[C]itizens must, after all, have some ideas of right and justice in their thought and some basis for their reasoning," Rawls observes, "And students of philosophy take part in formulating these ideas, but always as citizens among others" (Rawls 2005, p. 427). When we judge a particular conception to be the *most* reasonable one, we presumably have reasons for doing so—reasons that we can share with others, in the hope of convincing them also to endorse that conception. The original position thus can continue to serve as a tool by means of which some citizens can explain *why* they are committed to the conception of justice as fairness, and thus endorse (and support with public reasons) the political proposals that they think follow from that conception. Granted, it is not the *only* way to satisfy the criterion of reciprocity (Rawls 2005, pp. xlviii-xlix), but we think that it remains *an* important way, one that many citizens continue to find compelling, and consequently can draw upon in their political deliberations and activities.

Having revised the idea of the well-ordered society so that it is compatible with political liberalism, and in a way that we think expresses fidelity to Rawls's main commitments, we find it difficult to see the stark contrast that Gaus aims to draw between it and the Open Society. The Open Society, like the PL WOS, is deeply concerned with securing the support and compliance of its members. With respect to the public moral constitution of the Open Society, Gaus (2016, p.178) writes:

> The aspiration is for the various perspectives, each committed to its understanding of the nature of the social world and ideal justice, to find the public social

> world *endorsable*. If each perspective can make sense of the categories of the artificial social world and *endorse* their use […], we can have a *shared* artificial world without normalization. None of the perspectives that can relate to and endorse the artificial social world would find themselves normalized away, for each would be related to the public artificial world in a way that makes sense *to* that perspective.[27]

In the Open Society, then, there is a kind of overlapping consensus on the public moral constitution amongst that society's 'reasonable' perspectives on ideal justice (that is, those perspectives that do not insist on adopting the 'optimizing stance' with respect to their ideal of justice, and instead conform to a norm of reciprocity in their interactions with other perspectives). There is no *single* conception of justice shared by all members of the Open Society, of course, but the elements that make up the public moral constitution—such as individual liberties, rules concerning property, markets, and so forth—provide a framework for social cooperation and interaction based upon a norm of reciprocity.

But what can advocates of *particular* conceptions of justice do politically within the Open Society? On this matter Gaus writes:

> [W]e must recognize that ideals of distributive justice are part of particular perspectives on justice, and in the Open Society no perspective has a special claim to have its ideal legally instituted. Questions of distribution, like so much, are *matters of democratic politics*. A democratic polity in the Open Society must beware of undermining the moral constitution that renders a shared public life among diverse perspectives possible, but it has many tasks that go *beyond* maintaining this general framework (Gaus 2016, p. 202, italics added).

This position seems entirely compatible with our account of the political liberal version of the well-ordered society and the place of justice as fairness within it. All reasonable political conceptions of justice must satisfy certain criteria, but which particular laws and policies are adopted within the PL WOS—including laws and policies concerning economic distribution —is a matter to be decided by citizens working within the procedures of the democratic constitutional structure of their society. Rawlsians, of course, will push for laws and policies that they take to be justified by the two principles of justice as fairness. But as Rawls himself emphasizes, because there are no 'philosophical experts' to decide these political matters for all in a democratic society ('Heaven forbid!'), they may do so only as 'citizens among others.'

III. THE CHOICE? WHAT CHOICE?

Having established that even within Gaus's Open Society Rawlsian citizens should be free to push democratically for the laws and policies that they think will move their society closer to their ideal of justice, it may remain the case that they still face The Choice. In this final section we try to cast doubt on Gaus's claim that The Choice poses a real problem for Rawlsian citizens.

Recall that Gaus maintains that ideal theories, if they are to constitute a distinctive approach to thinking about justice, *must* confront The Choice (Gaus 2016, pp. 82-84). Roughly, The Choice holds that adherents of a conception of ideal justice often must decide between: (a) pursuing 'nearby' gains in justice, with the risk that such local improvements may take their society further *away* from ideal justice (the full or most complete possible institutional realization of their favoured conception of justice); or (b) forego certain local gains in justice in order to pursue—at some greater distance, likely outside the immediate 'neighbourhood' of their social worlds, and thus more difficult for them to ascertain or see clearly—ideal justice. Drawing on his reading of Sen, Gaus denies that ideal theory is distinctive or helpful *if* it presupposes that the 'peak' of ideal justice is in our neighbourhood and/or that the 'terrain' of relevant social worlds is 'smooth.' In other words, if we face what Gaus refers to as a 'Mt. Fuji' terrain of social worlds—with ideal justice as the peak of the mountain, and all social worlds accessible to us, other than the ideal and our own, located either higher up or below us on the slope—then we do not need an ideal theory of justice at all. Instead, we can simply focus on climbing 'up,' as *any* movement upwards will get us closer to the top of the mountain. In contrast, if we confront a 'moderately rugged' region of social worlds (think of the Himalayas, with various peaks and valleys), then ideal theorizing can make sense, according Gaus, since simply climbing 'up' may lead us to the top of a peak *other* than the highest one, that is, a social world that is only somewhat just (according to our favoured conception of ideal justice), rather than fully just (or the closest to ideal justice possible for us).

(As an aside, we do not think that it is the case that ideal theorizing plays *no* helpful role even within a Mt. Fuji region

of possible social worlds. Consider two rival conceptions of justice: classical utilitarianism and justice as fairness. It may be that the terrain between our current social world and a social world that realizes (as fully as possible) classical utilitarianism is a smooth slope: any institutional changes we make to our society will either move us 'up' the slope towards the peak of classical utilitarian utopia, or move us 'down' away from it. Imagine that the same applies to the relation between our current social world and the realistic utopia of justice as fairness. Assuming that these are the two accounts of justice that we think are the best available options, we are confronted with two possible routes: climb up towards the classical utilitarian utopia or up towards the justice as fairness one. While we may go *some* distance up both slopes,[28] there inevitably will come a point, probably quite early in our journey, where climbing 'up' the classical utilitarian mountain involves climbing 'down' the justice as fairness mountain (and vice versa). While there are a number of reasons why we might opt in favour of, say, the conception of justice as fairness over classical utilitarianism,[29] and thus begin our climb without thinking about the peak, surely *one* consideration that might inform our choice are pictures (perhaps only sketches) of the two peaks in question, that is, views of what societies that fully realize the two accounts of justice in question look like. Consequently, ideal theorizing may be helpful even on Mt. Fuji.)

One noteworthy feature of Gaus's discussion of The Choice is how abstract it is. Very little is provided by Gaus in terms of concrete examples of foregoing local improvements in justice for the sake of pursuing the more distant goal of ideal justice. The historical examples of groups making The Choice mentioned by Gaus (as far as we could find) are limited mainly to twentieth century communist movements.[30] While we do not quarrel at all with Gaus's use of these examples, we very much doubt that they are applicable to Rawls's conception of justice as fairness (and reasonable political conceptions of justice more generally). If The Choice is a general problem, applicable to *all* ideal theory accounts of justice, then we think that it would have been helpful to see some cases where pursuing the kinds of political and economic institutions that Rawls takes to be recommended by justice as fairness—say, those that characterize what Rawls calls a "property-owning democracy"[31] —involve clearly *foregoing* significant 'local' improvements in justice in existing welfare-state capitalist societies.[32]

For the remainder of this section we will outline a way of thinking about local justice that naturally leads to what may be called 'full ideal theorizing.'[33] Our account constitutes an approach to promoting Rawlsian justice for which The Choice is not (we believe) a problem.

Recall that Rawls's criterion of reciprocity underpins his ideal of public reason. When citizens give public reason justifications for their political proposals, they necessarily are committed to the acceptability of those justifications for other citizens. Moreover, as we suggested earlier, there is an important connection between the acceptability of a justification for, say, a law, and compliance with that law. In finding a justification for a law acceptable citizens are willing, ceteris paribus, to *comply* freely with that law, that is, citizens acknowledge the normative authority of that law for their behaviour.[34] This is what it is for reasonable citizens to be motived by reciprocity in their public political relations with others. One reason for citizens to advance justifications for laws that satisfy the criterion of reciprocity is to bring about (or make possible)[35] compliance with those laws in a manner compatible with the deliberative agency of the citizens in question.

Now when citizens advance a political proposal for their society's basic structure—say, a proposal that they think will improve the justice of their society—part of the process of justifying that proposal is indicating what their basic structure would look like should that proposal be implemented. The hope is that other citizens will find the picture of the revised basic structure to be normatively attractive, and thereby come to support freely the proposal in question. Part of a public reason justification for a political proposal is the description of an alternative social world in which that proposal, through the (adequate) endorsement or acceptance of other reasonable citizens, is realized. Thus a public reason justification for a particular political proposal involves what we might call 'local ideal theorizing': consideration of an amended basic structure following reasonable citizens' acceptance of, and consequent compliance with, the political proposal in question.

Finding a justification for a political proposal convincing, and consequently endorsing that proposal as the most just or best proposal for one's society, differs from finding a proposal and its justification 'merely' acceptable (we made this point earlier with respect to reasonable political conceptions of justice within the PL WOS). What is essential for our account is that public reason justifications for political proposals are at least acceptable to all reasonable citizens. In finding a political proposal acceptable, citizens have adequate reason to *comply* with it.[36] It is part of the nature of a public reason justification that a successful justification —a justification that is (at least) acceptable to the reason-

able persons to whom it is addressed—will motivate compliance. This is because a reasonable citizen *cannot* both (a) find a public reason justification for a law acceptable, and (b) be *un*willing to comply freely (ceteris paribus) with that law. To affirm both (a) and (b) reveals a citizen to not be committed to reciprocity in her relations with others, and hence to be unreasonable. Thus public reason justifications, as they are addressed to reasonable citizens, involve at least local ideal theorizing.

While public reason justifications for political proposals cannot be severed from local ideal theorizing, what is the relation between local ideal theorizing and full ideal theorizing? How might citizens go from evaluating particular political proposals supported by public reason justifications to evaluating entire basic structures? Here is our suggestion. Some citizens endorse *multiple* political proposals. Such citizens judge their basic structure to be in need of wide-ranging reform. And each of these proposals aims at being acceptable to other citizens by means of public reason justifications; hence each of these proposals involve local ideal theorizing. Yet the proposals also should be *mutually realizable*, that is, citizens who endorse multiple political proposals for their basic structure should do what they can to ensure that their various proposals, should they be implemented, do not undermine or conflict with each other.[37] It makes little sense, at least normally, for citizens to aim at political proposals x, y, and z, if x undermines or conflicts with y, y undermines or conflicts with z, and x undermines or conflicts with z.[38] In other words, a desideratum, if not a criterion, of a set of political proposals is that they all be (at least) mutually realizable.[39]

At the limit, in offering a wide range of 'local' political proposals, all of which aim at acceptance by, and thereby compliance on the part of, other reasonable citizens, citizens may end up describing a realistically utopian version of their society. More precisely, through the process of determining how their various political proposals fit together and can be supported adequately by public reason justifications, some citizens *may* find it necessary or helpful to engage in something like Rawlsian ideal theorizing.[40] According to the account sketched in this section, though, such ideal theorizing can begin with a concern with *local* political reforms. By proceeding in this way, it is not clear why The Choice poses a distinct problem for such Rawlsian citizens.

The above discussion was rather abstract. We now will sketch how that account may work with reference to our current political and economic condition. Citizens naturally may reflect critically upon their circumstances, the basic structure in which they find themselves. In the case of the United States, they might conclude that its basic structure manifests a number of features incompatible with the ideal of democratic equality, understood as fair social cooperation among free and equal citizens. For instance, wealthy citizens have exercised, and increasingly exercise, highly disproportionate influence within the American political system.[41] Moreover, this influence has altered the basic structure of the United States in ways that have dramatically *increased* economic inequality over the past four decades.[42] Not only is this growing inequality economically damaging to society overall,[43] it has not improved the absolute incomes of the 'least advantaged' within the United States during this period, that is, there has been no noteworthy 'trickle-down' of economic benefit to the least advantaged.[44] Moreover, recent research on the intergenerational elasticity of citizens' incomes suggests "that the United States is very immobile," and thus falls far short of realizing anything like a principle of equality of opportunity (Mitnik and Grusky 2015, p. 4).[45] And despite important changes to the legal structure of society over the past five decades, profound race-based and gender-based inequalities in income and wealth, economic opportunities, and political influence continue to persist.

