

10 Public Reason¹

For John Rawls, public reason is not one political value among others. It envelops all the different elements that make up the ideal of a constitutional democracy, for it governs “the political relation” in which we ought to stand to one another as citizens (*CP*, p. 574). Public reason involves more than just the idea that the principles of political association should be an object of public knowledge. Its concern is the very basis of our collectively binding decisions. We honor public reason when we bring our own reason into accord with the reason of others, espousing a common point of view for settling the terms of our political life. The conception of justice by which we live is then a conception we endorse, not for the different reasons we may each discover, and not simply for reasons we happen to share, but instead for reasons that count for us because we can affirm them together. This spirit of reciprocity is the foundation of a democratic society.

Public reason has emerged as an explicit theme in Rawls’s writings only after *A Theory of Justice* with his turn to “political liberalism” and the pursuit of a common ground on which people can stand despite their deep ethical and religious differences. But the concept itself has always been at the heart of his philosophy. It runs through his first book in the guise of the idea of publicity, playing an indispensable part in the theory of justice as fairness. The notion of fairness itself, so central to Rawls’s thought, denotes that mutual acknowledgement of principles which public reason demands and which forms the real import of the language of social contract he has used to articulate his conception of justice.

Rawls’s recent writings about public reason outline a complex model of deliberative democracy, as it is called today,² and I examine

his account in detail in Sections III and IV. But I begin by unearthing its roots in his earlier idea of publicity and by showing how fundamental is the dimension of Rawls's philosophy to which the idea of public reason gives expression.

I. PUBLICITY IN *A THEORY OF JUSTICE*

Readers of *A Theory of Justice* ought to wonder more than they do about the contractarian form in which Rawls presents his theory of justice as fairness. Even in the introductory chapter of the book, he does little to explain the need to think about justice in terms of a contract. His notion of an "original position" is meant, like "the state of nature" in the social contract tradition, to describe a situation in which free and rational beings determine the principles that will regulate their subsequent conduct. Yet as Rawls admits (more forthrightly than earlier contract theorists), this initial situation has never existed and never will. The "original position" is a condition in which we imagine choosing principles of justice, not one in which as real people we ever find ourselves. Something needs to be said, therefore, about the reasons to think of principles of justice as the result of an agreement that in fact we never make.

In one passage, Rawls remarks that to understand fair principles of justice as the object of agreement among free and rational persons entails seeing that "the theory of justice is a part, perhaps the most significant part, of the theory of rational choice" (*TJ*, p. 16/15 rev.). This formulation wrongly suggests that fairness derives from the rational pursuit of individual advantage, when in reality it forms an irreducibly moral notion. Of course, the Rawls of *A Theory of Justice* would scarcely have denied this point. Though the parties in the original position are described as deliberating in accord with the principles of rational choice, the conditions he imposes upon their choice (the "veil of ignorance" that denies them knowledge of their class position, assets, and abilities) constitute moral limits on the sorts of information it would be fair for them to utilize. Later Rawls will introduce a distinction between two capacities of reason, the "rational" and the "reasonable," to differentiate these two elements of the original position, and he will disclaim any attempt to derive the reasonable, or the disposition to seek fair terms of cooperation, from the rational (*PL*, pp. 48–53). Still, our question remains, What

useful purpose is served by the idea of an original contract, morally defined though it is, if it refers to an agreement which is never really made?

One might suppose that the structure of Rawls's theory would have been clearer had he not made use of this idea. In an insightful review of *A Theory of Justice*, Ronald Dworkin noted that a hypothetical contract, being strictly speaking no contract at all, can have no binding force on the people supposedly subject to it. It is also an idle notion. To claim that certain principles are valid because they would be the object of rational agreement is a roundabout way, he argued, of saying that they are valid because there is reason to accept what they assert. The two principles of justice favored by Rawls have their real basis in the fundamental right to equal concern and respect which they express.³

Dworkin's skepticism about contractarian terminology has my sympathy, and I agree that a principle of respect for persons undergirds Rawls's theory of justice. But though the idea of an original contract is, as Rawls will later say, just a "device of representation" (*PL*, p. 24), we need to attend to all the aspects of justice which it serves to represent. In fact, conceiving of the principles of justice as the object of a rational agreement comes to more than saying that each individual concerned has reason to accept them. The language of contract also points to the good in each individual's finding that reason in the reason that others have to accept them as well. This good lies at the core of the ideal which Rawls calls "publicity," and a virtue of the idea of contract, as he observes (*TJ*, p. 16/15 rev.), is that it gives expression to this ideal.

