The Complexity of a Diverse Moral Order

Gerald Gaus

1 THE NEW DIVERSITY THEORY

The late Russell Hardin observed that "the worst failing of contemporary political philosophy is its frequent irrelevance to actual and plausible conditions."[1] Contemporary moral philosophy too seems increasingly irrelevant to the world in which we live. Most moral philosophy blithely (or, alternatively, adamantly) adopts the Enlightenment presupposition that its aim is to uncover true moral beliefs — and since what is true is true for all — to discover universal true principles of morality, which will eventually be manifest to everyone who properly exercises their reason. While the modal moral theorist clings to this Enlightenment faith that free inquiry will lead to moral consensus, when we look about us we discern a far more complex pattern of emerging agreement about some matters such as the wrongness of racial discrimination, together with ever-deepening disputes about the place of humans in the universe, the roles and natures of the sexes, the role of the state, the relative importance of liberty and equality, and indeed the very nature of morality itself. "Western Judeo-Christian society" has not been transformed into a new universal rational secular order, but has dissolved into a complex pattern of religious, spiritual, and secular orientations — and each of these refracts into a spectrum of versions.[2] Yet for most moral philosophers all this is basically an irrelevant empirical annoyance pointing to the folk's ignorance or stupidity, certainly not a basic datum for philosophical moral inquiry. What could the mere fact that actual people, thinking about their lives and morality, come to very different conclusions, tell the moral philosopher — about Morality Itself?

---

[1] Russell Hardin, "From Bodo Ethics to Distributive Justice," *Ethical Theory and Moral Practice,* vol. 2 (1999): 399–413, at p. 412. Showing that one person's bug is another's feature, David Estlund embraces the conclusion that the knowledge a political theory may have "no practical value at all." See his "Prime Justice" in *Political Utopias: Contemporary Debates,* edited by Michael Weber and Kevin Vallier (Oxford: Oxford University Press, 2017), pp. 35-56 and "Human Nature and the Limits (if Any) of Political Philosophy," *Philosophy & Public Affairs* vol. 39 (2011): 207-35.

[2] Survey research indicates clear moral differences throughout the world. See, for example, The Pew Research Center's *American-Western European Values Gap*. [Downloaded at http://www.pewglobal.org/2011/11/17/the-american-western-european-values-gap/ August 6, 2017]; Richard Shweder, Manamohan Mahapatra and Joan G. Miller, "Culture and Moral Development" in *The Emergence of Morality in Young Children*, edited by Jerome Kagan and Sharon Lamb (Chicago: University of Chicago Press, 1987), pp. 1–83. See also Jonathan Haidt, *The Righteous Mind* (New York: Pantheon, 2012), Part II.

The great contribution of the Rawlsian project was to take, as its basic supposition, that the exercise of human reason under free institutions leads to "intractable struggles" and "irreconcilable" conflicts of "absolute depth" about the nature of the universe, the world, value — and perhaps even justice.[3] Although Rawls and his followers certainly distanced themselves from the Enlightenment's assumption that moral consensus is the natural outcome of the free use of human reason, their project nevertheless remained something of a damage limitation exercise — seeking to sort through the morass of disagreement to uncover a basic, shared conception of political justice.[4] Indeed, we might wonder, what else could remain for moral and political philosophy once reasonable moral diversity is taken as a permanent feature of free societies? Surely it cannot be that moral diversity, as it were, *all the way down,* is the permanent feature of a free society. While many types of diversity, such as demographic diversity, are valuable, deep moral diversity seems to almost everyone not only socially damaging but ultimately incoherent.[5] Morality, it is assumed, necessarily implies some universality, while moral diversity implies non-universality. The upshot of these two thoughts — that the exercise of free human reason leads to disagreement but justice must be endorsed by all reasonable people — led the Rawlsian project to marshalling its complex philosophical machinery to justify ever-thinner, rock-bottom consensus, conclusions.[6]

In this essay I explore what I call the "New Diversity Theory." Sections 2-5 draw together analyses from a number of works and authors to help articulate this new paradigm. The crux of this approach is to analyze moral diversity not as moral reasoning gone awry, or even as a feature of free societies to be managed, but as a

---

[3] John Rawls, *Political Liberalism,* expanded edition (New York: Columbia University Press, 2005), pp. xxvi, 4.

[4] This is not to say this project succeeded; as it evolved, at every juncture where it seemed that agreement could be secured, yet deeper diversity revealed itself. See Gerald Gaus and Chad Van Schoelandt, "Consensus on What? Convergence for What? Four Models of Political Liberalism," *Ethics,* forthcoming.

[5] This certainly was, at least at one point, the view of Haidt. See Jonathan Haidt, Evan Rosenberg and Holly Hom, "Differentiating Diversities: Moral Diversity Is Not Like Other Kinds," *Journal of Applied Social Psychology*, vol. 33 (January 2003): 1–36; Haidt, "Invisible Fences of the Moral Domain," *Behavioral and Brain Sciences*, vol. 28 (2005) 28: 552-3. My thanks to Piper Bringhurst for discussions on this point.

[6] In place of the original aspiration to justify justice as fairness *over* alternatives, by the end of the project what is left is a vague argument for a broad "family" of liberal conceptions, of which Rawls's two principles are but one member. See Gaus and Van Schoelandt, "Consensus on What? Convergence for What?" An alternative is to essentially exclude from the population of morally reasonable agents all those who do embrace the requisite controversial conclusions — and so agreement is secured. This path is pursued by Jonathan Quong, *Liberalism without Perfection* (Oxford: Oxford University Press, 2010).

fundamental moral phenomenon.[7] Moreover, the theme of New Diversity Theory is that moral difference is not simply a challenge to reasonably stable moral order (though it certainly may be), but a critical resource for free societies to discover better ways of living together under conditions of limited knowledge and an unpredictable environment. Section 2, then, presents the core idea of a moral perspective, at the heart of New Diversity Theory. We shall see that the idea of moral diversity is itself multifaceted. Sections 3 and 4 analyze just two of these facets: categorizations and predictive models. These two sections provide a partial answer to the commonsense worry that it is incoherent to even want moral diversity: they advance two lines of reasoning indicating why we can make better moral judgments when others see things differently. Section 5 continues this reply, showing how a moral system characterized by evaluative and other types of diversity can form a robust moral order. Having defended the benefits and coherence of moral diversity, section 6 examines a critical implication of New Diversity Theory: the social morality of a free society is, strictly speaking, a complex phenomenon. I briefly conclude (this rather long essay) with a few remarks (§7) about the tasks of moral, social and political philosophy in New Diversity Theory.

2 DIVERSE MORAL PERSPECTIVES

Recent analyses of diversity have employed the idea of a perspective.[8] For present purposes, I assume a model of a moral perspective of person i that has five elements.

I. A set of evaluative standards, values, judgments and/or, normative principles.

II. A method for arriving at, and modifying these standards, etc.

III. A set of categorizations of the natural and social world. These categorizations must:

---

[7] For valuable contributions to this new approach see Fred D'Agostino, *Free Public Reason* (Oxford: Oxford University Press, 1996) and "How Can We Do Political Philosophy?" *Cosmos + Taxis* (forthcoming); Ryan Muldoon, *Social Contract Theory for a Diverse World* (New York: Routledge, 2016); Julian Müller, *Polycentric Democracy: Making Use of Diversity* (New York: Routledge, forthcoming); Brian Kogelmann, *Agreement, All the Way Up: An Essay on Public Reason and Theory Choice*, Doctoral dissertation, University of Arizona, 2017.

[8] Muldoon, *Social Contract Theory for a Diverse World* and his "Perspectives, Norms and Agency," *Social Philosophy & Policy*, vol. 34 (2017): 260-76; Müller, *Polycentric Democracy;* Scott E. Page, *The Difference* (Princeton: Princeton University Press, 2007), chap. 1; Hélène Landemore, *Democratic Reason* (Princeton: Princeton University Press, 2013), chap. 4. I have employed the idea extensively in *The Tyranny of the Ideal: Justice in a Diverse Society* (Princeton: Princeton University Press, 2016).

(*i*) determine what are the morally relevant features of the world (e.g., states of affairs, agents, actions, character traits, etc.) for any

(*ii*) judgment/action-context, J_X. A judgment/action context (e.g., " φ is an action that involves harm to another, and so is of morally relevance;" "X is an institution of the basic structure, and so relevant to social justice") is itself a categorization.

IV. A set of options {α, β, γ…} in any given judgment-action context, J_X.

V. A set of predictive models, which take the set of options {α, β, γ…} and predicts for each, the social world(s) {S_1, S_2, S_3…} that would be produced (by each option).

VI. A function, which takes **I**, **III**, **IV** and **V** and applies them to the resulting social worlds {S_1, S_2, S_3…} yielding moral ranking of at least some worlds.

I certainly do not claim that every recognizable moral view can be modeled in this way: the model is by no means intended as a set of necessary and sufficient conditions for a view to constitute a person's "moral theory." And indeed, many of the conditions could be relaxed. For example, the "function" in **VI** might be no more than "an intuition" or a guess. The aim is to get clearer about the moving parts of a typical view, which: has a set of standards (and some idea about how to modify it); looks at the world, identifying contexts for moral decisions; in any given context, examines the options, considers the social world that each option would produce; and makes a decision/judgment about the best thing to do. I propose that a wide variety of moral viewpoints could be modeled along these lines.

Already we see a benefit from taking a more formal view: moral diversity can arise from differences on any of these elements. Too often real moral disagreement is thought to arise only from disagreements about feature **I**, our basic values, moral standards and principles.[9] This has been the traditional focus of pluralistic analyses of political morality.[10] Once we appreciate the diversity of sources of perspectival

[9] Sometimes disagreement about ultimate standards of evaluation are said to be "deep," as opposed to those arising from mere "empirical facts," which understood as "shallow" — as if different understandings of the world we inhabit and the way it works never go to one's critical scientific, metaphysical and religious commitments. In any event, I put aside claims about what is shallow and what is deep, matters which are themselves internal to a perspective. See my *Contemporary Theories of Liberalism* (London: Sage, 2003), pp. 129-31.

[10] See George Crowder, "Pluralism" in *The Routledge Companion to Social and Political Philosophy*, edited by Gerald Gaus and Fred D'Agostino (New York: Routledge, 2013), pp. 353-63. For a

diversity, we see that any hope to contain "reasonable disagreement about justice" to a small number of views looks quite forlorn. Suppose that on each of the six elements there are ten reasonable positions, and that all combinations of the reasonable are themselves reasonable: that yields 1,000,000 reasonable perspectives. If we make the very strong assumption of *complete agreement* on evaluative standards, we still have 100,000 distinct reasonable moral perspectives. Omit element **II**, and we still have 10,000 reasonable moral perspectives. Once the sources of moral diversity are recognized, it is well-nigh impossible to fulfill the hopes of so many Rawlsians that there would be only a modest "family" of reasonable perspectives on justice. As we shall see in section 5, this does not imply that consensus on basic rules is impossible, only that such consensus cannot plausibly be secured by supposing that the diversity of moral perspectives is modest.