In response to these features of their society, citizens of the United States committed to the ideal of society as a fair system of social cooperation amongst free and equal citizens might propose changes to their basic structure. Such changes might include, inter alia, public financing of election campaigns, reforms to the provision of basic education and the distribution of higher education (so that the distribution of education counter-acts, rather than reinforces, existing class- and race-based inequalities),[46] a guarantee of employment for all citizens, ensuring universal heath care for citizens,[47] and limiting the total amount of wealth that citizens can inherit in order to counter-act the intergenerational concentration of wealth within a small portion of the population.[48] Measures that aim at promoting greater racial[49] and gender equality[50] within society also can be justified as necessary for securing the freedom and equality of all.

In advancing these kinds of political proposals, citizens aim at securing other reasonable citizens' acceptance of—and, if implemented democratically, willing compliance with—them by providing mutually acceptable justifications. But in formulating such proposals, and in trying to justify them, citizens might also try to determine how, and indeed

whether, the proposals and their justifications *fit together*. In other words, citizens might ask: can these proposals be justified and organized via coherent and compelling underlying principles which, in turn, express or are based upon an ideal of fair social cooperation among free and equal citizens? This process of justification, reflection, and revision naturally may lead some citizens to identify what they take to be the most reasonable political conception of justice, the conception that, overall, they think best expresses the ideals of free and equal citizenship, and of fair social cooperation. And part of this process can include, for those citizens, trying to think about what their society would look like should all of their main political proposals be realized.

Consequently, beginning with critical reflection on the *local* injustices that citizens identify with their basic structure, and coming up with political proposals for addressing those injustices in a manner compatible with the criterion of reciprocity, some citizens may find themselves eventually trying to ascertain what a realistically utopian version of *their* society would look like. Through this process some citizens might end up supporting the general features of a property-owning democracy,[51] the form of society that Rawls thinks is both feasible and can realize successfully the ideal of free and equal citizenship. But citizens need not stop there. Their commitment to property-owning democracy is not unconditional, but rather is a revisable one. It may turn out that some features of 'welfare-state capitalism' also can be used in order to reform the basic structure so that it more adequately satisfies the principles of justice as fairness.[52] In any case, the processes of addressing *both* questions of local justice and overall (full) justice need not involve fixating on one set of questions at the expense of the other; instead, citizens' reflections on both sets of questions can (and we think should) inform each other.

One way to understand Rawlsian ideal theorizing, then, is to see it as *beginning* with a concern with local injustices. According to this account, citizens can move 'up' from a concern with local injustices to a realistically utopian conception of their society—and this process, of course, can involve moving back and forth (or 'up' and 'down') between local improvements in justice and reflections on overall justice. Drawing on this account, and focusing on the particular kinds of political proposals that this process might generate (e.g., commitments to the public financing of election campaigns, universal health care, a social minimum for all citizens, and so forth), it is not at all obvious to us that The Choice will confront the reasonable Rawlsian citizens we are envisioning in any stark way.

Certainly there may be some cases in which difficult choices have to be made by political actors, and the long-term consequences of any particular policy or legislative decision can never be fully foreseen. But these difficulties do not seem peculiar to the kind of ideal theorizing that we have described here. All of the proposals sketched in the above paragraphs, for instance, are not only 'local' improvements in justice (at least from the perspective of justice as fairness), but *also* seem to move society closer to the institutions characteristic of a property-owning democracy (or at least do not appear to move society further *away* from property-owning democracy). Perhaps we are wrong about this. But appreciating the gravity of The Choice—or whether it even applies to many or most significant political decisions—requires more detailed consideration of what promoting a democratic and liberal conception of justice, like justice as fairness, involves in *practice*. Based on what Gaus says in *Tyranny*, we cannot see why the kinds of political proposals that would improve the justice of existing welfare-state capitalist societies (like the United States) plausibly would involve moving those societies *away* from the Rawlsian ideal of property-owning democracy.

IV. CONCLUSION

To conclude, we remain unconvinced that one main target of Gaus's critique—Rawlsian political liberals—are operating under a *tyranny* of ideal theory. While it is fair to say that justice as fairness requires revision in light of the insights and commitments of political liberalism, it is hardly a jumbled mess ('in disarray'). Moreover, Gaus's insistence that Rawls or Rawlsians owe an account of *the* liberal theory of justice is simply strange in light of his recognition that a commitment to (reasonable) pluralism entails a rejection of any such single, unified account as *the* account of justice.[53]

Here we have formulated a version of the idea of the well-ordered society—the PL WOS—that is consistent with political liberalism's commitments and aims. The PL WOS avoids Gaus's critique of Rawls's original formulation of the idea. We also have argued that, as it turns out, the Open Society that Gaus defends is not very dissimilar from the PL WOS. And finally, we proposed that there is a way to understand the role of public reasoning and ideal theorizing within political liberalism such that political liberals do not face (a serious version of) The Choice. Rather than being subjugated under the tyranny of ideal theory, then, perhaps Rawlsians are operating under the *democracy* of ideal theory.

NOTES

1 Gaus's use of the term 'Open Society' is an homage to Popper (1962), a book Gaus thinks has been undeservedly neglected and underappreciated by political philosophers.

2 Contrary to some recent work that argues that ideal theorizing is unnecessary or even counterproductive with respect to the aim of improving the overall justice of society (e.g. Sen 2009), Gaus proposes that ideal theorizing—*if* its aims are properly reconceptualised—in fact can help improve society's overall understanding of the requirements of justice. Ideal theorizing may be tyrannical in its *current* form, as practiced by many contemporary political philosophers, but it is not *necessarily* so, and in fact can play a constructive social role.

3 For Rawls's explanation of "ideal" and "non-ideal" theory, "partial compliance" and "strict compliance" theory, and the ideas of a "well-ordered society" and a "realistic utopia," see Rawls 1999, pp. 4-5, 7-8, 215-16, 308-9; Rawls 2001, pp. 4-5, 13, 65-66. The main assumptions of Rawlsian ideal theory are "strict compliance" and "favourable circumstances" (Rawls 1999, p. 216; 2001, p. 101).

4 Rawls (2005) initially articulates the idea of 'a comprehensive doctrine' by contrasting political conceptions of justice with moral doctrines (p. 13). He understands moral doctrines to be "comprehensive views" insofar as they have wide scope, covering a wide range of subjects. As such, a moral conception is comprehensive "when it includes conceptions of what is of value in human life, and ideals of personal character, as well as ideals of friendship and of familial and associational relationships, and much else that is to inform out conduct, and in the limit to our live as a whole" (p. 13). In contrast, "a political conception tries to elaborate a reasonable conception for the basic structure alone and involves, so far as possible, no wider commitment to any other doctrine" (p. 13). (On the relation between 'comprehensive doctrines' and 'conceptions of the good,' see Rawls 2001, p.19.)

5 The reason why we refer to our position as 'Rawlsian' rather than 'Rawls's' should become clear in Sec II.

6 In Rawls's final formulation of justice as fairness (2001, pp. 42-43), roughly, the first principle specifies a set of 'basic liberties' that are to be secured equally for all citizens within the constitutional structure of society (these liberties include freedom of thought, liberty of conscience, freedom of association, the political liberties (including their 'fair value'), freedom of political speech, and the like). The second principle requires that any economic inequality in society must (a) benefit the 'least advantaged' citizens over time more than any other system of economic distribution, and (b) not undermine or violate the 'fair' equality of opportunity of all citizens to compete for positions of authority and responsibility. The first principle, moreover, enjoys 'lexical' priority over the second.

7 The 'basic structure,' roughly, consists of the main institutions of society understood as an overall system of social cooperation. These institutions apply to all citizens within society, unlike 'voluntary associations,' such as religious organizations, firms, unions, clubs, and universities. (See Rawls 200), Sec 4; Rawls 2005, Lecture VII.)

8 For an insightful discussion of this development, see Weithman 2010.

9 Rawls emphasizes this repeatedly. See, e.g., Rawls 2005, pp. 68, 136, 216.

10 Rawls 2005, pp. 446-47, xliv-xlv. The criterion of reciprocity also grounds the "liberal principle of legitimacy" (Rawls 2005, pp. xliv, 137, 446-47). In fact, the criterion of reciprocity expresses the "intrinsic (moral) political ideal" of political liberalism (2005, p. xlv).

11 While there is "a family of reasonable political conceptions" of justice, "[t]he limiting feature of these forms is the criterion of reciprocity" (Rawls 2005, p. 450).

12 The basic structure, recall, is made up of society's main political and economic institutions, understood as an overall system of cooperation (see note 7).

13 For Rawls's final account of how this might be achieved, see his "Reply to Habermas" in Rawls (2005). (The achievement of such integration may involve *revisions* to citizens' comprehensive doctrines.)

14 See Gaus (2010, pp. 233-75; 2011; Gaus and Vallier 2009, pp. 51-76).

15 It should be noted that some Rawlsians hold that the idea of public reason should apply to *all* political decisions (see Quong 2011, ch.9). The question of the appropriate scope of public reason is not central to the differences between Rawls's and Gaus's views that concern us here.

16 In his "Reply to Habermas" (Rawls 2005), Rawls refers to this as the 'pro tanto' justification of a reasonable po-

litical conception of justice. Weithman (2011) contends that by not employing the idea of a freestanding political conception of justice, the convergence view of public justification cannot realize citizens' political autonomy; instead, in a society governed by the convergence account of public justification, citizens are politically heteronomous.

17 Throughout the book Gaus makes use of the idea of 'social worlds,' which he takes from Rawls. See, e.g., Rawls (2001, p. 128; 2005, p. 53).

18 Gaus takes this metaphor from Simmons 2010.

19 Neither sex nor race were originally included among the kinds of information deliberators in the original position were deprived of behind the veil of ignorance, and representative deliberators were said to be 'heads of households.' English (1977) was the first to criticize the heads of household assumption. Okin (1989, especially Ch. 5) famously criticized Rawls for failing to include 'sex' among the characteristics of which representatives in the original position lack knowledge. With publication of *Political Liberalism* Rawls explicitly enumerates sex and race, among other traits, as characteristics about which representatives in the original position lack knowledge (see Rawls 2005, p. 25; 2001, p. 15).

20 Thus strict compliance is *assumed* only within the first stage of the original position, specifically, the stage at which the parties initially select which conception of justice should govern the basic structure of the society in which the citizens whom they represent will live out their lives. The second stage of the original position involves determining whether a fully just well-ordered society—a society with a basic structure that is organized in accordance with the conception of justice selected at the first stage —would be stable over time for the right reasons, namely, through the free compliance of its reasonable citizens (see Rawls 2001, Sec 54-55). Here compliance is *not* assumed but must be demonstrated to be feasible, given the kinds of psychologies and interests that citizens can be expected to have in the society in question.

21 On the ambiguity of the nature of Rawls's concern with stability in *Theory*, see Gaus 2014.

22 Rawls claims that any conception of justice that fails to include these three features cannot satisfy the criterion of reciprocity within a pluralist society. We think that this claim is correct, though we cannot defend it here.

23 There may be some citizens who find two or more conceptions equally plausible, or who find all reasonable conceptions 'acceptable' but do not endorse any particular one (we explain what it is for a citizen to 'find acceptable' a conception in this paragraph). Such complications do not affect our discussion.

24 Because of their commitment to reciprocity, then, reasonable citizens do not adopt the 'optimizing stance' with respect to their favoured conceptions of justice (Gaus 2016, pp. 215-18).

25 It may be that different reasonable political conceptions of justice 'overlap' in justifying the same kinds of institutions within the basic structure.

26 Gaus is referring to Rawls 1958.

27 Our emphasis on 'endorsable,' 'endorse,' and 'shared.'

28 Indeed, an important aspect of Sen's work is that citizens can improve the overall justice of their society even if they do *not* share the same conception of justice. But obviously this will not always be the case.

29 We assume that we need not rehearse Rawls's arguments against classical utilitarianism here.

30 E.g., see Gaus 2016, pp. 88, 143. (And even these examples are mentioned only in passing.)

31 Rawls emphasizes the distinction between a property-owning democracy and a welfare state in the following way: "One major difference is that the background institution of a property owning democracy, with its system of (workably) competitive markets, try to disperse the ownership of wealth and capital, and thus to prevent a small part of society from controlling the economy and indirectly political life itself. Property owning democracy avoids this, not by redistributing income to those with less at the end of each period, so to speak, but rather by ensuring the widespread ownership of productive assets and human capital (educated abilities and trained skills) at the beginning of each period, all this against a background of the equal basic liberties and fair equality of opportunity" (Rawls 1999, pp. xiv-xv.)

