

## Can Justice Be Based on Consent?\*

FRANK LOVETT

*Political Science, Columbia University*

THIS article will discuss a general class of theories that attempt to base a normative account of justice on consent-giving procedures. Contemporary political philosophers offer many different accounts of justice, but at present consent-based theories are probably the most popular. Notwithstanding their (often understandable) appeal, I will argue they cannot possibly succeed. All consent-based theories trade on a crucial ambiguity as to whether they are supposed to be understood as voluntarist accounts of justice on the one hand, or as rationalist accounts on the other. In the former case, consent-based theories fail because they are at best indeterminate and at worst incoherent; in the latter case, they fail because they are superfluous. I will refer to this fatal ambiguity as “Cassirer’s Dilemma.” Interestingly, Cassirer’s Dilemma is not at all new (although its relevance for contemporary consent-based theories does not seem to have been noticed previously): on the contrary, it may be among the oldest problems plaguing western moral and political philosophy. Faced with this insurmountable difficulty, I believe we ought to abandon proceduralism and focus instead on developing substantive accounts of justice.

These admittedly bold claims will be explained and defended in due course. Before going much further, however, it will be useful to clarify a few concepts as they will be employed in this article. Although the terms “just” and “unjust” can be applied to a wide range of political institutions, legal regimes, public policies, individual conduct, and so on, contemporary political philosophers tend to abstract from such particular applications so as to focus in a general manner on what are often called the *principles of justice*. One example might be John Rawls’s “difference principle,” according to which, roughly, social and economic inequalities should be arranged to the greatest benefit of the least advantaged.<sup>1</sup> J. S. Mill’s famous “harm principle” might be another.<sup>2</sup> By a *theory of justice*, therefore, political philosophers generally mean a normative account of what the fundamental principles of justice actually are, and not some more specific claim

\*I would like to thank Robert Amdur, Julian Franklin, Paul Sigmund, Paul MacDonald, Nick Toloudis, Christian Rostboll and the anonymous *Journal of Political Philosophy* reviewers for their invaluable comments on drafts of this article.

<sup>1</sup>See especially Rawls 1971, pp. 75–83.

<sup>2</sup>Barry 1995, pp. 86–8, discusses the harm principle as a potential theorem of what he calls “justice as impartiality.”

that a particular institution, law, policy, course of conduct, etc., happens to conform (or not) to such principles; thus, to continue with the first example, not a claim of the sort that market capitalism, or socialism, or something else is the most just economic system because it best realizes Rawls's difference principle.

We must be careful here to distinguish a theory of justice from a normative account of political legitimacy.<sup>3</sup> A theory of legitimacy is generally understood to be an account of the conditions under which it is morally acceptable for people to wield coercive powers over one another (together, possibly, with an account of the conditions under which one has an obligation to obey those wielding such coercive powers). One possible view might be that coercive powers are legitimate to the extent that they are wielded on behalf of just laws or policies; in this case, the distinction between a theory of legitimacy and a theory of justice would not be terribly important. But often theories of legitimacy are quite different from this. For example, it might be argued that so long as a particular law or policy were enacted and applied according to the appropriate procedures, then enforcing it with coercive force would be legitimate, notwithstanding the possibility that the law or policy in question happens itself to be unjust.

Like theories of justice, theories of legitimacy can be consent-based—if, for example, they argue the appropriate procedures of legitimation include consent-giving mechanisms. The distinction between consent-based theories of justice on the one hand, and consent-based theories of legitimacy on the other is important, however, because the arguments in this article are directed *only against the former*. Although consent-based theories of legitimacy face important limitations of their own, these will not be discussed except in passing at the conclusion of this article. I emphasize this point because arguments against consent-based theories of justice are not necessarily also arguments against consent-based theories of legitimacy.<sup>4</sup>

Having clarified these points, I will define a *consent-based* theory of justice as any procedural account of justice according to which we may determine what the principles of justice are by asking what people would (or would not) agree to, under appropriate consent-giving conditions. Defined in this manner, the class of such theories is quite broad. It includes contractualist theories of justice like that of Rawls and his followers. It also includes what might be called impartiality theories, like those advocated by T. M. Scanlon, Brian Barry and Thomas Nagel; and what might be called deliberative–democratic theories, like those of Jürgen

<sup>3</sup>Normative accounts of legitimacy are often confused with empirical accounts of the conditions under which people happen to believe their ruling authorities are legitimate. The question of whether a ruling regime is actually legitimate is conceptually separate from the question of whether those living under the regime believe it is (or not). It is the former question that is at issue here.

<sup>4</sup>Another distinction I regard as less important is that between consent-based theories of justice on the one hand, and consent-based moral philosophies on the other. A theory of justice is one part of a broader moral philosophy, and—so far as I am aware—no one advocating a consent-based moral philosophy excludes questions of justice from their theory. (Conversely, some political philosophers caution against extending their theories of justice into complete moral philosophies.)

Habermas, Joshua Cohen and others (when these are not narrowly interpreted as theories of legitimacy). While contractualist theories may have been the most influential in contemporary political philosophy, deliberative–democratic theories are probably the most popular at present. The general class of consent-based theories of justice will be discussed in more detail in section II below.<sup>5</sup>

This article has four sections. The first explains Cassirer’s Dilemma, and briefly traces some of its interesting intellectual genealogy. The basic effect of the dilemma is to force theories having a particular structure to choose one of two possible self-interpretations, which will be referred to as *voluntarism* and *rationalism* respectively. Consent-based theories of justice, it will be argued in section II, have this particular structure in question. Section III goes on to show the problems faced by consent-based theories when they are given a voluntarist construction, and section IV when they are given a rationalist construction. Consent-based theories of justice seem to escape these problems only so long as they remain ambiguously uncommitted: once forced to commit either to voluntarism or to rationalism, a theory attempting to base an account of justice on consent-giving procedures necessarily fails.

## I.

Although this article is ultimately concerned with contemporary political philosophy, in this section we retrace our steps somewhat, and focus on Jean-Jacques Rousseau. This is not necessarily because Rousseau is the original inspiration for consent-based theories of justice (though some might regard him as such), but rather because his doctrine of the “general will” highlights in a particularly acute manner the deeply-rooted problems facing all such theories.

Rousseau intends his doctrine of the general will to solve what he calls the “fundamental problem” of political philosophy: namely, to find a form of political association such that “each one, while uniting with all” for defense and protection, “nevertheless obeys only himself and remains as free as before.” The answer to this fundamental problem is given in the formula of the social contract, as follows: “each of us places his person and all his power in common under the supreme direction of the general will.”<sup>6</sup> What then is the general will?

On its surface, Rousseau’s answer may seem straightforward enough. While he does not offer an explicit definition of the general will—except to say that it is not merely the sum of private interests<sup>7</sup>—he does offer a reasonably clear method or procedure for determining where it lies in any given instance. Roughly

<sup>5</sup>Two groups of theories that will not be discussed are entitlement theories of a libertarian sort (see Nozick 1974), and neo-Hobbesian mutual advantage theories (see Gauthier 1986, or Binmore 1998). Sometimes these are also viewed as instances of consent-based theories of justice, but I would argue the former are best understood as theories of legitimacy, and the latter as theories of prudence; accordingly, they fall outside the scope of our discussion.

<sup>6</sup>Rousseau 1762, I.6.4, 9: p. 148.

<sup>7</sup>Ibid., II.3.2: p. 155; cf., II.1.3: pp. 153–4.

speaking, when the members of a political community assemble and vote on some matter of common concern, the outcome of their vote will correspond to the general will, provided that the following four conditions obtain: first, the vote of every member of the community is counted;<sup>8</sup> second, the voters are “sufficiently informed;”<sup>9</sup> third, each voter deliberates with himself alone, and does not communicate with the others;<sup>10</sup> and fourth, questions of common concern presented to the assembled community members for consideration are posed in the form of a general law to be accepted or rejected.<sup>11</sup> Provided these conditions obtain, laws accepted by the assembled members of the community constitute expressions of the general will.

