Between Discovery and Choice:
The General Will in a Diverse Society

Gerald Gaus

1 THE NORMALIZED SOCIAL CONTRACT

In his *Lectures on the History of Political Philosophy* Rawls tells us that "a normalization of interests attributed to the parties" is "common to social contract doctrines."[1] Thus in his account of Rousseau, Rawls characterizes the general will as a shared "point of view."[2] As Rawls reads Rousseau, as private individuals we have a variety of different interests, we are characterized by self-bias and selfishness. Such individuals can live together under freely endorsed common laws if they "share a conception of the common good."[3] Such a shared conception is generated by their fundamental interests and capacities, which derive from their shared human nature. On Rawls's reading, these shared fundamental interests allow the parties to abstract from their differences and occupy a shared, legislative point of view, based on a shared conception of the common good.[4] When occupying this shared view they all have the same basis for their deliberations, and so will the same laws: and that is why they freely legislate common laws.

If we accept this reading, Rousseau and the other social contract theorists appear to see diversity as a problem to be handled and contained by showing that, hidden within diversity is really a deeper agreement, and so in an important sense justificatory diversity is illusory. On the first look it may seem that we confront basic diversity, but then we see this is not so when we abstract to the perspective of a shared point of view of the common good. In this way it may well seem that Rousseau is in the end unable to come to grips true justificatory diversity. I aim in this essay to show why this interpretation of the social contract is unpersuasive in its own terms and, more importantly, that the social contract so conceived is unsuitable as the basis of a diverse society of free and equal persons.[5] It only sees diversity and difference as problems to be coped with, not resources to be employed, in justifying common rules. Indeed, I shall maintain that this normalization view is especially inapt as a reading of Rousseau: of all the great social contract theorists Rousseau provides the keenest

---

[1] John Rawls, *Lectures on the History of Political Philosophy*, edited by Samuel Freeman (Cambridge, MA: Harvard University Press, 2007), p. 226.

[2] Ibid., pp. 229ff.

[3] Ibid., p. 224.

[4] Ibid., p. 224.

[5] Ryan Muldoon's *Diversity and the Social Contract (*Ph.D. dissertation, University of Pennsylvania) defends much the same conclusion, partly on similar grounds, though in crucial respects his fascinating analysis differs from that offered here.

insights as to how diversity is not only consistent with, but conducive to, the free endorsement of common rules.[6] Properly interpreted, the idea of the general will provides the key to understanding how under conditions of deep diversity we can achieve a morality of free individuals that best promotes the good of all.

2 WHY NORMALIZATION IS THE PROBLEM, NOT THE SOLUTION

The entire normalization view of the social contract — and, more generally the contemporary Rawlsian public reason project — supposes that *if* we could only achieve normalization, if only the problem of social evaluation could be reduced to the reasoning based on a single perspective, than we would have solved the problem of uncovering the social rules (laws, basic structure, or whatever) that would promote the common good. To be sure, there may be costs incurred and controversial moves made along the way (i.e., the original position and the veil of ignorance), but these would be well worth it if they could identify a shared perspective for evaluation.

To "share a point of view" is not simply to share a set of fundamental interests, but involves also sharing (*i*) a basic conception of the nature of the social world and (*ii*) a way to map the common good requirement onto social worlds so that they can evaluated. These last two features are essential if the shared perspective is to succeed in its aim of producing evaluative agreement. It would be of no avail to share a set of interests if deliberators, say, employed different trade-off rates among the different dimensions of the common good (and, so, employed different mapping functions from the common good to the world). But even sharing mapping functions would not be enough: they must share the same perspective on the social world and what parts of it are relevant to the common good.[7] Is the common good to be applied to the public and not the private; and if so, what parts of the social order are public and which are private?[8] Is it be applied to the basic structure — and what does the basic structure encompass? Is the family part of it? Is the common good to regulate the political or social? A perspective thus identifies the aspects of the social world that are relevant to evaluation.[9]

---

[6] Rousseau's theory of the general will is, of course, open to multiple interpretations; I have defended elsewhere the sort of constructivist reading I pursue here. See "Does Democracy Reveal the Voice of the People? Four Takes on Rousseau," *Australasian Journal of Philosophy*, vol. 75 (June 1997): 141-162. I incline more here to what I called "individualistic public constructivism" in that essay.

[7] On perspectives, see Scott Page, *The Difference* (Princeton: Princeton University Press, 2007), pp. 30ff.

[8] This is a matter on which different viewpoints disagree. See S.I. Benn and G.F. Gaus, eds. *Public and Private in Social Life* (New York: St. Martins, 1983).

[9] Rawls seems to disagree, holding that a "sphere of life" is not "something already given apart from political conceptions of justice." John Rawls, *Justice as Fairness: A Restatement*, edited by Erin Kelly (Cambridge: Harvard University Press, 2001), p. 166. Habermas, in contrast, holds that Rawls's account presupposes a basic perspective identifying the public and private, or the political and non-political. Jürgen Habermas, "Reconciliation through Public Reason: Remarks on John Rawls's Liberalism," *Journal of Philosophy*, vol. 92 (March 1995): 109-131 at p. 129. I

