

CHAPTER 7

The Logical Bases of Deliberative Democracy: The Limits of Consensus

Much (but not all) of the deliberative democracy literature takes the idea of consensus as its logical basis. Some of this literature follows Habermas, and the idea that something can be justified if it results from universal unforced agreement in an ideal deliberative situation. Others follow Rawls, justifying principles and institutions on the grounds that they would have been agreed upon unanimously in a hypothetical deliberative situation. Both these theorists, and those who follow them, rely on the idea of unforced consensus. Unfortunately, unforced consensus is logically impossible in a political context, as Rae (1975) shows. Put simply, in a political context, a decision still has to be made in the event consensus is not reached. Thus there is always an implicit threat, and consensus is not unforced (even if none of the participants makes use of the implicit threat). This problem undermines not only a great deal of the deliberative democracy literature but also the social contract tradition.

However, it is possible to build a theory of deliberative democracy on a sounder foundation, namely, majority rule. Previous chapters have shown that majority rule uniquely satisfies the value of political equality, and that it is the rule most likely to encourage reasonable deliberation. Indeed, it can be argued that the use of consensus in Rawls and Habermas is motivated by considerations of equality. However, the idea that equality is realized in consensus is clearly incorrect, as was shown in chapter 5. Nevertheless, a theory of deliberative democracy based on majority rule actually produces results quite similar to those anticipated by Rawls and Habermas. There is also, of course, a tradition of deliberative democracy based on majority rule in representative bodies as opposed to consensus. Burke (1777/1963), Hamilton, Madison, and Jay (1788/1961), and Mill (1861/1993) are classical examples. Dewey (1927/1946), Barry (1995), Nino (1996), and Knight and Johnson (1994, 1996, 1999) are modern examples of this line of thought.

Section 1 considers the problem of consensus as outlined by Rae (1975) and its relevance for social contract theory and deliberative democracy. Section 2 considers the problem of consensus in the work of

Habermas and Rawls. Section 3 defends majority rule as a basis for deliberative democracy.

1. The Problem of Consensus and Social Contract Theory

Douglas Rae's article "The Limits of Consensual Decision Making" (1975) demonstrates that no decision rule can be completely consensual if a definite decision needs to be made. This result—which I will call the problem of consensus—undercuts social contract theory by showing that social decisions cannot be based completely on consent. Formally, Rae shows that no decision rule can give everyone the right of consent and be robust. Here, *robust* means that the decision rule has to return some decision. Rae argues that this quality is inescapable in a political context, because even if no decision is taken, some outcome must result. Given that a political decision rule must be robust, it cannot give everyone the right of consent. That is to say, everyone cannot have a veto over every outcome. If there is disagreement in society, an outcome has to be imposed that at least one person objects to. Thus, any decision-making rule is to some degree coercive. Consent is always conditional on the outcome that would have been imposed if agreement had not been reached.

The intuition underlying Rae's result is simple. In politics it is not possible to not make a decision. If no agreement is reached, and no formal decision is taken, then some outcome still results. This outcome may be to do nothing for the time being. However, this is still an outcome—indeed it may be some people's preferred outcome. Alternatively, the default outcome may be to keep doing what we were doing previously, maintaining the status quo. In any case an outcome results. If agreement was not reached, then this outcome is imposed on at least one person against his will. However, even if there is agreement, this agreement is not completely free. This is because the threat of the imposed outcome hung over the discussion. If people had not agreed, a certain outcome would have to have been imposed. Of course, this threat will often weigh more heavily on some people than others. People who like the imposed outcome, or at least can tolerate it, will have a considerably stronger bargaining position than those for whom the imposed outcome is intolerable. It is for this reason that Rae argues that consensual decision making has a strong bias toward the status quo. Consensual decision making implies that everyone has a veto and the power to impose the status quo (or some other default outcome) on everyone else.

Rae's argument is aimed first and foremost against Buchanan and

Tullock's *The Calculus of Consent* (1962). Buchanan and Tullock ask which decision rule individually rational people would choose for their constitution, given the (crucial) assumption that property rights have already been settled. They argue that the optimal rule would be unanimity, because this is the only rule that guarantees economic efficiency in the sense of Pareto optimality (it is not possible to make anyone better off without making anyone worse off). We would choose unanimity, it is argued, because if a government action made people better off as a whole, it would be possible to compensate the losers, and thus gain unanimous consent. (Buchanan and Tullock accept that unanimity increases decision-making costs associated with getting universal agreement, and therefore they argue that a rule less than complete unanimity may be preferable in practice.) By Rae's argument, however, this agreement is not truly consensual. Rather, it is only consensual conditional on the initial distribution of property. What unanimity essentially does is allow those who like the initial distribution to lock in their gains and impose this distribution of property on society, even if the vast majority of the population disapproves.¹

Similarly, Rae argues that unanimity implies a form of anarchism. This is because anyone can use their right of consent to block government action. Therefore it is impossible to use the government to protect yourself against another individual or group. Rae uses this argument against both Calhoun's (1850/1943) theory of concurrent majorities and Wolff's (1970) defense of anarchism, illustrating that logically similar arguments are made by people of apparently very different ideologies. Calhoun, of course, was insisting on the right of planters in the Southern United States to have a veto on any decision to abolish slavery, while Wolff was defending the right of the individual to oppose government actions they disapprove of. In both cases, however, the insistence on universal consent (applied only to groups in Calhoun's case) brings us back to the state of nature—individuals cannot be coerced by the government to do anything that they do not consent to. The cost of this is that individuals certainly can be coerced by each other. Unanimity is a rule very favorable to people who are powerful in the state of nature, whatever that is taken to be.