The point is that just as the validity of a contract does not turn solely on the terms agreed to, but also on the fact of agreement, so justice consists in more than the proper distribution of rights and assets. Principles of justice should also be public, each of us affirming them in light of the fact that others affirm them too. More is necessary (we may say to fix terms) than just a *scheme of distributive justice*, even one that each of us has reason to endorse. Equally important is the *publicity* of its defining principles – that our reason for accepting them turns on others having reason to accept them too. When a conception of justice enjoys this kind of common support, it figures in our thinking, as Rawls says a public conception will do (*TJ*, pp. 55f./48 rev., 133/115 rev.), exactly as though it had been the result

of an agreement. Even though no formal act of agreement is needed for us to base our reasons on those of others, a “just so” story about a hypothetical contract helps to highlight this public dimension of justice. The contractarian metaphor has the merit of combining in a single image two essential conditions which the principles of justice should satisfy – their justifiability to reason and their publicity. Together these two conditions define Rawls’s ideal of a “well-ordered” society, which not only advances the good of its members but does so in accord with a public conception of justice. It is “a society in which (1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles” (*TJ*, p. 5/4 rev.). Justice would not be all that it should be without this shared affirmation.

It is therefore unfortunate that Rawls does not adequately explain why publicity represents so preeminent a value. In *A Theory of Justice*, the “publicity condition” generally enters the discussion from the side, as though merely a further desideratum that principles of justice should possess. It receives no extended treatment of its own. One can easily overlook how central it is to Rawls’s very idea of justice, and the contractarian terminology can then appear as otiose as Dworkin claims. Regrettably, too, Rawls’s statements about publicity in this book generally equate it with public knowledge, as though it entailed only that citizens know the operative principles of justice and one another’s reasons for accepting them. The work he expects from it shows, however, that he has something more ambitious in mind. Publicity really amounts to the demand that the reasons each person has to endorse the principles be reasons the person sees others to have to endorse them as well. It requires that the principles of justice be grounded in a shared point of view.

Publicity’s true import becomes evident if we follow the role the concept plays in *A Theory of Justice*. It shapes the ideal of a well-ordered society, as we have seen. But it also acts as a crucial premise in the “stability argument” for the two principles of justice, and to that I now turn. Rawls begins by rejecting those “indirect” forms of utilitarianism that favor maximizing the general happiness by means of encouraging people to act on nonutilitarian principles. Such a system of justice would fail even to be an object of public knowledge (*TJ*, p. 181/158 rev.). However, he next invokes publicity in a deeper sense

as he argues that utility, understood as the explicit charter of society, would prove unstable since it would place too great a strain on individual self-esteem (*TJ*, Sec. 29, 69–77). Other things being equal, a conception of justice is better, he holds, the stabler it is, generating its own support so as to outweigh contrary motives. Citizens living under the institutional arrangements it recommends should tend to acquire a commitment to its principles. Stability obtains “when the public recognition of its realization by the social system tends to bring about the corresponding sense of justice” (*TJ*, pp. 177/154 rev.; also 454/398 rev.). The maximization of average utility is thus unlikely to generate its own support since it accords poorly with the facts of moral psychology. Its overriding devotion to efficiency will not inspire the allegiance of those whom it asks to give up their life prospects for the greater good of the whole. Only by calling upon improbable reservoirs of sympathetic identification can utilitarians hope that such a system of justice will endure. By contrast, Rawls argues, his own liberty principle secures the fundamental inviolability of each individual, and his difference principle ensures that everyone benefits from social cooperation. These two principles define a system of justice whose operation is more likely to engage the support of all, even of those who fare worst. Unlike utilitarianism, they exemplify the idea of “reciprocity” (*TJ*, p. 14/13 rev.), an idea rather undeveloped in *A Theory of Justice* but lying at the center of *Political Liberalism* and its doctrine of public reason.