Let us leave aside the many questions related to the practical feasibility or normative desirability of Rousseau’s procedural requirements, and simply focus on the structure of the theory itself. Now whereas “the general will,” Rousseau states, “is always right and always tends towards the public utility,” the outcome of any actual vote in some cases may be neither.<sup>12</sup> This is easily accounted for, however, by the fact that the enumerated procedural requirements are hypothetical ideals, and in practice will only approximately obtain. (It is the task of Rousseau’s “legislator” to engineer political institutions as close to these ideals as possible.) Thus we are able to save both the normative fact that the general will is infallible, and the descriptive fact that what people actually vote for is sometimes misguided. One might even add, perhaps, that perfection of the procedures and the outcome’s approximation of the general will are related, such that the closer our procedures are to the ideal requirements, the more assurance we have that the outcome of voting corresponds to the true general will.

So far, interpreters of Rousseau more or less agree. The difficulty begins once we try to answer the following question: Supposing for the sake of argument Rousseau’s ideal conditions for assembling and voting did obtain, what is it exactly that accounts for the identity of the actual outcome with the general will? One obvious suggestion is that Rousseau intends simply to define the general will as the collective will of the members of the political community, expressed in their voting under appropriate conditions. There is, in other words, no external standard for the common good apart from what the people genuinely desire—though, of course, one must add that what the people genuinely desire is often distorted by procedurally imperfect attempts to ascertain their will. This reading is natural for several reasons: First, the use of the term general *will* seems to encourage it.<sup>13</sup> Second, it seems an obvious extension of tendencies in early modern European

<sup>8</sup>Ibid., II.2.1: p. 154 (Rousseau’s footnote to this paragraph); cf., II.3.3: p. 156.

<sup>9</sup>Ibid., II.3.3: p. 156.

<sup>10</sup>Ibid., II.3.3–4: p. 156.

<sup>11</sup>Ibid., II.4.5–7: p. 161.

<sup>12</sup>“La volonté générale est toujours droite et tend toujours à l’utilité publique,” *ibid.*, II.3.1: p. 155. See also, *ibid.*, II.6.9, p. 162.

<sup>13</sup>In the passage quoted in n. 13 above, notice the difference it would make translating *droit* as “law” rather than “right”—a translation with some merit: see Waldman 1960, p. 223.

political thought, which often equated political authority with the will of the sovereign—Rousseau’s innovation thus consisting in the claim that only the whole body of the citizens can be a genuine sovereign (and indeed, he explicitly defines “sovereignty” as the general will when active).<sup>14</sup> For these reasons, this first interpretation—sometimes referred to as a “voluntarist” reading of the general will—was for a long time fairly standard, and continues to be common.

An alternate interpretation later emerged, however. While it is difficult to pinpoint its origin, one might reasonably cite Ernst Cassirer’s influential monograph, *The Question of Jean-Jacques Rousseau* (1932). In contrast to the voluntarist reading, Cassirer advanced what is sometimes termed a “rationalist” reading of the general will. The starting point for this interpretation is the same: first, the claim that assembling the members of the community and conducting votes yields the general will, so long as the necessary conditions obtain; and second, the claim that whereas the general will itself is never wrong, the outcome of actual votes may be in error due to imperfections in the procedure followed. The explanation for why these two propositions hold is quite different on the rationalist interpretation, however. The central claim is that there exists in some sense a general will or common good for the community prior to anyone’s actually knowing what it is. The point of assembling and voting on questions of common concern, then, is not that this process *generates* the general will or common good per se, but rather that it constitutes a sort of cognitive exercise in discovering what the general will or common good already is. To be sure, the result of any actual voting process may not yield the genuine general will or common good, but this is again due to imperfections in the manner in which this exercise is carried out.

Which interpretation best captures Rousseau’s own view? Unfortunately, the relevant textual evidence can usually be read either way. Consider, for example, the following well-known passage:

The constant will of all the members of the state is the general will; through it they are citizens and free. When a law is proposed in the people’s assembly, what is asked of them is not precisely whether they approve or reject, but whether it conforms to the general will that is theirs. Each man, in giving his vote, states his opinion on this matter, and the declaration of the general will is drawn from the counting of votes. When, therefore, the opinion contrary to mine prevails, this proves merely that I was in error, and that what I took to be the general will was not so. If my private opinion has prevailed, I would have done something other than what I had wanted.<sup>15</sup>

Now the first two sentences strongly favor a voluntarist interpretation, for indeed it seems that the mere fact that the members of the community will something

<sup>14</sup>Rousseau 1762, I.6.10: p. 149. The early modern European political thinkers I have in mind are Jean Bodin and Thomas Hobbes in particular, but the essentials of a conception of sovereignty were commonplace by Rousseau’s time.

<sup>15</sup>Rousseau 1762, IV.2.8: p. 206.

makes it the general will. But what does Rousseau mean by this curious argument that if “the contrary opinion to mine prevails, this proves merely that I was in error”? On the one hand, it might mean that by participating in the process of assembling and voting, I come to discover what the general will really is, and of course *ex post* recognize what I really want was other than what I initially voted for. One might imagine two archeologists both searching for the same ancient ruins, but at two different sites. When the second archeologist discovers the ruins that both seek, the first realizes she was mistaken (though of course no one could have known this without actually digging). This seems to give us a rationalist reading to the passage. But then again Rousseau might mean something rather different. Suppose what I really want is for the community to be joined in the collective desire to accomplish a united end or goal. In this case, each member votes for what she believes this united end or goal should be, but once the results are tallied those in the minority come around to the view that a true consensus should be formed around the opinion that won out. Now we have a voluntarist reading of the passage again.<sup>16</sup>

Scholarly opinion now seems to lean in the direction of a rationalist reading of the doctrine of general will, though in my view the truth is probably that Rousseau himself was never really reconciled to either position.<sup>17</sup> What is more interesting, however, is that this problem was by no means new, even with Rousseau—a fact emphasized by the choice of the terms “voluntarism” and “rationalism.” These terms (sometimes “intellectualism” being substituted for “rationalism”) are commonly employed with reference to a stubborn—and ultimately intractable—debate during the golden age of natural law theory, roughly from the time of Thomas Aquinas to that of Locke and Leibniz.

Aquinas’s great accomplishment was to systematize the originally Stoic doctrine of natural law, now embedded within Christian moral and political thought. His idea was roughly to equate the right reason of natural law with the will of God, and argue that this law is morally binding on grounds of the obligation human beings owe their creator. Although appealing in broad outline, the following dilemma was soon noticed by Aquinas’s successors: Are the precepts of the natural law what they are simply because God willed them, or did God will the particular set of precepts he did because they were the right ones to will?<sup>18</sup> The latter, rationalist (or intellectualist) view seems to entail a limitation on the omnipotence of God, for it suggests He is not perfectly free

<sup>16</sup>Additional passages perhaps encouraging a rationalist interpretation can be found at *ibid.*, II.1.1: p. 153, II.4.7: p. 158, and IV.1.4; others favoring a voluntarist interpretation at II.1.2–4: p. 153–4, II.4.8: p. 158, and II.6.6: p. 161. None of these are decisive, however.

<sup>17</sup>Voluntarist readings of Rousseau can be found in Waldman 1960; and Sreenivassan 2000. For examples of rationalist readings in addition to Cassirer, see Allen 1961; or Dagger 1981. The view that Rousseau was unable to resolve the voluntarist–rationalist tension can be found in Riley 1970.