The same argument applies very naturally to classical social contract theory. People do not choose the social contract freely. Rather they only give consent conditional on what outcome would be imposed on them if they fail to agree—that is, depending on what the state of nature is like. Thus the results of social contract theory are simply a function of the state of nature described by various theorists. Hobbes (1660/1972) creates a state of nature so horrible that the sovereign has extreme bargaining power because any outcome is better than the war of all against

all. In Locke (1690/1986), property rights exist in the state of nature prior to the social contract, so naturally the social contract protects these rights. In Rousseau (1762/1997), property rights do not exist in the state of nature, although morality does. As a result we get a social contract where property rights are protected but subordinate to the general will.

As a result, social contract theory loses much of its force. Social choice theory is based on the idea of consent. If everyone were to freely accept something, it is very difficult to argue against it. Unfortunately, the idea of freely given consent turns out to be illusory. Consent in classical social contract theory can only be conditional on the default outcome, the state of nature. Therefore the results depend on the state of nature, and, of course, the mythological states of nature given in classical social contract theory are not particularly compelling in any kind of literal sense. However, modern versions of social contract theory, such as the work of Rawls and Habermas, use the conditions under which deliberation takes place in a different manner. The conditions under which consent is to be reached are taken as artificial constructions, to be defended in explicitly normative terms. That is, the conditions under which the parties deliberate are chosen in order to make them deliberate fairly, rather than being chosen to represent some naturalistic situation. However, we will see that the problem of consensual decision making weighs just as heavily on these theories.

2. The Consensual Bases of Deliberative Democracy (or, Why the Ideal Speech Situation Is a Logical Impossibility)

The modern successors to social contract theory still rely on consensual agreement as the basis of legitimacy. Instead of considering agreement in a mythological or naturalistic setting, they either construct a situation so that parties are forced to argue in a moral fashion (Rawls), or they reconstruct the conditions under which an agreement reached would be morally legitimate (Habermas). However, as they still rely on the concept of consensual agreement, they are also subject to the problems raised by Rae—unconditional consensus is not possible if a decision needs to be made, as is always the case in politics.

Habermas and the Ideal Speech Situation

Habermas's philosophical project, at least since *The Theory of Communicative Action* (1984, 1987), has been to use the ideal of consensual agree-

ment achieved through discourse to provide a justification for the concepts of truth and validity that can survive the critique of skeptical, relativistic, and postmodern philosophies.² Habermas accepts that metaphysical justifications are no longer viable—it is not possible to argue convincingly that there is a transcendent truth that resides beyond human reason. However, if we engage in discourse in good faith, we must engage in what Habermas describes as communicative action. That is, we enter into communication with the intention of finding agreement with another person. We make statements that can be valid or not (they can be correct or incorrect; morally valid or invalid; aesthetically authentic or inauthentic), and we try to convince others of their validity. Doing this assumes that such agreement is possible, otherwise our speech would be futile and our action self-contradictory. Therefore we must believe that agreement (and thus validity) can be obtained, at least under ideal conditions. These ideal conditions comprise what Habermas (1984, 42) calls the “ideal speech situation,” in which unforced argumentation is conducted openly, without strategic manipulation and without time constraints. Thus the possibility of consensual agreement in ideal conditions provides the basis for the belief in truth and validity.

Habermas’s work since *The Theory of Communicative Action* has sought to apply this principle in more concrete settings. *Moral Consciousness and Communicative Action* (1990a), particularly the essay within it entitled “Discourse Ethics: Notes on a Program of Philosophical Justification,” applies the principles of communicative action to moral philosophy. Once again, it is the possibility of consensus through discourse that creates the philosophical basis for the validity of moral claims. Following Strawson (1974) he argues that moral claims can be seen as relative from the vantage point of a disinterested observer but not from the point of view of a participant. If we enter into an argument about what is right, we accept that our statements can be valid or not. This validity is provided by the discourse principle: “Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse” (Habermas 1990a, 66). It should be noted that Habermas’s (1990a, 211) goal here is quite modest. He does not claim to be making any substantive contribution to morality, only providing a refutation of moral skepticism. (See Heath 2003 for an excellent reconstruction of Habermas’s argument without some of the harder-to-defend assumptions, notably the assumption that engaging in discourse commits a participant to reaching agreement on the moral claims underlying norms.)

Between Facts and Norms (1996b) applies the discourse principle to the legitimacy of law and democratic decision making. Once again, it is

the possibility of agreement under ideal conditions that provides the basis for validity claims. Law differs from morality in that it is codified and enforceable by state coercion. Thus law facilitates social cooperation by ensuring that people know what their legal duties are, that others will in general comply with their duties, and that people's duties and rights are coordinated in an organized manner (114). The problem is to show how positively existing law can be justified in a moral sense. Habermas continues to hold that norms (legal and otherwise) can only be validated by agreement in a discourse including all involved. However, in *Between Facts and Norms* this discourse can take the form of decentralized social discourse that generates "influence" and "communicative power," which in turn is mediated by procedures in a representative parliament that produces binding decisions. Thus the people ideally adopt morally valid norms that they impose freely upon themselves. This, incidentally, gives equal and co-original status to the political right to self-govern and the liberal rights to individual self-determination: self-governance is only possible if individuals are free to enter into discourse; while the only legitimate way to decide the limits of individual self-determination is through discourse. Again, it should be noted that Habermas is only providing conditions under which laws generated by parliamentary procedures can be morally legitimate; he explicitly denies providing "anything original at the level of particular details" (444).