Now it is important that to share a "point of view" is not only to share a conception of the social world, but it is to share a perspective in the sense that some possible social worlds appear quite close and other distant. Think about Rawls's perspective, in which the object of evaluation is a basic structure of a democratic society. Rawls includes the family in the basic structure;[10] but we might also say that for a Rawlsian a social world in which the family is excluded from the basic structure is very close to the Rawlsian perspective, certainly closer than the world in which the economy is excluded from, or the Catholic Church is included in, the basic structure. A social world in which the common good only applies to the political, leaving the market a sort of morality-free zone, is very far from the social world as Rawlsians see it — and they may have a hard time even making sense of it. Indeed Samuel Freeman has argued that libertarianism is not really a liberal doctrine at all because it has no public sphere — it does not conceive of the social world in terms of a political-public sphere, and so he sees it as so distant as not to be a liberal view of the world at all.[11]

A shared perspective in this sense is needed if the normalization contract is to succeed — but it almost surely means that we shall fail in finding a social world that best satisfies the requirements of common good. Our deliberators (or, at this point, our single deliberator) confronts an optimization problem: given the normalized perspective P that relates different dimensions of the common good (the different fundamental interests), and maps them on to social worlds $S_1...S_n$, what social world (or basic structure, or set of laws, rules, etc.), optimizes over all the dimensions of the common good (i.e., all the fundamental interests involved)? If the dimensions of the common good are interdependent — if, for example, the dimension of economic well-being varies to some extent with absence of harm, the amounts of freedom and equality — then the evaluators are faced with the sort of complex decision problem recently analyzed by Scott Page, Fred D'Agostino and others.[12]

These can be exceedingly difficult problems to solve. Given our current perspective P, these problems often form rugged landscapes.[13] To see the problem, contrast the simple two-dimensional cases in Figures 1 and 2.

---

believe that Habermas is correct. See my *Contemporary Theories of Liberalism* (London: Sage, 2003), pp. 186ff.

[10] John Rawls, "Public Reason Revisited" in his *Law of Peoples* (Cambridge, MA: Harvard University Press, 1999): 129-180 at pp. 156-64.

[11] Samuel Freeman, "Illiberal Libertarians: Why Libertarianism Is Not a Liberal View," *Philosophy & Public Affairs*, vol. 30 (no. 2): 105-151.

[12] See Page, *The Difference*; Fred D'Agostino, "From the Organization to the Division of Cognitive Labor," *Politics, Philosophy & Economics*, vol. 8 (January 2009): 101-129 and his *Naturalizing Epistemology* (London: Palgrave, 2010). See also Michael Weisberg and Ryan Muldoon, "Epistemic Landscapes and the Division of Cognitive Labor," forthcoming *in Philosophy of Science*.

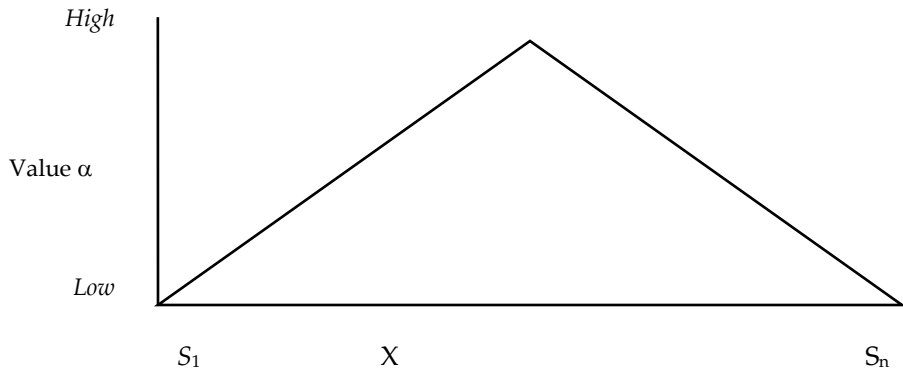[13] I am following here Page, *The Difference*, pp. 45ff.
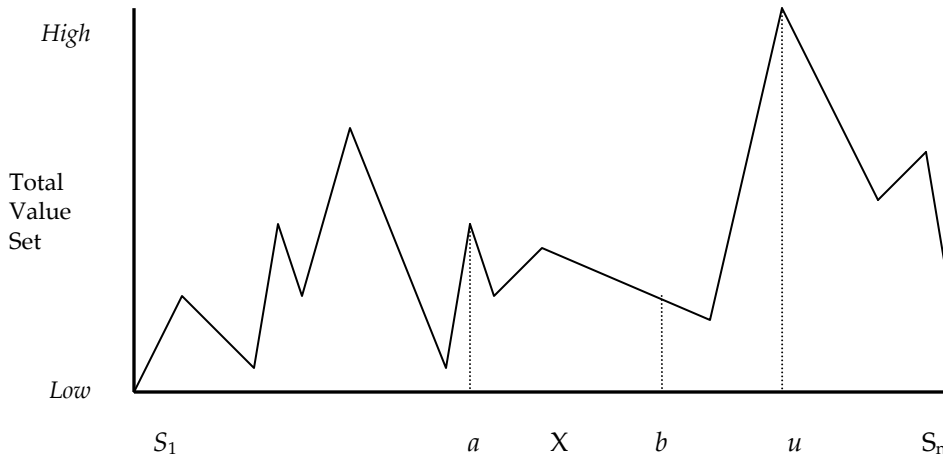
FIGURE 1: A MOUNT FUGI LANDSCAPE

FIGURE 2: A RUGGED LANDSCAPE

In Figure 1, the extent to which a social arrangement satisfies value α is mapped on the vertical axis. The horizontal axis employs a perspective, which ranges social arrangements in terms of their proximity. (How alike are these social arrangements, given this point of view?) For example, suppose the point of view is the traditional left-right perspective, in which social state S_1 is the far left and social state S_n is the far right, and suppose α is the value of individual freedom. Suppose that we are now at X. Note that in Figure 1, for any social state (except at the peak, the global optimum) there is a proximate social state which does better at satisfying α. In this case, the value promotion problem is easily solved: we only have to have knowledge of the social state near our present state of X, knowing this we know what direction to travel — what changes are recommended by α. And at each point along the path, we will be faced with the same happy problem: mere knowledge of the proximate social states is enough for us to make a decision, and the series of decisions will lead to the global optimum. If a society begins locally optimizing on a Mount Fugi landscape, it

ends up at the global optimum through a series of steps, each of which led to a better social state.