3 CATEGORIZATIONS

3.1 Diverse Categorizations

A moral or political philosopher resisting this conclusion is apt to deny that there are so many sources of respectable moral diversity. It would be tedious to canvass all the elements of a perspective to show their importance, but let us of focus on two that may seem contentious: diverse categorizations (this section) and predictive models (§4).

One cannot, I would venture, even begin to understand the intractable dispute about the ethics of abortion unless it is firmly kept in mind that the protagonists categorize the world in deeply different ways. Many opponents of abortion rights understand an embryo to be ensouled: their world contains the category "being with a soul" that, on their moral perspective, is intimately linked to rights of a person. Proponents of abortion rights insist that an embryo is not a person, employing a naturalistic conception of a person that has no room for the category "ensouled" (or a number of other related categories). Hence the utter futility of this long, very angry, dispute, which still would be intractable if all the parties accepted the same set of evaluative standards. Or consider the categorization "homophobe" employed by most "progressives." Conservative criticism of homosexuality is thus categorized as a mental disorder (a phobia), indicating that the criticism is irrational and outside bounds of rational disagreement. On the other hand, the conservative often invokes the category of an "impure" or "blasphemous" action — categories that have little or no place in a progressive's understanding of the world.[11]

---

nuanced discussion see Landemore, *Democratic Reason*, pp. 213-19.

[11] See Haidt, *The Righteous Mind*, chap. 6. When it comes to the environment and food consumption, notions of purity may, after all, be invoked by the progressive.

3.2 The Naïve Naturalist Response

While to some of us the importance of categorizations seems manifest, moral and political philosophers are apt to dissent, insisting that all such disagreements can be reinterpreted as familiar differences in evaluative standards or tractable disputes about the nature of the world. The opponents of abortion, it might be claimed, ascribe rights to "a clump of cells" while pro-choice advocates do not (a moral dispute); or "homophobe" can be said to describe a medical condition that further scientific inquiry will either confirm or not.

Now once we have arrived at this sort of disagreement, further progress is unlikely to be made by looking at specific cases, for the dispute is about how we are to interpret the cases. Going deeper, the rejection of categorization as an independently significant part of a moral perspective would certainly make sense if something along the lines *Naïve Naturalism* held:

1. The world is ultimately constituted by all the facts there are (the complete set of facts);

2. Science is the progressive uncovering of increasingly accurate representations of these facts; each new scientific discovery presents a set of facts that strictly dominates earlier versions (the later includes all what is correct in the earlier version, and adds to it).

3. Those who (*a*) accept the best current science describing the world, and (*b*) the same evaluative standards, will agree in their moral judgments.

On Naïve Naturalism, scientific progress secures, as it were, ever-better resolutions of the facts that compose the world. It is as if blurry and confused pictures become increasingly better representations of what *is*. If Naïve Naturalism held, disputes about categorizations would be disputes that science ultimately will resolve, and a reasonable person would have no grounds for not embracing science's current picture, since it is more accurate than previous ones. Perspectives that accept the best current science thus would have the best available picture of the world, and so have no disputes about the proper description of the natural or social world — only about the evaluation of the facts.[12] All sensible moral disputes would have to derive from their evaluative standards.

---

[12] It appears to also suppose (*i*) that that social facts are irrelevant to moral judgment, (*ii*) they are themselves subject to Naïve Naturalism's view of scientific discovery, or (*iii*) they are reducible to the natural facts subject to science. None seem plausible. Stanley Benn and I long ago explored the relation of social categorizations (social facts) and normative theory in our *Public and Private in Social Life* (New York: St. Martin's Press, 1983), chaps. 1 and 2. On the implausibility of (*iii*) see Emile Durkheim, *The Rules of Sociological Method*, edited by Steven Lukes, translated by W. D. Halls (New York: The Free Press, 1982).

Naïve Naturalism — though seldom stated quite this baldy — is, I would venture, at the root of a great deal of secular moral theory, and its almost obsessive attention to differences in evaluative standards. After all, if how we conceive of the world is really uncontroversial, isn't it only evaluative standards that could cause deep reasonable disagreements? I believe this radically misconstrues science, and this because it fundamentally misconstrues the relation between discovery, theory, and categorization. As Hayek stressed, all categorization depends on ignoring some differences (which become non-differences) and focusing on others.[13] At its most basic level, categorization is a many-to-one mapping; members of a set {*a, b, c, d, e*} are mapped onto categories *W* and *X*, such that {*a, d, e*} → *W* and {*b, c*} → *X*.[14] Note under this categorization *b* and *c* are the *same;* any difference between them is lost (an embryo *is* just a clump of cells, like a tumor). Yet on another set of categories, say *Y* and *Z*, where {*a, b, d*} → *Y* and {*e, c*} → *Z*, *b* and *c* are different (perhaps *entirely* different) phenomena (perhaps one has a soul and so is the height of creation).

One of the things that Kuhn meant to capture with his idea of a paradigm was relations of similarity and categorizations at work in a scientific community.[15] Symbolic labels (based on relations of similarity) are applied in a uniform way (e.g., in the form of equations) by the community to describe nature. Science often proceeds by analogies; categories in a known area are employed to understand an unfamiliar one. Following Mary Hess, in such scientific explanations 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 categories of (*B*). "Sound (primary system) is propagated by wave motion (secondary system)."[16] Metaphors are critical in suggesting analogies, such that the primary system can be understood in terms of the categories of some other system. For example, early electricians sought to understand electrical phenomena (*A*) under the category "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.[17] Different paradigms, often building on different metaphors thus, in a straightforward sense, categorize the world in importantly different, and partly incommensurable ways. What are the same phenomenon on one paradigm may be different ones in another. An upshot is that even when the community agrees

[13] F. A. Hayek, *The Sensory Order* (Chicago: University of Chicago Press, 1952), pp. 48-52.

[14] See Page, *The Difference*, chap. 3. This is entirely consistent with there being one, real, world that we all inhabit, and we all recognize this to be the case.

[15] See Thomas Kuhn, *The Structure of Scientific Revolutions*, 2nd edition (Chicago: University of Chicago Press 1970), pp. 107, 148, 167, and especially his "Second Thoughts on Paradigms" in his *The Essential Tension* (Chicago: University of Chicago Press, 1977), pp. 293-319.

[16] Mary B. Hess, *Models and Analogies in Science* (Norte Dame, IN: Notre Dame University Press. 1960), pp. 158-9.

[17] Kuhn, *The Structure of Scientific Revolutions*, p. 17.

that a new paradigm solves its problems better than an earlier one, there is loss of knowledge, in the sense that ways of describing nature that made enlightening distinctions are no longer available.

3.3 Diverse Categorizations and Similarity: Disagreement and Moral Gains

Naïve Naturalism notwithstanding, to conceive of the world is to categorize, and those who categorize differently, even if they share the same evaluative standards, can disagree in their moral judgments. Now the upshot of a system of categories is not simply so — categorical. A system of categories reveals meaningful structure in the world, and a critical component of this structure are relations of similarity. Once one has categorized a phenomenon, relations of similarity are generated. Suppose, for example, one's category of "punishment" is characterized simply by incarceration due to criminal behavior.[18] This categorization also implies similarity judgments about punishments: those that involve similar length of incarnation are similar punishments, while those that involve very different lengths are very dissimilar punishments.

In a world in which we are all searching for better options under conditions of severely limited information, individuals with different categorizations and so different understandings of similarity can sometimes assist each other in finding better moral options. In difficult choice situations — including moral choices — agents engage in some search routine, knowing the value of some options, and expending resources to see if they can find better ones. We do not have full knowledge of all the options confronting us: we explore some options, see how well they score, and see if modest modifications improve things or make them worse. When a person searches for a better option in the face of uncertainty, she starts from what she knows, and typically moves on to examine variations. Already knowing option α, looking at the value of similar option β is much more tractable; much of β is already captured by what one knows of α. Given the structure of the world, the obvious alternative to α is β (as opposed to, say, ω, which seems to have hardly anything in common with α — why would one investigate *that?*). On these *searching models* one keeps looking to see if variations of well-understood options improve matters.

---

[18] It is more plausible than one might think that categorizations are unidimensional (say, punishment only involves incarceration). While in their informal classifications people seem to employ family resemblances, subjects from around the world gravitate to single-dimensional classification systems when they are forced to think formally about their classifications. When asked to sort elements in a domain, participants repeatedly select some single common dimension and sort according to it. See *The Tyranny of the Ideal*, Appendix A.

Consider a mundane case: a manufacturer searching for innovations in product T. The advice to formulate an unordered list of possible variations of T, 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 T); because there is so little fiddling with the structure of the present T 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 optimal T). Think of SONY teams that were working on the ideal Betamax recorder. 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.[19] 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*.

Now suppose that we have teams that employ different categorizations of T such that one team orders the possible products (in terms of similarity), $\{T_1, T_2, T_3, T_4, T_5\}$ while another orders them $\{T_1, T_2, T_5, T_4, T_3\}$. The first team will investigate them roughly in order, so starting with T_1 they move on to T_2; if it improves upon T_1, they may move on to consider T_3. Suppose T_3 is a lot worse than T_2, and when they consider T_4 it is even worse than T_3. They may stop their search at this point, concluding that for now T_2 is best they can do. But the second team, given their different classifications, after looking at T_2, sees T_5 as a small variation, and finds T_5 a great improvement on T_2. *If* the two teams share the high evaluation of T_5, then the different categorization of the second team benefits the first, revealing a better option that they did not think it worthwhile to investigate.[20] Scott E. Page and Lu Hong can be understood as exploiting this insight in their innovative and important work,

---

[19] This, perhaps, is the role of ideal political philosophy. See my *Tyranny of the Ideal.*

[20] See Scott E. Page, *Diversity and Complexity* (Princeton: Princeton University Press, 2011), pp. 90-95, 217-27.

specifying the precise conditions under which different perspectives are guaranteed to aid each other in finding the best alternatives to a common problem.[21]

Once again, resistance comes from moral and political philosophers. Among the more sophisticated objections are those based on commitment to what might be called the *simple optimization model*.[22] This more familiar approach, informing neo-classical economics and rational choice theory, supposes that agents in a choice situation confront an unordered set of options, assigning an expected score (in terms of, say, justice) to each. If they can do all *that* they confront a relatively simple optimization exercise: choose the option with the highest expected justice. Now in some cases we can approximate this, and simple optimization models work well. However, when the justice of what I do depends on what others do, it becomes increasingly difficult to accurately predict the expected justice of many of my options, much less an entire feasible set.[23] As we have learned from Herbert Simon, real agents have limited cognitive resources, and seldom find the best solution.[24] They must employ resources searching for unknown, but possibly, better options; for them it is organizing the search, not the final optimizing selection, that is the critical problem. Some are more cautious, exploring only until further moves make things worse; others hoping for excellent unseen options, are willing to explore some worse options for a while before

[21] See Lu Hong and Scott E. Page, "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory,* vol. 97 (2001): 123–163; "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, No. 46 (Nov. 16, 2004): 16385-16389. The Hong-Page conception of a perspective is importantly different than that offered here: I have stressed that those with different categorizations may well disagree in their evaluations of an option, and different similarity judgments arise out of different categorizations. Hong and Page begin with a one-to-one mapping of objects on to a language; on their model those who concur on the other elements of a perspective but have different similarity judgments will always agree on the value of any specific option (See Page's *The Difference*, chap. 1). In *The Tyranny of the Ideal* (chap. 3) I have argued that this is too strong an assumption in most moral and political contexts; on the analysis offered here it is thus contingent to what extent those with different similarity orderings will concur on the value of an option. Page notes some of the differences between his and my notion of a perspective in "The Imperative of Complexity," *Cosmos + Taxis*, forthcoming. On some formal limits of the Hong-Page proofs, see John Weymark, "Cognitive Diversity, Binary Decisions, and Epistemic Democracy," *Episteme*, vol. 12 (2015): 497–511.