<sup>18</sup>Although this may have been the first discussion of this question in western European philosophy, one might note that nearly the exact same debate can be found in Plato’s *Euthyphro*, making this one of the truly ancient questions of moral and political theory.

to will any sort of natural law He thinks fit. This, of course, will not do. But the former, voluntarist view seems to entail an equally problematic and counterintuitive conclusion: namely, that God might equally well have willed murder, theft, adultery, and so on to be morally acceptable. How can this be? Despite many attempts to rework the basis of natural law theory through the years, this essential tension was never resolved.<sup>19</sup>

Interest in natural law theory began to wane with the coming of the Enlightenment. It is remarkable, then, that Rousseau's doctrine of the general will manages to resuscitate exactly the same difficulty, even after its originally theological basis had been cast aside. But this dilemma was not apparent for some time—not, perhaps, until Cassirer put the possibility of a rationalist reading of Rousseau seriously on the table. For this reason, I refer to the problem as *Cassirer's Dilemma*, as in “the dilemma exposed, if unintentionally, by Cassirer.” Stated in general terms, the dilemma is this: That any normative moral or political theory having a particular structure (about which more will be said shortly) will be open to two possible interpretations or constructions, one voluntarist and the other rationalist. This would not be a problem if one of these were more desirable or tenable than the other, but unfortunately it turns out both are seriously problematic. Rousseau's doctrine of the general will is an example of a theory having this structure, as is the doctrine of natural law in its classical formulation. And so too, as I attempt to show next, are contemporary consent-based theories of justice.

## II.

Let us begin by outlining the basic structure of a consent-based theory. To repeat what was said in the introduction, by a *theory of justice* political philosophers generally mean a normative account of the fundamental principles of justice. In other words, if we all consider the set of possible principles of justice,<sup>20</sup> a theory of justice is an argument to the effect that some subset of these are normatively valid. According to a consent-based theory of justice, any given principle *p* is a genuine principle of justice if and only if people would agree to it under suitable conditions.<sup>21</sup> Obviously, we must immediately ask: Which people, and under what conditions? It is principally in their various responses to these questions that consent-based theories of justice differ from one another. At the most general level, one might respond in two ways: on the one hand, we might consider actual

<sup>19</sup>Schneewind 1998, esp. chs. 2–8, 12, discusses the rise and fall of natural law theory in connection with the voluntarism–intellectualism problem.

<sup>20</sup>That is, the set of all propositions properly formulated such that they could be principles of justice. Descriptive statements, as for example “democracy is the most stable form of government,” cannot be principles of justice, and thus will not be members of this set. Other propositions might be excluded for various strictly formal reasons.

<sup>21</sup>I will in this essay ignore the possible difference between accepting *p* and not-rejecting *p*. Some argue this is an important distinction (see Scanlon 1982, pp. 111–12), but in my view it would vanish on a more precise specification of the problem, as Barry (1995, pp. 69–70) correctly points out.

people, under ordinary real-life conditions, whereas on the other hand, we might place restrictions on the sorts of conditions under which consent matters. In the former case, we are interested in *actual* consent, and in the latter case we are interested in *hypothetical* consent.<sup>22</sup> Let us consider each in turn.

Suppose we look to actual consent as the basis for our theory of justice. In this case, a proposition *p* is a principle of justice if and only if actual people in fact agree to it under ordinary, real-life circumstances. Now surely one must agree we can hardly expect unanimous consensus among real people under real conditions on any important normative principle of moral or political theory, to be very common. (Consensus might be more common if people could deliberate with one another as long as they liked, but of course this is impossible: under real political conditions, decisions have to be made under the constraints of time.) Therefore, let us consider what is probably the normal case first—namely, situations short of consensus.

If people do not agree on what the principles of justice are, how would a theory based on actual consent respond? First, it might say there is nothing more to be done: justice is simply what people believe it is, and if that happens to be different for different people, so be it. But this would leave a patchwork of principles of justice in any political community—an entirely unsatisfactory result. Suppose, for example, we ask whether it is a principle of justice to respect one another's basic security in person and possessions. Now anyone who believed they were strong enough to protect themselves, and moreover thought they might gain some advantage over others by not restricting themselves under such a principle, could simply reject the principle, and on that basis not be regarded as accountable to it. Clearly, this route is entirely unworkable for any political theory other than anarchism.<sup>23</sup>

Our next option is to introduce some decision rule so as to resolve the discordant opinions into a uniform set of principles of justice for everyone. Majority rule would be an obvious candidate, but of course there are other possibilities. Let us leave aside the serious difficulty (considered below) that decision rules—majoritarianism included—will not always yield coherent results.<sup>24</sup> A more pressing question is, Which decision rule should we adopt, and

<sup>22</sup>The latter sort of consent is hypothetical at least so long as the conditions in question do not perfectly obtain. Some argue for a middle ground: once the necessary conditions obtain, according to this view, consent must *actually* be given by real people and not merely hypothesized (for example, see Habermas 1983, pp. 68–76). Let us term this “counterfactual” consent, to distinguish it from “hypothetical.” To clarify the alleged issue: perhaps no one actually consents to *p* under present conditions, but on the counterfactual view *if they actually did* under conditions *C*, then *p would be* a principle of justice, whereas on the hypothetical view *if they would* under conditions *C*, then *p is* a principle of justice. Since no one placing conditions on consent claims those conditions ever actually obtain, this seems to me a distinction without much difference.

<sup>23</sup>Traces of this difficulty are noticeable in Nozick's tortuous attempt to justify the state's authority over anarchist holdouts (1974, pp. 51–118).

<sup>24</sup>I also leave aside the difficulty that all decisions rules are vulnerable to manipulation by strategic voting. This is a problem only for actual consent theories, and so I will not consider it in the later discussion.



on what grounds? One answer might be, Whichever rule people agree to. But this obviously leads to a regress: if people do not agree on substantive principles, they can hardly be expected to agree on a decision rule for selecting substantive principles, in which case they will need another decision rule to decide the first, and so on. Of course we assume here that people are always clever enough to figure out which decision rules are likely to yield which results, and this is not always the case. (Though often it is: consider the many arguments presented before the U.S. Supreme Court that, with varying degrees of success, mask substantive claims in procedural garb.) As a practical matter, then, the decision-making regress may be unlikely. But now the argument in favor of decision rule *R* seems to be that no one knows *ex ante* what results *R* will yield—a strange argument that equally well supports, say, the decision rule of tossing a coin. Clearly we need some additional argument to the effect that one particular decision rule is best on independent normative grounds. Unfortunately, this creates new problems. For example, suppose we defend majority rule because its results are most likely to maximize happiness. Now we are selecting decision rule *R* (majority rule) merely because it instrumentally secures outcomes *X* desired on utilitarian grounds; utility maximization has replaced consent as the standard for determining what the principles of justice are. In other words, we have abandoned exactly what we are supposed to be advocating—a consent-based account of justice.<sup>25</sup>

Faced with these difficulties, perhaps we should return to the idea of holding out for consensus, for this is the only situation in which a decision rule is unnecessary. But as we remarked above, a genuine consensus of real people under real conditions is a rare bird. Indeed, it is so rare on any important question of moral or political theory that we have strong reasons to be suspicious whenever it appears. Is it not more likely that opposing viewpoints are somehow being suppressed or ignored? Or perhaps that people's preferences are being somehow manipulated? As Ian Shapiro pointedly states, "one person's consensus is often another's hegemony."<sup>26</sup>

Thus viable consent-based theories of justice are inexorably driven to rely on *hypothetical* consent; this handily eliminates the problems discussed above by attributing them to imperfections in securing the appropriate consent-giving conditions.<sup>27</sup> Not surprisingly then, all serious contemporary consent-based theories employ hypothetical rather than actual consent.<sup>28</sup> On all such theories,

<sup>25</sup>For this argument, I am indebted to Manin 1987, p. 342.