The problem is much more difficult in Figure 2. Here the vertical axis maps the total satisfaction of a complex bundle of interrelated values (their relation is not simply additive).[14] Again, the horizontal axis arranges social arrangements according to some perspective. Here, at social state X we are at a local optimum; moving in either direction from X will mean a decrease in overall satisfaction of some our values α…ω. This might occur, for example, if at our state of X value β is highly satisfied, but in ways that preclude the satisfaction of many others (perhaps we have an anarchistic sort of liberty); given the perspective employed, in the near social states there would be less satisfaction of, say, value β, but there would not yet be significant gain in other values, resulting in a net loss of total value satisfaction in social states adjacent to X.

If under our highly normalized contract where we all share the landscape of Figure 2, if we are now in social state X, and if we are boundedly rational, and so only have firm ideas of the evaluation of relatively proximate social worlds, we may well conclude that our present social world is optimal: moving in any direction to proximate social worlds will satisfy the requirements of the common good less well than in our present social world. We are then caught at a local optima, which is considerably short of the global optima. Suppose instead that we are a less boundedly rational, having the ability to adequately evaluate social states between points *a* and *b*. We will then decide that the common good requires us to move to state *a*, which, alas, moves us further from the global optimum. We thus see how real individuals who really share a common perspective and have identical epistemic traits are apt to land at a social contract stuck at a local optimum.

3 MINIMAL DIVERSITY: ALLOWING DIFFERENT SEARCH STRATEGIES WITHIN THE SAME PERSPECTIVE (THE ADVENT OF IDEOLOGIES)

A society seeking to implement a social contract furthering the common good — that is, to best articulate the general will — can do better if it takes seriously Rousseau's insight that discovery of the general will is more effectively seen as a social rather than an individual discovery, even when we do all share the same perspective. Let us continue to suppose we all share the same perspective, and so occupy the same rugged landscape in Figure 2: all agree that on the way the core values should be mapped on to social worlds, and we concur on which social arrangements are proximate to our present one, and which are very different

---

[14] We can depict the relation of these values as itself forming a rugged landscape; our aim is to search the terrain for the optimal combination. See D'Agostino, "From the Organization to the Division of Cognitive Labor;" Weisberg and Muldoon, "Epistemic Landscapes and the Division of Cognitive Labor."

(distant). But now suppose that the individuals have different search strategies: they use their boundedly rational capacities to explore different parts of the space of possible social arrangements. Some people better appreciate different social possibilities. For example, conservatives know the current social state well (they "intimate its traditions") but tend to get stuck local optima; reformers specialize in searching for better optima near the current state; utopians specialize in exploring the possibilities of far-flung social states (but are not uusally very good at showing us how to get there). Notice that this extent of ideological diversity would exist even under a normalization contract, in which we share the same perspective (the same way of mapping our common values onto a shared viewpoint on social worlds).

Of course there is still the problem of communication of these discoveries. Suppose the utopian comes back and tells us of the high peaks she has witnessed at point u. If only we institute this realistic utopia, the values constituting the common good will reach to the sky! We may be skeptical. Moreover because real ideologies are not just epistemic proclivities (about what social worlds are worth investigating) but reflect value differences (e.g., risk preferences, relative valuing of fairness and order), it seems that once we allow ideological diversity we increase our store of normative knowledge, but we almost certainly introduce differences in value perspectives as well. The conservative might not simply be skeptical about reports of the great height at u, but just because it is so far away, from his viewpoint it really is lower than in the utopian's perspective. Although admitting different epistemic search strategies implicit in ideologies adds social knowledge, it seems that we have confirmed the fear of the advocate of the normalization thesis: we have reduced the possibility for advancing the common good by introducing diversity.

4 HOW A DIVERSITY OF PERSPECTIVES CAN MAKE DIFFICULT PROBLEMS EASIER

Our problem, it will be recalled, is that we are trying to find the global optimum in a rugged landscape. If we all had a God's eye perspective, and so could scan the entire landscape, we may know what to do, but boundedly rational individuals cannot survey all possibilities. That is the attraction of Mount Fugi landscapes: we can climb to the top through a series of steps, each one of which is better than the one before (so there is no need to drop into a crevasse in order to eventually get higher). We now come to the crucial point. As opposed to the geographical world, in the social world whether we are climbing a Mount Fugi or a rugged landscape is not a brute feature of the world but a feature of our perspective. Our perspective is what determines whether social world S_n is near or far from X. On a left-right spectrum they may be distant; but need we see the world through the left-right spectrum? How *should* we see it?