[22] David Wiens and I discuss the relative merits of these two approaches; see his thoughtful essay "The Tyranny of a Metaphor" and my "Political Philosophy as the Study of Complex Normative Systems" in *Cosmos + Taxis,* forthcoming.

[23] The simple optimization model can accommodate lack of information in terms of risk — what we know we do not know — but has great difficulty making sense of deep uncertainty about what the options and their value — what we don't know we don't know. See section 6 below. On uncertainty, see Frank Knight, *Risk, Uncertainty, and Profit* (New York: Augustus M. Kelly, 1964), chaps 7-8.

[24] See Herbert Simon, *The Sciences of the Artificial,* third edition (Cambridge, MA: MIT Press, 1996), chap. 2, pp. 119-124.

quitting — at the risk of great wasted time. But in a world where we seldom have a clear idea of the set of options or most of their values, uncovering better solutions implies searching the unknown. And it is collective searching that diverse agents can do well.

This is critical. In contrast to traditional reasonable disagreement analyses, New Diversity Theory stresses the ways in which diversity not only leads to problems of securing a just social order, but the conditions under which it can induce better moral outcomes for all.[25] This is not to be Pollyannaish; diversity certainly can be a barrier to a just social order as well as a resource. But once we recognize the complex ways in which moral diversity works, the task of moral and social philosophy becomes more nuanced, investigating what patterns of diversity impair moral order and improvement, and which encourage it.

4 PREDICTIVE MODELS

4.1 All Sensible Perspectives Employ Predictive Models

Let us (this time more briefly), consider another unfamiliar element of our model of a moral perspective — predictive models (§2, feature **V**). Now the political philosopher resisting the multi-dimensionality of diversity might insist that, while "consequentialist" theories must employ such models, "deontological" ones do not. This is surely too simple. Consider Rawls's interpretation of Kant's universal law formulation of the categorical imperative, which he models in terms of a four-step "CI procedure." The first three steps on the CI procedure are fairly straightforward:

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

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

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

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

(3) Everyone always does X in circumstances C in order to bring about Y (as if by a law of nature).[26]

At the fourth step we are to consider the "perturbed social world" that would result from the addition of this new law of nature; we seek to understand the new "equilibrium state" on which this perturbed social world would settle. We are then to ask ourselves whether, when we regard ourselves as a member of this new social

---

[25] See also *Muldoon, Social Contract Theory for a Diverse World,* chap. 3.

[26] John Rawls, "Themes in Kant's Moral Philosophy" in *Rawls: Collected Papers*, edited by Samuel Freeman (Cambridge, MA: Belknap Press of Harvard University Press, 1999), pp. 497-528 at p. 499.

world, we can "will this perturbed social world itself and affirm it should we belong to it."[27] Manifestly, on Rawls's analysis of Kant's "deontology," our endorsement of a moral imperative takes place against the background of a predictive model of the social world that would result from introducing a maxim. To understand what principles are just, we must predict the results of the new social world that takes them as a law of nature.

It is hard to imagine a plausible moral perspective that makes no use of predictive models: such a perspective must insist that it is *never* morally relevant what the effects of a moral judgment, moral principle, or action will be. "All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy."[28]

4.2 The Diversity of Perspectives and Predictive Models

That predictive models are required by a sensible perspective does not imply a diversity of them. Recent political philosophy has tended to embrace a distinction between the analysis of "desirability" and "feasibility." On the stark version of this view, political philosophers investigate desirability (a normative question); they then hand off their "findings" to social scientists who investigate the "feasibility" of their desirability (or other normative) judgment. [29] In principle, all political philosophers could employ the same social scientist, employing "the best" predictive model, to "check for feasibility."

Now the division of cognitive labor should certainly not be belittled — nevertheless, the common idea that a political philosopher can perform his normative work and then hand it off to *the* expert social scientist to check for feasibility typically manifests Naïve Naturalism: there is just one uncontroversial social world to be investigated, and the best social scientist has the best understanding of it, which all should use. But to any perspective, what constitutes the best predictive model depends on the way the perspective categorizes the world. If, for instance, a perspective conceptualizes a certain problem as pertaining to mutual benefit and exchange, it is apt to employ economic or rational choice models; the features of the world economic models pick out as variables in prediction will be features deemed important by the moral perspective. In contrast, a perspective that views what some

---

[27] Ibid., p. 500.

[28] John Rawls, *A Theory of Justice*, revised edn. (Cambridge, MA: Belknap Press of Harvard University Press, 1999), p. 26.

[29] For a recent sophisticated analysis that is skeptical of starker versions of this view, while still friendly to the general idea, see John Hamlin, "Feasibility Four Ways," *Social Philosophy & Policy* vol. 34 (2017) 209-31. For an analysis that calls for political philosophers to engage with social science, see David Wiens, "Demands of Justice, Feasible Alternatives, and the Need for Causal Analysis," *Ethical Theory Moral Practice*, vol. 16 (2013): 325–338.

may see as "the same problem" as concerning power and domination may well employ sociological models because they, not economic models, pick up on the relevant categories. These economic and sociological models have different features and strengths — the moral or political philosopher does not simply go shopping for the unequivocal "best model."[30]

I do not think that this observation is restricted to normative theories or the social sciences (thus it is not closely related to the long and inconclusive debate on whether the social sciences are "value free"). Even natural science is based on a complex interaction between theory, and the categories and processes it postulates, and the construction of instruments to observe these categories. What constitutes an observation is theoretically-informed; the theory predicts what will happen in a world with certain entities. Kuhn observes:

> After about 1630, for example, and particularly after the appearance of Descartes's immensely influential scientific writings, most physical scientists assumed that the universe was composed of microscopic corpuscles and that all natural phenomena could be explained in terms of corpuscular shape, size, motion, and interaction. That nest of commitments proved to be both metaphysical and methodological. As metaphysical, it told scientists what sorts of entities the universe did and did not contain: there was only shaped matter in motion. As methodological, it told them what ultimate laws and fundamental explanations must be like: laws must specify corpuscular motion and interaction, and explanation must reduce any given natural phenomenon to corpuscular action under these laws. More important still, the corpuscular conception of the universe told scientists what many of their research problems should be.[31]

A prediction that referred to an "occult quality" — such as gravity[32] — did not enter these models, just as corpuscular action was not predicted in later models.

4.3 The Benefits of Predictive Diversity

Even without these deeper sources of diversity, in the rare case in which two perspectives agree on all other aspects of their perspectives, they may well employ different predictive models. Some models may be good at predicting long-run tendencies, others better over the short-term; some are excellent in more-or-less stable environments but cannot accommodate the possibility of rare events. Page's work is notable in consistently calling our attention to the collective benefits of multiple predictive models. According to the Diversity Prediction Theorem, the collective error

---

[30] See Page, *The Difference,* p. 286; *Tyranny of the Ideal*, pp. 138-9.

[31] Kuhn, *The Structure of Scientific Revolutions*, p. 41.

[32] Ibid., p. 105.

of an aggregation of predictions is always less than the average predictive error.[33] If we take the average of how wrong each model is, this is greater than the predictive error of the aggregation of predictions. The collective prediction will always beat the average of individual predictions. This is an important result: even if our predictive models are not very good, a diverse society can draw on diverse predictive models. Moreover, given this result there will often be a loss of collective predictive power if everyone adopts the same predictive model — even if it is the best in the set.[34] And there certainly will be no advantage to everyone adopting the average prediction rather than each using her own preferred model and then drawing on the results of others. In other contexts as well, Page has shown the conditions under which it is better to employ diverse models than *for all of us to embrace the best model*.[35] Again, the slogan of the New Diversity Theory: a diverse population can achieve better answers than a homogeneous population, even if the homogenous population has adopted the best available model or perspective.

5 MORAL ORDER OUT OF DIVERSITY

5.1 A Modern American Parable

Having articulated some of the commitments and claims of the New Diversity Theory, I turn to the incoherence worry. Whatever the advantages of diversity in other aspects of human life, moral diversity has consistently been thought to be antithetical to social cooperation and, indeed, to the very idea of morality. This not a concern to be lightly dismissed. Under what conditions does moral diversity threaten disastrous outcomes?

Moral diversity runs the risk of disaster if individuals in a population morally deliberate only according to:

> *The "I conclude we ought" implies "I ought" View:* As a competent moral agent, if (*i*) Alf conscientiously deliberates and concludes that, given what he takes to be the correct normative premises and relevant empirical information, one ought to φ (ought not φ, or may φ)[36] under conditions *C*, where this does not require taking

---

[33] See Page, *The Difference*, pp. 205-12; *The Tyranny of the Ideal*, Appendix B.

[34] See Hélène Landemore and Scott E. Page, "Deliberation and Disagreement: Problem Solving, Prediction, and Positive Dissensus," *Politics, Philosophy & Economics* (2014): 1-26. DOI: 10.1177/1470594X14544284. A truly excellent predictive model can beat the collective prediction, but the collective prediction will beat the average of diverse individual predictions.

[35] Scott E. Page, "Not Half Bad: A Modest Criterion for Inclusion" in *Complexity and Evolution: Toward a New Synthesis for Economics*, edited by David S. Wilson and Alan Kirman (Cambridge, MA: MIT Press, 2016): 319-26.

[36] Henceforth I shall not state these alternatives, assuming that they are implicit.

account of the conclusions of the deliberations of others and (*ii*) he reasonably concludes that morality instructs that we all ought to φ under conditions *C*, then (*iii*) he ought to φ in circumstances *C*, even if others fail to do as they ought.

It is important that on the *"I conclude we ought" implies "I ought"* View, Alf does not typically assert that we all ought to φ in *C because he* has concluded that we ought to φ: Alf may believe that "we ought to φ" in *C* because it is a moral truth that we ought to, or that an impartial spectator would approve of our φ-ing. The important point is that once Alf conscientiously comes to the belief that one ought to φ in *C* — it is, we might say, his best judgment about the morally best thing to do — then, as a competent moral agent, he will justifiably φ in circumstances *C*, and indeed insist that we all do so, for that is what we ought to do.[37] And, as I have stressed, none of this necessitates (though it may be epistemically recommended) that Alf factors into his moral deliberation the moral conclusions of others.