Now this stability argument relies on the inner meaning of publicity, though to see it we must look at the argument somewhat differently than Rawls does himself. One might suppose that it requires principles of justice to be public simply so that everyone may know that they are in force and see what their institutions stand for. However, more must be involved. For stability is said to obtain when the “public recognition of [their] realization”, thus the knowledge that others too affirm these principles, fosters everyone’s conviction that they are valid and worthy of support. Yet Rawls does not fully explain why people might be moved to espouse certain principles because others espouse them as well. Indeed, he presents this argument in terms of the good which each person will discern in the liberty and difference principles from his own point of view – their guarantee of individual inviolability and their assurance that all will benefit from social cooperation (*TJ*, pp. 177ff./154ff. rev.). But

the stability argument has to be different in character. It must show that each person can find reason to embrace these principles in the fact that others embrace them too. It should therefore indicate the good that the public affirmation of the principles may embody.

Halfway through his account of the argument, Rawls takes up a line of thought that suggests what that good is. "The public recognition of the two principles," he writes (*TJ*, p. 178/155 rev.), "gives greater support to men's self-respect and this in turn increases the effectiveness of social cooperation." Respect is indeed the good in question. But note that Rawls's statement asserts not so much that the principles express respect as that their public recognition does so. Thus, the self-respect each person finds confirmed in them has to be part of a mutual respect which their common affirmation displays. Though Rawls is not as clear on this matter as one might wish, his discussion of "the natural duty of respect" here and later in the book entails that the good of mutual respect lies in there being a shared basis for the determination of principles of justice. We respect others as ends in themselves, he holds, when in regard to their claims and interests we act on reasons that we are prepared to explain to them in the light of mutually acceptable principles (*TJ*, pp. 179/156 rev., 337–8/297 rev.). We try to see things as they do, taking our bearings from a point of view that we can all endorse together. Respect for persons implies allegiance to principles that we affirm in the light of others having a reason, indeed the same reason, to affirm them too.

Naturally, respect can mean many things, but in the sense just mentioned it makes up the true nature of the publicity condition. When citizens adopt certain principles of justice for reasons they understand one another to acknowledge, their joint endorsement of the principles amounts to showing one another respect. Their grounds for embracing them do not lie solely in their own, but in a shared point of view. The mutual respect demonstrated by their allegiance to this common basis is then a good which they can regard themselves as having achieved, and that is why the scheme of justice gains in stability. Their society illustrates Rawls's claim that "a desirable feature of a conception of justice is that it should publicly express men's respect for one another" (*TJ*, p. 179/156 rev.). What the publicity requirement really comes to, therefore, is that each person's adherence to the principles of justice should turn on reasons that he understands others to have to affirm them as well. This point

remains largely implicit in *A Theory of Justice*, but in several essays published shortly afterwards Rawls spells it out in full detail (see Section II).

Once its full meaning is laid bare, we can better understand why Rawls should attach so great a value to the ideal of publicity. Only principles of justice which citizens affirm on a common basis are ones by which they can show one another respect as persons. The idea that political community should rest on this sort of mutual respect belongs to the heart of Rawls's philosophy. It underlies one of the most telling ways he has of contrasting his view of justice as fairness with utilitarian conceptions (*TJ*, pp. 23–7/20–24 rev., 187–190/163–6 rev.). "Utilitarianism," he observes, "does not take seriously the distinction between persons." It proposes that we adopt for society as a whole a form of practical reasoning appropriate for the single individual: just as the prudent person evaluates his possibilities with an eye to achieving the most good overall, accepting some losses for a greater gain, so a just society regards persons as different lines for an allocation of benefits and burdens that will maximize the net balance of satisfaction as judged by a sympathetic observer. To heed the separateness of persons, by contrast, is to seek principles which they can freely acknowledge before one another – principles, that is, which each can see that others have the same reasons to endorse as he. This mutual acknowledgment of principles is the very essence of what Rawls means by fairness as a conception of justice, though he brought it out better in his earlier and foundational essay "Justice as Fairness" (1958) than in the book, which cloaks it in the language of an original contract.⁴ The idea of fairness explains the value of publicity and embodies what Rawls himself calls the natural duty of respect.