<sup>26</sup>Shapiro 1999, p. 14. See also Scott 1990, and Moon 1993, on the suspiciousness of consensus.

<sup>27</sup>This essay will ignore Dworkin's (1973, pp. 17–21) famous objection that hypothetical consent cannot give rise to real obligations. I more or less agree with Stark 2000, that this problem is not as serious as is sometimes thought.

<sup>28</sup>One possible exception might be Walzer 1983, to the extent that his theory of justice is interpreted as consent-based. I will not discuss Walzer in detail because it seems to me that his response to reasonable pluralism is wholly inadequate: he argues (p. 313) that "a given society is just if its substantive life is lived . . . in a way faithful to the shared understandings of the members."

a proposition *p* is a principle of justice if and only if people *would* agree to it under conditions *C* as, let us suppose, given by a list. The actual list of conditions varies widely from theory to theory, as will be discussed; what is immediately relevant, however, is that we are now in a position to see that all theories of justice based on hypothetical consent have a structure quite similar to that of Rousseau's doctrine of the general will, and as such they all face Cassirer's Dilemma.

In the introduction, I identified three main groups of contemporary consent-based theories. The first are *contractarian* theories, as exemplified by Rawls's account of "justice as fairness."<sup>29</sup> For Rawls, the appropriate conditions for securing consent are given in the specification of the "original position," and these include, for example, formal requirements on the principles admissible for consideration (they should be general, universal, etc.), cognitive and informational requirements (persons should be rational, subject to a veil of ignorance, etc.), a few assumptions about the nature of reasonable preferences, and so on. The argument is then roughly that the principles of justice are those given by a hypothetical social contract all persons in the original position would agree to and, he goes on to argue, these are his well-known "two principles of justice." Clearly this argument can be given either a voluntarist or a rationalist interpretation. According to the former, we know the two principles are the correct account of justice because they are the principles persons in a suitably constructed original position would select. Rawls inclines towards this interpretation whenever he emphasizes that "justice as fairness is able to use the idea of pure procedural justice from the beginning," and specifically that "the aim is to characterize this situation [the original position] so that the principles that would be chosen, *whatever they turn out to be*, are acceptable from a moral point of view."<sup>30</sup> By contrast, according to a rationalist interpretation we know the original position is suitably designed because in it people would correctly select the two principles of justice (known to be valid on independent grounds). Rawls inclines towards this interpretation when discussing the merits of his characterization of the original position. "For each traditional conception of justice," he correctly notes, "there exists an interpretation of the initial situation in which its principles are the preferred solution." The correct interpretation is the one that "leads to a conception [of justice] that characterizes our considered judgments in reflective equilibrium" as to what justice really is.<sup>31</sup>

When people disagree, he adds only that justice requires "the society be faithful to the disagreements," without saying much regarding what this is supposed to entail.

<sup>29</sup>See esp. Rawls 1958 and 1971. In later works, Rawls additionally developed what is arguably a consent-based theory of legitimacy, but as indicated above, this falls outside the scope of discussion here. For an extension of contractarianism to moral philosophy in general, see Milo 1995.

<sup>30</sup>Rawls 1971, p. 120, emphasis added; cf. p. 136. Rawls is even more emphatic on the pure proceduralism of justice as fairness in 1993, pp. 72–3.

<sup>31</sup>Rawls 1971, p. 121. One might argue Rawls's "reflective equilibrium" is precisely an attempt to overcome Cassirer's Dilemma by having it both ways. If so, I believe the attempt fails, for the reasons given in n. 52 below.

The second family of consent-based theories are *impartiality* theories, such as those advocated by Scanlon, Barry and Nagel.<sup>32</sup> The structure of these theories is much simpler than that offered by Rawls. Roughly, the argument is that the principles of justice are those which “no one could reasonably reject as the basis for informed, unforced general agreement.”<sup>33</sup> The requirements that the persons be “informed” and that their agreement be “unforced” and “general” are straightforward enough; considerably more work is performed, no doubt, by the “reasonableness” provision. But the important point, again, is that we can still ask whether the fact that people could not reasonably reject something is what makes it a principle of justice, or rather that asking what people could not reasonably reject is simply a device for determining what the principles of justice (already) are. Nagel clearly suggests the former view when he says the agreement in question “is not the kind of ideal unanimity that simply follows from there being a single right answer which everyone ought to accept because it is independently right,” but rather, “a unanimity which could be achieved among persons in many respects as they are, provided they were also reasonable.”<sup>34</sup> Barry inclines towards the opposite view, noting that if consent were “little more than a device for talking about what is fair, on a certain fundamentally egalitarian conception of fairness,” then he “would not regard that as a devastating criticism” of the theory.<sup>35</sup>

Finally, a third family of consent-based theories, such as those of Habermas, Cohen and many others, might be called *deliberative–democratic*. Sometimes these theories are offered as accounts either of justice or moral philosophy generally, sometimes they are offered as accounts of political legitimacy, and sometimes they are simply ambiguous. To reiterate what was said in the introduction, I am for the moment only interested in these theories so long as they are interpreted as accounts of either justice or moral philosophy (assuming the latter will include the former).<sup>36</sup> At their core, deliberative–democratic theories do not differ substantially from other consent-based theories. But whereas all consent-based theories are in some sense procedural, this last group may be regarded as procedural in a particularly strong sense. This is because they tend to place a strong emphasis on the transformative effects of engaging in the deliberative or democratic will-formation process itself. The idea is roughly to point out the fact that while people may (or may not) enter the deliberative

<sup>32</sup>See Scanlon 1982 and 1998; Barry 1989, esp. part 3, and 1995; and Nagel 1991. Note that Scanlon offers a general moral philosophy (including, ultimately, a theory of justice), whereas Barry and Nagel are interested more narrowly in a theory of justice alone.

<sup>33</sup>Scanlon 1982, p. 110; 1998, p. 153 and passim.

<sup>34</sup>Nagel 1991, p. 33.

<sup>35</sup>Barry 1995, p. 113.

<sup>36</sup>There are too many deliberative–democratic theorists to cite comprehensively, but see especially (and here I leave it to the reader to decide whether these are intended as theories of justice or theories of legitimacy): Habermas 1983, 1992 and 1993; Cohen 1989 and 1996; Benhabib 1996; Estlund 1997; Young 2000. Deliberative–democratic theorists working from a different tradition include: Ackerman 1980; Dahl 1989; Gutmann and Thompson 1996.

process with preferences for this or that possible outcome, their preferences at the end of the process will be transformed—one hopes—into something better than they were before.

Deliberative–democratic theories do not on this account escape Cassirer’s Dilemma. As representative, consider Cohen’s formulation. The deliberative–democratic idea, he writes, “is rooted in the intuitive ideal of a democratic association in which the justification of the terms and conditions of association,” or—as Rawls would say—the account of justice underlying the basic structure of society, “proceeds through public argument and reasoning among equal citizens.”<sup>37</sup> Certainly we may grant people’s opinions will change as a result of this process of public argument, but what ensures that these transformed preferences constitute expressions of the principles of justice? It might either be that justice is defined as whatever the outcome happens to be (provided the necessary debates take place), or else that (having been transformed) those preferences are now more likely to express what justice was all along. Cohen provides no clarification on this point, but instead simply outlines what an “ideal deliberative procedure” would look like.