In *Between Facts and Norms*, Habermas (1996b, 179) does allow for the legitimate use of majority rule and bargaining, but this remains subordinated to consensual agreement and communicative action. Majority rule produces legitimate norms, but only in a provisional sense. It represents a "caesura" in discourse—we have not reached the agreement necessary for full validity, but the majority-rule result represents a fallible interim result of discussion that is presumed to continue. Furthermore, the majority-rule decision is only morally valid if the discussion that led to it was rationally motivated (informed and oriented toward reaching agreement, rather than toward strategic advantage). Habermas also accepts the role of bargaining (that is, strategic negotiation as opposed to argument seeking agreement) in politics. Indeed, he accepts (282) that most of politics consists of bargaining. However, bargaining is subordinated to communicative discourse in two ways (167). First, the terms under which bargaining takes place can only be morally validated using the discourse principle; second, only particular interests can be involved, because generalizable interests require moral discourse. These concessions to practical decision making are not particularly problematic from a philosophical point of view—consensus through rationally motivated discourse is intended as a counterfactual ideal, after all, not a description

of reality. However, we will see that the ideal of consensus is problematic for more intrinsic reasons.

The problem is that Rae (1975) shows unconditional consensus is logically impossible when a decision needs to be made. If this is so, then the ideal speech situation is a logical—not just a practical—impossibility in these cases. As a result, it is not possible to conceive—even as an ideal—a completely free situation in which people reach agreement by means of consensus. This undercuts the principle on which Habermas bases the legitimation of moral and legal norms. It should be noted that the problem here is that the ideal speech situation is *logically* impossible. If it were merely impossible as a practical matter, it could still serve as a counterfactual ideal. Indeed Habermas himself emphasizes that the ideal speech situation is intended as a counterfactual. As we have seen, in *Between Facts and Norms* he argues that majority rule is legitimate in political institutions precisely because it represents a provisional and temporary stopping point on the road to the ideal of consensus. However, if the ideal of unforced consensus is illusory and impossible, this kind of argument is no longer viable. It is no longer possible to argue that unforced agreement is the source of legitimacy, because unforced agreement can never exist even under ideal conditions.

Thus the ideal speech situation is logically impossible for precisely the type of practical questions that we encounter in politics. The ideal speech situation may still be possible for theoretical questions. If we do not have consensus on some point of (say) physics or chemistry, then we can simply say that this question is not yet resolved. Indeed the same is true of moral questions, as long as these remain on a theoretical level—we can say that there remain unresolved questions as to which moral principles are valid. We hope that as more research is done and better arguments are made, we will move toward the agreement that currently eludes us. However, we cannot do this in the case of practical questions of the type “What are we to do?” In these cases, not making a decision is effectively a decision to do nothing (at least for the time being) or some other default outcome. Furthermore, in these kinds of practical situation, we cannot choose the question—rather the question presents itself. If we cannot agree whether a man is guilty or not, we cannot substitute a more abstract question, such as whether acts of type x are in general culpable. (This is not to say that considering the more abstract question may not help in our deliberations.) Rather we have to decide whether to change a policy, whether to enact a constitution, whether to go to war, or whether to punish a defendant. For these kinds of decisions—which are precisely the kind of decisions we are concerned with in politics—an ideal speech situation is logically impossible.

We should note that the problem here is not simply the lack of time. Rather it is that a decision has to be made, which makes unconditional consensus impossible. Habermas acknowledges the problem of time and accepts that in practical settings majority rule may have to be used instead of consensus when a decision needs to be made within a limited time. However, time is not really the issue here, as illustrated by the following thought experiments. Suppose that we were able to get unlimited time for argumentation by means of a time machine. That is to say, a decision has to be made at time x , but we can go back in time as far as we like in order to give us enough time to deliberate. (Assume also that we have indefinite life spans and patience.) This would not solve the problem of unconditional consensus. We would still have to make a decision at time x , and thus the threat of the imposed outcome would still hang over us. Suppose, alternatively, that we had a technology that allowed us to talk and think at an infinite speed. We would then have effectively unlimited time to deliberate. However, this would not help us either. We would still have to make a decision at time x , with the implicit threat of the imposed outcome. The problem does not result from a lack of discursive resources (time, goodwill, etc.) but from the threat implicit in the fact that a definite decision of some kind has to be made.

It is interesting that Habermas (1987, 310–11) accepts an argument very similar to the one made here when considering organizational behavior. Consider the following passage.

Members of organizations act communicatively only with reservation. They know they can have recourse to formal regulations, not only in exceptional but in routine cases; there is no necessity for achieving consensus by communicative means. Under conditions of modern law, the formalization of interpersonal relations means the legitimate demarcation of scopes for decision making that can, if necessary, be utilized in a strategic manner. Innerorganizational relations constituted via membership do not replace communicative action, but they do disempower its validity basis so as to provide the legitimate possibility of redefining at will spheres of action oriented to mutual understanding into action situations stripped of lifeworld contexts and no longer directed to achieving consensus.

People in organizations can behave communicatively—that is, in a manner oriented toward reaching consensus. However, they always have the fallback option of being able to resort to bureaucratic rules and procedures. This strips the communicative action of its “validity basis,” because the participants are acting communicatively only “with reservation,” knowing that they do not have to reach a consensus by communicative

means. This situation, however, applies to all practical political matters, and not just to members of bureaucracies. There is always an outcome that will be imposed if agreement is not reached, so consensus can only be conditional.

Thus the logical impossibility of the ideal speech situation in the case of practical political questions undermines the applicability of Habermas's approach to politics. However, I will argue that a considerable amount of Habermas's project is still viable, if we replace consensus with the principle of majority rule. However, before turning to this, let us consider the implications of the problems of consensus for the work of Rawls.