One reason for his reluctance to present his theory in these terms may well be the many different meanings "respect" can have. At the end of *A Theory of Justice* (*TJ*, pp. 585–6/513 rev.), he declares that he has not derived the principles of justice from respect for persons because the very notion of respect calls for interpretation, which only a conception of justice can provide. The hermeneutic point is well taken. But it does not rule out the possibility that respect, in a specific sense we grasp perhaps only in the light of his theory as a whole, is a value on which that theory rests. And so, as Rawls goes on to admit, respect for persons plays two roles in his conception of

justice. It shapes the two principles themselves with their emphasis on the inviolability of the individual – the role which Dworkin was concerned to lay bare. It also figures in the demand that persons be treated “in ways that they can see to be justified” (id.). That is the role of respect underlying the ideal of publicity.

II. FROM PUBLICITY TO PUBLIC REASON

In several essays published after *A Theory of Justice*, the notion of publicity receives more systematic attention, and not by accident. Its greater prominence reflects the new direction in Rawls’s thinking that leads to *Political Liberalism*. In the course of this transformation the ideal of publicity grows into his doctrine of “public reason.”

For instance, Rawls acknowledges more clearly that the importance of publicity in a well-ordered society is not simply a matter of its principles of justice being known to all. They should also be principles that citizens affirm on the basis of a shared rationale. Such is the intent of the distinction introduced in his *Dewey Lectures* (1980) (*CP*, pp. 324–6) and contemporaneous writings (*CP*, p. 293) between three “degrees” or “levels” of publicity.⁵ A conception of justice satisfies the “full” publicity condition when its acceptance is not only an object of public knowledge, and not only based upon beliefs to which everyone can assent, but also thereby justified in a manner which all can embrace. In *A Theory of Justice*, Rawls generally used the term “publicity” in a sense equivalent to the first of these levels, the other two being tacitly at work in the way the stability argument capitalizes upon public knowledge of the operative conception of justice. Now the virtue which principles of justice have in being affirmable from a common point of view is made part of the very idea of publicity. Principles public in this strong sense should be our goal, he argues (*CP*, p. 325) because a well-ordered society rests upon fair terms of cooperation to which free and equal persons could agree. Thus, also for the first time, Rawls connects publicity directly (and not just via the metaphor of contract) to the ideal of fairness so that its centrality to his conception of justice comes through more conspicuously than before.

Rawls offers a further argument for the full publicity condition (*CP*, pp. 325–6). Principles of justice should draw on common ground because they apply to institutions having a deep and durable effect

on people's lives – not least through the machinery of legal coercion. This transparency in which people can acknowledge before one another the basis of their common life is “a precondition of freedom.” Plainly Rawls must mean political freedom or self-rule, where the use of force is involved. For moral principles outside the domain of justice need not, he adds, be public in this strong sense, though their effects on adherents and others alike can be equally profound. Why does Rawls thus limit the scope of the publicity condition? One tacit reason is that coercion differs so significantly from other forms of social influence (namely by its irresistibility) that it ought to be grounded in consensus. Elucidating this moral assumption would require developing the implicit idea of respect, of what it is to treat others as ends in themselves, and this Rawls has never been inclined to do (at least in his political philosophy).

But another reason for the limitation is presented explicitly in the *Dewey Lectures* (CP, p. 326). Moral notions distinct from the principles of justice often belong to religious, philosophical, or ethical doctrines on which people in modern societies are unlikely to agree, even as they can find a shared basis for settling questions of political justice. Publicity aims at a freedom of self-determination which citizens can exercise together despite their abiding disagreements. To enjoy this identity-in-difference, they must observe therefore a certain self-discipline, bringing to their deliberations about issues of justice only those convictions which can form part of a common point of view. “In public questions,” Rawls writes, “ways of reasoning and rules of evidence for reaching true general beliefs that help settle whether institutions are just should be of a kind that everyone can recognize” (CP, p. 326). Here is a first statement of the theory of public reason formulated in response to the doctrinal diversity which will be the chief preoccupation of his emerging political liberalism.

In the transitional essays of the 1980's, Rawls often describes this public form of reasoning in terms of a distinction between justification and proof.⁶ Justification is not merely “valid argument from listed premises.” Instead, it “is addressed to others who disagree with us, and therefore it must always proceed from some consensus, that is, from premises that we and others publicly recognize as true” (CP, p. 394; also pp. 426–7). The contrast is overdrawn, for justification can take many forms, depending on the purpose at hand; sometimes

it only consists in showing people how our assertion follows from our own beliefs. But the point Rawls has in mind is obvious. In a well-ordered society, citizens do not determine basic matters of justice by announcing to one another the conclusions they each have derived from their own first principles and then resorting to some further mechanism, such as bargaining or majority voting, to resolve the conflicts. They reason from what they understand to be a common point of view; their aim is to adjudicate disagreements by argument. As we have seen, a public life founded on mutually acknowledged principles is what fairness entails.