According to what Page calls the Savant Existence Theorem: "For any problem there exist many perspectives that create Mount Fugi landscapes."[15] There always is some arrangement of the options that create a mount Fugi landscape; if we find one that makes sense to us, what was distant can become near, and our problem of how to optimize a complex common good is transformed from the very difficult into the tractable. Reconceptualizing the problem via a new perspective can make the difficult problem of rugged landscapes into the much easier Mount Fugi problem. Perhaps liberalism itself was a such reconceptualization. At one point western societies faced the problem of which religions to tolerate. Think of Locke's "Letter on Toleration." While he thought it would promote the good of the commonwealth to tolerate Protestants, extending toleration to Catholics decreased the common good (an England that tolerated Catholics was very far from his own), as did extending toleration to atheists (perhaps an even further social world).[16] Locke was pushing towards a Mount Fugi liberal landscape in which each additional right of conscience and speech advanced the common good, but there still were ravines. Eventually — with much help from Locke — the early modern problem of which false creeds to tolerate was transformed into the problem of securing freedom of thought and belief. The options were arrayed in something much closer to a Mount Fugi landscape.

5 OUR PROBLEM THUS FAR

This line of reasoning has the great benefit that it analyzes the diversity of perspectives not as a problem to be accommodated but as a resource to be exploited. And this is quite true. A rigid sharing of the same perspective almost ensures that we will be caught at a local optimum that falls short of the global optimum. Yet, we ought not get carried away with the ability of diversity to produce consensus. As we saw above, once we introduce a realistic diversity in perspective — allowing in ideological differences — we inevitably allow in genuine value disagreements.

Value disagreement will occur even when our perspectives are partly normalized, in the sense that we share a list of fundamental interests but we disagree about the way this list is to be mapped on to the evaluation of social worlds.[17] Some differences will be essentially epistemic, where we disagree about how the shared conception of the common good will fare in a certain social worlds, but others will occur because differences in the way different citizens trade-off values. We all will agree that a social world that optimizes along all

---

[15] Page, *The Difference*, p. 47.

[16] See John Locke, "A Letter Concerning Toleration" *in Two Treatises of Government and A Letter Concerning Toleration*, edited by Ian Shapiro (New Haven: Yale University Press, 2003): 215-54 at pp. 245-46.

[17] In the end, I believe that Rawls adopts this sort of quasi-normalized perspective in *Political Liberalism*, paperback edn. (New York: Columbia University Press, 1996), p. 225.

dimensions is to be chosen over one that optimizes over fewer but when deciding between options *A* and *B*, where *A* beats *B* on the α dimension and *B* beats *A* on the β dimension — that is, where neither *A* nor *B* weakly dominates the other — real rational individuals will disagree on what constitutes the better solution. Evaluative perspectives not only rearrange possible social worlds, but disagree about what constitutes a peak and what constitutes a valley. The best boundedly rational and good-willed persons can expect is that on the basis of their collective reasoning about a shared understanding of the common good, we can eliminate some possible social worlds as clearly suboptimal, leaving us, at any given time, with a set of social worlds that neither weakly dominate, nor are weakly dominated by, others. To be sure, the set of social worlds that we evaluate will be dynamic: as the process of social discovery proceeds we may sometimes find ourselves converging on Mount Fugi landscapes where it is clear that the next optimizing move is also one that will take us to the best social state. But we should expect that the norm will be, as it were, an incomplete convergence on the nature of the terrain, leaving us with a number of proposals in the nondominated set of generally acceptable social worlds, with deep disagreement about which towers over the others.

So here is our problem. If we insist on rigid normalization so that all *really* share the same perspective, we are almost certain to get stuck at local optima. The general will thus would endorse a social world in which the common good is not optimally furthered. If we allow a diversity of perspectives, we can make progress in climbing our rugged landscapes, and may even find the terrain smoothing out. But at the same time, it seems that by admitting a diversity of ideological perspectives as an engine of discovery, we have precluded any hope that boundedly rational people converge on a single common general will.

If this is right, the best a Rousseauean-inspired conception of the general will can hope for is one that arises out of a process of social interaction and discovery, which helps us avoid being stuck in clearly sub-optimal local optima, but which must forgo all hope of arriving at a unique general will. It would seem that we are left with a set of social worlds, which will be evaluated differently by different normative perspectives.

The idea of such a non-normalized conception of the general will must confront two worries. First, we may be worried that the discovery process unleashed by allowing diverse perspectives will be undermined by ideological dispute. This, we might conjecture, was the impetus behind the normalization thesis in the first place: if we allow diversity, whatever benefits it may produce will be swamped by ideological dispute in which each party vetoes the favored social worlds of others, leaving us with an empty set of mutually acceptable social worlds. Call this the *problem of the null set by veto.* The other worry arises if this first problem can be solved. If a general will based on the normative diversity of perspectives is consistent with a set social worlds acceptable to all, once we abandon the idea that the process of discovery can, at least in principle,

lead us to a single agreed upon global optimum, we will be left with no determinate general will at all. If this is so we are them faced with moral indeterminacy. Call this *the problem of the indeterminate general will*.

6 THE PROBLEM OF THE NULL SET BY VETO

6.1 Three categories of social worlds

The obvious worry is that, if we grossly weaken normalization assumptions and so allow a true diversity of ends and values, there will be no common good. Alf may veto all social worlds that do not further his religion, and Betty may veto all social worlds that allow religion, and so the set of social worlds that satisfy the requirement of their joint common good is null. Now at the outset it is important to realize that our concern is not a bargain in the sense that allows bluffing or strategic behavior. As with Rousseau, our interest is with moral persons and citizens who are competently and sincerely seeking to evaluate social arrangements (how norms, laws, policies and institutions promote the good of all) but now we allow that each citizen evaluates them on the basis of whether they advance her own good. Each citizen reflecting on various social worlds must distinguish (*i*) that which is optimal on her perspective (*ii*) those which are acceptable and (*iii*) those that that are unacceptable. Certainly once we have abandoned the normalized version of the social contract one cannot reasonably equate (*i*) and (*ii*) — one cannot say that only the optimal is acceptable — for such a stance precludes common acceptance of a body of norms or laws under conditions of diversity.[18] Various economists and philosophers have offered formal answers to the contours of (*ii*) — the set of acceptable social worlds — in terms of various bargaining solutions, or the preferred ways to aggregate judgments into social welfare functions. But these devices are themselves one more point of controversy. It is of no avail to solve the problem of substantive disagreement by resorting to controversial procedures. As Nozick reminds us, "When sincere and good persons differ, we are prone to think they must accept some procedure to decide their differences, some procedure they both agree to be reliable and fair.... [But] this disagreement may extend all the way up the ladder of procedures."[19]