The fact that all theories attempting to base an account of justice on consent-giving procedures can be interpreted as either voluntarist or rationalist (and this is all I have shown so far) does not by itself constitute a problem. The problem is that once clearly understood as either one or the other, consent-based theories necessarily fail. Things appear otherwise only so long as consent-based theories remain ambiguously uncommitted to either view. In what follows, I will try to vindicate these claims.

### III.

On a voluntarist construction, consent-based theories argue that the fact that people would agree under conditions *C* that *P* are the principles of justice is what makes them so. Voluntarism has a certain important attraction, given today’s philosophical climate. Its rival, rationalism, may seem to demand the belief that certain normative principles of moral or political theory exist “out there” in the world, independently of our beliefs about them, somewhat like natural facts. Many people are skeptical that this is the case.<sup>38</sup> When given a voluntarist interpretation, consent-based theories handily dispense with such notions, for they tell us that the subject of moral philosophy is simply what people would or would not agree to. As this is probably less demanding metaphysically than the alternative, it makes sense to consider voluntarism first, and turn to rationalism only once it is clear the former cannot work.

<sup>37</sup>Cohen 1989, p. 72.

<sup>38</sup>In my view, it does not demand this, but a discussion is well beyond the scope of this article.

For reasons discussed in the previous section of the article, voluntarist attempts to base an account of justice on consent-giving procedures should be understood as theories of hypothetical, rather than actual, consent. This means all our theories impose conditions on when a person's consent is to be regarded as authentic. As we have seen, it is mainly in their specification of these conditions that consent-based theories differ from one another. Broadly speaking, the various sorts of conditions that have been proposed may be divided into two types. On the one hand, there are *cognitive* requirements, as for example that people think rationally, report their views sincerely, do not make mistakes of logic or inference, and so on. All hypothetical consent-based theories of justice impose cognitive requirements on the consent-giving process. On the other hand, some consent-based theories impose *non-cognitive* requirements in addition. These might include, for example, the requirement that certain sorts of substantive information, preferences, or beliefs not be allowed. Suppose some people happen to hold racist beliefs or are particularly selfish: while not logically false in the narrower sense, we might still want to discount or even ignore the consent of a person who happens to hold these views (or else, try to imagine what he *would* agree to *if* he did not). Other sorts of non-cognitive requirements might be that the process of consent-giving itself follow certain sorts of procedures. Rousseau's requirement that people deliberate only with themselves, and not with each other, in coming to a decision would be an example of this. Consent-based theories relying on type-I cognitive restrictions alone will be termed *unstructured* theories, whereas those relying on type-II non-cognitive restrictions in addition will be termed *structured* theories. Let us consider these in turn.

First, suppose we ask people what the principles of justice are, eliminating only outright cognitive errors as an unstructured theory would do. There are two possibilities: either there would be unanimous consensus that *P* are the principles of justice, or there would not. One reason there might be a unanimous consensus is that there exists one and only one (not cognitively erroneous) answer to the question, What are the principles of justice? But of course if we believed this we would not be voluntarists at all, but rationalists, and rationalism will be discussed in the next section. Apart from this possibility, we have no better reasons to expect a unanimous consensus from unstructured hypothetical consent than we do from actual consent, and indeed—as previously noted—some good reasons to be suspicious should a unanimous consensus seem to arise. We must focus then on the second possibility, namely, that people will disagree.

In this case, however, we face some well-known problems of social choice theory. Although these problems have been discussed extensively, not all moral and political philosophers have sufficiently appreciated their significance. Suppose we ask three people, What are the principles of justice? After allowing them as much time as is reasonable to debate the issue with one another, they report the following:

	Person 1	Person 2	Person 3
1st choice	$x$	$z$	$y$
2nd choice	$y$	$x$	$z$
3rd choice	$z$	$y$	$x$

This, of course, is an example of Condorcet's voting paradox. Now some might argue that this type of result would be impossible if sufficient deliberation were allowed; others might argue that using certain decision procedures our participants could avoid the voting paradox. Both concerns will be addressed shortly. For the moment, however, simply observing the result without prejudging its possibility, how can we say what the true principles of justice are? Person 1, let us suppose, believes utilitarianism ( $x$ ) is the most likely candidate for the true principles of justice, person 2 believes Rawls's two principles of justice ( $z$ ) are, and person 3 believes some form of liberal perfectionism ( $y$ ) is. None of these views is cognitively erroneous (in the narrow sense), and therefore not excluded by an unstructured theory; this being the case, there is no *ex ante* reason to believe some people might not hold the beliefs described.<sup>39</sup>

Now consider two typical objections to this example. First, it is sometimes complained that social choice theory incorrectly assumes people must have self-regarding or self-interested preferences and beliefs. This objection is obviously false: nothing in the above example relies on such an assumption. Person 1 might endorse utilitarianism because she expects she will do well under such principles, but she just as well might support utilitarianism on the beneficent grounds that it is the most sensitive to the wellbeing of everyone. In any case, restricting people to self-interested preferences—or, for that matter, excluding self-interested preferences—would make our theory a structured, not unstructured, theory; structured theories will be considered below. A second, and more sensible objection is that social choice theory regards preferences and beliefs as fixed and exogenous, and so does not consider the fact that in the process of deliberating with one another people will revise their beliefs.<sup>40</sup> This, of course, is true, so far as it goes. Unfortunately, it does not solve the social choice problem of Condorcet's voting paradox, for (except in a few special cases) no amount of preference revision eliminates the possibility of cycling preferences. This can easily be seen in a more technically advanced presentation of social choice theory which, unfortunately, space does not permit in an article of this scope.<sup>41</sup>

When these objections are seen to fail, some have resort to the clear empirical fact that we do not as a general rule observe groups paralyzed by inconsistent

<sup>39</sup>It is always possible as a contingent empirical fact that people will just happen not to have these beliefs; but it would be odd for a theory of justice to depend in this way on the mere coincidence of some actual distribution of beliefs among some actual set of people.

<sup>40</sup>This objection can be found in Cohen 1989, pp. 81–2, for example.

<sup>41</sup>In what is called a "spatial" presentation of social choice in two or more dimensions (see Austen-Smith and Banks 1999, chs. 5–6), this conclusion with respect to the inefficacy of preference revision should be obvious. That said, I am not aware of anyone bothering to point it out specifically in print.

preferences or beliefs in real life; this being the case, it must be that the formal model is wrong in some respect. Again, this may be true to some extent, but it entirely misses the point. Obviously we do not observe these situations in real life, because in real life decision-making is rarely, if ever, unstructured. All this objection shows is that we should be considering structured, rather than unstructured, hypothetical consent-based theories. Accordingly, let us turn to these.

Structured theories impose type-II non-cognitive requirements on the consent-giving process, in addition to type-I cognitive requirements alone. Generally speaking, these requirements are of three different sorts. First, there might be (type-IIa) restrictions on the sort of information persons are allowed to have access to. Rawls's "veil of ignorance" is an example of this sort of requirement. Second, there might be (type-IIb) restrictions on the sorts of preferences people are allowed to have. For example, some want to insist that we consider only the consent of those persons who have a genuine desire to cooperate with one another, or who are not too excessively self-interested, and so on. Third, there might be (type-IIc) restrictions on the procedures to be followed in arriving at an acceptable result. These might simply be decision rules (like majority rule), or they might be rules about the agenda-setting process, or about the manner of the deliberations, etc., or some combination of these. Bruce Ackerman's requirement that only arguments "neutral" between conceptions of the good be allowed in deliberation is an example of this last sort of requirement.<sup>42</sup>

Even the simple example from above can demonstrate how type-II requirements might solve the problem of Condorcet's voting paradox and engineer determinate results. Consider, for example, a seemingly innocuous agenda rule that "teleological principles will be considered against other teleological principles, and non-teleological against non-teleological, before teleological principles are compared against non-teleological principles." Utilitarianism ( $x$ ) and liberal perfectionism ( $y$ ) are teleological, so this vote will take place first; since persons 1 and 2 prefer utilitarianism to liberal perfectionism, the former wins, and then loses against Rawls's two principles ( $z$ ) in the second vote. This rule, therefore, engineers the victory of option  $z$ , and eliminates any indecision. Other results clearly could be engineered with other procedures. Alternatively, determinate results could be obtained by excluding certain sorts of preferences, or by excluding certain sorts of information, or by requiring that deliberation proceed according to certain rules, and so on. In short, structured consent theories easily avoid the problems associated with social choice theory. Unfortunately, they do not do this without cost.