This idea of consensus underlies the different notion of an “overlapping consensus,” which makes its appearance in his writings of this period.⁷ Principles of justice, he argues, ought to be the object of an overlapping consensus among citizens otherwise divided by their comprehensive ethical, religious, and philosophical doctrines. Rawls’s point has often been misunderstood. Many have supposed that he means to abandon the claim that his theory of justice is true or correct. If the nature of justice is to be defined by reference to what a society’s members happen to agree upon, how can there be any room to argue that current opinion is wrong? And why should we believe that in these matters there is much of substance that people agree upon at all? However, our earlier discussion of publicity explains why these worries are ill-conceived. The basic sense in which principles of justice ought to be the object of consensus is that each person should have both sound and identical reasons to embrace them, for only then does their publicity give expression to mutual respect. Consensus so understood is therefore hardly identical to the extent of agreement about justice that actually obtains in a society. Yet an important question is whether this shared perspective, rooted as it must be in reasons which citizens can acknowledge only by abstracting from their divergent visions of the human good, nonetheless coheres with the comprehensive conceptions to which they are attached. Only if the consensus shaping their public reasoning about justice also forms an overlapping consensus, a common element in their otherwise different points of view, is the structure of their political life likely to endure. The notion of overlapping consensus serves therefore to connect a conception of justice already arrived at, and already marked by a more fundamental kind of consensus, to the question of its stability.

Rawls himself speaks in this regard of two “stages” in his theory of justice as fairness (*PL*, pp. 64, 140ff.). In the first stage the theory aims to describe fair terms of cooperation among citizens, while in the second it considers whether such principles can prove stable. The notion of overlapping consensus comes into play only at this subsequent stage. Clearly principles of justice are not being fixed by appeal to the common denominator of existing opinion. At the same time we should not overlook the idea of consensus that does figure in the initial determination of these principles. Publicity requires that they draw on reasons which all can acknowledge. As Rawls observes (*PL*, p. 64), public reason is a value which the first stage of his argument seeks to respect.

Overlapping consensus became a central notion for Rawls in the 1980s as he realized how much broader is the range of moral outlooks congruent with a commitment to justice as fairness than he had assumed. He became far more alert than before to the fact that in their comprehensive philosophical and religious conceptions of the human good, people have a natural tendency to diverge, not because of prejudice or inadvertence, but because of what he called “the burdens of reason” (*CP*, pp. 475–8) or later “the burdens of judgment” (*PL*, pp. 54–8). The complexity of the evidence, the necessity of weighing together different sorts of considerations, the need for judgment in applying key evaluative concepts, the variety of life experiences in modern society – all these factors conspire to make agreement about the nature of the good life improbable. To be sure, some comprehensive ideals deny the importance of fair terms of social cooperation, and their adherents cannot be expected to endorse Rawls’s two principles of justice. But there remain a great many different ethical and religious ideals which share a commitment to fairness. Their proliferation Rawls calls “reasonable pluralism” (*PL*, pp. 36, 63f.), since by reasonableness he means, as I have noted, precisely such a commitment.⁸ Reasonable pluralism is the condition we should expect to thrive under free institutions, where in the absence of state power enforcing any particular doctrine the burdens of judgment drive people’s thinking in different directions. The principles of justice which citizens embrace from a sense of fairness can therefore prove stable only if they cohere with the various elements of this diversity.

A Theory of Justice did not itself approach the problem of stability in this pluralist spirit. Part III of that book laid out a single ethical conception based on regarding an individual's good as the object of a rational plan of life, in the light of which the citizens of a well-ordered society would be moved to act justly. Moreover, this conception, often in so many words but sometimes explicitly (*TJ*, p. 572/501 rev.), displayed the hallmarks of the Kantian ideal of individual autonomy, according to which all our principles of conduct (not just those of justice) should be ones that free and equal rational beings would choose under the ideal conditions of an original position. Such a strategy embodied too narrow a view of the possibilities. Indeed, the multiplication of reasonable views of the human good is something which a modern constitutional democracy is bound to encourage, and thus Rawls's initial solution of the stability problem is caught in an internal contradiction. The way out, he came to see, lies in recognizing that in a free society many disparate comprehensive views of life can still overlap in a public understanding of justice.