To many the "I conclude we ought implies I ought" View expresses the core of morality and articulates an ideal of the sanctity of conscience.[38] The moral agent thinks through how we should act (say, by employing the Four Step CI-Procedure (§4.1)), comes to a judgment about what morality or justice requires, which entails that we all ought to do whatever that is, which in turn entails that I (the agent) ought to do that — *even if no one else does*. Thus, Cristina Bicchieri insists, morality is "unconditional." "Under normal conditions," she writes, "expectations of other people's conformity to a moral rule are not a good *reason* to obey it. Nor is it a good reason that others expect me to follow a moral rule."[39] On this view, then, morality calls for an action that one can secure on one's own — one does not need the cooperation of others to do the moral thing (which is not to deny it would be even better if they too did the moral thing). Bicchieri thus likens morality to brushing one's teeth: one can secure the point of the action without the participation of others, even when the judgment is about what we all should do.[40]

---

[37] I assume here that conditions *C* are so defined that typical justifications for not φ-ing (duress, etc.) would show that *C* was not met. Recall that *C* cannot include the deliberations of others about what is moral in this circumstance. It can, though, take account of what Alf expects others to *do*.

[38] I criticize this view in "Private and Public Conscience" in *Reason, Value, and Respect: Kantian Themes from the Philosophy of Thomas E. Hill, Jr.,* edited by Mark Timmons and Robert Johnson (Oxford: Oxford University Press, 2015), pp. 135-56.

[39] Cristina Bicchieri, *The Grammar of Society* (Cambridge: Cambridge University Press, 2016), p. 21.

[40] Ibid., p. 20. I focus on Bicchieri rather than an orthodox moral philosopher as her own work on social norms stresses the necessity of social cooperation, but when it comes to morality even she conceives of it as inherently personal. I am not denying that there are some parts of morality that are indeed inherently personal, such as seeking some of the moral virtues, where the aim of the agent is that *she* is courageous, trustworthy, or pure. As I have argued, no single theory encompasses all of what we call "morality." See *The Order of Public Reason* (Cambridge: Cambridge University Press, 2011), Appendix A.

Individuals wedded to the "I believe we ought implies I ought" View will end up playing the Go Your Own Way Game depicted in Display 1.

Betty

	α	β
α	1st · · · · 4th	2nd · · · · 2nd
β	3rd · · · · 3rd	1st · · · · 4th

Alf (α, β rows)

Display 1: Go Your Own Way Game

In the interaction of Display 1, each orders the outcomes: (1st) we act on (my view of) justice (α for Alf, β for Betty); (2nd) I act on (my view of) justice, the other fails to; (3rd) I act on the erroneous view and the other acts justly (at least someone does!); and (4th) no one does the just thing. In this game the sole equilibrium is that Alf acts on his view (α), and Betty acts on her view (β), of justice. At either of the coordination solutions (when both play α or both play β), one of the parties would do better by changing his or her move, and acting on his or her favored interpretation of justice. So here even if the other does not do as you have concluded "we" ought, you still ought to do it. In this game each goes their own way.

Display 1 captures a common philosophical understanding of moral disagreement. As one contemporary philosopher has (somewhat approvingly) put it, Alf and Betty are "fighting" it out as to who is morally correct.[41] If this is the typical interaction among morally diverse agents, a society would be morally disordered. Each stays true to her own conscience (i.e., acting on her own "I believe we ought" judgment), but each sees her neighbors as doing the unjust thing. Such a society is one of moral conflict.[42] Now perceived moral violations typically trigger punishments, if only in the form of gossip, ostracism, and (as we see today), shouting and ridiculing.[43] If Alf and

---

[41] David Enoch sees such fighting as a distinct improvement on my account in "The Disorder of Public Reason," *Ethics,* vol. 124, (October 2013), pp. 141-176 at p. 175. Others have found my argument for reconciliation in the next section to be "squeamish." I leave aside the objection from manliness.

[42] See Kevin Vallier, *Must Politics Be War? In Defense of Public Reason Liberalism* (Oxford: Oxford University Press, forthcoming)

[43] "We do not call anything wrong, unless we mean that a person ought to be punished in some way or other for doing it; if not by the law, then by the opinion of his fellow creatures; if not by opinion, by the reproaches his own conscience. This seems the real turning point between morality and simple expediency." John Stuart Mill, *Utilitarianism* in *The Collected Works of John Stuart Mill*, edited by J.M. Robson, vol. 10 (Toronto: University of Toronto Press, 1965), pp. 203-60, at pp. 246. This phenomenon of "moralistic aggression" is as ancient as morality itself, and especially prominent in matters of justice. For a psychological analysis, see Michael T. McGuire, "Moralistic

Betty take this step, they are all too likely to end up in actual conflict. Evidence indicates that when a person is "punished" for an act she does not deem wrong, she is apt to counter-punish, leading to cycles of conflict.[44] Alf screams "Evildoer!" at Betty, who does not see herself as doing anything wrong, so she retaliates against what she perceives as Alf's naked aggression, who responds to her new act of outrage with yet more punishment. Fox News supports Alf, *The New York Times* rallies to Betty: *The Times* publishes Op-Ed essays on "Why Alfites are so Immoral" while Fox questions whether Betty is really an American at all.

5.2 Reconciliation

Such moral disorder was the primary concern of the classical social contract theorists. In one way or another, Hobbes, Locke, Rousseau and Kant believed that individuals' "private judgments" about morality or justice diverge, and because of this individual private judgment is an inappropriate ground — or at least an inappropriate sole ground — for demands of justice.[45] Kant famously insisted that, even if we imagine individuals "to be ever so good natured and righteous," when each does what "seems just and good to him, entirely independently of the opinion of others" they live without justice.[46] This apparently paradoxical conclusion — that a world of competent moral agents who acted only on their own sincere convictions about justice would live without justice[47] — derives from two claims: (*i*) It is taken as given that reasoned private judgments of justice inevitably conflict. This is partially because of self-bias, but only partially. The path-dependency of reasoning,[48] psychological foundations of judgments,[49] emotional differences,[50] and of course differences of categorizations (say,

Aggression and the Sense of Justice," *American Behavioral Scientist*, vol. 34 (1991): 386–40; for a path-breaking ethnographic account, see Christopher Boehm, *Hierarchy in the Forest* (Cambridge, MA: Harvard University Press, 1999).

[44] See Astrid Hopfensitz and Ernesto Reuben, "The Importance of Emotions for the Effectiveness of Social Punishment," *The Economic Journal*, vol. 119 (October 2009): 1534–1559; Herbert Gintis and Samuel Bowles, *A Cooperative Species* (Princeton: Princeton University Press, 2011), p. 26.

[45] For a comprehensive and detailed defense of this claim, see the essays by S.A. Lloyd, Gerald Gaus, Christopher Bertram and Oliver Sensen in *Public Reason in Political Philosophy,* edited by Piers Norris Turner and Gerald Gaus (New York: Routledge, 2018). See also my "Public Reason Liberalism" in *The Cambridge Companion to Liberalism*, edited by Steven Wall (Cambridge: Cambridge University Press, 2015), pp. 112-40.

[46] Kant, *The Metaphysical Elements of Justice,* 2nd edition, edited and translated by John Ladd (Indianapolis: Hackett, 1999), p. 116 [§43]. Emphasis added.

[47] I have defended this paradox in some depth in "The Commonwealth of Bees: On the Impossibility of Justice-through-Ethos,*" Social Philosophy & Policy,* vol. 33 (2016): 96-121.

[48] See Fred D'Agostino, *Naturalizing Epistemology* (New York: Palgrave, 2010) and my *Order of Public Reason*, pp. 235-51.

[49] See Haidt, *The Righteous Mind*, Part II; Joshua Greene, *Moral Tribes* (New York: Penguin, 2013). For a classic ethnographic analysis that points to psychological differences, see R. B. Brandt's *Hopi*

due to different metaphysical views),[51] and so on (and on) drive disagreement. (*ii*) Furthermore, it is assumed that a critical role of justice in our social lives is to adjudicate disputes about our claims and so coordinate normative and empirical expectations. For Kant the problem of universal private judgment was that "when there is a controversy concerning rights (*jus controversum*), no competent judge can be found."[52] Each, thrown back on her own reasoning, ends up in conflict, and ultimately unjust relations, with others. Understood thus, a necessary role of justice (or social morality) is to provide an interpersonally endorsed adjudication of conflicting claims.[53] Securing justice is inherently something we do together. If no other good-willed and conscientious moral agent accepts that in circumstances *C* justice demands φ, Alf's φ-ing will not secure just social relations. Justice is understood as a property of social relations, not simply of individual actions — and that is why the simple unconditional view of morality is seen as deeply flawed.

Surely, though, this is just one more perspective on justice: some view it as essentially a property of personal action, others as a property of social relations. True enough. We certainly cannot say that views of justice that see it as purely a quality of personal action are unintelligible. We can, though say that those perspectives that view justice as, at least to a large extent, a property of social relations are able to secure the good of moralized social cooperation — they are able to employ justice as a source of cooperation, not primarily of conflict. Outside of contemporary philosophy, almost all analyses of human justice understand it as, first and foremost, the basis of human cooperation, and so critically involves social recognition and endorsement of its requirements.[54]

Ethics (Chicago: University of Chicago Press, 1954), esp. chap. 21.

[50] This was the focus of Edward Westermark's classic *Ethical Relativity* (Westport, CT: Greenwood, 1970). See Jesse J. Prinz, *The Emotional Construction of Morals* (Oxford: Oxford University Press, 2007).

[51] For example, beliefs in the supernatural. See Brandt, *Hopi Ethics*, pp. 332ff. Sorcery is the offense most frequently met with capital punishment in hunter-gatherer societies. See Christopher Boehm, *Moral Origins* (New York" Basic Books, 2012), pp. 84-7.

[52] Kant, *The Metaphysical Elements of Justice*, p. 116 [§43]. Emphasis added.

[53] See John Rawls, "An Outline of a Decision Procedure for Ethics" in *Collected Papers*, pp. 1-19.

[54] This idea is everywhere. See Gintis and Bowles, *A Cooperative Species;* Boehm, *Moral Origins;* Philip Kitcher, *The Ethical Project* (Cambridge, MA: Harvard University Press, 2011). The idea that justice is at least partly a social practice that involved coordinated expectations, though now perhaps a minority view in philosophy, was dominant in the middle of the last century. Rawls's 1958 "Justice as Fairness" (*Collected Works*, pp. 47-72) explicitly supposed that justice was a social practice; Kurt Baier's the *Moral Point of View* (Ithaca, NY: Cornell University Press, 1958) was centrally concerned with relation of social morality to mores. See also R. B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979), chap. 9. Thus Hayek's remark on the idea of "social justice" — "one would have thought that all justice is a social phenomenon!" "Kinds of Rationalism" in *The Market and Other Orders* in *The Collected Works of F. A. Hayek*, vol. 15, edited by

Those who suppose that justice is a social relation and so cannot be secured without the cooperation of others will play not the "Go Your Own Way Game" (which is apt to degenerate into a moralistic aggression game),[55] but rather the Reconciliation Game of Display 2:

Betty

		r_1	r_2
Alf	r_1	2rd / 1st	4th / 3rd
	r_2	3rd / 4th	1st / 2rd

Display 2: The Reconciliation Game

r_1 and r_2 here designate rules. Once justice is understood as a social, cooperative relation, the focus becomes rules, practices, or institutions that provide for relatively fine-grained mutual expectations of what others normatively expect of me, and what I expect of them. Critically, rules provide shared public categorizations that allow for coordination.[56] And these in turn ground a practice of accountability, in which I know for which actions others will hold me accountable, and for which I will experience guilt and regret should I perform them.[57] Under these conditions moralistic responses to the violations of others stabilize rather than destabilize the order of cooperation.