32 For Rawls's discussion of five different 'ideal types' of socio-economic systems ('laissez-faire capitalism,' 'state socialism,' 'welfare-state capitalism,' 'liberal-democratic socialism,' and 'property-owning democracy'), see Rawls 2001, Sec 41-42, 49. Further discussion of the idea of a property-owning democracy can be found in Krouse and McPherson (1988) and O'Neill and Williamson (2012). (Most contemporary liberal societies, such as Canada, Denmark, Japan, and the United States, are welfare-state capitalist societies in Rawls's sense.)

33 This discussion draws upon Sec 2-3 of Neufeld (2017).

34 This claim concerns the way in which public reason justifications are meant to function in citizens' deliberations about fundamental political issues. We do not mean to presuppose anything controversial about the nature of 'reasons' per se (say, some form of 'reasons' or 'judgement' 'internalism').

35 Of course, compliance may often simply be the result of habit or deference to authority. What is important is that such compliance *could* be justified or, if necessary, be *brought about* via the rational agency of citizens.

36 It may be that finding acceptable or even endorsing a public reason justification for a law or institution is not always sufficient to bring about adequate compliance on the part of citizens with that law or institution. For instance, citizens may be subject to foreseeable akrasia even with respect to their compliance with laws and institutions that they support. In such cases, though, insofar as citizens accept the justification(s) for the law or institution in question, and thus agree that their compliance is warranted, they will accept (via instrumental reasoning) the use of those institutional mechanisms necessary to bring about and ensure their own compliance over time. (If the costs of such institutional mechanisms are quite high, though, reasonable citizens may reconsider their acceptance or endorsement of the laws or institutions in question.)

37 Citizens may find acceptable incompatible political proposals as alternatives, but in doing so they of course acknowledge that only one of these alternatives can be realized within their basic structure.

38 It would be rational, however, for citizens to promote political proposals *x*, *y*, and *z*, even if those proposals conflict with or undermine each other, *if* the citizens in question believe that (a) only one (at most) of those proposals has a chance of being implemented, (b) the implementation of any one of those proposals would improve the overall justice of their society, and (c) they do not know which one (at most) of *x*, *y*, or *z* will (possibly) be implemented. Such unusual cases can be put aside for our purposes here.

39 It also may be desirable that the proposals support or reinforce each other. Here we focus on the weaker claim that they should at least be mutually realizable.

40 This is not to say that in endeavouring to ensure the coherence of their various local political proposals citizens *must* turn to Rawlsian ideal theorizing (with its focus on the basic structure as an overall system of social cooperation, and so forth), only that it is *one* plausible strategy for doing so. (Thanks to an anonymous commenter for pressing us to clarify this point.)

41 See Gilens and Page (2014).

42 See Hacker and Pierson (2010). On the overall increase in income inequality within the United States in recent decades, see: Congressional Budget Office 2009, 2011.

43 On the economic harms of high levels of inequality in income and wealth, see: Stiglitz 2012 and Galbraith 2014.

44 The main reason for this, according to Kenworthy (2010), has been government policy decisions.

45 See also Isaacs, Sawhill, and Haskins (2008).

46 Among such educational reforms would be those that promote racial integration (see Anderson 2010).

47 The proposals concerning campaign financing, employment, and heath care are mentioned in Rawls (2005, pp. xlvi-xlvii).

48 On the importance of minimizing the intergenerational accumulation of wealth within a small number of citizens, and the consequent role of taxing bequeathments and gifts, see Rawls (1999, pp. 245-246; 2001, pp. 160-161. Piketty (2014) recently has documented the long-term tendency of capitalist societies toward what he terms 'patrimonial capitalism.' A patrimonial capitalist society, roughly, is one in which the members of that society's economic elite enjoy their privileged position primarily as a consequence of inheritance, not innovation or entrepreneurship. Simplifying greatly, the reason for this tendency is that returns to capital ('r') generally grow at a higher rate than the overall economy ('g'). Consequently, the already wealthy within society tend to become wealthier at a much faster rate than anyone else, and, moreover, pass this advantage on to their descendants. This economic elite becomes largely a class of rentiers. The members of this class also are able to employ their wealth to influence the political decision-making processes of their society, thereby undermining the democratic equality of citizens. Piketty's research appears to support Rawls's more speculative worries about the long-term tendency of capitalist societies toward growing inequality, decreasing political freedom for most citizens, and hence injustice.

49 On extending Rawls's account of justice to address racial inequality, see Shelby (2004; 2013). (The account of ideal theory that we advance here is, we believe, broadly compatible with Shelby's position.)

50 For public reason justifications for laws and policies that promote gender equality, see: Baehr (2008),

Brettschneider (2007), Hartley and Watson (2010); Lloyd (1998); and Neufeld (2009).

51 See n.31.

52 Krouse and McPherson (1988) suggest this with respect to the role of a progressive income tax in promoting justice as fairness. O'Neill (2012) recommends that we regard the institutions of both property-owning democracy and welfare-state capitalism as comprising a kind of 'toolkit' that egalitarians (citizens committed to justice as fairness) can draw upon when trying to improve the justice of their society.

53 Gaus (2016, pp, 153-154, n. 8) writes, anticipating one kind of political liberal reply, "Some might argue that political liberalism is concerned with legitimacy, not justice. Even if so, this would not show that a coherent *theory of justice* remains. […] 'It's only about legitimacy' is not a magic phrase that can make these issue disappear. What *is* the liberal theory of justice?"

REFERENCES

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

Baehr, A. (2008). Perfectionism, Feminism, and Public Reason. *Law and Philosophy* 27:193-222.

Brettschneider, C. (2007). The Politics of the Personal: A Liberal Approach. *American Political Science Review* 101:19-31.

Congressional Budget Office (2009). Data on the Distribution of Federal Taxes and Household Income.

Congressional Budget Office (2011). Trends in the Distribution of Household Income between 1979 and 2007.

English, J. (1977). Justice Between Generations. *Philosophical Studies* 31:91-404.

Galbraith, J. (2014). *The End of Normal: The Great Crisis and the Future of Growth.* New York: Simon & Schuster.

Gaus, G. (2010). Coercion, Ownership, and the Redistributive State: Justificatory Liberalism's Classical Tilt. *Social Philosophy & Policy* 27: 233-75.

— (2011*). The Order of Public Reason.* Cambridge: Cambridge University Press.

Gaus, G. (2014). The Turn to a Political Liberalism. In: D. Reidy and J. Mandle (eds.) *The Blackwell Companion to Rawls.* New York: Wiley-Blackwell, pp. 235-50.

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

Gaus, G. and Vallier, K. (2009). The Roles of Religious Conviction in a Publicly Justified Polity: The Implications of Convergence, Asymmetry, and Political Institutions. *Philosophy & Social Criticism* 35: 51-76.

Gilens, M. and Page, B. (2014). Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. *Perspectives on Politics* 12: 564-81.

Hacker, J. and Pierson, P. (2010). *Winner-Take-All Politics*. New York: Simon and Schuster.

Hartley, C. and Watson, L. (2010). Is a Feminist Political Liberalism Possible? *Journal of Ethics and Social Philosophy* 5:37-54.

Isaacs, J, Sawhill, I, and Haskins, R. (2008). Getting Ahead or Losing Ground: Economic Mobility in America (The Economic Mobility Project, the Brookings Institution and the Pew Charitable Trusts).

Kenworthy, L. (2010). Rising Inequality, Public Policy, and America's Poor. *Challenge* 53:93-109.

Krouse, R. and McPherson, M. (1988). Capitalism, 'Property-Owning Democracy,' and the Welfare State. In: A. Gutmann (ed.) *Democracy and the Welfare State*. Princeton: Princeton University Press, pp.79-105.

Lloyd, S. (1998). Toward a Liberal Theory of Sexual Equality. *Journal of Contemporary Legal Issues* 9:203-24.

Mitnik, P. A. and Grusky, D. B. (2015). Economic Mobility in the United States. The Pew Charitable Trusts and the Russel Sage Foundation.

Neufeld, B. (2009). Coercion, the Basic Structure, and the Family. *Journal of Social Philosophy* 40: 37-54.

— (2017). Why Public Reasoning Involves Ideal Theorizing. In: K. Vallier and M. Weber (eds.) *Political Utopias: Contemporary Debates*. Oxford: Oxford University Press, pp. 73-93.

Okin, S. M.(1989). *Justice Gender and the Family*. New York: Basic Books.

O'Neill, M. and Williamson T. (eds). (2012). *Property-Owning Democracy: Rawls and Beyond*. Oxford: Wiley-Blackwell.

Piketty, T. (2014) *Capital in the Twenty-First Century*. Cambridge MA: Harvard University Press.

Popper, K. (1962). *The Open Society and Its Enemies*. London: Routledge and Kegan Paul.

Quong, J. (2011). *Liberalism without Perfection.* Oxford: Oxford University Press.

Rawls, J. (1958). Justice as Fairness. *The Philosophical Review* 67: 164-94.

Rawls, J. (1999). *A Theory of Justice*. Revised Edition of 1971. Cambridge, MA: Harvard University Press.

— (2001). *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

— (2005). *Political Liberalism*. Expanded Edition, Original edition 1993. New York: Columbia University Press.

Sen, A. (2009). *The Idea of Justice*. Cambridge MA: Harvard University Press.

Shelby, T. (2004). Race and Social Justice: Rawlsian Considerations. *Fordham Law Review* 72:1697-1714.

— (2013). Racial Realities and Corrective Justice: A Reply to Charles Mills. *Critical Philosophy of Race* 1:146-62.

Simmons, A. J. (2010). Ideal and Nonideal Theory. *Philosophy & Public Affairs* 38:5-36.

Stiglitz, J. (2012). *The Price of Inequality: How Today's Divided Society Endangers Our Future*. New York: W.W. Norton.

Weithman, P. (2010). Why Political Liberalism? On John Rawls's Political Turn. New York: Oxford University Press.

— (2011). Convergence and Political Autonomy. *Public Affairs Quarterly* 25/5: 327-48.

Political Philosophy as the Study of Complex Normative Systems

GERALD GAUS

Email: jerrygaus@gmail.com
Web: http://www.gaus.biz

It is an honor and a treat when innovative social theorists and philosophers take time out of their important work to think about one's own. The set of papers being published in COSMOS + TAXIS are especially flattering. Fred D'Agostino, Blain Neufeld, Scott E. Page, Kevin Vallier, Lori Watson and David Wiens all constructively engage *The Tyranny of the Ideal* and open up new issues to be explored. I am especially grateful to Ryan Muldoon for organizing this symposium (and, I should say, for all his work from which I have learned so much). I am also delighted that the symposium appears in COSMOS + TAXIS because the main theme of the book, which I hope to emphasize here (and which D'Agostino and Page bring out wonderfully in their essays) is that the subject matter of social philosophy is complex systems, something Hayek (1964 [2014], chap. 9) was one of the first to stress. Until, as Page puts it, "the imperative of complexity" is appreciated by political philosophers their work will remain what Hayek warned against—constructivist fantasies in which the critical problems of evaluative diversity, path-dependency, uncertainty, and interconnectedness are assumed away.

My comments focus on three main themes. I begin (Sec I) by taking up some questions of method. To a large extent *Tyranny* (as I shall call it, hopefully sans unfortunate self-reference) is adamant that political philosophy greatly benefits from the rigor of more formal ways of thinking. This, unfortunately, is one of the features of the book which causes the most resistance, as political philosophers are generally deeply averse to abstract model thinking, and often dismissive of "metaphors." I then turn (Sec II) to discussing some aspects of the rugged landscape model I employ, and what I think it tells us about the nature of normative thinking. Section III turns to what D'Agostino calls my "constructive intention," the account of the Open Society. Many issues arise here, of which only a few can be addressed. In these remarks I make no effort to respond to all the ideas and, yes, criticisms, in these thoughtful essays, but I do believe that many fundamental points will be addressed—hopefully in a way that enlightens readers who are not especially concerned whether *Tyranny* is bullet proof (it is not).