It was by taking to heart the fact of reasonable pluralism and seeing the error in his earlier solution to the stability problem that Rawls went on to develop his new theory of "political liberalism" (*PL*, p. xlii). In the classical liberalisms of Kant and Mill, the account of justice had been presented as part of an all-encompassing moral philosophy, and *A Theory of Justice* followed their lead. Yet in fact neither the moral ideal of individual autonomy nor an experimental attitude toward life is an essential ingredient in the rationale for his conception of justice as fairness. What is necessary, Rawls announces in the first essay marking this turn in his thought, is "to apply the principle of toleration to philosophy itself."⁹ Justice as fairness, along with the reasons making up the public understanding of its basis, should be regarded as a "freestanding" conception, which people who see a greater value in tradition and belonging than the Kantian and Millian philosophies allow can still embrace.

Political liberalism is not "political" in the sense that, forsaking principled argument, it reduces justice to a compromise among given interests or to the common denominator of existing opinion (*CP*, p. 491). That should now be plain. But it does seek principles of political association which citizens have reason to affirm together despite

the religious and philosophical disagreements setting them apart. Moreover, their reasons for embracing the principles must not spring simply from their different perspectives but must also draw upon a common point of view. Only so, as we have seen, can these principles represent fair terms of cooperation that express mutual respect. The shared understanding of principles of justice must therefore be at once reasoned and neutral with regard to the comprehensive conceptions of the good on which citizens disagree. Obviously, it is no small task to work out the character which this common language should have. In the essays of the 1980's and then in *Political Liberalism*, the "full publicity condition" is expanded into a detailed and sophisticated account of "public reason." The idea of public reason has its roots in the notion of publicity employed in *A Theory of Justice*, but Rawls's new concerns have moved this theme from the periphery to the center of his attention.

III. THE DOMAIN OF PUBLIC REASON

Rawls gives the idea of public reason two extended treatments. The first occurs in *Political Liberalism*, chiefly in Chapter VI, and the second in an essay of 1997, "The Idea of Public Reason Revisited." I center my account around the first while noting the significant revisions in the later essay.

Public reason, Rawls writes (*PL*, p. 217), is an ideal of democratic citizenship, a "duty of civility," which governs the way in which citizens should deliberate together about the fundamental questions of their political life. In seeking to draw up fair terms of cooperation, they should reason from premises which they can all acknowledge. The exercise of public reason will not as a rule mobilize their full thinking about the problems before them, since their comprehensive conceptions of the good and the right are bound to entail distinctive views about other aspects than those having to do with justice. On these matters they may find themselves in deep and irresolvable disagreement. But such differences are set aside when citizens committed to fairness decide questions having to do with the "basic structure" of society – questions which in Rawls's view concern both "constitutional essentials" (the general form of government and the fundamental rights of citizens) and basic matters of social and economic justice (*PL*, pp. 227–30). Similarly, people may well

continue to understand these decisions in the light of their various comprehensive doctrines. The demand is only that they see the need for a common perspective and be able and ready to justify their decisions within its terms (*PL*, pp. 241–43).

Why, one might ask, should the domain of public reason be limited to these fundamentals instead of extending to all the political decisions which a community must make? Rawls does not give a clearcut answer to this question (*PL*, p. 215). On the one hand, he suggests that the restriction might eventually be lifted: the sense in focusing on fundamentals is that if the demands of public reason do not apply in this case they can scarcely hold more broadly.¹⁰ On the other hand, he ends his discussion with the thought that citizens might sometimes be right to settle these further issues in a more particularist spirit. Though he gives no example of what he has in mind, one possibility is the present system in the province of Quebec, which guarantees basic rights for all while also giving special protection and support to the use of the French language despite the existence of a sizable Anglophone minority.¹¹