In the Reconciliation Game, Alf still engages in "I believe we ought" reasoning, but it does not follow that this always provides sufficient reasons for him to act. On the basis of his reasoning he concludes that a practice based on r_1 best conforms to the demands of justice, Betty disagrees, endorsing r_2 as the best social instantiation of justice. But in the judgment of both, just social relations can only be secured by joint action — again, neither can secure these alone. So Alf holds that if both endorse and act upon r_1 their relations will be fully just, for Betty joint action on r_2 meets full

---

Bruce Caldwell (Chicago: University of Chicago Press, 2014), pp. 39-54 at p. 40. The idea that justice is a property of social relations can be traced back even further — to *The Republic*.

[55] An all-too-familiar game — the Prisoner's Dilemma.

[56] See *Tyranny of the Ideal*, pp. 177-207.

[57] I have made this argument in some depth in *The Order of Public Reason*, chap. 4. Note that unless Alf and Betty morally endorse the same rule, even if they somehow managed to coordinate their actions, they could not maintain a practice of accountability; Betty might, for reasons of self-interest act on r_1, but that would not be sufficient for her to hold herself accountable for acting on it. Without recognition of mutual accountability, punishment will be typically seen as mere moralistic aggression, invoking grievance and perhaps counter-punishment (§5.1)

justice. However, because justice is a joint action they are forced to consider whether the joint action on the other's favored rule better satisfies justice than failure to coordinate on a joint action. They both might agree that there would still be some justice in unilaterally acting on his or her "I believe we ought judgment" (in Display 2, both rank this as their 3rd choice), but unilateral action simply cannot do much to secure justice. Alf (and Betty) asks himself (herself) two critical questions:

(Q1) On my moral perspective, how important is the achievement of just social relations, as opposed to unilateral action on my "I believe we ought" judgment?

(Q2) Given my "I believe we ought" judgment, how far is the other's preferred rule from mine — how much of a moral gap is there — in terms of the justice of the two rules?[58]

It is their answers to these two questions that determine whether they play the Go Your Own Way Game or the Reconciliation Game. The greater the importance of just social relations, and the smaller the gap, the stronger the tendency to play the Reconciliation Game.

5.3 The Social Contract

The possibility of those with diverse moral perspectives coming to shared rules (practice, institutions) via the Reconciliation Game is at the heart of the social contract response to moral diversity. We might distinguish two routes to a social contract that expresses reconciliation, the *state of nature* and *reasonable population* versions. The first is best manifested by Locke, who stresses that the conflict engendered by each acting on natural law or justice as she sees it leads to such a morally disordered state that each sensible person will abandon her acting on her private ("I believe we ought") judgment about justice to be ruled by common law. Thus, like Kant, Locke's answer to Question 1 above stresses the moral disorder and injustice produced by each unilaterally acting on her private judgment — this is the importance of the model of the state of nature, showing why all sensible agents would be driven to seek reconciliation. On Locke's view, "*all private judgment of every particular Member being excluded*, the Community comes to be Umpire, by settled standing Rules, indifferent, and the same to all Parties; and by Men having Authority from the Community, for the execution of those Rules, *decides all the differences that may happen between any Members of that Society concerning any matter of right*…."[59] Locke is sensitive to the limits of reconciliation; these common rules of justice must be within the bounds of

---

[58] I explore this idea of the gap in "The Commonwealth of Bees."

[59] Locke *Second Treatise of Government* in *Two Treatises of Government*, edited by Peter Laslett (Cambridge: Cambridge University Press, 1960), p. 342 (§87). Emphasis added.

reasonable disagreement; each must see the laws as sufficiently just to merit endorsement (Question 2 above).[60] Should those legislating and enforcing the common rules overstep those bounds, individuals are justified in employing their private judgment and rebelling. Specifying this limit can only be internal to a perspective (each makes his own appeal to heaven), but Locke is clearly confident that only if most individuals believe this limit is surpassed will the common system of rules be undermined. [61]

On this route to reconciliation, the description of the disorder of the state of nature is meant to show why sensible moral agents seek reconciliation within some bounds of reasonable disputes about justice. An alternative route is to *begin* by identifying a population of agents who have answered our two questions in "reasonable" ways. It is not uncommon to understand Rawls's original position as a sort of "primordial" state of nature like baseline,[62] but Rawls is clear in rejecting this — the no agreement point in justice as fairness is generalized egoism, not the state of nature.[63] Now an egoism specification of the baseline is appropriate if justice is purely dependent on joint action, for then "no agreement" would entail "no justice" and so the result would be simply generalized self-interest. But once we have distinguished "I believe we ought implies I ought" reasoning from reconciliation, we see that the default need not be egoism. In the Reconciliation Game of Display 2, the default is the (r_1, r_2) outcome, in which Alf and Betty each acts on his or her preferred rule.[64] For those who believe that justice is thoroughly social in the way we have examined, this default could be equivalent to egoism (no justice), while to others it might represent the inherent justice of their personal action on their "I believe we ought" judgments.

Now if we understand the baseline as a no coordination point (leaving behind whether this means no justice at all), the contractual analysis will be sensitive to how much each party gains from coordination. If we had individuals who gained little from it (think of the Go Your Own Way Game), they will have little incentive to reconcile. Each may give the others an ultimatum, and all end up walking away.

---

[60] In contemporary political philosophy this is typically understood as an aspect of a legitimacy — specifying when a law is "too gravely unjust." Rawls, *Political Liberalism*, p. 428.

[61] See further my "Locke's Liberal Theory of Public Reason," pp. 163-83.

[62] Amartya Sen, *The Idea of Justice* (Cambridge, MA: Harvard University Press, 2009), p. 10.

[63] Rawls, *A Theory of Justice,* p. 127.

[64] The language of preference is, unfortunately, usually a red flag to moral philosophers. It is important to always keep in mind that a preference is simply a binary relation according to which one state of affairs or action is ranked as superior to another; it is not a reason or explanation as to *why* it is so ranked. To say that Alf holds that "α is preferred to β" because that "is his preference" is a tautology, not an explanation. A utility function is a mathematical representation of a consistent preference structure. As Rawls recognized (*Political Liberalism*, p. 332n), even deontological theories such W.D. Ross's can be faithfully represented in terms of standard cardinal utility measures.

Rather than deriving the incentive to reconcile from an analysis of the state of nature, this version of the social contract assumes that the parties have a strong moral interest in reconciliation on common rules. Thus Rawls restricts the parties to those who are "ready to propose principles and standards of fair terms of cooperation and to abide by them willingly, given the assurance that others likewise will do so."[65] This supposition entails that all are playing a version of the Reconciliation Game: each strictly prefers some *coordination* on fair common rules to going it alone. Other assumptions, such as the concern of each for their basic interests in pursuing a plan of life and the strains of commitment, ensure that not every possible coordination point is preferred to no agreement,[66] so the parties are truly looking for a way to reconcile that improves (given their moral perspectives, I would say) on the no agreement point.

We thus have a different pair of answers to Questions 1 and 2 than Locke's. To Q1, each has a strong interest in securing just common rules on which we coordinate; people who are inclined to play Go Your Own Way simply are not parties to the agreement. And to Q2, justice as fairness answers that each is willing to accept rules that improve on their no agreement point. Rawlsian-inspired theories have different outcomes once they start varying their answers to these questions. As I read him, Jonathan Quong answers Q2 very differently (at least than Rawls did in *Theory*) indicating that the parties' acceptable gap between their preferred conception and the reconciliation point is restricted by robust liberal commitments.[67] Their no agreement point would be not general egoism, but any non-liberal outcome.

5.4 Must We Commence by Normalizing? The Self-Organization Insight

My main point in briefly sketching these two social contract approaches is to highlight how, since the seventeenth century, liberal political philosophy has endeavored to show that moral diversity does not imply moral disorder. Coordination on moral rules need not be based on full moral homogeneity. Here an acquaintance with elementary game theory helps a good deal; coordination and diversity are by no means incompatible. Indeed, at the heart of the division of labor is the insight that

---

[65] Rawls, *Political Liberalism*, p. 49.

[66] Rawls, *A Theory of Justice*, p. 126.

[67] Quong, *Liberalism without Perfection*, chaps. 1, 5 and 6. There are of course problems here in relating the central place of the contract in *Theory* to the more peripheral role it (and justice as fairness) occupies in *Political Liberalism*. I set aside these issues here. On these complex matters, see the fine work of Paul Weithman, *Why Political Liberalism?* (New York: Oxford University Press, 2010) and Samuel Freeman, *Justice and the Social Contract* (Oxford: Oxford University Press, 2007), chap. 7. See also my "The Turn to a Political Liberalism" in *The Blackwell Companion to Rawls*, edited by David Reidy and Jon Mandle (New York: Wiley-Blackwell, 2014), pp. 235-50.

coordination often is valuable simply *because* we are so diverse.[68] Given that, in addition, moral diversity can improve the moral perspectives of everyone (§§3.3, 4.3), there is nothing at all incoherent about wishing for a society with an array of diverse moral perspectives (§1).

This is not to say that the social contract — or indeed, any — approach accommodates unlimited diversity. "A normalization of interests attributed to the parties" Rawls observes, is "common to social contract doctrines."[69] As we have seen, contract theories assume some level of interest in reconciliation (e.g., all are devoted to finding fair terms of cooperation) and some assumption of a state of nature or nonagreement baseline, which identifies a common view of a minimally acceptable agreement. Although such assumptions render the contractual project tractable, they limit at the outset of the analysis the degree of diversity to be accommodated. The very beginning of the analysis identifies some group *G*, that disagrees on matter *M*, but has a shared interest *L* in reconciliation, given outcomes that improve on a baseline *B*. Every contract normalizes these four variables by specifying the range of acceptable values, and in so doing limits at the outset the moral diversity to be accommodated. Not only must it exclude some from the justificatory community from the get-go,[70] but it must assume away critical sources of diversity. The social contract theorist insists that because our "I believe we ought" reasoning leads us to disagree about the demands of morality and justice (*M*), we need to reconcile with others (*L*); yet the disagreement in individual moral reasoning (*M*) leads us to disagree on the specification of the *L* values as well: *to what extent should we reconcile with others*? Having reasonably accepted that we disagree in our first-level moral judgments, the contractualist seeks to resolve this problem through a philosophical procedure that supposes reasonable people do not have deep reasonable disagreements about the morally desirable degree of reconciliation. In this sense it supposes a normalized interest in reconciliation, and so avoids, rather than responds to, a fundamental challenge of moral diversity.

In various works I have sought to relax this normalization, considering whether the contractual approach can be supplemented by one focusing on self-organization. The point of the idea of a *socially eligible set* in *The Order of Public Reason* is to allow

[68] For an application of this insight to moral diversity, see Ryan Muldoon, Michael Borgida and Michael Cuffaro, "The Conditions of Tolerance," *Politics, Philosophy & Economics*, vol. 11 (2011): 322–344.