I. MODELS

Abstractness and Modeling

An explicit aim of *Tyranny* is to model a long-standing problem in political philosophy in a more rigorous and abstract way, which I believe alerts us to critical features of political theorizing that have hitherto gone unnoticed. When we model a phenomenon we always abstract from some of its features to better understand the working of others. All modeling is a selection process. We construct a simplified, abstracted, analysis to get insight into critical features that are obscured by more detailed descriptions. It does not follow that the features from which we abstract are unimportant; in another context we may construct a model to better understand *them*, perhaps putting aside the very features our first model highlighted. As Michael Weisberg (2013, p. 13), points out, philosophers generally prefer models that are rich in description, and this perhaps leads them to so often object to a model because it has left something out or is "blind" to it (as, for example, *Tyranny*'s model quite intentionally abstracts from the "logic" of justification advanced by a particular perspective—more on that anon). The aim of *Tyranny's* modeling approach is to abstract from many features that have traditionally been in the foreground to discover those thus far overlooked. And the feature that I wished to stress—which I tried to show is implicit in the presuppositions of many "ideal theories"—is that seeking ideal justice involves optimization in a certain type of structured complex system. Once we understand that ideal theories are committed to analyzing the pursuit of justice as an optimization exercise in, strictly speaking, a mathematically complex system, we see them and their problems in an entirely different light. Indeed, we see political philosophy in a new light. I believe that this rather startling insight justifies abstracting from so many of the features and issues that have hitherto been the concerns of political philosophy.

In a way my intention was to present political philosophers with an unfamiliar type of analysis, for it is through confrontation with the alien that we appreciate our presuppositions. That, indeed, is the theme of the entire book.

Tyranny advances an analytic result. If ideals are characterized by institutions, and institutions are "coupled" and so have "interactions" or "interdependencies" in their resulting justice, then as Page nicely summarizes in his essay, under the conditions I specify we are almost certainly confronted with "The Choice" between pursuit of local optima (local improvements in justice) and the global optimum (ideal justice). This is an inescapable conclusion. Of course one can dispute the assumptions or the applicability of the model, but not the well-nigh inevitability of The Choice given them. To be sure, as Neufeld and Watson note, this makes the argument abstract (given the norm in political philosophy), but I believe that is a good thing, for it helps us to achieve distance from our ideological convictions and intuitions to see the relations between our commitments in ways that may surprise and enlighten us (Johnson 2014). And as we abstract from the details, we open up ourselves to what Page calls "transdisciplinary" insights—that the logic of the problem in, say, politics, is similar to problems in management or evolutionary theory (Lane 2017, chap. 1). Thus *Tyranny* argues that Kauffman's *NK* and the Hong-Page models provide insights into what we might have thought was a distinctive feature of "ideal political theory." To be sure, we must proceed with care to ensure that our model captures the fundamental features we are interested in—thus my significant modifications of the Hong-Page model in understanding ideal political philosophy. Still, the aim is to abstract and see if we can distill the problem to its essentials, and so provide a general result. So like Rawls (2005, p. lx), I do not apologize for the abstract nature of the analysis.

Of Metaphors and Models

As Page explains in his essay (and in his path-breaking and, if I may say, often stunningly innovative work over the last two decades) different models can be employed to understand complex systems. I focused on what have been deemed "rugged landscape models." My idea was this. Some political philosophers have hit upon the metaphor of mountain climbing with two core dimensions: climbing up (achieving more justice) and moving laterally (getting closer to the arrangement that characterizes the most just social state, the ideal). The thought is that to move closer to the ideal on the lateral dimension may sometimes require first moving up (getting more just) but then, like an intrepid climber, going down for a while before forging up the next slope. Thus at times we are decreasing justice (altitude) in order to make our society more like the ideal (a closer latitude). (We can make the model more complex by adding other dimensions (*Tyranny*, pp. 258-9) but two were enough for my purposes).

Wiens's essay focuses on metaphors and models: what are we doing when we employ "metaphors," "models" (and "theories")? There is no canonical view; in her classic work Mary B. Hesse (1966) argued that models just *are* metaphors. I offer a somewhat different account here, according to which metaphors are typically basic or initial models. In scientific explanation we can distinguish primary and secondary systems; the primary system is the phenomenon (*A*) to be explained, the secondary system is the explanation of *A* in terms of the metaphor/model (*B*) (Hess 1966, p. 158). "Sound (primary system) is propagated by wave motion (secondary system)" (Hess 1966, pp. 158-9). Metaphors are critical in suggesting analogies, such that the primary system can be understood in terms of the working of some other system—which it patently is not. "For the conjunction of terms drawn from the primary and secondary systems to constitute a metaphor it is necessary that there should be patent falsehood or even absurdity in taking the conjunction literally. Man is not, literally, a wolf, gases are not in the usual sense collections of massive particles" (Hess 1966, p. 160). To take, say, navigating a mountain range (*B*) as a metaphor for pursuit of the ideal social state is to suggest that we can begin to understand the primary system *A* (pursuing the ideal) as having similar relations to navigating a mountain range (*B*). It is important, as Hess notes, that metaphors are not mere similes: we do not know ahead of time in just what way *A* is like *B*, or the ways in which it functions analogously. When we think about *A* as a *B,* we begin to think about *A* as acting as we know *B* does, and we look for familiar features of *B* to see if they are in *A* as well. We can use *B* as the basis of analogies that help us understand some of the puzzling workings of *A*, and provide the basis of further investigations.

This already is a model; we focus on the features of the primary system that are revealed by our more thorough grasp of the metaphorical secondary system. At this stage *B*'s dynamics and features are critical in picking out and understanding *A's*; still employing analogies we search for parts of *A* that seem a lot like *B*'s, and use our knowledge of the way *B* works with those parts to model how *A* must work too. For example, early electricians sought to understand electrical phenomena (*A*) as a liquid (*B*); and since *B*

could be bottled, so should we be able to bottle *A*—hence their successful effort to develop the Leyden jar (Kuhn 1970, p. 17). In my view this metaphor-analogy stage is often the first step in building a rigorous model of *A* that ultimately jettisons the metaphorical *B* as the base comparison and, so, no longer employs metaphors or analogies. The model is stated in axioms and equations, and variations in these, and anomalies encountered by our current model, suggest further developments.

Perhaps even after a formal model is developed the original metaphor may be used as a basis for speculating what variations of the formal model are worth exploring, though often (and this is my point) at some juncture the original metaphorical base, *B*, may become a hindrance rather than a source of further discovery, even if we continue to keep some of the labels suggested by the now-discarded metaphor. Thinking of conflict as a game blossomed in Prussia in the nineteenth century in the fad of "Kriegspiel" (literally, "war game")—a board game of conflict, played both by the public and the general staff. Indeed, in 1825 the German chief of staff proclaimed "It is not a game at all. It's training for war!" (Poundstone 1992, chap. 3). Young John von Neumann and his brothers played their own version of Kriegspiel. Game theory built on this metaphor; and we still have terms like "players," "moves," "strategies," and so on. But von Neumann and others developed "game" theory such that now its categories and relations are strictly defined within it, and so reference back to the way board games function is not apropos. It has its own, well-defined, concepts and mechanisms—in Wiens' terms its own "mathematical objects"—and any lingering game terms are simply for ease of exposition or to help beginners by invoking in their minds the now-discarded metaphor, which can help them to begin to see their way around the fully formalized model. (On the other hand, the atavistic labels can be an impediment, as when neophytes are told that they are "playing a game" and so, going back to a "game frame," play iterated prisoner's dilemmas not to maximize their own outcomes but to ensure that they "beat" the other "player"). In my view, however, game theory is not a metaphor: it is a formal model. We explain the primary system (people in interdependent actions) in terms of a formalized secondary system.

Hence my basic idea. The mountain range metaphor recurring in the political philosophy of ideal justice is a basic model that gives us a clue to some important features, but this metaphor has been superseded by fully formalized and developed models employed in fields such as evolutionary biology and complexity theory. The development from model-as-metaphor to a fully axiomatized (which is not to say fully developed) model has already been made, and was awaiting exploitation by political philosophers. As Page explains in his essay, I relied extensively on Stuart Kauffman's *NK* model. In Kauffman's model, all the terms are fully specified and the relations mathematically determined. There are Boolean nodes (*N*) that may be linked to *K* other nodes each of which can be "turned on or off" by their connected nodes. If we are modelling genes, the state of each node (gene) affects the fitness of the organism. When *K* is greater than 0, some genes are interconnected, and so the overall fitness of the organism will not simply be an additive function of the fitness of each gene, but of their number and degree of interaction.[1] This implies that when *K*=0, an organism *O*' that is a one gene variant of *O* will have a fitness highly correlated with the fitness of *O*. When *K*=*N*-1, the fitness of *O*' will not be correlated with *O*. For ease of exposition we can label the former a "smooth optimization landscape" (when we graph the fitness of variants they will increase or decrease in smooth lines) and the latter "maximally rugged" (fitness values may jump all over the range from one variant to the next). On all but the smoothest such optimization "landscapes" there will be local optima (any one gene variant will be less fit), which are not the global optimum. Each optima can be called a "peak." For now the critical point is that once the model is specified "landscape," "ruggedness," "peaks," "height" and so on are fully determined by the value function (e.g., adaptiveness, justice), the structure of the domain (the measure of genetic or world variations) and the *NK* dynamics. These terms are no more metaphorical than "player" and "strategy" are in game theory; they can be replaced by purely formal notation and nothing would change.

I have spent perhaps too much space on the relation of metaphors and models, but I have repeatedly encountered political philosophers who, confronting the rugged landscape model, conclude "well, that's a nice (or bad) metaphor, but let's get beyond metaphorical talk and do some *real* political philosophy" (e.g., contemplate our intuitions or the perennially-fascinating question of whether "ought" implies "can"). I hope it is clear that this is a basic misunderstanding of the place of both metaphors and formalized models in inquiry.

II. COMPLEXITY AND OPTIMIZATION

Pursuit of an Ideal and Ruggedness

Let us build on Page's {Market, Bureaucracy, Democracy} model to better see how the complex optimization model works. He writes:

> In the model, a society must allocate resources and opportunities across a set of domains. Within each domain, the society chooses among three pure institutional types: a market (M), a bureaucratic organization (B), or a democratic mechanism (D). For example, to select a construction firm to build roads, a society could hold an auction among qualified firms (M), it could construct a bureaucracy that develops criteria for selecting a firm (B), or it could hold a vote among elected representatives for the winner (D). If there exist ten such domains, then the set {X} consists of all vectors of length ten whose entries belong to the set {M,B,D}. Though a simplified characterization of the world, the model allows for a combinatorial explosion of social arrangements—59,049 distinct possibilities to be precise.

We see how quickly the number of distinct possible social worlds expand. To simplify, suppose we have only a four-domain world, any of which can be organized on market, bureaucratic or democratic institutions. Suppose:

> Domain 1: Decisions about supply of public goods
> Domain 2: Decisions about the supply and distribution of private goods
> Domain 3: Decisions about income distribution
> Domain 4: Decisions about the distribution of employment

Suppose we have a certain market socialist ideal, {DMDD} (public goods, democratic; private goods, market; income distribution, democratic; employment allocation, democratic). We are now at {BMMM}. One thing we might do is simply list the justice of all 81 possibilities, but it is hard to know, say, how the {MBDM} world would function: what would a social world be like where public goods are determined by the market, private goods are distributed by a bureaucracy, income distribution is democratically voted upon, but allocation of employment is via the market? Hmm. The ways all these mechanisms would interact are, *Tyranny* claims, extraordinarily difficult for us to model and predict, as our social science is based on understanding market provision of private goods and market (and some bureaucratic) distribution of incomes.