6.2 What it means to veto a norm in social world

To make our problem more precise, let us suppose for now that we share this much of a perspective: we are evaluating social arrangement S_i in terms of how the common good maps on to a set of social rules or norms, assuming certain

---

[18] Once we understand the problem of searching rugged landscapes, we will see that it is always unreasonable to insist that only the optimal is acceptable, for we seldom know where the global optimum is. On the way in which insisting on one's optimal option precludes the possibility of solutions promoting a common good, see my "On Two Critics of Justificatory Liberalism," *Politics, Philosophy & Economics*, vol. 9 (May 2010): 177-212, pp. 200ff.

[19] Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), p. 98.

background social facts (about our tendency to cooperate and cheat, etc.). Our question is what it means to say that a person vetos S_i. Again, to make the problem more tractable, suppose a person is evaluating a single norm in S_i. Taking some cues from Rawls's explication of Kant's first formulation of the categorical imperative, suppose one asks oneself whether one can will that this be a norm which all accept and endorse in this social world.[20] It is important that one's concern is not simply that we all conform to the norm, but that each makes it her own norm, a norm that each endorses and internalizes. To say that a person internalizes the rule is to say that, if confronted with the possibility of violating it without detection, and so with no fear of punishment or censure, she wills acting on it, and she will experience guilt should she fail to do so.[21]

This test is not arbitrary or contrived: only rules and norms that can pass it provide the basis for a cooperative and stable social order. For example, David Schwab and Elinor Ostrom argue that there are compelling reasons to conclude that mere convergence of interests is insufficient to sustain free market transactions.[22] If we suppose that individuals are solely devoted to their own private ends, the development of market exchange — which depends on trust — is difficult to explain. As Hobbes so effectively showed, individuals who are solely devoted to their private ends will be sorely tempted to renege on "covenants": if the other party performs first and so gives the second party what she wants, there seems to be no incentive for the second party to perform her part of the bargain.[23] Rather than exchanging, she will be tempted to snatch the goods and flee.[24] Given self-interested utility functions, she may often do better by snatching: she gets the good without paying for it. However, when individuals internalize norms — when they feel guilt for violations even when undetected — norm-governed action, and so social cooperation, is stabilized. In short, people don't snatch even when they can get away with it. A large body of work indicates that extended market societies are only possible when norms are moralized in this way: individuals come to conceive the basic norms of social cooperation as morally authoritative.[25] Participants largely refrain from cheating

---

[20] John Rawls, *Lectures on the History of Moral Philosophy*, edited by Barbara Herman (Cambridge: Harvard University Press, 2000), pp. 167-68. I am side-stepping the issue of whether this judgment must be holistic, considering whether one can will this norms in a system of norms, or whether (as I believe is the case), one can evaluate norms much more locally.

[21] I provide a rather more subtle version of this test in *The Order of Public Reason* (Cambridge: Cambridge University Press, forthcoming), chap. V.

[22] David Schwab and Elinor Ostrom, "The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies" in *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul Zak (Princeton: Princeton University Press, 2008): 204-227.

[23] Hobbes actually thinks a person has some reason to perform second, but this is usually too weak to outweigh her selfish passions. See *Leviathan,* chap. 14.

[24] On the game of snatch, see Schwab and Ostrom, "The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies," pp. 205ff.

[25] See, Peter J. Richerson and Robert Boyd, "The Evolution of Free Enterprise Values," in *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul Zak (Princeton: Princeton University Press, 2008); "The Vital Role of Norms and Rules in Maintaining Open Public and

and exploiting others, not primarily because they are worried about, say, their reputations or punishment, but because they believe it to be wrong to violate rules that are generally followed by others.[26]

A person may well conclude that some social norm or rule does not adequately promote her good in some social world. Given the costs of adopting such a rule, the overall set of evaluative standards implicit in his perspective is set back by adopting the norm as moral. In such a case if others conform to it, he may also still follow it as a descriptive norm, but he will accord it no authority over him. If we allow a diversity of evaluative perspectives to come into play, different people will reach this decision at different points and for different reasons. But in all cases it is a grave decision. A person who, as it were, opts out of a system of authoritative norms undergoes a change in his moral status. Having refused to internalize the norm himself, he cannot reasonably demand that others do so, and so it is no longer appropriate for him to experience the Strawsonian reactive attitudes such as resentment and indignation in response to violations, or expect others to feel guilt. Whether or not they conform to the norm no longer is his business: with such norms one follows them as long as it is one's interest to do so — one does not owe obedience to others.[27]

6.3 The problem of the null set by veto reconsidered

Given this understanding of what is involved in vetoing a norm, let us consider whether allowing a diversity of evaluative perspectives into our conception of the common good makes the account especially vulnerable to a null set by veto. It should be clear that a person would only refuse to embrace a norm if, given her understanding of the costs of internalization and the benefits to her good through generalized norm-regulated interactions, the costs of the norm exceed the benefits. In this case, looking at all she considers important and relevant, she does not have adequate reason to endorse the norm. Now in this case, the Rawslian normalization-abstraction contract will be also fail in securing effective endorsement. As Rawls explicitly points out, the argument from a shared, abstract, normalized, point of view only yields a "pro tanto," or "as far as it goes," justification. For what Rawls calls "full justification," a citizen must consult her overall perspective, and she may discover that the pro tanto justification is "overridden" once she tallies up "all values."[28] In this case she would not find the rule or norm fully justified and so she would not have an all-things-considered reason to endorse it. If a person's total set of values really

---

Private Economies;" Daniel Friedman, *Morals and Markets: An Evolutionary Account of the Modern World* (New York: Palgrave Macmillan, 2008), especially chap. 3.