Rawls and the Problem of Consensus

Rawls, like Habermas, uses consensus reached through discourse as a means to provide a justification for political and moral norms without having to resort to metaphysical foundations. In *A Theory of Justice* (1971/1999), the concept of justice derived is moral, covering all aspects of life. In the series of articles and lectures that culminate in *Political Liberalism* (1993/1996), however, Rawls makes the conception of justice more strictly political and relevant only to the consensus reached in certain modern, liberal societies. However, although *A Theory of Justice* has a broader scope, I will argue that it is considerably more rigorous than *Political Liberalism* at least in terms of how it deals with the problem of consensus. In *A Theory of Justice*, the problem of consensus is recognized and "solved" by making everyone identical in the hypothetical original position in which the terms of society are negotiated. This, of course, opens the author up to other criticisms, such as that the results claimed are just an artifact of the author's precise construction of the original position and the solution concept he imposes on it. *Political Liberalism* attempts to deal with such criticism but, I will argue, fails to deal with the problem of consensus.

In *A Theory of Justice* (1971/1999) the problem is to specify just terms for social cooperation between people. To do this he considers the agreement that people would come to if placed in a situation where they are deprived of information that is morally irrelevant to choosing just terms, such as information about their personal interests and how various terms would affect them. Thus Rawls's participants are forced to deliberate in an "original position" where they decide the terms of social cooperation behind a "veil of ignorance"—they know enough about society to be able to conceptualize the issues that are relevant, but not enough to be able to tailor the terms in their favor. According to Rawls, in this position people would use the maximin principle

(choose the terms that gives you the highest guaranteed outcome, given that the worst event for you happens) as their solution concept. Rawls argues that this would lead to the adoption of “justice as fairness,” which is comprised of Rawls’s two principles of justice, with the first taking absolute priority.³ With these principles of justice established, the veil of ignorance is lifted slightly in successive stages, and people receive more information about the specific character of society. People are then able to consider what constitution would best serve the principles, and after another relaxation of the veil, consider legislation and its application. If the resulting principles of justice, constitution, legislation, and application can then be accepted by everyone who was socialized in this society with full knowledge, then we have stability and the society is “well-ordered.”

It is apparent that Rawls (1971/1999, 120–21) considers the problem of consensus, as the following passages illustrate.

To begin with, it is clear that since the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, each is convinced by the same arguments. Therefore, we can view the agreement in the original position from the standpoint of one person selected at random. If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached.

Thus there follows the very important consequence that the parties have no basis for bargaining in the usual sense. No one knows his situation in society nor his natural assets, and therefore no one is in a position to tailor principles to his advantage.

The veil of ignorance makes possible a unanimous choice of a particular conception of justice. Without these limitations on knowledge the bargaining problem of the original position would be hopelessly complicated. Even if theoretically a solution were to exist, we would not, at present anyway, be able to determine it.

The problem of consensus is resolved by making argument monological. In the original position everyone is assumed to have the same information and to be equally rational. Therefore everyone is convinced by the same arguments, and we can consider a single person chosen at random. We achieve unanimity by the complete absence of differing interests or perspectives. Thus there is no real bargaining (or discourse), but simply the reflection of a single hypothetical individual. Rae (1975), as we have seen, shows that this complete lack of differing interests is the only way to achieve unforced consensus.

This consensus, however, is bought at a considerable cost—the cost of making the argument monological. This exposes Rawls to the argument that the results he claims are not the result of agreement reached through discourse but simply artifacts of the solution concept⁴ and the original situation imposed by the theorist. As a consequence the results cannot claim the legitimacy that comes from consensus. For example, Rawls’s argument for the difference principle (allow inequality only if it benefits the least advantaged) over utilitarianism with a basic social minimum relies on the information constraints of the original position preventing the participants from knowing what the basic minimum would be (see Rawls 1971/1999, 278–85). Habermas (1995) argues that the construction of the original position requires the theorist to decide what information is morally relevant and what information will lead participants to partial judgments. This logically requires a theorist to make impartial judgments on what information leads to partial judgments. (Habermas is critiquing *Political Liberalism* here, but the argument applies even more to *A Theory of Justice*.) Of course, it could be argued that a particular set of information constraints is justified in terms of ensuring impartial decisions. However, this merely shifts the argument up one level of abstraction. To argue that the decision is consensual, we have to argue that there is consensus about what conditions are appropriate. Furthermore, this consensus would have to be among concrete people, not the constructions of a theorist, as we are arguing precisely about what the appropriate constructed situation should be.

It should be noted that Rawls prior to *A Theory of Justice* did not rely on such a monological construction. In the article “Justice as Fairness” (1958) he justifies principles of justice very similar to those in his later work on the grounds that they would be adopted by a group of individuals deciding on how complaints about the terms of social cooperation should be handled. These individuals are not assumed to be ignorant about their situations. The only thing forcing them to argue reasonably is the knowledge that principles they adopt may work against them in the future. Of course, it is clear why Rawls moves from this position to the identical (and thus monological) deliberators of *A Theory of Justice*. As the passages quoted show, Rawls realized (correctly) that without identical agents it was impossible to draw any definite conclusions about what autonomous agents would agree on.