[69] John Rawls, *Lectures on the History of Political Philosophy*, edited by Samuel Freeman (Cambridge, MA: Harvard University Press, 2007), p. 226. See further my "Is Public Reason a Normalization Project? Deep Diversity and the Open Society," *Social Philosophy Today*, vol. 33 (2017): 27-55.

[70] I believe Quong's analysis is one of the most radical, excluding all nonliberals at the outset. See my "Sectarianism Without Perfection? Quong's Political Liberalism," *Philosophy and Public Issues*, vol. 2, No. 2 (Fall 2012): 7-15.

each moral agent to identify, given her own perspective, both the set of rules on which she has sufficient reasons to reconcile, and her own baseline, i.e., rules that are ineligible as a shared rule of justice on her perspective.[71] Each moral agent thus identifies a set of rules on which, given her moral perspective, she has moral reason to reconcile (and she orders the acceptability of these rules in light of her moral perspective). A person's eligible set indicates when she is willing to play the Reconciliation Game. She then searches for others with whom she can play the Reconciliation Game to achieve her preferred moral outcomes. The critical point is that, instead of these decisions being made by the moral or political philosopher at the level of initial theory specification, the theory seeks to include the decisions as elements of each person's moral perspective.[72] I see this as a fundamental shift in the concept of a moral theory, placing individual moral perspectives more firmly at the very heart of the analysis.[73] In more recent work, I have tried to understand some of the dynamics of self-organized convergence among those who disagree on both their "I believe we ought" judgments of justice and the extent to which (on their perspective) justice provides reasons to seek reconciliation.[74] Under what conditions does a diverse group of moral agents converge on a common rule, and when are their differences so deep that they form allegiances to different moral rules and networks? On this approach, we seek to *discover* what types of moral diversity are consistent with moral coordination, rather than the philosopher specifying it at the outset. Interestingly (to me, at any rate), this analysis provides some initial grounds for supposing that under a surprising array of conditions, agents with diverse — sometimes even polarized — moral views can organize themselves under common rules.[75] And, most interestingly, sometimes more diverse populations may have a stronger tendency to morally coordinate than less diverse populations. An uptake of

---

[71] *The Order of Public Reason*, §16.

[72] This is also a feature of Muldoon's *Social Contract Theory for a Diverse World.* For an application to a specific category of disagreement see Kevin Vallier, *Liberal Politics and Public Faith* (New York: Routledge, 2014).

[73] For some reflections on what this means for moral method, see my "Social Morality and the Primacy of Individual Perspectives," *The Review of Austrian Economics,* vol. 30 (2017): 377-396.

[74] See my "Self-organizing Moral Systems," *Politics, Philosophy and Economics*, DOI 10.1177/1470594X17719425. I certainly do not wish to give the impression that I am alone in exploring this idea. See Ryan Muldoon, "Exploring Tradeoffs in Accommodating Moral Diversity," *Philosophical Studies*, vol. 174 (2017): 1871–83. DOI 10.1007/s11098-016-0825-x

[75] The most basic reason for this is that reconciliation on a moral rule manifests an increasing returns dynamic: the more others adopt a rule, the greater its reconciliation benefits. This induces cascades to widely shared rules, often eventually joined by those who rather strongly prefer another rule and who do not greatly value reconciliation. High levels of diversity tend to encourage cascades that are not interrupted and so end in fixation on one rule. See "Self-organizing Moral Systems."

this idea would constitute a fundamental change in the orientation of the public reason project, which has thus far supposed that diversity is the problem, and commonality the sole route to sharing. So far from heterogeneity always being an impediment to convergence on a shared rule of justice, some configurations of diversity, I have tried to show, can help secure agreement. We may have made a fundamental error in asking "do we agree enough to live together?" We should, perhaps, be inquiring whether the overall pattern of homogeneity and heterogeneity induces coordination on common ways of living together.

Still, deep moral polarization certainly blocks reconciliation. Everyone has their limits, and in our increasingly deeply polarized societies these may thwart the reconciliation of everyone on many matters.[76] The focus of the social contract on a reconciliation in a unitary national state is unfortunate, for this is precisely the arena that requires, as Milton and Rose Friedman put it, "conformity without unanimity;"[77] a single answer is necessary regardless of dissension. The majority is thus tempted to push on playing an unfortunate version of "Go Your Own Way" in which the minority is not allowed to go its own way, i.e., all must play "We Go My Way." In a highly morally diverse society, when political decision-making pushes beyond maintenance of basic rights and liberties to the legal codification of deeply controversial conceptions of justice, hostility and contempt for the law is apt to be triggered.[78] Politics is ill-equipped to cope with such deep moral disagreement. Each faction, 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 "Now We Go *My* Way!" aggravating mistrust and, in the end, inducing hatred.

Although we are 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

---

[76] See D'Agostino, "How Can We Do Political Philosophy?"

[77] Milton and Rose Friedman, *Free to Choose* (London: Secker and Warburg, 1980), p. 66.

[78] As I argue in "The Open Society and Its Friends: With Friends Like These, Who Needs Enemies?" *The Critique* (January 2017), <accessed August 22, 2017 at http://www.thecritique.com/articles/open-society-and-its-friends/> and in "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). See also Paul Robinson, "Why Does the Criminal Law Care What the Layperson Thinks Is Just?" *Virginia Law Review*, vol. 86 (2000): 1839-1869; William J. Stuntz, "Self-Defeating Crimes," *Virginia Law Review*, vol. 86 (2000): 1871-1899; Gerry Mackie, "Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience," in *Cultural Agents Reloaded: The Legacy of Antanas Mockus*, edited by Carlo Tognato (Cambridge, MA: The Cultural Agents Initiative at Harvard University).

networks of individuals seeking to live together to solve their real social problems, the greater the prospects for reconciliation. As Elinor Ostrom has taught us, self-organizing rule networks are most apt to arise when they correspond to individuals sharing common problems that require cooperative solutions.[79] The idea of a polycentric moral order — in which some fundamental rules are shared by all, but other rules are restricted to sub-communities — helps us to see our way to free reconciliation in a vibrant moral order.[80]

5.5 Avoiding Moral Monoculture

Notice that the Reconciliation Game combines equilibrium with moral disagreement: both (r_1, r_1) and (r_2, r_2) are possible equilibrium solutions, yet both manifest continuing disagreement. In the strong sense of the idea, a society based on either would not be what Rawls called "well-ordered."[81] Although in equilibrium Alf and Betty accept the same rules of justice, they disagree about the justice of those rules. Most moral and political philosophers would not see this as a happy result. Moral disagreement continues, so full confidence in the justice of our fellows is lacking. Further, although in a simple (two strategy) game either rule is an equilibrium result, if we add additional strategies (such as Betty seeking to covert others to r_2, even when they are now playing (r_1, r_1)), the result may not seem dynamically stable. And, if nothing else, the current public reason literature is nearly obsessed with showing how the stability of a just society can be threatened and, with a great deal of work, apparently secured.[82] To many, it is manifest that our ultimate goal should be a society in which it is public knowledge that all agree on the best conception of justice, and that our current rules best satisfy this conception.[83]

This, though, is not so if we accept:

---

[79] See Elinor Ostrom, *Governing the Commons* (Cambridge: Cambridge University Press, 1990).

[80] On polycentrism, see Elinor Ostrom, "An Agenda for the Study of Institutions" in *Choice, Rules and Collective Action*, edited by Fillippo Sabetti and Paul Dragos Aligica (Essex, UK: ECPR Press, 2014), pp. 97-119; Vincent Ostrom, "Polycentricity (Part 2)" in *Polycentricity and Local Public Economies*, edited by Michael D. McGinnis (Ann Arbor: The University of Michigan Press, 1999), pp. 119-38; Paul Dragos Aligica, *Institutional Diversity and Political Economy* (Oxford: Oxford University Press, 2014).

[81] Rawls, *A Theory of Justice*, pp. 397-8.

[82] This arose from Rawls's analysis of overlapping consensus in *Political Liberalism* (Lecture IV), and was greatly spurred by Weithman's, *Why Political Liberalism?* Given the strong conformity bias of humans, moral stability is, I think, better seen as the default rather than a hard-to-achieve result. On the strong influence of conformity to the present set of rules, see Peter J. Richerson and Robert Boyd, *Not by Genes Alone* (Chicago: University of Chicago Press, 2005), 120-4, 162-63; Natalie and Joseph Henrich, *Why Humans Cooperate* (Oxford: Oxford University Press, 2007), pp. 11-13, 22-30.

[83] See the very demanding conditions for "full publicity" in *Political Liberalism*, pp. 66-71. A well-ordered society is said to satisfy them.

The Relevance of Environment: That moral rule *r* is part of the best conception of justice for society *S* at *t* partly depends on the overall social and environmental conditions characterizing *S* at *t*.

These environmental factors include population density, cultural norms, technological factors, forms of economic cooperation, common social categories (do some categorize others by race and ethnicity?), forms of family organization, and so on. If a significant subset of these are relevant to the claim that *r* is best for *S* at *t*, then even if *r* is indeed the best or optimal rule for *S* at *t* it would be unfortunate for all in *S* to take *r* as the best and drop allegiance to competing rules. For this would make *S* a moral monoculture in relation to *r*, but as we know monocultures are fragile. They can do wonderfully in some environments, but collapse in others. In contrast, diverse systems have less variance in their performance given environmental change.[84]

Moral diversity functions in many ways like biodiversity. When the environment changes, in a diverse system some elements that did well under previous conditions become less functional, while other elements, which may have been ill-adaptive to the previous system, now are ready to be exploited and assume a more important place given the new conditions. In a world without effective contraception and antibiotics, moral rules against pre- and extra martial sex may have promoted a host of morally important goals, albeit at costs to human freedom. That dissidents, holding different conceptions of justice, placing more weight on freedom, argued for greater sexual freedom in the nineteenth century meant that these views were already represented and defended, so that morality could quickly adjust to environmental conditions, such as advances in medicine and greater economic opportunities for women. Societies that had a greater moral monoculture on this issue — perhaps some Catholic societies — had a significantly greater difficulty in adjusting to the new environment, experiencing high transition costs (for example, higher rates of untreated sexually-transmitted diseases and so unnecessary deaths and social disruption).

As Hayek noted, "we are as little able to conceive what civilization will be, or can be, five hundred or even fifty years hence as our medieval forefathers or even our grandparents were able to foresee our manner of life today."[85] Because of this continuing moral disagreement is by no means regrettable.[86] It spurs moral

[84] See Page, *Diversity and Complexity,* chaps. 6 and 7. Cf. Hayek: "A change in the environment may require, if the whole is to survive, a change in the order of the group, and therefore in the rules of conduct of individuals." "Notes on the Evolution of Systems of Rules of Conduct" in his *Studies in Philosophy, Politics, and Economics* (Chicago: University of Chicago Press, 1967), pp. 66-81 at p. 71.

[85] Hayek, *The Constitution of Liberty* (London: Routledge, 1960), pp. 23-4.

[86] See further my *Tyranny of the Ideal*, chap. 4, and "The Role of Conservatism in Securing and Maintaining Just Moral Constitutions" in *NOMOS LVI: American Conservatism,* edited by Sanford V. Levinson, Joel Parker, and Melissa S. Williams (New York: New York University Press, 2016), pp. 256-291.

discussion, learning and adjustment to a constantly changing, and almost always surprising, world. Given this we also come to better to see how diversity is distinct from simple disagreement, for a diverse system has a wide range of views: views are not simply different, but far apart (variance is high on many issues),[87] providing the moral resources for the next era of our society, whatever that may be. Lots of small differences do not for diversity make.