[26] This qualification is important: people tend to follow rules that are generally followed — they tend to ignore normative exhortations to follow rules when this is not backed up by actual general conformity.

[27] On the way moral compliance is everyone's business, see Kurt Baier, *The Moral Point of View*, abridged edn. (New York: Random House, 1965), pp. xviii–xix.

[28] Rawls, *Political Liberalism*, p. 387.

oppose endorsing the rule it will not help to show that she would endorse it from some abstract, normalized, perspective that excludes much of what she cares for; unless her overall perspective endorses it she does not in the end have sufficient reason to adopt and internalize it. Given this, the problem of the null set by veto is pushed aside, not solved, by the abstraction-normalization strategy.

We now can see that with respect to the problem of the null set by veto, allowing diversity of normative perspectives into our conception of the common good can do no worse than the normalized social contract procedure, and may do better. To see this, consider two cases.

> I. When reasoning on the basis of a shared viewpoint, all have reason to endorse norm *R*, but when each considers her full perspective, some do not.

> II. When reasoning on the basis of a shared viewpoint, all do *not* have reason to endorse *R*, but when each considers his full perspective, they *all do*.

We have explored Case I in which, to employ Rawlsian terms, the pro tanto argument supports *R*, but it is not fully justified. I have argued that, in the end, both the normalized and diverse social contracts must accept that some citizens will not internalize the norm as morally authoritative. But the normalized contract does not allow that in Case II *R* is justified, because the normalization view insists that only that which is endorsed on the basis of normalized interests is truly justified, and in Case II normalized interests do not endorse *R*. But in Case II there is convergence on *R* from different viewpoints. For example, we each may have quite different reasons for supporting an environmental norm: some may appeal to basic human interests, others to our convictions about stewardship of the planet, others to ecological values, and yet others to the sanctity of property rights (environmental harms are, after all, almost always negative externalities). If we restrict ourselves to a common core of normalized shared reasons, in Case II there is no set of acceptable norm at all.

It is with stressing that the diverse perspectives and interests conception of the common good makes the discovery of the general will an inherently social process. There is no possibility of a philosopher sitting in her study, abstractly considering what our shared interests are, proclaiming the common good. Until we actually engage in social interaction with diverse interests, we cannot know whether there will be a convergence.

7 THE GENERAL WILL AS A SOCIAL EQUILIBRIUM

Seeing that the diverse contract view is not especially subject to the problem of the null set, we must confront the other objection — that of the indeterminate general will. Let us suppose that our deliberators have considered at least proximate social arrangements: that is they consider their current social world S_x

and at least proximate others arrangements that each sees as available alternatives. Because we have allowed diversity of perspectives, we can assume that a wide range of social worlds will be canvassed; what is a proximate world for one perspective may be relatively far off for another. This diversity, as we have seen, will promote social discovery, allowing citizens to identify poor local optima. If the process of social discovery reveals that social world S_i Pareto dominates S_j — if on all perspectives S_i is a higher peak in the value satisfaction landscape — S_j will be eliminated from consideration. Our diversity of perspectives will thus yield what I shall call an *optimal eligible set* of social worlds; $\{S_i...S_k\}$: evaluating the set of norms in each world, all citizen-deliberators agree that every member of the set falls into categories (*i*) or (*ii*) (§6.1), i.e., each holds either that S_i is optimal or acceptable, and is not Pareto-inferior to any other social world. For present purposes we can suppose that each person ranks the alternative social worlds in this set based on her perspective. To recall: if we had been able to justify some uniquely rational bargaining solution every good-willed citizen would share, all would share the same ranking the choice worthiness of the options; but given that each citizen does not even know the perspectives of others, this is rather a lot to expect. So we suppose simply that each evaluates the set on the basis simply of her own perspective. More formally, let us say each has an ordinal ranking of $\{S_i...S_k\}$.

We can conceive of each person having an overall utility function divided into two parts: (*i*) a part based solely on his ranking of $\{S_i...S_k\}$; he puts value on living in a social world he ranks highly (he gets his preferred result) but (*ii*) he also values society converging on a common social arrangement (that there is a result). The point of thinking about the common good, after all, is for us to live a shared social life as free and equal persons. The aim of the social contract is for us to freely accept a common system of governance, in which, as Kant and Rousseau indicate, each is the legislator while each also obeys herself.[29] So we all prefer to coordinate on a common social world than failing to do so, but we have differences about which world is optimal. In the two-person case this yields an impure coordination game, as in Figure 3.

FIGURE 3: COORDINATING ON A SOCIAL WORLD

---

[29] See Rousseau, *The Social Contract*, p. 176 (Book I, chap. vii); Immanuel Kant, *Foundations of the Metaphysics of Morals*, edited and translated by Lewis White Beck (Indianapolis: Bobbs-Merrill, 1959), p. 52 [Akademie, 433-434].