The structures imposed on the consent-giving process can be relatively sparse or relatively dense. The more dense they are, the more likely they are to guarantee a particular result. Here, then, is the crux of the problem for voluntarist theories

<sup>42</sup>See Ackerman 1980, p. 11 and *passim*.

in general: On the one hand, we might adopt relatively sparse structures, in which case no particular result is ensured, and indeed (as social choice theory suggests) the results may even be incoherent. Our consent-based theory of justice is then indeterminate, for it fails to tell us what the principles of justice actually are. On the other hand, we might adopt structures sufficiently dense as to yield a determinate result, and thus avoid the problems of indecision or incoherence. But then, with the structure itself engineering the outcome, it is hard to see why the selection of one particular structure does not amount in the end simply to the selection of a particular set of principles of justice, since the adoption of some other structure would have resulted in a different outcome.<sup>43</sup> The procedural structures end up doing all the normative work, leaving nothing for the element of consent itself, and undermining the whole point of a voluntarist consent-based theory of justice to begin with.

#### IV.

The failure of consent-based theories of justice when given a voluntarist construction leads us to reconsider the viability of a rationalist construction. The advantages of rationalism should be clear in light of the difficulties raised in the previous section. It does not matter, for example, that a theory attempting to base an account of justice on consent-giving procedures is structured so as to avoid the problems of social choice theory, because the principles of justice themselves now supply the necessary independent standards: we know we have the right consent-giving procedures, so the argument goes, precisely when they yield the right results.

Rationalist consent-based theories of justice face a new set of objections, however. On the voluntarist view, the principles of justice are defined as whatever people would agree to under appropriate conditions; the need for consent-giving procedures is thus obvious. But if, on the rationalist view, we have independent grounds for knowing some procedural structure that yields a determinate result is the right one, it is difficult to see what normative work is being done by the consent-giving process at all. Some might suggest that asking what people would agree to under appropriate conditions is merely a metaphorical way of illuminating more clearly or persuasively those principles already known to be the true principles of justice. If this is the case, however, then of course we do not really have a consent-based theory at all, for consent adds nothing to the strength of the underlying argument on behalf of the particular principles of justice in

<sup>43</sup>This might not be the case if (a) our grounds for adopting some procedure were independent and non-normative, and (b) if this procedure happened also to ensure a determinate result with respect to the principles of justice. Habermas's argument (see 1983, pp. 82–94, for example) that the appropriate procedures are those presupposed by the deliberative process itself plausibly might satisfy (a), but as he himself recognizes (*ibid.*, pp. 94, 103, etc.) the procedures remain well short of satisfying (b).



question. The consent-giving process turn out to be merely an empty rhetorical device.<sup>44</sup> In order to avoid this result, the rationalist needs to argue that figuring out what people would consent to happens to be the best or most appropriate *method*—or perhaps, an important component of the best or most appropriate method<sup>45</sup>—for determining what the principles of justice (already) are.

This might be clarified by the following analogy. Suppose the principles of justice are things akin to natural facts like, say, the gravitational constant or the atomic weight of chlorine. We cannot simply look and see what the gravitational constant or the atomic weight of chlorine really is—rather, we must try to figure these out via some appropriate method of scientific research. Now, of course, even the best scientific research cannot determine either exactly, though perhaps we may get closer and closer with a continued refinement of technique. The point, however, is that there is not some other direct or shortcut method against which the results of our best scientific research might be compared: the *only* account available to us is precisely whatever that research happens to turn up. Now, analogously, some sort of consent-giving process might be a part of the best or only possible method for discovering the principles of justice, even if it is not perfect. In this case, the consent-giving process would not itself *create* justice, any more than scientific research creates the gravitational constant or the atomic weight of chlorine,<sup>46</sup> but regardless, there is no better method for discovering what the principles of justice really are. On this view, asking what people would consent to under appropriate conditions is simply the best possible mode of moral reasoning (at least when it comes to figuring out what the principles of justice are).

Suppose we accept something like this view. The difficulty will now be to explain *why*, exactly, this is the case. In other words, we need something akin to what philosophers of science would call a “reliability argument”: although we cannot directly test the results of consent-based moral reasoning against independently known conclusions (any more than scientists can directly test the results of their research against independently known natural facts), we can still ask for an argument purporting to show why this particular mode of moral reasoning is better than any alternative. Moreover, the rationalist must supply a reliability argument that does not reduce consent to an empty rhetorical device on the one hand, or fall back into voluntarism on the other.<sup>47</sup>

Let us suppose that under the appropriate conditions *C* persons would agree that *P* are the principles of justice. Since the fact that they agree is not itself what

<sup>44</sup>Dworkin 1973, pp. 18–19, suggests this as an interpretation of Rawls.

<sup>45</sup>This may be Rawls’s view: seeking a reflective equilibrium between the original position argument (a consent-based theory) and an argument from moral intuitions, we work on the problem “from both ends” as he put it (1971, p. 20). More on this below.

<sup>46</sup>Some would argue scientific research actually does create the gravitational constant, the atomic weight of chlorine, etc.; such philosophy of science questions are obviously beyond the scope of this article.

<sup>47</sup>Stark 2000, pp. 326–8, seems to fall into the latter trap.

makes *P* the principles of justice, we must ask the rationalist what other information is being conveyed here. One response might be the following: Suppose we believe that ordinary (English-speaking) persons already possess an intuitive grasp of the correct principles of justice. If so, then it would seem a reliable method for determining what justice is would be to simply ask what people think it is, because ordinary language speakers obviously have privileged access to this sort of information. Thus, we may understand the moral philosopher's role as refining and clarifying the conception of justice already latent in ordinary language. Consent-giving procedures serve as a conceptual device—the political philosopher's analogue to a standard sort of analytic linguistic philosophy. An alternative response might begin by pointing out that any theory of justice must include an account of moral motivation. In other words, unless it can be shown that the principles of justice connect somehow with the desires or interests of ordinary persons, those principles will lack the motivational force necessary if they are to be effective. Consent-giving procedures might then be understood as providing the required moral motivation by showing what rational persons might in fact agree to on the basis of their desires or interests.<sup>48</sup> Surely, however, neither of these can be what is intended by a consent-based theory of justice, for the simple reason that neither yields an account of justice at all. The former offers only an account of what a particular group of ordinary language speakers *mean* when they use the term “justice”—in other words, a descriptive theory of meaning, and not a normative theory of moral philosophy. Similarly, the latter offers no more than a psychological account of motivation.<sup>49</sup>

Assuming we want something more than just a theory of meaning or human psychology, our reliability argument must take some other form. The rationalist view must be that asking what people would agree to under certain conditions reveals somehow which principles are in fact reasonable and which are not. Accordingly, let us consider the different sorts of reasons people might have for selecting certain principles of justice over others. Of course, not any sort of reason will do. Suppose a group of people have been situated under the appropriate conditions *C*, and they are asked to select the principles of justice. If they select principles *P* on the basis of faulty reasoning, or misinformation, or because they are impatient to get on with their lives, etc., we would reject their selection as meaningless. Likewise if they decided to make their decision on the basis of a throw of dice. (And if we refuse to second-guess their selection then

<sup>48</sup>I am grateful to an anonymous *Journal of Political Philosophy* reviewer for pointing out this possible argument.