Another question is whether the discipline of public reason applies to every kind of political deliberation in which citizens may engage, or only to those deliberations which form part of the official process for arriving at binding decisions that will have the force of law. Certainly when citizens take part in decision making by voting in elections or exercising public office as legislators and officials, Rawls holds that they must base their decisions (again, where matters of basic justice are involved) upon reasoning rooted in a point of view which all can share. Thus, the American Supreme Court, charged as it is with settling questions of constitutional principle, counts as an exemplary organ of public reason (*PL*, pp. 231 ff.). He also emphasizes that in “the background culture” – as members of the particular associations (churches, universities, and professional groups) making up civil society and as adherents of different philosophical and religious conceptions – citizens may discuss among themselves political questions, even of a fundamental sort, according to their own “non-public reasons” (*PL*, p. 14; *CP*, p. 576). Political debate rightly shows a greater mix of voices in areas of society other than the circumscribed realm of public reason, and it would be wrong to suppose that Rawls’s theory of public reason means to encompass the “public sphere” in this broader sense, which was the topic, for example, of

a widely influential study by Jürgen Habermas.¹² Much misdirected criticism has arisen from this confusion.

But to return to our question, Can citizens or particular associations address their comprehensive conclusions about political issues, not only to like-minded souls, but to everyone in the community whatever their persuasion? Can the Catholic bishops, for example, direct their religiously inspired arguments for regarding abortion as murder to believers and nonbelievers alike? Or does the ideal of public reason require that citizens participating in the political debates of society as a whole hold back and speak only in the regimented terms it provides even if they are engaged not in making binding decisions but only in the back and forth of argument? Should they reserve their full-scale views for intramural use?

It may seem that Rawls believes they should. For neither in *Political Liberalism* nor in "The Idea of Public Reason Revisited" does he note the difference between two forms of public debate – *open discussion*, where people argue with one another in the light of the whole truth as they see it, and *decision making*, where they deliberate as participants in some organ of government about which option should be made legally binding. To say, as he does, that public reason concerns the "kinds of reasons [citizens] may reasonably give one another when fundamental political questions are at stake" (*CP*, p. 574) fails to discriminate between the two. Yet the distinction is plain and important. Rawls remarks that there exists a kind of political discourse he calls "declaration," in which citizens make known to one another their comprehensive understandings of the right and the good (*CP*, p. 594; *PL*, p. 249). But he has in mind only their showing one another how their conceptions support the common viewpoint of public reason. Such exchanges do promote mutual trust, as he observes, but they are not the same as a free and open discussion of political questions. Moreover, several times in *Political Liberalism* Rawls places "political advocacy in the public forum" (*PL*, pp. 215, 252) among the activities regulated by public reason. Many have therefore taken him to be claiming that all political debate in society at large, at least when it bears on fundamentals, should not depart from the common ground citizens share.¹³

Now such a view would be unappealing for several reasons. First, it is essential for us to know the different convictions our fellow citizens hold about controversial issues and not only because we need to be reassured that they can nonetheless find in them reason to

embrace a common standpoint for political decisions. We also gain a firmer appreciation of the value of that standpoint, seeing how without it so much would tend to drive us apart. Second, unbridled public discussion has the obvious virtue that through it we can come to change our mind. We can find ourselves persuaded by the way some initially unattractive opinion is defended. We can also be impelled to think through more carefully than before our own comprehensive commitments. In fact, the community as a whole may be moved to give a deeper or more nuanced articulation to the common principles by which it orders its political life.¹⁴

It should be observed, however, that a strait-jacketed view of political debate does not follow from the justification which Rawls himself presents for the ideal of public reason. Its basis, he writes, is “the liberal principle of legitimacy,” which holds that “our exercise of political power is proper and hence justifiable only when it is exercised in accordance with a constitution the essentials of which all citizens may reasonably be expected to endorse in the light of principles and ideals acceptable to them as reasonable and rational” (*PL*, p. 217). This principle captures the thesis advanced in earlier writings, namely, that the terms of political association must form part of a public consensus because of their essentially coercive character. And that thesis in turn, as I mentioned before, gives expression to the value of respect for persons which inspires his thought as a whole, particularly his cardinal ideal of fairness. Rawls has never made clear, it is true, the foundational role of respect, which structures even the freestanding conception of justice he attributes to political liberalism. But that is not our present concern.¹⁵ The pertinent fact is that the ideal of public reason, based as it is upon the principle cited, ought to be understood as governing only the reasoning by which citizens – as voters, legislators, officials, or judges – take part in political decisions (about fundamentals) having the force of law. Rightly conceived, it does not thwart the uninhibited political discussions which are the mark of a vigorous democracy. We can argue with one another about political issues in the name of our different visions of the human good while also recognizing that, when the moment comes for a legally binding decision, we must take our bearings from a common point of view.