On *Tyranny's* analysis, ideal theory presents a perspective on justice that *orients* our quest for perfect justice by locating the ideal in relation to our current social state. A perspective (*i*) identifies what social states are similar to others (critical to orienting us in the quest for justice), and (*ii*) assigns a justice score to social states given their expected functioning. Considering just the first function, in our case we might have, say, the following perspective:

{BMBM}—{**BMMM**}—{DMMM}—{DMBM}—{DMDD}
(ideal)

On this perspective we are presently at {BMMM}; public goods are allocated by a bureaucracy, all others by the market. According to this democratic socialist perspective the ideal is a condition where public goods are allocated democratically, private goods by the market, while incomes and employment opportunities are decided democratically {DMDD}. Now on this perspective a world where public goods are decided by a democratic rather than a bureaucratic mechanism—{DMMM}—is pretty close to ours, and we can estimate how it would function and its justice. After all, for a democratic vote to replace a bureaucratic decision does not seem a huge jump in the social space. Bringing in now the second (value) function of a perspective, suppose this perspective judges that we are presently at 50, and the ideal is 100. What about {DMMM}? Here is a distinct possibility: while moving from {BMMM} to {DMMM} makes our society's structure closer to the ideal {DMDD}, it could decrease justice. In {DMMM} public goods are decided by vote but incomes by the market. Many social democrats such as Rawls believe that great wealth corrupts democracy, and so the provision of public goods in {DMMM} might produce a less just society, where the wealthy control their provision more than at present (so justice goes down to 40). On the other hand {BMBM}, instead of democratizing decisions about income distribution, puts them under bureaucratic decision-making. This, plausibly, leads us away from "bourgeois democracy" and its ideal of self-government by instituting an expert bureaucratic elite who end up exercising great control over the economy (Schumpeter 1950, 296ff; Levy and Peart 2017). Yet, it may lead to more distributive justice, and perhaps is superior in overall socialist justice to {DMMM}, say 60. Hence "The Choice" in Figure 1.

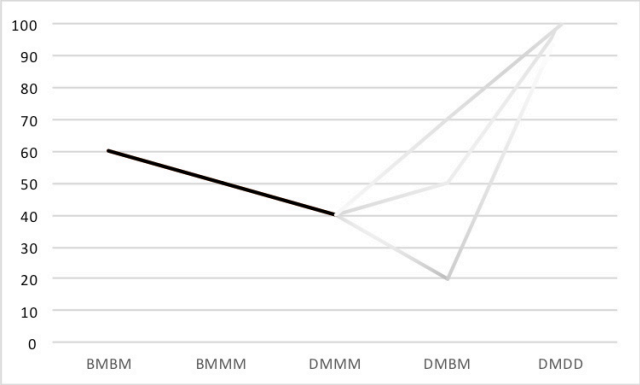

Figure 1.

At our present {BMMM} we have a choice between moving to {BMBM} with a justice of 60, but at the cost of making our society's institutional structure less like the ideal (we move away from it). If we seek to make our society conform more closely to the institutional structure of the ideal, we will suffer a loss in justice. After that (as indicated by the multiple light grey lines), we are uncertain as to the justice of the next step, as we are modeling a world where the bureaucracy controls income distribution but not employment allocation. The ideal theorist need not always choose to move toward the ideal, but if the ideal theory is to provide significant guidance, she must often choose to pursue the ideal and forgo local improvements in justice. The Choice is really a choice.

It might be wondered, if we know so little about {DMBM}, how do we know so much about {DMDD}, the ideal? And, of course, that is *the* question: how can we be sure of the functioning, and so justice, of a set of social arrangements that are so different from our own? Should we ever get to {DMBM}, it is almost certain that we will drastically re-evaluate the ideal, {DMDD}. As we approach the ideal, it changes before our eyes, perhaps receding into the distance (or perhaps, alas, we spot it in our rear-view mirror).

As the number of domains, dimensions of evaluation, or distinct institutional structures involved in an ideally just society increases (N), and as the justice-relevant interdependencies among them increases (K), it becomes essentially certain that The Choice will arise. Nuefeld and Watson are certainly right that institutions must be "mutually realizable" but that is not enough for their evaluation given the inevitable coupling in their functionings: some mutual realizations lead to excellent interactions (from the perspective of justice) while others interact in detrimental ways. This is the root of the complexity and uncertainty that confronts all attempts to move us from one institutional scheme to a very different one.

I think it is important to stress that this is not a just-so modeling story. In their extensive fieldwork on real institutions Elinor and Vincent Ostrom stressed that institutions are composed of numerous rule configurations; the constituent rules have strong interdependencies, both with each other and with environmental conditions. "A change in any one of these variables produces a different action situation and may lead to very different outcomes" (E. Ostrom 1986 [2014], p. 111). When we talk about social states such as a "property-owning democracy" we are referring in a loose way to a myriad of interconnected rules and behavioral tendencies that constitute the working of the set of institutions that are summed up by this moniker. Building these institutions inevitably leads to problems of complex optimization and searching.

The Fundamental Diversity Insight

Vincent Ostrom (1972 [1999], p. 125) once remarked in respect to his empirical findings, "[t]he complexity of relationships … is such that mortal human beings can never observe the 'whole picture.' Anyone who attempts to 'see' the 'whole picture' will 'see' only what is in the 'eyes' or 'mind' of the beholder." This insight is formalized in the Hong-Page model, where diverse perspectives, each seeing (coding) the problem in different ways, each see possibilities to which the other is "blind." For example, one might wonder about Figure 1: why is BMBM a move *away* from the ideal, rather than step toward it? The Choice would not be confronted by the perspective:

{DMMM}—{**BMMM**}—{BMBM}—{DMBM}—{DMDD}
(ideal)

The Hong-Page model shows us how different perspectives on a problem can help each other in their searches for the best outcome. We can benefit from interactions with those who see the world differently: what is a tricky problem for me might, given your perspective, be an easy one. This is a fundamental insight, and *Tyranny* spends a good deal of time evaluating its applicability to the search for ideal justice. Page's essay nicely summarizes both the Hong-Page model's resources, and the reasons why I conclude that it has a restricted applicability to the problem of ideal theory. For now, I merely stress that I strongly endorse (stated approximately)

The Fundamental Diversity Insight: Any given perspective Σ on ideal justice is apt to get stuck at poor local optima; other perspectives can help by reinterpreting the problem or applying different predictive models, showing better alternatives to be in Σ's present neighborhood.

Wiens believes the arguments for this and related diversity-relevant conclusions do not "rely much" on the rugged landscape *(NK)* analysis. I do not agree. It is just because we are navigating this sort of problem that our perspectives on justice are so apt to get stuck on poor local optima, from which we confront The Choice. Critically, as the Hong-Page model taught us, other perspectives can—sometimes—point the way forward.

A Tale of Two Models

Wiens's essay is, to a large extent, a contrast between his optimization model and the model we have been examining. Trying to be a little less formal, Wiens's (as I shall call it) "simple optimization model" seems characterized by:

1. A set of possibilities, *X* (options, states of affairs, social worlds, etc.). This is the domain of the value function on which it operates. In *Tyranny* they are alternative social worlds.
2. A binary value function (which I'll call *V*): the value relations among members of *X* are built up through binary comparisons. Letting (*x, y, z*) be elements of *X* (i.e., particular social worlds), *xRy* if and only if *x* is ranked by *V* at least as high as *y*. If *xRy* and *yRx*, *x* and *y* are ranked the same; if *xRy* and not *yRx, x* is ranked above *y*.
3. It also seems required that the binary relations generated by *V* be transitive among all triples in *X*; (*xRy*) & (*yRz*) implies *xRz*. Another requirement seems to be completeness, i.e., for all members of the domain *X* (*x, y*), either *xRy* and/or *yRx*. Other conditions on the ordering are allowed.
4. There is an additional set of constraints. These allow us to partition the domain, *X,* of possible social worlds, into those that meet these constraints and those that do not. Call X_C the subset of *X* that meets some given constraints *C*.

As I understand Wiens (and I'm sure I have not grasped all the intricacies of his model, so apologies in advance), conditions 1-4 imply that the only structure among the elements of *X* that are not based on *V* (the binary value relation) is the partition of *X* into the subset X_C and the rest of *X*. Consider, then, simply the X_C partition. For all the social worlds in X_C the only structure relating them is that yielded by *V*. Until the binary ordering is applied to the members of X_C, they are an unordered set of social worlds that meet certain constraints.

Perhaps the most fundamental difference between simple optimization and *NK*-optimization is that the latter models a structure among the members of the domain (possible social worlds) that is not generated by the value function. As Page notes in his essay, we can make this point by saying that the simple optimization model's theory of justice *T* only has a measure of the range of *T* (the ordering produced by the value function) while *Tyranny*'s model has independent metrics for the range and domain of *T*. Without these two distinct metrics, as we have seen, The Choice does not arise. Rather than confronting an unordered set of possibilities to which we apply our value (justice) function, *Tyranny* assumes that all the members of *X*—"social worlds"—have certain justice-relevant features and these generate a structured domain, to which we apply our value function. (We should refrain from calling this domain metric simply "descriptive," as it is generated by the similarity of worlds' justice-relevant features, so they are, we might say, normatively loaded descriptions.) And in most of these landscape models—e.g., those in evolutionary biology—the structure of the option set is the similarity of the constituent features of the options, such as genotypes. If *x* is almost but not quite identical to *y* in the relevant respects (and, as always, "relevancy" is defined by the perspective or theory being employed, see below), then before any value function is applied, *x* is located close to *y*; if the defining features of *z* share very little with *x*, *x* and *z* will be located far apart. In a simple evolutionary biology model, *y* might be a one gene variation from *x*, and so we can entirely correctly say—before we know their adaptiveness (value), *x* and *y* are close (though it could turn out that in terms of adaptiveness they are not).

What Model Should We Use?

So which model should we use? A model that only looks at the range of value (i.e., the justice of the options) or one that includes a meaningful structure (i.e., the institutional similarity) of the elements of the domain? As soon as we phrase the question so bluntly, we see how misguided it is. The Fundamental Diversity Insight indicates that different models often provide different insights, so we are all apt to

benefit when multiple models are pursued (Page 2016). The model we employ depends on our perspective on the primary system and our theoretical aims. As we know from economics, in many cases a simple optimization model, which only orders according to the value function (i.e., utility), elegantly analyzes the problem, say of most consumer choice. But as Harold Hotelling (1929) showed in his analysis of the location of shops, sometimes we get more insight by also including a structure of the option set—in Hotelling's case, the geographical location of shops along a street. Anthony Downs (1956) saw that this "spatial" model could be extended to include "ideological space" (as well as utility values defined by the value function) and so began spatial models of politics. Downs's model was not metaphorical, even though it was developed from a "geographical" model: ideological space was well-defined.

Hotelling's and Downs's models included a structured domain (as well as a value function), but were not *NK* models. I have tried to indicate why I believe the ideal of orienting our quest for justice by a fully just social state is well modeled by a rugged landscape in which our reliable knowledge is confined to our current neighborhood. Given the large number of institutions and background conditions that constitute an ideal social state, and the myriad of couplings between them that result in vastly different social states of different degrees of justice, setting out to pursue the ideal is an exercise in Knightian uncertainty (Knight, 1921 [1964], chaps. 7-8). For the most part, we only have useful assignments of probabilities within our neighborhood, so we simply do not have the information necessary for a fruitful simple optimization exercise.

Consider a mundane case: a manufacturer searching for innovations in product *P*. The advice to formulate an ordered list of possible variations of *P*, and then take the best choice is not of much help; the manufacturer cannot assign values to many of the options. Indeed, she doesn't yet even have the blueprints for many variations. So here is an entirely sensible approach to the development problem. Have most of the research teams work on near improvements (slight modifications of the technology underlying *P*); because there is so little fiddling with the structure of the present *P* we'll often find the new versions slightly better or slightly worse, and can further build on the slightly better ones, and then build on some of those, etc. This is a conservative "climb the gradient" heuristic, always seeking local improvements. But we may get stuck at a product for which no small improvements could be made (yet is not the ideal *P*). Think of SONY teams that were working on the best Betamax recorder.[2] So the manufacturer may also wish to invest in an R&D department that has some teams working on more radical innovations (say, laser discs), some of which could result in really high values, but it is almost impossible at this point to make sound judgments about whether they will pan out. Here we are dealing with hunches, hopes and dreams—not probability assignments. In the case just described, we are searching the value of options with a certain structure, and these structures are the very properties that yield valuable products. This structured space thus *orients* the product search. It is important that in this case one's optimization problem *starts* from a location (a place in the structure), and we are thinking of how to best move *given where we already are*.