<sup>49</sup>Perhaps a theory of moral philosophy is implied in these proposals—roughly, that there is no such thing as “justice” apart from what ordinary people mean when they use the word, or from what ordinary people would agree is ultimately in their interest. Either would be a version of moral skepticism. However, neither an account of what people mean by the term “justice,” nor an account of what people would agree is ultimately in their interest, is an argument for or against moral skepticism, and in any case skepticism is not at issue in this article.

we slip back into voluntarism, which regards the fact of consent—regardless of its basis—as the source of justice.) In short, the selection of *P* by people under conditions *C* only conveys something meaningful if it is based on the *right* sorts of reasons.<sup>50</sup>

Suppose, therefore, the consent-givers indeed select *P* as the principles of justice on the basis of a set of good reasons *R*. For example, these reasons might be: because *P* will maximize the expected wellbeing of the participants in the selection process; or, because *P* will best realize a particular conception of the good for human beings; or something else. It does not matter for the purposes of the argument here what *R* happen to be. What does matter is that there is no substantive difference between saying, “under conditions *C* people would select *P* as the principles of justice on the basis of reasons *R*,” and saying, “*P* are the principles of justice for reasons *R* (and anyone would agree too, if subject to conditions *C*).” The former is simply an opaque and roundabout version of the latter.<sup>51</sup> So why not offer the argument for *P* directly?

Perhaps the appeal of the roundabout version has something to do with the manner in which it effaces the role of the political philosopher presenting the argument; it is, in other words, a somewhat more modest (if also more confusing) manner of presentation. Be that as it may, it should be clear by now that there are no grounds for regarding the consent-giving process as the best or only appropriate mode of moral reasoning, for the simple reason that it is not a mode of moral reasoning at all. Rather, it is merely a mode of presenting an argument of moral reasoning. An argument for *P* on the basis of reasons *R* can always be semantically recast in the form “under conditions *C*, people would select *P* on the basis of reasons *R*.” Imagine two competing sets of arguments, which rely on reason *R*<sub>1</sub> and *R*<sub>2</sub> respectively: recasting both arguments in the consent-giving form would not resolve the issue between them, since it would just be another way of asking which argument has the stronger set of reasons going for it. Thus, on the rationalist interpretation, the consent-giving process turns out to be no more than a superfluous heuristic device after all.<sup>52</sup>

## V.

On either the voluntarist or the rationalist interpretation, theories attempting to base an account of justice on consent-giving procedures ultimately fail. In the

<sup>50</sup>This argument follows to some extent that of Nagel 1973, p. 5.

<sup>51</sup>And, what is worse, the former brings along with it unwanted conceptual baggage. For example, consent-based arguments expose themselves to boundary problems: How do we define the community of consent-givers? How are we to understand justice for persons who cannot possibly be members of the same community (e.g., future generations)? And so on.

<sup>52</sup>Since a rationalist consent-based argument alone adds nothing, it follows that operating as one part in a more elaborate method (see n. 45 above), it will derive whatever force it has from the other parts. Thus I agree with Barry (1995, pp. 61–7) that the real normative work in Rawls is being done by the “strains of commitment” argument, and not the argument from the original position.

former case, they fail because they are indeterminate and possibly incoherent; in the latter case, because they are superfluous. Considered individually, not all of the criticisms discussed here are new, but their cumulative effect when considered together has not been clearly pointed out. What we find is something akin to an impossibility result: if a voluntarist consent-based theory can be defended only by a retreat into rationalism, and a rationalist consent-based theory only by a retreat into voluntarism, it is difficult to see how any attempt to base an account of justice on consent-giving procedures could possibly succeed.

Consider, for example, Scanlon's development of a consent-based theory—a clear instance of someone trapped on the first horn of Cassirer's Dilemma. In its generic form, as noted earlier, the theory is ambivalent: the right principles are those “that no one could reasonably reject as a basis for unformed, unforced general agreement.”<sup>53</sup> It might be that these principles are right just because no one could reasonably reject them, but in order to avoid the indeterminacy of such a view Scanlon steadily retreats in the other direction. “Deciding whether an action is right or wrong,” he writes, “requires a substantive judgement on our part about whether certain objections to possible moral principles would be reasonable.”<sup>54</sup> The substantive judgements in question turn out simply to be our direct moral intuitions concerning which principles are reasonable and which are not; thus, consent has become merely an empty heuristic. Habermas provides a clear example of someone trapped on the second horn of Cassirer's Dilemma. Recognizing that an introduction of substantive moral judgements into the theory will ultimately eviscerate any authentic role for consent, he finds himself driven in the opposite direction, towards extreme voluntarism: “All contents, no matter how fundamental the action norm involved may be, must be made to depend on real discourses,” he concludes. “The moral theorist may take part in them as one of those concerned,” but to the extent that “a moral theory touches on substantive areas . . . it must be understood as a contribution to a discourse.”<sup>55</sup> Here consent becomes indeterminate and, as I have argued, vulnerable to incoherence on well-known grounds of social choice theory. Other consent-based theories may seem stronger than these two only because they are more successful at remaining ambiguous.

The problems addressed in this article are not merely the transient features of a particular contemporary approach to questions of justice. On the contrary, Cassirer's Dilemma is a recurring, deep-rooted problem at the very core of moral and political philosophy. What ties together the problems faced by classical natural law theory, Rousseau's doctrine of the general will, and contemporary consent-based theories of justice (and other theories I have not discussed), is a heavy reliance on notion of *the will* as a foundational concept in moral philosophy. It does not matter particularly whose will is regarded as important

<sup>53</sup>Scanlon 1998, p. 153, and passim.

<sup>54</sup>Ibid., p. 194.

<sup>55</sup>Habermas 1983, p. 94; cf. p. 103.

(thus, a shift from the will of God to the will of the people does not change the terms of debate as much as one might expect), because the problem lies precisely in the irreducibly contingent nature of the will as such. Given the age-old intractability of this problem, it seems to me very unlikely that a moral or political philosophy based on such unstable grounds could ever overcome it.

Before concluding, let me briefly consider one other apparent rationale for developing a consent-based theory of justice. The argument goes roughly as follows: Suppose we believe *P* are the correct principles of justice, but that we also happen to believe that people should not be forced to accept *P*, but rather should be convinced on the basis of good arguments to accept *P* freely. (Note that we must be careful not to say, “*P* are the principles of justice *because* they are the principles we could get ordinary people to agree with,” for this would throw us back to voluntarism.) On this view, a consent-based theory of justice is a way of emphasizing the second belief. In other words, by asking what people would agree to, we force ourselves to find arguments for *P* that can be accepted voluntarily by ordinary people (and not, perhaps, philosophers alone). In a well-known article, Jeremy Waldron argues that something like this belief constitutes the theoretical core of liberalism as a political philosophy.<sup>56</sup>

There is nothing wrong with this argument, but it does not amount to a theory of justice: rather, it is only a theory of legitimacy (albeit a consent-based one). The attacks I have directed here against consent-based theories of justice do not necessarily apply to consent-based theories of legitimacy. That said, consent-based theories of legitimacy face important limitations of their own, one of which is the following: Suppose we agree in general that people should not be forced to accept principles they would not under the right conditions voluntarily agree to. But what are the right conditions? Presumably, our answer to this question will in turn be bound up in a general account of why consent matters in the first place. For example, we may believe that only by securing voluntary consent can we respect the equal moral worth of all human beings. This, of course, constitutes a substantive claim of moral or political philosophy, something a theory of legitimacy cannot produce by itself.<sup>57</sup> In other words, theories of legitimacy must ultimately be parasitic on theories of justice; we do not escape Cassirer’s Dilemma merely by ignoring it.<sup>58</sup>

What sort of theory could take the place of an attempt to base a normative account of justice on consent-giving procedures? One obvious answer would be some version of utilitarianism. Let me briefly suggest an alternative I find more

<sup>56</sup>Waldron 1987.