Thus *A Theory of Justice* clearly recognizes the problem of consensual decision making but avoids it by making the “consensus” be among identical, hypothetical agents. This, of course, gives up much of the claim to validity that would result from freely reached agreement among actual people. *Political Liberalism*, on the other hand, considers the possible

grounds of consensus among actual people but does not consider the problem of consensus rigorously. There are actually two positions taken in *Political Liberalism*, one in the original hardback edition and a second in the introduction to the paperback edition and Lecture IX (“Reply to Habermas”). The “Reply to Habermas,” originally published in the *Journal of Philosophy* in 1995 in response to Habermas’s review of *Political Liberalism*, amends Rawls’s position considerably; the paperback introduction seems to take into account Rawls’s response to Habermas’s objections. The original text of *Political Liberalism* argues that in a plural society, it is not reasonable to expect that any single moral doctrine will be universally accepted. Therefore, all society can be expected to agree on is a political conception of justice, which leaves people to hold any moral doctrine compatible with political justice. However, justice as fairness is still conceived as the appropriate political conception of justice (together with a slightly modified version of the two principles outlined in *A Theory of Justice*), and the original position is taken as the appropriate representation people should use to reason politically. The “Reply to Habermas” and the paperback introduction, however, accept that there are many reasonable political conceptions of justice, and that justice as fairness is simply Rawls’s favored solution. Similarly, it is accepted that the original position is merely one possible representation, and that it needs to be defended in public discourse. This brings Rawls far closer to Habermas’s position that validity comes from agreement reached between actual people. It also exposes him to the problem of consensus, in the same way that Habermas is exposed.

Political Liberalism considers what kind of political conception of justice might be approved by an “overlapping consensus” in a free, plural society. The conception of justice proposed is a version of justice as fairness that is very similar to that in *A Theory of Justice*, including the two principles (slightly modified). In *Political Liberalism*, however, Rawls accepts that in a modern, plural society, we should not expect agreement on which comprehensive moral doctrine is correct. Nevertheless, Rawls argues that there may still be a consensus on a political conception of justice. Various reasonable comprehensive doctrines (*reasonable* is defined in terms of seeking fair terms of social cooperation) may agree on the basic issues of political justice, although they may disagree on issues as to what the good life is, and indeed why the consensus on political justice is justified. Thus there are different comprehensive doctrines (that may include religions, comprehensive philosophies, and political ideologies), but there is agreement on the key issues required for these doctrines to coexist. However, it must be noted that the overlapping consensus Rawls envisions is not a *modus vivendi* or compromise between various existing

comprehensive doctrines. Rather, it has to be justified as a “freestanding” conception independent of comprehensive doctrines, and Rawls terms this process that abstracts from specific comprehensive doctrines “public reason.” The device of representation Rawls advocates for this is the original position, as used in *A Theory of Justice*.

However, it is not clear whether the political conception of justice is really justified by the reasoning from the original position, or whether it is justified by the actual achievement of an overlapping consensus of actual people. This presents a dilemma. If justification is through reasoning from the original position, then the argument is subject to the same objections made against *A Theory of Justice*. The reasoning is monological and conducted under conditions constructed by a political philosopher. Thus it cannot claim the legitimacy that would come from an agreement freely reached by actual people. In this case, the achievement of an overlapping consensus is not necessary for the validity of the “freestanding” conception of political justice, but only important to its empirical acceptance and stability. On the other hand, if justification is through the actual achievement of consensus, then there is no guarantee that actual people will agree on the political conception of justice that John Rawls presents—they could just as well agree on something else. This is the basis of Habermas’s (1995) objection that Rawls does not distinguish sufficiently between the acceptance and the validity of the political conception of justice. Habermas, of course, advocates leaving the actual specification of justice to actual participants, with the philosopher merely providing the terms and procedures that can make any agreement legitimate.

Rawls’s (1995/1996) “Reply to Habermas” and the introduction to the paperback edition of *Political Liberalism* take the second course. He accepts that a political conception of justice is justified by the agreement of actual people, which brings Rawls’s position far closer to Habermas’s. Rawls (382) argues that justice as fairness as a political conception of justice and the original position as a device of representation are to be judged by debate in the public sphere, as the following passage illustrates.

From what point of view are the two devices of representation to be discussed? And from what point of view does the debate between them take place? Always, we must be attentive to where we are and whence we speak. To all these questions the answer is the same: all discussions are from the point of view of citizens in the culture of civil society, which Habermas calls the public sphere. There, we as citizens discuss how justice as fairness is to be formulated, and whether this or that aspect of it seems acceptable—for example, whether the details of the set-up of the original position are properly laid out and

whether the principles selected are to be endorsed. In the same way, the claims of the ideal of discourse and of its procedural conception of democratic institutions are considered.

Thus these concepts are justified by actual agreement obtained in discourse. Furthermore, it is emphasized that the philosopher does not have a privileged role in this discussion (1993/1996, 382–84), and that it is denied that the original position makes reasoning monological, because the original position is itself subject to debate by citizens (1993/1996, 383ff.). Of course, a consequence of this is that justice as fairness and the original position lose their privileged position—people could just as easily agree on other conceptions of justice or devices of representation. In the introduction to the paperback edition of *Political Liberalism*, Rawls (1993/1996, xlviii–xlix) merely states that justice as fairness is the conception of justice that he believes is most reasonable, but he does not deny that there are other conceptions that are also reasonable. Indeed any conception that accepts the fallibility of human reason and the principle of reciprocity is taken as a candidate.

Thus Rawls's final position is that the political conception of justice is justified through agreement reached through discourse in the public sphere. This is a position very similar to Habermas's and brings with it the same problems. Most notable for us is the problem of consensus. As Rae (1975) shows, completely unforced agreement is a logical impossibility in a political matter, because there is always the threat of what will happen if agreement is not reached. Therefore we cannot base legitimacy even on the counterfactual ideal of unforced consensus. To the extent that Rawls appeals to agreement reached in the public sphere, this applies as much to his argument as it does to Habermas's. *A Theory of Justice* recognized the problem of consensus and avoided it by relying not on a genuine consensus but on the monological reasoning of a single, constructed individual. *Political Liberalism* relies (at least after the exchange with Habermas) on consensus between actual people and thus is subject to the problem of consensus.