The claim of *Tyranny* is that an interesting class of ideal theories (I never say "all") have much in common with this case, except that these ideal theorists believe that they have already developed the blueprint for the perfect (or at least truly excellent) "product" but we can't build it right now (the blueprint may require components not yet developed, like public-spirited folk). Our job is, given our present location in the domain, to begin to develop structures that are more like those that generate this great result. That is what it means to *build* an ideally just society. To build is to assemble the components. Because the task of such ideal theories is to work from our given structure to the ideal one (and thereby achieve the perfect value score), an optimization model that includes structure (which is not simply generated by the value function) is necessary.[3] It is in this sense that, in a structured domain, the ideal "orients" improvements in justice in a much more complex sense than does the simple injunction "given some set of unordered social worlds, choose by maximizing the value function subject to constraints."

Does a Simple Optimization Model Capture Ideal Theory?
Although I favor an ecumenical approaching to modeling, I am not a model nihilist who supposes that one modeling choice is as good as the next. *Tyranny* adopts a theoretical perspective on analyzing theories of ideal justice which, I think, is in many respects superior to a simple optimization model based on a binary value relation. Deriving "perfect" or even "good enough" from a value function that can only yield judgments about what is "better" is, I think, a job and a half. Theories based on a binary value function can only say that an option is the best in the sense that it is better than all the options in some set of options. So to say that a certain social state is ideally just is, essentially, to say that

it is better than all other social states in the domain X, or more realistically in some X_C partition. This makes it both too easy and too hard to find the ideal. It is too easy, because when a theory has some non-empty option set and can identify a best element, the theory then has apparently located an "ideal"—essentially every optimization exercise ends up as an ideal theory (which is why Wiens can see the Open Society as an ideal, since I think there is a set of "devices" that are maximal choices, even if no set is the best choice, to accommodate diversity). On the other hand, it is too hard to identify the ideal, as we need an exhaustive enumeration of all elements in the set X_C of possible worlds to identify a "best" that is not simply dependent on the choice of the comparison set.[4] Thus we can only be confident that i is the ideal world if we are confident i is not merely the best in some subset of X_C, but remains best when the set comprises all of X_C: i is better than each and every element in X_C. In formal terms, it must be the unique choice set from X_C. Showing that is a daunting task, unless we very tightly specify the constraints that define X_C. I don't think this is the usual way that ideal theorists have reasoned in the history of political thought. On my view, *The Republic,* More's *Utopia* and Bacon's *New Atlantis* are paradigmatic. Construct first an imaginary social world in which humans would relate to each other with full justice, etc., and *that* is the ideal. We score *it* as fully just, without a complete ordering of all worlds in X_C. That then becomes our inspiration, even if we do not know of many of the other worlds in X_C. Indeed, the common utopian theme that the ideal is a far-off land to which we have yet to find a path-to-construct-it implies that there are many unknown options between it and where we now are.

Of course to inspire it does have to be absolutely perfect—we might construct a world that is within the human horizon of workability and yet is essentially but not ideally just. For Rawls identifying such a "realistic utopia" was a critical task of political philosophy. At various times he agonizes over the question whether a reasonably just society is within the grasp of humans. "The wars of this century with their extreme violence and increasing destructiveness, culminating in the manic evil of the Holocaust, raise in an acute way the question whether political relations must be governed by power and coercion alone. If a reasonably just society that subordinates power to its aims is not possible and people are largely amoral, if not incurably cynical and self-centered, one might ask with Kant whether it is worthwhile for human beings to live on the earth" (Rawls 2005, p. lx). It is hard to see how showing some social world is "ranked first in our option set" could possibly assuage such worries. Rawls seeks to find out whether the *best* among our options is *truly just*—and if there is one that is truly just it is to orient our long-term endeavors at reform.

Perspectives, Normalization and Generality

As we have seen following (though, as he points out, substantially modifying) Page's framework, I argue that a *perspective* provides an orienting structure to the complex optimization problem. A perspective on justice includes, as it were, all the elements needed to generate an ideal theory. It includes a set of evaluative criteria (perhaps liberty and reciprocity, perhaps sanctity and respect for authority), an identification of what parts of a social world are relevant when evaluating it in the light of these criteria and a model of the way these features interact to provide an overall social order, which then can be scored in terms of its justice. In addition, a perspective must have some view of how similar worlds are given their justice-relevant features; it must be able to say world x is almost identical to y in terms of the institutions, rules, motivations and so on that define the worlds. Now fundamental to *Tyranny* is that how all this is accomplished is *internal to a perspective*: in Wiens's language these *must* be "black boxes" the contents of which the model is "blind" to. Political philosophy is normally devoted to explicating the correct perspective, giving accounts of what the correct evaluative criteria are, how they should be combined and what institutions are relevant to justice. As D'Agostino puts it, this is part and parcel of the "legislative" stance in political theorizing. In addition, an ideal political philosophy seeks to show what the ideal is, and how far we are from it. Call this *a fully normalized view of ideal justice*. Sometimes, as with social contract theory, a set of perspectives is identified as "correct enough" because of their similarities in evaluative standards or their basic agreement on the relevant institutions; call this a *partially normalized view of ideal justice*. It was Rawls (2007, p. 226) who notes that all social contract theory supposes some normalization—all take a peek inside the box and seek to give us some idea how it ought to function.

Of course we often do care about exploring evaluative criteria or, for Wiens, the logic of justification. When we do so we seek to develop a fully or partly normalized theory—one that commences by identifying a set of correct or minimally acceptable perspectives. However, the aim of *Tyranny* was to present a fully general model of ideal theory, one which analyzed ideal theory *qua* ideal theory, and not *qua* liberal ideal theory, or *qua* ideal theories that embrace a view of

the logic of justification, or even ideal theories that accept a standard secular understanding of social reality. Two aims were regulating here. First, as I have been stressing, the aim was a fully general theory that uncovered the logic of ideal theorizing itself, as far as possible freed from familiar substantive commitments (a "pure theory of the ideal," we might say) and, secondly, to explore how a society with a truly radical diversity of perspectives, each committed to its own theory of justice, might not only live together, but learn from each other. It was thus important not to start out by imposing any normalization on perspectives.

Still, as Wiens perhaps suggests, isn't the model itself a form of normalization? It says that a perspective must have evaluative criteria, must identify features of the social world, must be able to score them and so on. Isn't this just another normalization? We might call this the "all theory implies normalization" objection. And this raises one of the most perplexing issues for what we might call ecumenical theories of a phenomenon (see Gaus, 2017b). Think of the difference between an objective theory of value (say, the labor theory) and a subjective one (such as Carl Menger's). On the one hand, a subjective theory seeks to allow that value is, in a basic sense, up to the agent, and she can place value where she wills. In that sense it is ecumenical about what can be "correctly valued." Yet, because it is a theory *of value* it must delimit its inquiry—valuing isn't the same thing as sleeping (even if sleeping is valuable). To give a theory of *X*, one must identify a class of *X*-phenomenon. Unless we characterize *X* in *some* way we can't even begin to talk about "it," and unless our characterization has enough structure we can't begin to analyze it. So, yes, categorizations and specifications cannot be avoided in theorizing. If all categorization implies normalization, so be it. The important point for *Tyranny* is that, having identified the basic elements of a perspective, the analysis does not proceed to employ additional criteria to distinguish "good" *v.* "bad" or "reasonable" *v.* "unreasonable" variants (though each perspective may make such judgments of the others, as we are about to see).

III. THE OPEN SOCIETY

The Fundamental Diversity Dilemma

As I said above, *Tyranny* relies on Hong-Page reasoning to endorse *The Fundamental Diversity Insight*. A point of departure from their model is my analysis of, and emphasis on, the *Fundamental Diversity Dilemma*. A radically different perspective from Σ may provide important insights to Σ, but adherents of Σ will have great difficulty making sense of this radically different perspective. As Σ sees them, they categorize the social world in very odd ways, and employ strange evaluative standards. This should be familiar to everyone. Think of the radical atheist's claim that religious perspectives suffer from a cognitive dementia, or the religious view that the atheist is infected by evil. Under these conditions both see the other as saying barely intelligible things, and they certainly do not find each other's "searches" for the ideal of value. In a diverse society, sometimes we can learn a lot from other perspectives, sometimes we can use some of their discoveries, and sometimes we will dismiss them as lunatics.

As D'Agostino has argued in his marvelous *Naturalizing Epistemology* (2010), the trick is to "get it together," to assemble the insights of our different research programs in a way that leads to mutual enlightenment. Adapting D'Agostino's analysis of scientific communities, I employed his idea of a "republican community" to designate a community of moral inquiry, i.e., one that shares sufficient standards, problems and concerns such that the results of the searches of some in the community can be taken up by others. Of course such "republican communities" come in all varieties, from the thick who share many deep commitments, to those who sometimes find some results of the others as somewhat enlightening. Employing a "small world model" *Tyranny* argues that a society composed of diverse groups, some of whom entirely dismiss each other's insights, might, nevertheless, be one in which everyone learns from everyone else. The basic idea is straightforward. Our Radical Atheist proposes committing the Evangelical Christian to an asylum, while the Evangelical Christian dismisses the Atheist as vigorously. They will never directly share insights. But, say, a Roman Catholic Scientist may engage the Evangelical and, in turn, what we might call the Broadminded Secular Scientist is willing to engage the Roman Catholic Scientist; so there is a line of engagement from the Evangelical Christian to the Broadminded Secular Scientist. If the Radical Atheist engages the Broadminded Scientist, the network of mutual influence is complete. The Evangelical Christian and the Radical Atheist may end up enlightening each other. As we multiply the number of perspectives, diverse, crisscrossing networks of various "republican communities" of inquiry will arise. Such networks mitigate the Fundamental Diversity Dilemma. The Open Society, I argue, provides a framework that allows these networks to develop, and so a framework that allows each view of justice to better cope with its own internal challenges.

Communities of Moral Inquiry v. Moral Communities

Vallier's essay raises the important question of the relation of "republican" communities of inquiry to moral rule networks—people who share moral expectations and demands. Let us call a "moral community" a set of individuals who (*i*) share very similar perspectives on justice, (*ii*) interact on moral rules grounded on these perspectives and (*iii*) share very few moral rules with those outside of (*ii*). In the history of political philosophy, many have thought that an ideal way to cope with moral diversity would be for us to divide up into numerous like-minded moral communities (Nozick 1974, chap. 10). *Tyranny* tries to show why that is an error. In a network in which the moral rules reflect only similar perspectives, members are able to act on views of justice they take as superior, but at the same time they lose resources (interaction with other perspectives) that have the real potential to enlighten them about justice. What constitutes a "republican community of inquiry" is in constant flux; as Σ develops, perspectives that Σ previously saw as beyond the pale of sensible inquiry become intelligible, while some that it previously deemed sensible may now appear antiquated. *Tyranny*'s concern is with those who are devoted to knowing what real justice is, and that cannot be secured in like-minded moral communities. Thus the crux of *Tyranny*: the subjects of my inquiry—those who are deeply devoted to knowing what a just society is—have powerful reasons to embrace the diversity of the Open Society.

One important difference between *The Order of Public Reason* (2011) and *Tyranny*, then, is that the latter is focused on those who are convinced there is a notion of perfect justice, and are committed to knowing what it is. Such folks would seem to be an especially hard case for a theory of public reason as given in *The Order of Public Reason*, as their primary concern is *getting justice right*. Why would they, of all people, want to share a moral order with people who are getting it *wrong*, much less accommodate them in some way? *Tyranny* (the book, that is), is an extended answer. Surprisingly (at least it was surprising to me), if you want to really get justice right, you must live with, and learn from, those who get it wrong—*if* knowing the perfectly just society is a complex problem.

Polycentrism and Accountability

A second difference concerns the model of moral relations. As Vallier points out, *The Order of Public Reason* analyzed morality in terms of the social rules endorsed in some group, *G*. It was supposed that the persons in *G* share a network of moral rules. As I noted, groups come in many different sizes, but my concern was typically with the largest interacting group, what Hayek called "the Great Society." *Tyranny* develops a more nuanced account—polycentric moral networks. What I have called "moral communities" are typically confused with what Elinor and Vincent Ostrom called "polycentrism." The Ostroms argued that a "highly fragmented" system in which different groups were largely confined to their own "jurisdictions" (or communities) is apt to result in conflict and institutional failure (Ostrom and Ostrom 1977 [1999], p. 96). They thus modeled a successful polycentric order in terms of many crisscrossing and overlapping jurisdictions and norm networks. *Tyranny* relies on polycentrism thus properly understood. In *Tyranny* "the group" dissolves into complex and overlapping moral rule networks. On some fundamental matters, the moral rule network is coextensive with the Great Society of strangers, the group that I had in mind in *The Order of Public Reason*; but on many other matters "the group" disaggregates into moral networks that are subsets of *G*. As the Ostroms stressed, whether moral rules and institutions are needed in the eyes of some individuals depends on what sort of problem they are facing; some problems can only be solved with the participation of almost everyone, others have much more restricted scope.