<sup>57</sup>Unless, of course, there exists one and only one set of principles *P* of justice that people could ever come to voluntarily accept. In this case, our commitment to the consent-based theory of legitimacy would force our hand, and render a theory of justice moot. This, however, seems implausible—particularly when we consider the potential for education and social pressures to shape what people might consent to.

<sup>58</sup>In his later work, Rawls often tries to do this. See for example his strikingly evasive discussion of political constructivism (1993, pp. 90–9).

appealing: Suppose we adopt some broad ethical account of the good as human flourishing. Without having to specify in detail exactly what constitutes human flourishing, we might nevertheless be able to give a reasonable account of the most serious and most common obstacles to achieving it. For example, these might be (tentatively in order of importance): material deprivation, domination or oppression, and a lack of individual autonomy. Justice can then be understood negatively as the reduction or minimization of these obstacles to human flourishing.<sup>59</sup> Obviously, developing an account of justice along these lines would be a difficult and challenging task. Consent-based theories of justice (indeed, procedural theories in general) may be popular precisely because they provide an apparent means of evading the difficult task of actually developing a substantive theory of justice. Moral and political philosophers would do better to stop trying to avoid the issue, and start trying to address it directly.

## REFERENCES

- Ackerman, Bruce A. 1980. *Social Justice in the Liberal State*. New Haven, Conn.: Yale University Press.
- Allen, Glen O. 1961. La volonté de tous and la volonté générale: a distinction and its significance. *Ethics*, 71, 263–75.
- Austen-Smith, David, and Jeffrey S. Banks. 1999. *Positive Political Theory I: Collective Preference*. Ann Arbor: University of Michigan Press.
- Barry, Brian. 1989. *Theories of Justice*. Berkeley: University of California Press.
- Barry, Brian. 1995. *Justice as Impartiality*. Oxford: Oxford University Press.
- Benhabib, Seyla. 1996. Towards a deliberative model of democratic legitimacy. In Benhabib (ed.), *Democracy and Difference*. Princeton, N.J.: Princeton University Press.
- Binmore, Ken. 1998. *Game Theory and the Social Contract, vol. 2: Just Playing*. Cambridge, Mass.: MIT Press.
- Cassirer, Ernst. 1932. *The Question of Jean-Jacques Rousseau*, trans. Peter Gay. New York: Columbia University Press, 1954.
- Cohen, Joshua. 1989. Deliberation and democratic legitimacy. In James Bohman and William Rehg (eds), *Deliberative Democracy*. Cambridge, Mass.: MIT Press, 1997.
- Cohen, Joshua. 1996. Procedure and substance in deliberative democracy. In James Bohman and William Rehg (eds), *Deliberative Democracy*. Cambridge, Mass.: MIT Press, 1997.
- Dagger, Richard K. 1981. Understanding the general will. *Western Political Quarterly*, 34, 359–71.
- Dahl, Robert A. 1989. *Democracy and Its Critics*. New Haven, Conn.: Yale University Press.
- Dworkin, Ronald. 1973. The original position. In Norman Daniels (ed.), *Reading Rawls*. Stanford, Calif.: Stanford University Press, 1989.
- Eslund, David. 1997. Beyond fairness and deliberation: the epistemic dimension of democratic authority. In James Bohman and William Rehg (eds), *Deliberative Democracy*. Cambridge, Mass.: MIT Press.
- Gauthier, David. 1986. *Morals By Agreement*. Oxford: Clarendon Press.
- Gutmann, Amy, and Dennis F. Thompson. 1996. *Democracy and Disagreement*. Cambridge, Mass.: Harvard University Press.

<sup>59</sup>This sort of theory of justice is “negative” in the sense that it does not say anything directly about what people can or should aspire to be. Doing this would be the task of utopianism (in a non-pejorative sense), which can be seen as complementing a theory of justice.

- Habermas, Jürgen. 1983. *Moral Consciousness and Communicative Action*, trans. Christian Lenhardt and Shierry Weber Nicholsen. Cambridge, Mass.: MIT Press, 1990.
- Habermas, Jürgen. 1992. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, trans. William Rehg. Cambridge, Mass.: MIT Press, 1996.
- Habermas, Jürgen. 1993. *Justification and Application: Remarks on Discourse Ethics*, trans. Ciaran Cronin. Cambridge, Mass.: MIT Press.
- Manin, Bernard. 1987. On legitimacy and political deliberation. *Political Theory*, 15, 338–68.
- McKelvey, Richard D. 1976. Intransitivities in multidimensional voting models and some implications for agenda control. *Journal of Economic Theory*, 12, 472–82.
- Milo, Ronald. 1995. Contractarian constructivism. *Journal of Philosophy*, 92, 181–204.
- Moon, Donald. 1993. *Constructing Community*. Princeton, N.J.: Princeton University Press.
- Nagel, Thomas. 1973. Rawls on justice. In Norman Daniels (ed.), *Reading Rawls*. Stanford, Calif.: Stanford University Press, 1989.
- Nagel, Thomas. 1991. *Equality and Partiality*. New York: Oxford University Press.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Plott, Charles R. 1967. A notion of equilibrium and its possibility under majority rule. *American Economic Review*, 57, 787–806.
- Rawls, John. 1958. Justice as fairness. *Philosophical Review*, 67, 164–94.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Rawls, John. 1993. *Political Liberalism*. New York: Columbia University Press.
- Riley, Patrick. 1970. A possible explanation of Rousseau's general will. *American Political Science Review*, 64, 86–97.
- Rousseau, Jean-Jacques. 1762. On the social contract. In *The Basic Political Writings*, trans. Donald A. Cress. Indianapolis, Ind.: Hackett Publishing, 1987.
- Scanlon, T. M. 1982. Contractualism and utilitarianism. In Amartya Sen and Bernard Williams (eds), *Utilitarianism and Beyond*. Cambridge: Cambridge University Press.
- Scanlon, T. M. 1998. *What We Owe To Each Other*. Cambridge, Mass.: Harvard University Press.
- Schneewind, J. B. 1998. *The Invention of Autonomy: A History of Modern Moral Philosophy*. Cambridge: Cambridge University Press.
- Scott, James C. 1990. *Domination and the Arts of Resistance: Hidden Transcripts*. New Haven, Conn.: Yale University Press.
- Shapiro, Ian. 1999. *Democratic Justice*. New Haven, Conn.: Yale University Press.
- Shepsle, Kenneth A. 1979. Institutional arrangements and equilibrium in multidimensional voting models. *American Journal of Political Science*, 23, 27–59.
- Sreenivassan, Gopal. 2000. What is the general will? *Philosophical Review* 109, 545–81.
- Stark, Cynthia A. 2000. Hypothetical consent and justification. *Journal of Philosophy*, 97, 313–34.
- Waldron, Jeremy. 1987. Theoretical foundations of liberalism. *The Philosophical Quarterly*, 37, 127–50.
- Waldman, Theodore. 1960. Rousseau on the general will and the legislator. *Political Studies*, 8, 221–30.
- Walzer, Michael. 1983. *Spheres of Justice*. New York: Basic Books.
- Young, Iris Marion. 2000. *Inclusion and Democracy*. Oxford: Oxford University Press.