3. A Nonconsensual Justification for Deliberative Democracy

It has been argued in the previous two sections that the problem of consensus undermines both social contract theory and theories of deliberative democracy based on consent, including the projects of Rawls and Habermas. If there are people with differing interests, it is logically im-

possible for consent to be unforced, because some outcome has to be enforced in the absence of consensus. The only way to get around this is to limit deliberation to identically constructed agents and thus not have divergent interests, as Rawls does in *A Theory of Justice*. This, however, gives up most of the moral force of the idea of consent. I will argue that it is possible to construct a theory of deliberative democracy on a non-consensual basis. This involves going back to the basic value of political equality. If we start with political equality, we can show that this implies majority rule and not consensus as a decision rule. Indeed, a great deal of the existing deliberative democracy literature assumes majority rule rather than consensus (Dewey 1927/1946; Knight and Johnson 1994, 1996, 1999; Barry 1995; Nino 1996).

Actually, it is the value of equality that underlies the justification of consensus. As Dworkin (1975, 53) argues, equality is the basic value that motivates the construction of the original position in *A Theory of Justice*. The original position is a device of representation (to use a term from Rawls 1993/1996) that ensures that the interests of all people are treated equally. Similarly, Habermas's ideal speech situation is premised on the idea that everyone affected by a norm has an equal right to consent to it or not. The problem is that both Rawls and Habermas assume that equality implies consensus, and that consensus does not involve coercion. Chapter 5 has shown that consensus is anything but egalitarian, while Rae (1975) has shown that unforced consensus is a logical impossibility. Instead I have argued that the decision rule correctly derived from equality is majority rule.

Of course, it is possible to argue along constructivist lines (Rawls 1971/1999; Scanlon 1982; Beitz 1989) that equality should be applied to agents in hypothetical choice situations, as opposed to actual participants. I have already rejected this argument in chapter 2. The most telling argument against this approach is the argument made by Habermas (1995) against Rawls. If we create a hypothetical situation of choice under equality as a representational device, then different people may come to different opinions as to what the hypothetical participants would choose. We are then faced with the question of how to arbitrate between them. There has to be some procedure for coming to an outcome, and this procedure needs to respect the values represented in the hypothetical choice situation, or the final choice is arbitrary. Thus the conditions for an acceptable choice situation have to apply to actual participants, not just hypothetical deliberators.

Whereas Habermas criticizes Rawls's approach for being monological, Habermas's definition of the ideal speech situation is open to the charge that it is dialogical, when it should be polylogical. The standard

case for Habermas is a conversation between two people seeking agreement. For Habermas, adding participants does not seem to change the nature of the interaction. However, a polylogical decision situation is qualitatively different from a dialogical one. If we have two people, the only way they can resolve their differences (without one enforcing their will by coercion) is by mutual agreement. Indeed, in a two-person situation consensus and majority rule are equivalent—the only possible majority is for both people to agree. However, if we have a third person, we have other possibilities. If two people are unable to agree, they may attempt to persuade the third person that their proposed action plan is reasonable. As such the third person can act as an audience or judge. The situation is further complicated by the fact that the third person may also have interests that need to be negotiated with the other two players. It is at this point that social choice theory becomes relevant. Rae (1975) has shown that when a decision is required, unforced consensus is a logical impossibility. The only decision rule that is fair in the sense of treating everyone equally is majority rule. Therefore the best we can hope for is a fair and reasonable decision process—that is, decision by majority rule with sufficient time for deliberation.

We can compare the ideal of a fair and reasonable decision process to the ideal speech situation defined by Habermas. Habermas requires that the ideal speech situation be without coercion, strategy, and time limits, so the outcome is determined only by the “forceless force of the better argument.” However, as Rae (1975) shows, all decision rules where a decision must be made are to some degree coercive. A fair and reasonable decision process based on majority rule does not claim to be exempt from coercion. Rather, it is equally coercive to all participants—the coercion is distributed evenly. For this decision-making process to be reasonable as well as fair, some other conditions are required. It is necessary that the participants have sufficient time to deliberate, and that any question can always be raised again. As we have seen, unlimited time does not solve the problem of consensus. However, sufficient deliberation to recognize one’s interests and engage in negotiation is necessary for a decision to be reasonable. Furthermore it is necessary that the participants be granted such basic rights as are necessary to realize what their interests are and to argue for them.

It may be asked how much this formulation varies from Habermas’s in practice. After all, Habermas allows the legitimacy of norms validated by majority rule, on the grounds that majority rule serves as a “caesura” on the road to consensus. In our derivation majority rule provides the basis for a fair and reasonable decision process. In Habermas’s formulation, majority rule serves as a stand-in for agreement reached in an ideal

speech situation for decisions that have to be made in a timely way. Although the justification is different, the practical prescription is the same. However, there are two significant differences. First, the fair and reasonable decision process recognizes the plurality of reason—it does not imply that there is a single true solution to a political question, which we could find if only we had enough time and goodwill. Second, it forces us to view deliberation and negotiation—communicative action and strategic action—as intrinsically linked, not separate modes of action, as Habermas argues.

Let us first consider the plurality of reason. The ideal speech situation implies that if we had enough time and could argue openly enough, we would come to an agreement, and this would be *the* true, right, or authentic outcome. The fair and reasonable decision process makes no such promise. Majority rule does not produce a single, best outcome. Rather it produces cycling. This is not to say that some outcomes are not far more defensible than others, or that majority rule produces arbitrary results. Indeed, as was argued in chapter 4, majority rule is an extremely effective way of eliminating unreasonable outcomes. However, it does not produce a single outcome that can claim to be uniquely valid. This is appropriate when dealing with questions of social cooperation. As Rawls (1971/1999, 1) argues, social cooperation typically makes everyone better off than a situation of no cooperation, but some terms of cooperation benefit some people more than others. Thus the argument over the terms of social cooperation is partially distributive. As argued in the previous chapter, such arguments do not admit a single, correct answer.