6 MORAL DIVERSITY AND MORAL COMPLEXITY

6.1 The Impossibility of a Non-normalized Social Contract

We have seen, then, (*i*) that moral diversity does not entail moral disorder. Social contract theory provides at least two routes from moral diversity to moral coordination — the state of nature and original position set-ups. (*ii*) Although the social contract is the most developed philosophical device for generating moral coordination in the face of moral diversity, it does so through putting constraints — normalizing assumptions — on admissible diversity. The social contract, then, is an important, but limited, moral method for securing reconciliation in the midst of diversity. (*iii*) I have sketched an alternative where each moral agent identifies a ranked set of rules on which she can reconcile (i.e., her perspective's eligible set) to secure justice. She then searches for others seeking justice-through-reconciliation, playing reconciliation-type games with them. The appropriate method here is not a collective agreement *a la* the social contract, but agent-based modelling of agent's diverse moral choices, and the conditions under which coordination on common moral rules will occur. The advantage of this unorthodox approach is that the theorist does not, at the outset of analysis, identify the range of diversity consistent with a moral order that can be endorsed by all.

That all social contracts normalize does not imply they must. So, one might wonder, can the contract device accommodate wide diversity?[88] To answer this query, suppose the task of the diversity contract is to take as inputs an array of diverse of sensible perspectives $\{P_A...P_N\}$, and consider alternative sets of rules (institutions)[89]

---

[87] Thus diversity can be measured in terms of the degree of variation. See Page, *Diversity and Complexity*, chap. 3; Martin L. Weitzman, "On Diversity," *The Quarterly Journal of Economics*, vol. 107 (May, 1992): 363-405.

[88] Suppose again (§2) that we have ten degrees of difference on each of our six elements; we would have perspectives $(I_1, II_1, III_1, IV_1, V_1, VI_1)$ and $(I_{10}, II_{10}, III_{10}, IV_{10}, V_{10}, VI_{10})$. These are apt to be *very* different, disagreeing the maximal allowable extent on every element of a perspective, perhaps not even seeing each other as fully intelligible. It does not take strong assumptions to generate a high degree of diversity: it takes strong assumptions to preclude it.

[89] On Elinor Ostrom's analysis, institutional arrangements are composites of rules. Larry L. Kiser and Elinor Ostrom, "The Three Worlds of Action" in *Polycentric Games and Institutions,* edited by

$\{R_X\}$, $\{R_Y\}$ and to find which satisfies some endorsement test of $\{P_A...P_N\}$, and so would specify the terms of the contract. Consider P_B, the perspective of contractor Betty. She asks herself: how well does set of rules $\{R_X\}$ satisfy the demands of her perspective? To answer this she needs to consider, given a large set of perspectives held by a vast number of other agents, the justice of a social order regulated by $\{R_X\}$.

The first tack might be to disaggregate the set $\{R_X\}$ into its constituent rules r_1, r_2, r_3 and so on, and evaluate them one-by-one. If the rules in the set $\{R_X\}$ displayed no interactions this would seem feasible, i.e., if rule r_1 secured justice of degree a and r_2 justice of degree b, where the overall justice of both was $a + b$. Betty could then simply evaluate the justice of each rule, and aggregate her results. However, the assumption of the independence of each rule's justice is implausible: systems of rules and institutions display interactions.[90] A rule that invokes heavy penalties for stealing may have very different implications for justice if there is also a rule that requires assistance to those in need, or when the rules of property are vague. This is not a philosophical conjecture. In their extensive fieldwork on actual 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."[91]

We need not assume maximal interconnectedness:[92] clusters of rules will form systems reasonably independent of the rest of the set. But identifying such clusters will itself be a matter of discovery for Betty. This is not to say that she cannot make tentative judgments on the justice of any specific rule (we all clearly do), but these really must be tentative: she must constantly investigate which rules are interdependent and which are relatively independent in relation to each other. This is not a formal part of the rule; its interactions with other rules is often unexpected and can change as circumstances vary. We could not have anticipated, for example, that in reaction to a rule prohibiting the consumption and sale of alcohol people may become more skeptical of legal regulation, more prone to approve of alcohol, and less suspicious of organized crime.[93] Betty thus must engage in an ongoing iterative process, observing a rule, evaluating the justice of the resulting social world, but trying to see whether this justice was linked to other rules, and seeing how modifying them affects the justice of the resulting social world, and so on. Because perspectives

Michael D. McGinnis (Ann Arbor: The University of Michigan Press, 2000), pp. 56-88 at p. 56.

[90] Hayek, "Notes on the Evolution of Rules of Conduct," pp. 70-1.

[91] Ostrom, "An Agenda for the Study of Institutions," p. 111.

[92] Which is to the well and good, since this would be apt to produce chaos. See *The Tyranny of The Ideal*, pp. 67-72.

[93] See the references in note 78 for analyses along these lines.

are diverse, they will not all reach the same conclusions about what rules are interconnected, so that cannot be part of the shared contractual set-up. If there is a contract here, it is a highly unusual one, in which each individual is constantly changing her views about the justice of social rules in response to other people changing their views.[94] Indeed, this is precisely the self-organizing picture (§5.4), in which individuals are constantly searching for reconciliation, better understanding what rules serve their moral ends.

The problem is, however, deeper. Suppose we focus on only one rule, say a rule allowing abortion in first and second trimesters. Suppose Betty is evaluating the justice of this rule in a highly morally heterogeneous population. Applying Rawls's four-step CI procedure (§3.1), she asks herself whether she can will that this rule be a law of nature in her social world, given her perspective. But the moral upshot of this rule will depend on not simply the degree of conformity to the rule's deontic imperatives, but on how those with other moral perspectives react to the rule, and the morally-relevant options they take.[95] Rules, after all, seldom mandate specific actions: they generally permit and prohibit actions.[96] So given a rule that says "one may abort in the first trimester" she cannot know what people will do in response, even with the "idealized" assumption of universal compliance. Some may abort, approve of it, and encourage others to abort; some may abort with moral uncertainty; others may never think of having an abortion (and so comply), but observing some of the ill-effects of the permission (and most rules have some ill-effects) may come to disapprove of those who abort, even while allowing their right to.[97] This last group may come to see those who have an abortion as morally flawed, and so move some measure from reconciliation to conflict with them on other issues, leading to increased moral disorder (§5.1). All this is part of the social world produced by a rule allowing abortions, and there is no possibility that anyone could reliably predict the results when the rule is being proposed.

This is even more obvious with prohibitory rules, compliance with which can bring about radically different results.[98] Consider a moral rule that prohibits religious arguments in the political deliberations about basic justice.[99] Suppose that societies S_1,

---

[94] This is critical to Muldoon's iterative social contract. *Social Contract Theory for a Diverse World*, chap. 7.

[95] Recall that what options are relevant is itself part of a moral perspective.

[96] See Gerald Gaus and Shaun Nichols, "Moral Learning in the Open Society," *Social Philosophy & Policy*, vol. 34 (Summer 2017): 79-101.

[97] There is a reasonable conjecture that this occurred in the case of abortion. See my "The Open Society and its Friends."

[98] This is a central claim of Gaus and Nichols, "Moral Learning in the Open Society."

[99] Rawls, with a rather complex proviso, endorses this as a moral duty. See his "Public Reason Revisited" in *Political Liberalism*, pp. 440-90.

S_2 and S_3 all full comply with the rule. In S_1 many are religious, and while they comply, their moral perspectives lead them to retreat from the public sphere, where they cannot appeal to what they consider the fundamental basis of their convictions, leaving political matters to their secular brethren. In S_2 religious people tend to have a much stronger devotion to civic engagement, and so participate actively in political debate while complying (though perhaps with some misgivings) with this duty. In S_2 the secular citizens, interpreting this moral duty as confirming their conviction that religious arguments are bogus and are unworthy of admittance into public debate (their scientific arguments *are* admissible, after all), become even more dismissive of religious comprehensive doctrines. All three societies perfectly comply; the emerging moral relations between citizens are vastly different. Now imagine the multitudinous combinations of simply these three moral responses and the different moral orders that might result from the diversity of perspectives regarding this single moral rule. Then add other interacting rules and institutions. Now — employing one's preferred predictive model(s) — try to predict the justice of the emergent overall order of moral relations from each set of rules and perspectives. Such systems are characterized by complex feedback relations; each agent's reactions to the last interaction become inputs for the next.[100]

It would be impossible for intelligent parties to the contract — or even the contractualist philosopher — to say with any confidence what the moral upshot of a set of rules would be in comparison to alternative prima facie attractive sets. Indeed, attempts to mathematically model them quickly become incalculable. Betty might have many conjectures, but these would be constantly revised as she learns about other's moral perspectives, and as others revise, and she learns about their revisions and they respond to hers. The only way to identify a reasonably firm contractual output on a set of rules or institutions, again, is to control the level of diversity: to suppose agents have fairly homogeneous aims, concerns and moral outlooks, and so we can accurately predict what they will do in a rule or institutional setting.[101] That is, to normalize.

---

[100] "Complexity … asks how individual behaviors might react to the pattern they together create, and how that pattern would alter itself as a result." Brian Arthur, "Complexity Economics," SFI Working Paper: 2013-04-012. See Hayek, "The Theory of Complex Phenomena;" Paul Lewis, "Purposeful Behaviour, Expectations, and the Mirage of Social Justice: The Influence of Cybernetics on the Thought of F.A. Hayek," SSRN-id2790169; John H. Miller and Scott E. Page, *Complex Adaptive Systems* (Princeton: Princeton University Press, 2007), pp. 50-3.

[101] Many working in the broadly Rawlsian tradition maintain that the moral duty of civility described in the previous paragraph promotes civic trust and assurance. Although I admire their theoretical sophistication, those who seek to formally model the dynamics of assurance games based on this duty miss the crucial point — there is no good reason to think that assurance is the relevant emergent property: the same moral duty could easily breed contempt for, or the passivity of, the religious. Only by strongly normalizing the relevant perspectives can this result be avoided.

6.2 How Emergence Undermines Traditional Political Philosophy

As Ruth Lane observes, on the dynamic we have been exploring "every agent is constantly reacting to its neighbors; this eliminates any reference to an outside governor; and the reaction is in terms of *basic rules*…."[102] Once we have arrived at this juncture, our moral system — whether or not it is described in the language of contracts — is self-organizing and gives rise to the overall moral order as an emergent property of countless diverse individuals responding to each other's moral decisions. Having built in widespread diversity, considerable interconnectedness and individual moral choices according to rules, the moral system is, in the technical sense, complex.[103]

Hayek was one of the first to recognize that under these conditions of complexity, the overall moral order is an emergent property of the system of rules employed by diverse agents.[104] Emergent properties are sometimes distinguished from mere "resultant" properties on the grounds that, while a resultant property is the expected consequence of an underlying set of properties, emergent properties are very often novel and unexpected. In perhaps the earliest analysis of such systems, John Stuart Mill considered a system, say, *S*, composed of elements (e.g., rules) $\{r_1…r_n\}$ and an overall resulting order *O*.[105] Mill proposes three features of property *O*:

(1) *O* is not the sum of $\{r_1…r_n\}$;

(2) *O* is of an entirely different character than $\{r_1…r_n\}$;

(3) *O* cannot not be predicted or deduced from the behavior of the members of $\{r_1…r_n\}$ considered independently (i.e., apart from their interactions in *S*).