Suppose that S_a and S_b are alternative social arrangements in the eligible set $\{S_i...S_k\}$. The numbers in the matrix refer to ordinal utility, with high numbers indicating highly ranked options. The uncoordinated outcomes indicate no shared normative arrangement. Looked at ex ante, Betty's perspective gives her reason to accept S_b over S_a; Alf's lead him to accept S_a over S_b. Ex ante, Betty does not have reason to accept S_a over S_b, nor does Alf have accept S_b rather than S_a. They do, however, have reason to coordinate on either of the two social worlds rather than none at all (this is the importance of the second part of their utility function). Should Alf and Betty find themselves in social world S_a neither would have reason to change his or her action. Given each of their evaluative standards, they have the most reason to endorse social world S_a. Should they instead find themselves in social world S_b, each will then have most reason (given his or her conception of the good) to endorse it. Note that in neither case is any party induced by some external consideration to endorse a social world: consulting simply his or her own perspective, each has decisive reason to freely endorse whichever social world on which they have coordinated. In social world S_b Betty can demand that Alf conform and, consulting only his own evaluative standards, he will have a reason to conform; and in social world S_a Alf can demand that Betty conform to its demands and Betty will have sufficient reason to do so. And this even though, from the initial deliberative perspective, neither had reason to endorse the other's preferred social world. In this case, so long as S_a and S_b are both in the eligible set, both are Nash equilibrium solutions.

Let us make the model more adequate. First, of course we need to suppose that we have a large N game; second, we should allow that there will be great diversity in the weight different people put on the two parts of their utility function: some greatly care that the result is their preferred result, others have only a modest preference for one social world over another, but do care deeply that we all accept the normative arrangement. Now given these two steps towards realism, we are confronted with an increasing returns model.[30] In large N-person impure coordination games with multiple equilibria a bandwagon effect takes over. Suppose at time t_0 we have an absence of coordination on either S_a or S_b. Some play S_a and some play S_b; some greatly prefer S_a to S_b and vice versa, some have weaker preferences, and some are indifferent. Suppose that there is a chance event — perhaps simply the popular impression that S_a is more favored. In this case those with a weak preference for S_b over S_a, but a strong preference for coordination, will be apt to switch to S_a, as the best bet for coordination. Next those with a slightly stronger preference for S_b over S_b will observe that, despite their preference for S_b, an increasing number of the population are playing S_a; despite their preference for S_b, they too will see S_a as

---

[30] I have more formally modeled this in *The Order of Public Reason*, chap. VII. The path-breaking work on increasing returns was done by W. Brian Arthur. See his *Increasing Returns and Path Dependency in the Economy* (Ann Arbor: University of Michigan Press, 1994).

the better option for satisfying their coordination preference and their overall utility function. As more and more people switch to S_a, even those with a strong preference for S_b will conclude that there is no chance of a general coordination on S_b; to insist on playing S_b would give them their 0 payoff from absence of coordination. The cascade towards S_a will continue until everyone, even those with a very weak preference to converge and a very strong preference for S_b over S_a switch to S_a as the only possible coordination point. Figure 4 summarizes the dynamic.
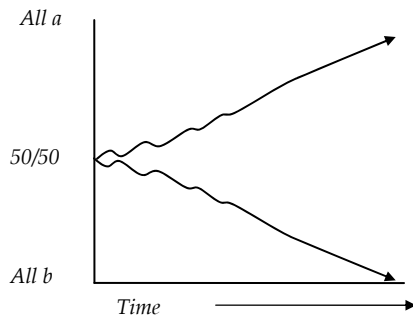

FIGURE 4: AN INCREASING RETURNS DYNAMIC[31]

As we can see, starting out with a population evenly split between advocates of social world *a* and world *b*, random events can lead the population to all S_a or all S_b equilibria. Which equilibrium emerges will be path-dependent: at time zero there is no reason why one or the other should emerge as the unanimously-selected choice. But once we have arrived at such a convergence, each citizen, consulting only her own evaluative standards, will freely accept the chosen social world. For our purposes what is crucial is that the contingent and accidental way in which large groups can come to coordinate on a common social world is no bar to there being a determinate general will that all can endorse given their different perspectives.

　　This selection process is an actual collective social choice. In this sense, choice by a society takes up where discovery leaves off. To be sure, this is not an intentional "we-choice"; it is a social choice that arises out of a multiplicity of individual choices. Neither is it an abstract choice from some impartial Archimedean perspective outside our real social world. It is a collective choice that arises out of the social nature of individual choices: each person choosing to do what his perspective recommends given what others are doing.

---

[31] Adapted from Arthur, *Increasing Returns and Path Dependency in the Economy*, p. 3.

8 HOW SOCIAL INTERACTION ALERTS US TO OPPRESSION

We thus have grounds for concluding that the free interaction of good-will persons will lead them to converge on a social arrangement. Now we have seen how diversity of perspectives can lead us to investigate various social worlds; and the process of social discovery can uncover Mount Fugi landscapes where we see the possibility of (switching now to economic language) of a series of Pareto-improvement which lead us to a global optimum. We always interested in social improvements, in finding ways of organizing society that are better for all (i.e., moving to better social worlds), and I have argued that diversity is an important engine of such discovery.

On the other hand, we live in a certain social world, and although we do not wish to get caught at poor local optima, much of our moral concern is, reasonably, rather more short-sighted: we want to know whether the social world we now inhabit really expresses a general will. We need to know whether our social world is really one that all can will, or whether it is simply a way in which the powerful control the behavior of others based on the ends of the powerful.