This, incidentally, provides a justification for Rawls's (1993/1996) use of the word *reasonable* rather than *true* to describe fair terms of social cooperation in a plural society. Habermas (1995, 119–26) criticizes Rawls's use of *reasonable*, because he thinks that Rawls is actually raising truth claims. However, I would argue that the use of the word *reasonable* is appropriate, although for reasons quite different from Rawls's. Rawls defines the reasonable as being that which is oriented toward fair terms of social cooperation, as opposed to the rational, which is oriented toward the fulfillment of some goal. This definition of *reasonable* does not imply that there is a single, reasonable outcome—more than one set of terms can be fair. Describing an outcome as reasonable does not make the claim that it is the one, true outcome. In the case of outcomes that are partially distributive, we would not want to make such a claim, so Rawls's use of the word *reasonable* is more appropriate than describing these terms as *true*.

The fair and reasonable decision process forces us to treat the rational and reasonable—or strategic and communicative action—as

inseparable. Rawls (1993/1996, 48–54) does argue just this, but Habermas (1984, 286) argues that strategic and communicative action should be treated as quite separate modes of action. This, of course, has led to criticism, notably from Johnson (1991) who argues that the two concepts cannot be coherently separated even in Habermas's work. If people engage in argument, then they are attempting to influence others. Therefore the action is strategic. Indeed the concept of action oriented solely toward the reaching of agreement is problematic for the reasons given by Elster (1983). It makes little sense for the reaching of agreement to be the final goal—reaching agreement can only be valuable as a side effect. If two people aim primarily to reach agreement, then the agreement will be trivial—the one makes a proposal and the other agrees primarily in order to get an agreement.⁵ Agreement can only be valuable if the parties have something to argue about and thus have (strategic) interests. Or, as Rawls puts it, the reasonable assumes the rational. We can only reach a reasonable agreement if we have interests to come to an agreement or compromise about.

A fair and reasonable decision process involves the final decision being made by majority rule. In majority-rule bargaining it is not possible to completely distinguish strategic and communicative action. In order to further one's strategic goals, one has to seek allies. This means seeking agreement with others and thus acting communicatively. However, seeking agreement with others has the inevitable side effect of improving one's strategic situation. Furthermore, it should be noted that in a multi-lateral, majority-rule bargaining situation, a participant being too accommodating does not necessarily produce a fairer outcome. A participant giving ground too easily may encourage other participants to behave in a predatory manner, and harm the interests not only of the accommodating participant but of other participants as well. Therefore we need to avoid simplistic judgments about communicative action oriented to agreement necessarily being morally superior to strategic action. Finally, given that we are not going to have a final "correct" solution to most political matters, compromise is inevitable. Compromise here is not a second-best we resort to when agreement fails; rather it is an essential part of being reasonable. When dealing with distributive matters, all reasonable outcomes are to some degree compromises.

Other Majority-Rule-Based Theories of Deliberative Democracy

The argument that has been developed in the last two chapters bears many similarities to other theories of deliberative democracy that are

based on majority rule, such as the work of Dewey (1927/1946), Knight and Johnson (1994, 1996, 1999), Barry (1995), and Nino (1996). The arguments are generally compatible with one another, differing more in emphasis than in practical implications. There are two main differences. First, Barry and Nino both accept the value of consensus as an ideal representing unforced, uncoerced agreement. However, they reject unanimity (and indeed supermajoritarian rules) as an empirical decision rule for practical reasons. This chapter, however, has argued that the idea that consensus is not coercive is logically flawed, and that majority rule is the most impartial rule because the (inescapable) coercion is equally distributed.

Second, and more significant, our argument starts from a procedural conception of political equality, derives institutions from this, and then argues that these are optimal for facilitating reasonable discourse. Thus the value of political equality has primacy over deliberation, although it appears that there is unlikely to be an empirical trade-off. The procedural arguments from political equality are valid regardless of the behavioral question about what institutions best promote deliberation. Furthermore, following Dewey, Rawls, and Habermas, the procedural conditions of political equality constitute in part what is meant by social reason. This is not to say that politically egalitarian institutions always produce reasonable outcomes (the previous chapter outlines the conditions in which they do not) but merely that there is no external standard for defining reason in a social context. Thus deliberation under egalitarian conditions is the most reliable means for finding a solution we may presume to be reasonable.

This difference in emphasis can be illustrated by considering a thought experiment. Imagine we had a supremely wise and good facilitator-king. This monarch's objective is to make social reasonable decisions based on inclusive deliberation. (Let us ignore the [im]possibility of finding a guardian of sufficient wisdom and goodwill.) The monarch could hold extensive consultations before making decisions, or realizing his own fallibility with regard to knowledge of the interests of others, could even hold straw polls or referenda. If the goal is to promote reasonable deliberation, then such a monarch could provide as high a quality of deliberation as any truly democratic system. (Note that the monarch requires participation, and that this participation affects outcomes, so Mill's [1861/1993] argument that a benevolent despot promotes servility does not apply.) However, this situation could not be described as a fair and reasonable decision process, in the sense outlined earlier. The monarch may provide incentives for deliberation, make good use of the information he receives, and even make reasonable substantive decisions; however, the situation is not democratic, in that the monarch has

the final say. The point is that democracy has value as an intrinsically fair procedure for distributing power, not just as a mechanism for facilitating reasonable deliberation.