Within each moral network—i.e., people who have shared normative and empirical expectations about what is to be done in some circumstances—each is accountable to others in his network for violating the rule.[5] This is not the case in relation to outsiders: one is not answerable to outsiders for one's failure to conform, nor can they hold one accountable. So if, in a moral network of vegetarians, Alf defects and has *foie gras*, he is not answerable to me (a devoted carnivore), unless we begin to tell a story with more detail that draws me into the matter. But he is accountable to others who share the rule, and have well-grounded normative and empirical expectations about Alf's eating habits. In the Open Society, of course, Alf can withdraw from many such moral networks, as others can join them. This is an engine of moral change.

This is a critical point. Philosophers usually see morality as homogenous—and each is accountable to everyone else for failing to act as morality requires. When we look around us, we see the moral world is not at all like this. A normal university professor participates in a wide array of moral networks. At the university as a scientist she participates in networks that have high expectations about evidence and impartiality; a student who flouts them will be the subject of intense moral criticism and, perhaps, severe punish-

ment. But as a member of a church she does not hold others accountable to them: it would be inappropriate for her to start condemning her pastor for lax evidential standards in his Sunday sermon. She may be a vegetarian and hold her university circle of vegetarians to high standards, but she would display outrageous behavior if she walked into a Subway and commenced to berate customers' choices. This is not to deny that the professor might believe that her pastor really ought to pay much more attention to evidence, or that everyone ought to refrain from eating meat. Her *judgments* about morality may or may not be universal. But this *is* to say that the conditions to hold the pastor or the Subway customers accountable do not obtain, and so their actions are not her business—that is, *they are not accountable to her* for their violations. We navigate these various networks of accountability constantly—so unconsciously that we may be surprised that is what we are doing.

Trendsetter Networks

Although *Tyranny* is critical of the idea that republican communities of inquiry should form "moral communities" (as I have defined them), we must distinguish a moral community from a group of individuals, perhaps who are joined in a republican community of inquiry, who seek to establish a new rule on some matter (Bicchieri 2016, chap. 5). In a polycentric moral order this innovative activity will be restricted to a single moral rule or a small set. Those adopting this new rule will be enmeshed in many other moral rule networks, and so constantly confronting diverse perspectives. They will not form a moral rule community.

On some matters a polycentric order can contain competing moral rules. Given such competition a trendsetter group (say, university students in the '60s), begin experimenting with a new rule (say, concerning sexual morality), withdrawing (usually very publicly) their allegiance to the old. In the case of sexual morality the new networks expanded but did not go to fixation—in many towns and rural areas, and in much of the south of the United States, the old rules of sexual conduct held pretty firm. Moral innovation on some rule of social morality certainly can, then, occur by spreading out from a trendsetting network (that may also be a republican community of inquiry). As Robert Boyd and Peter J. Richerson (2005) have argued, under some conditions group beneficial rules can spread very fast. Sometimes all of society will cascade to a new rule; at other times this competitive process leads to a polycentric order in which different sub-networks adopt different rules (Gaus, 2017c).

Vallier's essay is especially valuable in stressing how moral innovation requires moral space—a protection from universal accountability—for moral diversity and experimentation. So very often today those advocating moral change take their role as requiring that they browbeat others to conform to the moral rules they and their core network are convinced are correct (Gaus 1996, 123ff). They hold the world accountable for not living up to their convictions. An aggressive moral self-righteousness is often seen as mandatory for anyone who seeks a more perfect justice. Anything else is "relativism" or, to again harken back to the '60s, liberal "repressive tolerance." That, however, is to ossify one's current understanding of justice and to undermine the social conditions for knowing a more perfect justice. We are not at the end of history—and that includes our knowledge of a perfectly just social state.

The Fundamental Rules of the Moral Constitution

Some rules, however, are so fundamental to cooperative social life that it is very difficult to have sustained interactions with those who do not adhere to the same ones we do. If some reject our rules about harm to bodily integrity, property, truthfulness, etc. it will be immensely difficult to share a social life with them. These fundamental rules require a different analysis of moral reform than the competing networks account. When it comes to the rules of sexual morality, university students could practice their own rules freed of the hang ups of the rest of the society (and eventually convert a good deal of it). Sexual relations are certainly social, but even in the'60s they weren't especially large-scale social phenomena.[6] Some can go their own way. But, leaving aside retreating to a commune, it proved impossible for the '60s trendsetters to informally change property rules. Given our own deep commitments, we have strong reason to share a basic framework of such moral rules, but given the diversity of our moral perspectives we disagree about what those rules should be.

The Order of Public Reason introduced the idea of a socially eligible set of rules, which *Tyranny* employs. The basic idea is straightforward. In a large diverse social network (the Great Society, for example), individuals have clashing views about the specific form these moral rules should take (what should be the rules about harm? what should be the rules about property?). However, these types of rules are so fundamental to an ongoing scheme of moral accountability—which itself is fundamental to effective social life—that almost all individuals are prepared to endorse and adhere to specific versions that fall a long ways short of their most fa-

vored formulation. The socially eligible set, for any specific matter to be governed by a moral rule in a particular social network, is the set of all the rule variations being advanced that everyone (or, as near as possible, everyone, a point to which I shall return) in the network has sufficient reason to conclude are at least minimally worthy of endorsement and adherence. Since for these fundamental rules we really need to coordinate on a single, shared, rule, the socially eligible set identifies the set of possible rule variations all can understand as grounding a shared practice of moral accountability. For every rule in the socially eligible set, each person would endorse and adhere to it should it be the rule that their society (network) has hit upon.

How, then, can an Open Society change such rules? As I argued in *The Order of Public Reason* democratic government reforms the fundamental rules of our society through legislation. When the law moves us within the socially eligible set, the result is still a rule that everyone holds is a basis of genuine moral accountability, but many believe is a better rule (given their moral perspectives) than the one it replaced. Democratic decision making is, I think, indispensable is reforming our truly fundamental rules. In *Tyranny*, however, I explored informal, social mechanisms by which social movements can seek to move us to what proponents see as a superior basis for accountability. One interesting avenue I consider is to exploit rule ambiguity. All rules are ambiguous in many places, and typically in such situations others will accept several alternative actions as plausibly fitting the rule. In these circumstances we can nudge the rule in directions we morally favor (in the eligible set), without denying the validity of the normative expectations of others. For example, when I was growing up the basic rules about physical violence toward children allowed corporeal punishment, but not too much violence. The ambiguity about what constituted unacceptable violence toward children allowed a change in the basic rules, as some began to employ increasingly stringent interpretations of what "violence" was and when it was unwarranted. This was extraordinarily effective in changing the moral rule. When I was a lad seeing a mother smack her child in the supermarket was not especially rare or noteworthy. Now it is clear violation of a basic rule.

The Limits of Moral Space

In an earlier paper D'Agostino (2013) worried that, instead of finding an "eligible set" of acceptable moral rules, the "null hypothesis" will hold—there will not be *any* rule all deem eligible—worthy of endorsement. His paper in this symposium develops this worry: there seems to be precious little chance of an eligible set in a society split into opposing perspectives that "are increasingly likely to treat those whose adopt different social ideals as pariahs, unworthy of moral regard." The obvious case here is the great animosity in the United States between so many Republicans and Democrats.

D'Agostino is surely correct that this dismissal of the moral status of others is one of today's most serious threats to the Open Society. However, while recognizing the danger this poses to the Open Society, I also think we should be aware that to a large extent this a political problem more than a generalized moral one (reflect on the obvious example). As Milton and Rose Friedman (1980, p. 66) long ago pointed out, political decisions require "conformity without unanimity" whereas self-organizing systems (like the market) produce "unanimity without conformity." In a highly morally diverse society, when political decision-making pushes beyond maintenance of core rights and liberties to the legal codification of deeply controversial conceptions of justice, hostility and contempt for the law is apt to be triggered (Gaus 2017a). Politics is ill-equipped to cope with deep moral disagreement (i.e., where the null hypothesis holds). Each party, hopeful that a majority win in the next election will allow it to institute true justice, simply sets the stage for the next iteration of mistrust. To argue that democratic societies need to develop more trust, while seeing them as a struggle about which controversial conception of justice will be imposed on the appalled minority, is ultimately incoherent. One cannot make politics a justice jihad and hope to induce trust (see Vallier, forthcoming).

The idea of a polycentric moral order helps to show how we can secure "conformity without unanimity."[7] Although we are currently understandably obsessed by the hatred underlying so much American politics, we should not forget that Democrats and Republicans share a myriad of rules about bodily integrity, property, gender equality (yes, though they disagree on the policies to pursue it). They cooperate in neighborhood organizations (my own neighborhood, for example, has about an equal distribution of hybrids and pickups, yet an active neighborhood organization). The more our moral rules track networks of individuals seeking to live together and solve their social problems, the less the null hypothesis will be a worry.

I do not want to seem Pollyannaish. The debate about the status of abortion rights is a deep moral disagreement that inevitably flows to political dispute. Even about this vexed issue, however, I think a more decentralized politico-legal

system that is responsive to the differences of locality and region could at least mitigate the depth of hostility (Gaus 2017a; Kogelmann 2017, chap. 4). A morally diverse society is bound to have deeper political conflicts than one that gives the appearance of moral homogeneity (because some perspectives have been silenced or have been dismissed as unreasonable). Politics amplifies disputes, and these can obscure well-functioning informal networks on which an extended order of cooperation is built. Perhaps the main message of my last two books has been the importance of the social and informal, which contemporary political philosophy almost wholly ignores.

Still there are limits. I agree with Rawls in holding that the moral space of endorsed cooperation is always limited. Some will simply refuse to live with others on terms that ground mutual accountability and mutually endorsed expectations. Recall, though, that Rawls (2005, pp. 197-8) credits Isaiah Berlin with this thought. Whereas Neufeld and Watson see it as bound up with the Rawslian ideas of reasonableness and reciprocity, I see it as simply following from the recognition of Berlin's core theme of deep diversity. Given a deep enough diversity of moral perspectives, it is inevitable that some won't be able to see their way to endorsing the rules of cooperation that almost all others employ to structure their social-moral lives. Neufeld and Watson, as Rawlsians, would tend to describe these folk as "unreasonable." Of all the Rawlsian categories, this is the most vexed. While I do not believe it is hopelessly vague or confused, I certainly do not wish to employ it in any formal way. As I see it, typically these "excluded perspectives" are those who value their own purity (or, perhaps we should say, less controversially, their integrity) over reconciliation and living with others. I have recently modeled them as "maximum integrity agents" (Gaus, 2017c). These may be devoutly religious folk, or Kantian moral philosophers. I am not prepared to say the maximum integrity stance renders one unreasonable, but it does tend to make one unfit to live with many others. One need not embrace reciprocity to find a path to reconciliation with the moral perspectives of others: that is one route, but there are many—yet some may fail to find any of them.

However, we must remember that, *pace* the social contract tale, living with others on moral terms is not a single decision, such that one is "either in or out." That social-moral life is so clearly *not* like that should lead us to question the entire social contract approach. No one except a sort of cartoon nihilist (more plausibly, a psychopath) is a "holdout" on social morality per se, but rather one "holds out" on this or that moral rule in this or that network. Some who insist that the very idea of living with diverse others is an insult to their integrity will find themselves with a shrunken and impoverished moral space, but they will participate in some networks. However, they will not be able to reap the benefits of the Open Society. And the rest of us will have to beware of them and probably guard against them, as they may seek to undermine the basis of our public moral world.