Because social worlds within the eligible set are Nash equilibria, it follows that a free rational moral person, expecting that others will follow and have internalized its rules, will also do so if and only if they are in the eligible set. Alf will choose to coordinate on S_i if (*i*) others have endorsed it and are acting on it and (*ii*) he has not vetoed the norms of this social world given his perspective. Unless (*ii*) is the case, the set of rules on which others have coordinated will not be a Nash equilibrium; Alf will do better by unilaterally defecting (e.g., playing snatch rather than exchange when he can get away with it). This leads us to confirm one of Hayek's insights: rule violation can itself be an engine of social discovery about the justifiability of our current social arrangement.[32] Generally high levels of norm violation, or sustained norm violation by some social groups, is strong indication that we have not equilibrated on a member of the eligible set. In a free society the level of compliance with social rules is probably the best indication of whether its social arrangement satisfies the requirements of the common good. The last fifty years have witnessed a widespread rejection of traditional norms regulating sexual relations; an early indication of the unjustifiability of these norms was widespread violation.

While high levels of norm violation are reasonable indicators that we have failed to settle on a moral equilibrium, the inference from high levels of observance to confidence that we have found such an equilibrium is far less certain. Moral rules, like all social norms, require some level of punishment to sustain them, and it is always difficult to determine whether norm observance is due simply to fear of punishment or the fact that it is an eligible norm. Excessive

---

[32] See F. A. Hayek, *The Constitution of Liberty* (London: Routledge, 1960), p. 63; *The Political Order of a Free People* (London: Routledge, 1979), p. 161

punishment is a reasonable indicator that the norm is simply sustained by fear. In these cases a powerful group may enforce highly oppressive norms through severe social sanctions. (The Draconian American drug laws come to mind.)

9 THE GENERAL WILL AS A SOCIAL FACT AND THE POLITICAL PROCESS

On the account sketched here, the general will is the outcome of a social process, partly a process of discovery and partly an outcome of social choice. A true general will occupies the space between the conception of morality as simply the discovery of the truth and morality as what all happen to will.[33] This primarily social rather than political interpretation of the general will was advanced by the British Hegelians. T.H. Green understood a society's morality as constituting a recognition of claims based on a common good.[34] Strikingly, Bernard Bosanquet understood the general will as the outcome of something very much like an invisible hand process, in which each individual, following her conception of the good, helps to produce an overall social order and institutional structure willed by all.[35] What is important on this view is the way in which morality arises out of real social processes: it not a social contract in the form of an imaginary agreement, but a dynamic process of social discovery and choice that creates an actual social and moral fact — a social world that all will. As Rousseau correctly saw, a society that reflects the general will is one that all free and equal persons endorse. Each, consulting only her own perspective, chooses to act its requirements.

First and foremost, then, the general will is a characteristic of a free society rather than a democratic government. This is not at all to say that the ideal of the general will is irrelevant to political action. Social and technological changes may outpace the ability of society to equilibrate on norms; some matters requiring regulation (e.g., copyright) are too technical to be a matter for informal social norms. And we may see that there is a way to climb to a better optima, but lack the ability to coordinate without employing political authority. And, perhaps most importantly, through the use of punishment, societies can become stuck on social arrangements that are outside the eligible set. In such cases of oppression, democratic governments seeking the common good have often freed oppressed social groups. Yet, we also must remember that in itself majoritarianism has little tendency to advance the common good; a majority can easily move us away from

---

[33] On this account the general will is coherently both substantive and voluntarist — it thus constitutes a reply to the charge that the concept of the general will is an inconsistent muddle of both. See Patrick Riley, "A Possible Explanation of Rousseau's General Will," *American Political Science Review*, vol. 64 (March 1970): 86-97.

[34] See T.H. Green, *Prolegomena to Ethics* (Oxford: Clarendon Press, 1890), Book III

[35] Bernard Bosanquet, *The Philosophical Theory of the State* in *The Philosophical theory of the State and Related Essays*, edited by Gerald F. Gaus and William Sweet (Indianapolis: St. Augustine Press, 2001), and Gerald F. Gaus, "Bernard Bosanquet's Communitarian Defense of Economic Individualism" in *The New Liberalism: Reconciling Liberty and Community*, edited by David Weinstein and Avital Simhony (Cambridge: Cambridge University Press, 2001): 136-158.

an oppressive equilibrium, but it also can lead us to one. It is useful to remember here that Rousseau himself was no simple majoritarian. "The more concert reigns in the assemblies, that is, the nearer opinion approaches unanimity, the greater is the dominance of the general will. On the other hand, long debates, dissensions and tumult proclaim the ascendancy of particular interests and the decline of the State."[36] A political order of diverse perspectives guided by the general will must balance two considerations: great differences of rankings within the optimal eligible set conjoined with significant consensus that options within the set are superior to those outside. Recall that in his discussion of the assemblies of the people Rousseau has the people answering two general questions: should the form of government be preserved, and should the current administration be retained?[37] In a similar vein we might imagine a legislative process that first asks "should the law be changed on this matter?" (where we would expect supermajorities) and only then ask "what shall the law be? (where we might expect far less agreement). This points to a fascinating convergence between Rousseau's account of a polity based on the general will and contemporary public choice theory, which also stresses the importance for supermajorities at various stages of the legislative process. However, having advanced enough controversial theses about the general will, I will conclude without exploring this one.

Philosophy
The University of Arizona

---

[36] Rousseau, *The Social Contract*, p. 249 (Book IV, chap ii).
[37] Ibid., p. 245 (Book III, chap xviii).