In a much quoted passage, Dewey (1927/1946, 206–7) argues that the main value of majority rule is that it forces deliberation that allows the public interest to be defined. However, Dewey (146–47) also argues that the details of the political institutions under which this happens are only of secondary importance. The primary problem is that of how a diffuse public can recognize itself and define its interests. Once this problem was solved, appropriate democratic forms would emerge. This contrasts with our emphasis on the extreme importance of the details of institutions in both determining outcomes and determining how people deliberate.

Knight and Johnson's recent work on Deweyan pragmatism puts far more emphasis on institutions. They argue that pragmatism implies a commitment to radically democratic institutions (1996), and that rational choice theory is a useful tool in determining which institutions may be satisfactory in this respect (1999). In this light, our theoretical justifications of proportional representation and majority rule can be seen as a contribution to this agenda.

Barry's (1995) approach is very similar to that taken here in that the theory of deliberative democracy is embedded in a more general theory of justice as impartiality. As was argued in chapter 2, if we consider equality in liberal (that is, individual) terms, impartiality and equality are equivalent. Barry, however, does not proceed directly from impartiality to political institutions. Rather he defines impartial justice in contractualist terms, using the criterion proposed by Scanlon (1982).⁶ As with Rawls (1971/1999), justice is defined as what would be agreed by reasonable people in a hypothetical choice situation. Although this does not reduce all individuals to identical agents as Rawls (1971/1999) does, this construction inherits all the problems of contractualism that have been discussed in this chapter. In particular, it is monological in that it relies on the opinion of the particular philosopher conducting the thought experiment as to what can be reasonably rejected.

Barry (1989) recognizes that the Scanlonian procedure admits a great deal of uncertainty, and he proposes supplementing the consideration of hypothetical deliberators ("the a priori method") with the consideration of actual deliberation under conditions that approximate impartiality ("the empirical method"). Barry is also well aware of the problems of consensual rules—indeed he has been critical of supermajoritarianism at least since *Political Argument* (1965). He argues (1995, 104–5) that the best empirical approximation of the Scanlonian situation where anyone

can veto a proposal on reasonable grounds is not a situation where anyone can veto a proposal on any grounds, reasonable or not; rather the best approximation is majority rule. It is interesting that the countries that Barry considers best examples of the “circumstance of impartiality” (the small, “consensual” democracies of Europe) are precisely the countries that I argue best satisfy the procedural requirements of political equality. Barry’s (1989, 347–48) argument that these countries satisfy the conditions of impartiality seems to be mainly sociological—there exist multiple, organized interests with access to government, political candidacy is not determined by access to money, there is a culture of accommodation and debate. The arguments in this book provide strict institutional criteria for judging which countries best satisfy the conditions of impartiality. They also provide a far more direct route for getting to Barry’s conclusions—instead of arguing that impartial institutions are the best approximation of a hypothetical (and logically flawed) choice situation, we can directly apply the value of impartiality to the political institutions themselves.

Nino (1996) likewise argues that a parliamentary system with majority rule and proportional representation is the most reliable set of institutions to promote deliberative democracy, although he views representation as at best a necessary evil and is sympathetic toward direct democracy (132). Indeed many of the arguments in our previous chapter arguing that politically egalitarian institutions can be expected to promote deliberation were taken directly from Nino’s *The Constitution of Deliberative Democracy*. Nino, however, does not argue that these institutions are justified in terms of being intrinsically fair. Rather, they are justified as promoting deliberation, which he argues is the most reliable way of arriving at moral truth, a position he refers to as epistemic constructivism. Nino (117–18) argues that unanimity is the situation best capturing impartial decision making, because it involves unforced consent. However, when unanimity is not possible for pragmatic reasons, majority rule is the rule most likely to lead to the reasonable conduct of deliberation. (Nino refutes the argument that the next best thing to unanimity is something close to unanimity—such a situation may be extremely partial, as in a situation where the majority imposes a large cost on a minority of one.)

Summary

This chapter has considered the logical basis of theories of deliberative democracy built on the idea of consensus, as well as classical social contract theory. They are all problematic for the same reason, which I call

the problem of consensus. Rae (1975) shows that no decision rule can give everyone the right of consent if a definite decision needs to be made. Thus unforced consensus is impossible—if everyone does not agree, then some agreement needs to be imposed. Furthermore, even if agreement is reached, it was not unforced—the threat of the imposed outcome was hanging over the procedure. Thus Habermas's ideal speech situation is a logical—not just empirical—impossibility and cannot even serve as a counterfactual ideal. Similarly, any social contract agreement reached in a state of nature cannot be completely consensual. Rawls recognizes the problem of consensus in *A Theory of Justice* but overcomes only by making the reasoning monological—it is not based on actual consensus, but on hypothetical consensus choreographed by the philosopher. To the extent that Rawls's *Political Liberalism* ceases to be based on monological reasoning, it is vulnerable to the problem of consensus.

The problem of consensus, however, does not undermine the idea of deliberative democracy or the deliberative justification of democracy. On a practical level, democratic deliberation takes place within the context of majority rule. Indeed much of the deliberative democracy literature is based on majority rule rather than consensus (Dewey 1927/1946; Knight and Johnson 1994, 1996, 1999; Barry 1995; Nino 1996). However, majority rule should not be regarded as a practical stand-in for the utopian ideal of consensus. Consensus seems to offer the possibility of free, uncoerced choice, but this is illusory. All social decision rules involve some degree of coercion. The best we can do is to distribute coercive power equally. Thus we need to replace Habermas's ideal speech situation with the idea of a fair and reasonable decision process. As has been shown previously, the only social decision rule that treats all people equally is majority rule. Thus the theory of deliberative democracy is compatible with the theory of procedural democracy developed in the rest of this book.