[102] Ruth Lane, *The Complexity of Self Government* (Cambridge: Cambridge University Press, 2017), p. 10.

[103] See Melanie Mitchell, *Complexity* (Oxford: Oxford University Press, 2009), p. 12. Mitchell's book is an excellent introduction to complexity. For helpful accounts, see also Page, *Diversity and Complexity*; John H. Holland, *Complexity* (Oxford: Oxford University Press, 2014). For analyses stressing the relation between complexity and agent-based models, see Robert Axelrod, *The Complexity of Cooperation* (Princeton: Princeton University Press, 1997); Miller and Page, *Complex Adaptive Systems*.

[104] See, for example, Hayek, "The Theory of Complex Phenomena" (pp. 22-42) and "Notes on the Evolution of Systems of Rules of Conduct" in his *Studies in Philosophy, Politics, and Economics*. See also my "Hayekian 'Classical' Liberalism" in the *Routledge Handbook of Libertarianism*, edited by Jason Brennan, Bas van der Vossen, and David Schmidtz (New York Routledge, 2018).

[105] John Stuart Mill, *A System of Logic* in *The Collected Works of John Stuart Mill*, vol. 7, edited by J. M. Robson (Indianapolis: Liberty Fund, 2006), pp. 370-73; 438-40. For a helpful analysis see Sunny Y. Auyang, *Foundations of Complex-systems Theories in Economics, Evolutionary Biology and Statistical Physics* (Cambridge: Cambridge University Press, 1998), section 22.

Later analysis of complexity, such as Hayek's, add that given the novel *O*'s that may arise given the same underlying set of rules, we can, at best, only predict general patterns that may characterize the emerging orders, not their specific features.[106]

Now since Plato political philosophy has assumed that justice is simply a resultant property of a set of institutions, and thus a critical task is to specify the institutions of a society that yield a just moral order. In this vein Rawls examines five politico-economic systems "(a) laissez-faire capitalism [the classical system];[107] (b) welfare-state capitalism; (c) state socialism with a command economy; (d) property-owning democracy; (e) liberal (democratic) [market] socialism."[108] Rawls, and many of his followers (as well as critics), insist that useful judgments of the overall justice of such systems can be made, clearly treating the overall justice as resultant property of the sets of institutions. "Here is the set of rules, {*R*}, and here is the system of social relations *S* that will result — let us evaluate its justice." The assumption is that the relation of these institutions to overall justice is essentially linear: we can predict in a relatively confident way that, say, property owning democracy will be more just than welfare-state capitalism. But 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, and under conditions of extensive moral diversity these have the characteristics of a complex system. The overall justice of the same set of institutions could vary immensely, even under the assumption of full compliance. Moreover, given the interconnections between the functioning of the institutions, we cannot suppose that the closer we get to the ideal institutions of property-owning democracy the more just our society will be. Given complexity, it is easily the case that even if the ideal version of property-owning democracy is exceptionally just, getting *very close* to that social state (say {*R**}, which differs from the ideal {*R*} in only one rule) may have an overall justice that is far from {*R*}; many welfare-state capitalisms could be more just that {*R**}.[109] This is not simply "an intuition" that confident supporters of property-owning democracy can reject on the basis "commonsense" reflection.[110] Given the conditions we have articulated,

---

[106] Hayek, The Theory of Complex Phenomenon."

[107] "Laissez-faire" is a misnomer; as Rawls describes this system it includes a "rather low social minimum." His description suggests the classical system in political economy, which was most definitely not a laissez-faire system — laissez-faire was characteristic of the Manchester School and French Physiocrats. It is unfortunate that, despite the efforts of historians of political economy, this confusion is still common. See Lionel Robbins, *The Theory of Economic Policy in English Classical Political Economy* (London: Macmillan, 1952).

[108] Rawls, *Justice as Fairness*, p. 136.

[109] See *The Tyranny of the Ideal*, chap. 2 for an analysis of how this occurs.

[110] As, I think, do Blain Neufeld and Lori Watson, "The Tyranny—or the Democracy—of the

complexity, nonlinearity and emergence follow, and so the justice of institutions cannot be a resultant property.

7 COMPLEX MORAL DIVERSITY AND THE TASKS OF SOCIAL PHILOSOPHY

We have arrived at seemingly dire conclusions for a moral, social and political philosophy of diversity. Without substantial normalization, the social contract cannot identify the best rules of justice (§6.2); and even if it could identify a common vision of the just society, it cannot recommend institutional schemes to achieve it (§6.3). It is worth stressing that these conclusions arise from grappling with extensive moral diversity. If a theorist is willing to normalize enough, the modeled agents will be sufficiently homogenous in their anticipated justice-relevant choices that parties to the agreement can develop an institutional plan for justice. And if a philosopher is, as it were, willing to max out on normalization, she can simply defend one perspective as the correct one, and theorize on the assumption that everyone shares this perspective.[111] The very foundations of these familiar approaches to social and political theorizing are corroded by the New Diversity Theory.

It does not follow that New Diversity theory has nothing to say about institutional design. Although the complexity of a diverse moral system precludes institutional blueprints for the just society, or the identification of the overall pattern of just social relations, as Hayek insisted the general principles on which complex systems operate can be discovered and inform our political theorizing.[112] It is sometimes claimed that Hayek's thought is, at bottom, contradictory: he insists on our ignorance of social processes but, out of his analysis of ignorance, he generates prescriptions about what we should do.[113] If we know enough to say why, for example, socialism will not work, then we must have good enough insights into the economic order to intervene to promote social goals. This, though, is mistaken. On a complex systems analysis we can know quite a lot about the principles on which complex orders operate, and this theoretical knowledge can allow us to say that some system states cannot be achieved and that some ways of organizing social cooperation are more apt to accommodate

---

Ideal?" in *Cosmos + Taxis*, forthcoming.

[111] Although in relation to G. A. Cohen's egalitarianism, I argue in "The Commonwealth of Bees" that even if all share the same egalitarian ethos, reconciliation is still necessary.

[112] This, of course was one of Hayek's main — and often badly misunderstood— themes. See his *The Mirage of Social Justice* (Chicago: University of Chicago Press, 1976). It is hard to decide whether our understanding of Hayek's work has been set back more by the resolutely unsympathetic readings of hostile critics, or the banal readings of knee-jerk libertarians.

[113] Geoffrey M. Hodgson, *Economics and Evolution* (Ann Arbor, MI: University of Michigan Press, 1993), p. 183.

coordination by diverse agents than others, while acknowledging that we are unable to predict the course of, or control, the complex order itself.

In my work, for example, I have defended the following "devices" as structural principles that promote coordination and moral improvement under conditions of extreme moral diversity:[114]

1. Information and ideas should flow freely, providing opportunities for (*i*) individuals to interact with those who share their perspective or allied perspectives and (*ii*) with those who have more distant moral perspectives, but who face common problems, so that they can find common ways to tackle common practical problems.

2. The substantive normative injunctions of the set of rules {*R*}, whatever it might turn out to be, should be primarily in the form of prohibitions rather than permissions. As Nichols and I have shown, systems of rules that primarily take the form of prohibitions tend to lead participants to endorse "the principle of natural liberty," a meta-level interpretive principle according to which whatever is not (morally) prohibited is (morally) permitted.[115] If some new act-type is not to found on the list of prohibitions, it is permitted; thus moral experimenters need not first convince themselves that a new action falls under a previous permission before exploring new ways of coordinating and acting morally.

3. The set of moral rules {*R*} should extensively employ "jurisdictional rights" in which each is assigned a social space in which her perspective has authority. Such rights tend to render the system emerging from {*R*} less complex by decoupling perspectives: what Alf does in his jurisdiction is not a source of moral claims by Betty. This allows high levels of change in some perspectives without affecting coordination on {*R*}.

4. Markets must be a primary modality of social coordination, as they provide bridges between the valuings and categorizations of different perspectives; even though two parties do not concur on the same value or categorizations of the objects traded, they can agree to trade on mutually acceptable terms.[116]

5. Polycentric systems can accommodate extreme diversity by allowing some to forge some moral networks that are not shared by all. Feminists, libertarians or

---

[114] I have greatly benefited from exchanges with Fred D'Agostino about these issues, as well as from reading his insightful essay, "How Can We Do Political Philosophy?" The points sketched here are defended in *Tyranny of the Ideal*, pp. 177-208.

[115] These claims are based on our experimental results with ordinary learners. Gaus and Nichols, "Moral Learning in the Open Society."

[116] This is a point stressed Muldoon, Borgida and Cuffaro, "The Conditions of Tolerance."

vegetarians may, on some issues, develop shared rules of reconciliation that are not endorsed by the entire moral order. All are not forced to play "We All Go My Way" games (§5.4). At times, as with sexual morality in the latter half of the twentieth century, widespread moral change emerges out of these narrower networks, eventually reconciling a larger moral order.

Note that some of these (points 1 and 3) are recastings of familiar liberal commitments of freedom and rights, while point 2 stresses the importance of the formal features of rules, which are of long-standing importance in the liberal tradition, though perhaps less appreciated in current political philosophy.[117] Similarly, point (5) revives traditional liberal themes about the importance of multiple levels of moral life.[118]

In this context I intend this list as simply illustrative of types of "design" analyses that are consistent with the complexity insight. Other New Diversity theorists also stress institutional devices for learning, experimentation, and exchange.[119] These converge with recent analyses of democracy that stress its ability to learn, and to draw on diverse perspectives, to secure better results for all.[120] Self-organization ("bottom-up") and self-governance ("top-down") are different sources of order; complexity precludes some grander projects of self-governance, but even complex systems employ mechanisms of self-governance and collective experimentation.[121] The type of self-governance consistent with moral complexity is itself a critical topic for inquiry.

That the philosopher abjures constructing institutional blueprints and specifying end states thus by no means sounds the death knell of social and political philosophy. Rather than philosophy as advocacy of a preferred set of perspectives, it becomes an investigation into the nature of free, open, and dynamic diverse moral orders.

Philosophy and Moral Sciences
University of Arizona

---

[117] See F. A. Hayek, *The Constitution of Liberty*, chap. 10.

[118] See Jacob T. Levy, *Rationalism, Pluralism and Freedom* (Oxford: Oxford University Press, 2015).

[119] Muldoon, *Social Contract Theory for a Diverse World*; Müller, *Polycentric Democracy.*

[120] See Jack Knight and James Johnson, *The Priority of Democracy* (Princeton: Princeton University Press, 2011); Landemore, *Democratic Reason.*

[121] This is a critical theme of the work of the Ostroms. For a broader and insightful analysis of the relation between self-organization and self-governance, see J. T. Ismael, "Self-Organization and Self-Governance," *Philosophy of the Social Sciences*, vol. 41 (2011): 327-351.