IV. THE IMPERATIVE OF COMPLEXITY AND THE NEW PROGRAM FOR POLITICAL PHILOSOPHY

In 1971 Rawls's *Theory of Justice* revolutionized political philosophy, taking as its subject principles for a society characterized by enduring rational, normative disagreement. Today, in a society in which basic facts are in dispute and perspectives face each other with mutual incomprehension, the idea that all rational moral agents would agree on the difference principle[8] as regulating social and economic inequalities seems rather quaint. Rational moral agents disagree on the good life, but not about social justice! We can only try to remember when that seemed like deep disagreement. To his great credit, Rawls continually explored the basis of disagreement about the justice of our society, leading him deeper and deeper into the problems of social organization under diversity. Yes, I believe that at the end his political liberalism project was in disarray, in the sense that its unity and organization broke down (I would certainly not say, as Neufeld and Watson think I would, that it was a "jumbled mess"). As Chad Van Schoelandt and I (2017) argue, this disarray is a testimony to the protean nature of Rawls's project. He was constantly inventing new terms and advancing new analyses to capture evolving insights. I see no point in freezing an ongoing project at a moment in time, and insisting that it was finished. It manifestly was not.

What Rawls entirely failed to appreciate, however, was that as we make a system of interaction increasingly diverse, not only does the basis for consensus become increasingly thin, the system becomes increasingly complex. Diversity and complexity are intimately related (Page 2011). To accept that the subject of political philosophy is a system characterized by the dense interaction of diverse moral agents leads to the conclusion that its subject is in the formal sense a complex phenomenon. It is this "imperative of complexity" that is at the heart of the new program in political philosophy that D'Agostino announces in his essay. As Hayek

insisted throughout his career, constructivist blueprints for complex systems often look wonderful on the drawing board but simply cannot be built. The new program accepts the imperative of complexity, and so focuses elsewhere: on the perspectives of diverse moral agents and under what conditions they can organize themselves into a fruitful and cooperative social life that all endorse as moral—given their heterogeneous understandings of that contested concept. As I understand it, this new program switches focus from the moral convictions and plans of the philosopher to those of the agents who form the self-organizing and self-governing moral order. Once they abandon the legislative and planning tasks, political philosophers have much to contribute, especially if they are willing to engage in trans- and inter-disciplinary inquiry. In place of conceptions and plans for justice the new program—or at least my version of it—seeks to identify "devices" of public reason, to analyze how markets, democracy, polycentrism, liberty rules and jurisdictional rights provide the framework for diverse moral perspectives to form moral and political orders that not only accommodate, but improve, their disperate understandings of justice in an open society.

NOTES

1 For useful summaries, see Page's essay and Mitchell 2009, pp. 281-6.

2 I write here as a great fan of Betamax, being among the last to abandon it.

3 Sometimes simple optimization models seek to provide a non-*V* based structure among the members of *X* through appeal to feasibility; it is offered as a way to say what options are "close" and "far" in a sense independent of the value function. *Tyranny* argues that while feasibility is an important and often critical idea, it does not have the characteristics to define coherent and well-behaved structural relations among the members of *X*. We cannot build a better coherent structure by asking, at each moment, what improvements on the current structure are the most feasible.

4 More formally, I think we would want our binary-based theory of the ideal to satisfy some version of expansion and contraction consistency. See Sen 2017, pp. 317-23.

5 Thus the idea on which Vallier focuses: that a violation of a moral rule in one's network is "everyone's business." To be sure, I used this as something of a motto rather than a strict necessity, as a network can develop moral rules in areas like marriage (e.g. fidelity), but interests in privacy may lead to the conclusion that violations are only the business of the family unit. Nevertheless, without further considerations the general statement holds within any given moral rule network. In economic terms, rule violation is a public bad, so all are concerned when violations occur. Except for the special case of human rights, I did not, however, hold that conformity is the business of those outside the network (in *The Order of Public Reason*, outside of *G*). As I stressed, other groups may have different moral rules, and in most cases (human rights aside) one cannot hold them accountable for failing to conform to our social morality. So, yes, within some given *G*, violations are almost always a concern to all.

6 Although according to what my daughter calls "the source of champions," *Wikipedia*, up to 100,000 people participated in the summer of love.

7 Brian Kogelmann (2017) has rigorously analyzed the proposal that the political system draw on polycentrism to secure this.

8 That is, social and economic inequalities should be arranged so that they maximally benefit the least well off class.

REFERENCES

Bicchieri, C. (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms.* New York: Oxford University Press.

Boyd, R. and Richerson, P.J. (2005). Group-Beneficial Norms Can Spread Rapidly in a Structured Population. In: *The Origin and Evolution of Cultures*. Oxford: Oxford University Press, pp. 227-40.

Colander, D. and Kupers, R. (2014). *Complexity and the Art of Public Policy: Solving Society's Problems from the Bottom Up.* Princeton: Princeton University Press.

D'Agostino, F. (2010). *Naturalizing Epistemology: Thomas Kuhn and the 'Essential Tension.'* London: Palgrave-McMillan.

—(2013). The Orders of Public Reason. *Analytic Philosophy*, vol. 54 (March): 129–155.

Downs, A. (1956). *An Economic Theory of Democracy.* New York: Harper and Row.

Friedman, M. and Friedman, R. (1980). *Free to Choose*. London: Secker and Warburg.

Gaus, G. (1996). *Justificatory Liberalism*. Oxford: Oxford University Press.

— (2011). *The Order of Public Reason*. Cambridge: Cambridge University Press.

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

— (2017a). The Open Society and Its Friends: With Friends Like These, Who Needs Enemies? *The Critique* <http://www.thecritique.com/articles/open-society-and-its-friends/>. Accessed May 23, 2017.

— (2017b). Social Morality and the Primacy of Individual Perspectives. *Review of Austrian Economics*. vol. 30: 377-396.

— (2017c). Self-organizing Moral Systems: Beyond Social Contract Theory. *Politics, Philosophy and Economics*. DOI: 10.1177/1470594X17719425

— (2018). The Priority of Social Morality. *In Morality, Governance, and Social Institutions: Reflections on Russell Hardin*, edited by Thomas Christiano, Ingrid Creppell and Jack Knight. New York: Palgrave McMillan, 2018.

— and Van Schoelandt, C. (2017). Consensus on What? Convergence for What? Four Models of Political Liberalism. *Ethics*.

Hayek, F. A. (1964 [2014]). The Theory of Complex Phenomena. In: *The Market and Other Orders*, edited by Bruce Caldwell. Chicago: University of Chicago Press, pp. 257-266.

Hess, M. B. (1966). *Models and Analogies in Science*. Norte Dame: Notre Dame University Press.

Hotelling, H. (1929). Stability in Competition. *Economic Journal*, vol. 39: 41-57.

Johnson, James (2014). Models Among the Political Theorists. *American Journal of Political Science,* vol. 58 (July): 547-560.

Knight, F. (1921 [1964]). *Risk, Uncertainty, and Profit*. New York: Augustus M. Kelly.

Kogelmann, B. (2017). *Agreement, All the Way Up: An Essay on Public Reason and Theory Choice*. Doctoral dissertation, University of Arizona.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*, second edn. Chicago: Chicago University Press.

Lane, R. (2017). *The Complexity of Self Government: Politics from the Bottom Up*. New York: Cambridge University Press.

Levy, D. M. and Peart, S. J. (2017). *Escape from Democracy: The Role of Experts and the Public in Economic Policy*. Cambridge: Cambridge University Press.

Mitchell, M. (2009). *Complexity: A Guided Tour.* Oxford: Oxford University Press.

Nozick, R. (1974). *Anarchy, State and Utopia*. New York: Basic Books.

Ostrom, E. (1986 [2014]). An Agenda for the Study of Institutions. In: *Choice, Rules and Collective Action*. Eds. Fillippo Sabetti and Paul Dragos Aligica. Essex: ECPR Press, pp. 97-119.

Ostrom, V. (1972 [1999]). Polycentricity (Part 2). In: *Polycentricity and Local Public Economies*. Ed. Michael D. McGinnis. Ann Arbor: The University of Michigan Press, pp. 119-38.

Ostrom, V. and Ostrom, E. (1977 [1999]). Public Goods and Public Choices. In: *Polycentricity and Local Public Economies*. Ed. Michael D. McGinnis. Ann Arbor: The University of Michigan Press, pp. 75-103.

Page, S. E. (2011). *Diversity and Complexity*. Princeton: Princeton University Press.

— (2016). Not Half Bad: A Modest Criterion for Inclusion. In: *Complexity and Evolution: Toward a New Synthesis for Economics*. Eds. David S. Wilson and Alan Kirman. Cambridge, MA: MIT Press, pp. 319-26.

Poundstone, W. (1992). *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*. New York: Random House.

Rawls, J. (1971). A *Theory of Justice.* Cambridge, MA: Harvard University Press.

— (2005). *Political Liberalism*. Expanded edn. New York: Columbia University Press.

— (2007) *Lectures on the History of Political Philosophy*. Ed. Samuel Freeman. Cambridge, MA: Harvard University Press.

Schumpeter, J. A. (1950). *Capitalism Democracy and Socialism*. Third edn. London: George Allen and Unwin.

Sen, A. (2017). *Collective Choice and Social Welfare*. Expanded edn. Harmondsworth: Penguin.

Vallier, K. (forthcoming). *Must Politics Be War? In Defense of Public Reason Liberalism*. Oxford: Oxford University Press.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World.* Oxford: Oxford University Press.

Editorial Information

AIMS AND SCOPE

COSMOS + TAXIS takes its name and inspiration from the Greek terms that F. A. Hayek famously invoked to connote the distinction between spontaneous orders and consciously planned orders.

COSMOS + TAXIS publishes papers on complexity broadly conceived in a manner that is accessible to a general multidisciplinary audience with particular emphasis on political economy and philosophy.

COSMOS + TAXIS publishes a wide range of content: refereed articles, unrefereed though moderated discussion articles, literature surveys and reviews.

COSMOS + TAXIS invites submissions on a wide range of topics concerned with the dilemma of upholding ethical norms while also being mindful of unintended consequences.

COSMOS + TAXIS is ecumenical in approaches to, and not committed to, any particular school of thought and is certainly not a talking shop for ideologues of any stripe.

COSMOS + TAXIS is not committed to any particular school of thought but has as its central interest any discussion that falls within the classical liberal tradition.

SUBMISSIONS

COSMOS + TAXIS only accepts digital submissions:
david.andersson@rmit.edu.vn

Submitting an article to COSMOS + TAXIS implies that it is not under consideration (and has not been accepted) for publication elsewhere. COSMOS + TAXIS will endeavor to complete the refereeing process in a timely manner (i.e. a publication decision will be made available within three months).

Papers should be double-spaced, in 12 point font, Times New Roman. Accepted papers are usually about 6,000-8,000 words long. However, we are willing to consider manuscripts as long as 12,000 words (and even more under very special circumstances). All self-identifying marks should be removed from the article itself to facilitate blind review. In addition to the article itself, an abstract should be submitted as a separate file (also devoid of author-identifying information). Submissions should be made in Word doc format.

1. Submissions should be in English, on consecutively numbered pages. American, Canadian and UK spellings and punctuation are acceptable as long as they adhere consistently to one or the other pattern.
2. Citations should be made in author-date format. A reference list of all works cited should be placed at the end of the article.

   The reference style is as follows:

   Author, A. B. (2013). Title. *Journal*, 1(1): 1-10.
   Author, C. D., Author, B., and Author, C. C. (2013). Article Title. in *Title*. City: Publisher, pp. 1-10.
   Author, J. E. and Author, B. (Eds.) *Title*. City: Publisher, pp. 1-10.
   Author, E. F. (2008). *Title*. Place: Publisher.
3. All notes should be as end notes.
4. No mathematical formulae in main text (but acceptable in notes or as an appendix).

Please consult the latest issue of COSMOS + TAXIS to see a fully detailed example of the Journal's elements of style.

CONTACTS

COSMOS + TAXIS welcomes proposals for guest edited themed issues and suggestions for book reviews. Please contact the Editor-in-Chief to make a proposal:
david.andersson@rmit.edu.vn

All business issues and typsetting are done under the auspices of The University of British Columbia. Inquiries should be addressed to the Managing Editor: leslie.marsh@ubc.ca

http://cosmosandtaxis.org

Books for review should be sent to:

Laurent Dobuzinskis
Department of Political Science
Simon Fraser University
AQ6069 - 8888 University Drive
Burnaby, B.C.
Canada V5A 1S6

Design and typesetting: Claire Roan, UBC Studios,
Information Technology, The University of British Columbia.

COSMOS + TAXIS