Rawls never puts things in this way, and so one cannot be sure that he would agree. But it is what the logic of his position entails. By “political advocacy in the public forum” perhaps he meant (as the

context suggests) only the terms in which a candidate for office seeks support, and indeed political campaigns no less than the votes they solicit should adhere to the canon of public reason when constitutional essentials and basic matters of economic justice are at stake. Clarity would have been better served had Rawls given a more complete picture of the kinds of political discussion in a constitutional democracy. Generally, he contrasts the “background culture” with what he calls the “public political culture” (*PL*, pp. 13–14) or the “public political forum” (*CP*, pp. 575–6). Though he defines these latter terms as referring to the institutions and traditions in which citizens or their representatives authoritatively settle fundamental questions of justice, the terms themselves suggest a wider range of political discussion where the rules of public reason do not rightfully apply. I think that at one point Rawls himself may have been misled by this terminology. But to see what I mean, we need first to look more closely, now that the scope of public reason has been defined, at the way that political argument in its jurisdiction ought to proceed.

IV. THE RULES OF PUBLIC REASON

Obviously, many questions of an ethical or religious character, important though they are to people’s self-understanding, will have to be set aside when citizens go about determining the political principles by which they will live, for such questions cannot receive any commonly acceptable answer. But it would be wrong to suppose that for Rawls issues are to be removed from the political agenda just because there exists widespread disagreement about their solution. Public reason does not demand the blanket avoidance of deep-seated conflict as though its highest value were civil peace. On the contrary, public reason embodies the ideal of fairness, and so questions having to do with the fair terms of social cooperation – in other words, matters of basic justice – belong on a society’s program of political deliberation, however disputed they may be. Rawls could not be more explicit on this score (*PL*, p. 151). It is hard to see, therefore, how he would have been obliged, as some have charged,¹⁶ to side with Stephen Douglas in the famous Lincoln–Douglas debates of 1858 and regard the issue of slavery as too controversial to be the object of political decision.

Slavery and its abolition constitute one of the formative experiences in American political life. To regard civil war as the worst of political evils and to suppose that differences should always be papered over by a *modus vivendi* is not a view likely to impress any American thinker, though Europeans of a Hobbesian persuasion often espouse it. One of the benchmarks not just of Rawls's conception of public reason but of his political philosophy as a whole is that basic justice takes precedence over civil peace or, perhaps better put, that it is a precondition for any civil peace worthy of the name. However, the American abolitionists along with the more recent civil rights movement have also inspired a more specific feature of his theory of public reason. William Ellery Channing argued for the emancipation of the slaves just as Martin Luther King, Jr., argued against racial segregation by appealing to the belief that all human beings are equally God's creatures. Clearly they did not do so simply to indicate where they stood personally and to persuade others to share their faith. Their aim was to encourage others to take this religious view to heart as they dealt with those questions in their capacities as voters, legislators, officials, and judges. (Note that I have made clear the involvement of both forms of public debate, which Rawls's own depiction of these movements unfortunately does not distinguish). Did Channing and King therefore overstep the boundaries of public reason? On a straightforward understanding of that concept, they did. But this "exclusive" interpretation is not, Rawls argues (*PL*, 247–54), the only or proper way to think about public reason.

In a well-ordered society, where all citizens affirm together just principles of social cooperation, no one would need in public debate to look outside this common point of view to settle what justice requires. But the situation is different when a society is deeply at odds with itself about constitutional essentials. Then, there exists no generally accepted language of public reason. In such cases, citizens may base their decisions upon comprehensive views that are themselves unlikely ever to form part of public reason, provided they believe or could have believed that thereby the ideal of public reason would be strengthened in the long run (*PL*, pp. 247, 251). Such is the "inclusive" interpretation of public reason as Rawls presented it in the first edition of *Political Liberalism* (1993). Subsequently, however, he has revised his account, though not so as to switch to the opposite position. Rather, he has concluded that even the conditions

he had imposed on the appeal to comprehensive views need to be relaxed. Instead of holding that citizens may reach beyond public reason only when their aim is to steer a profoundly unjust society toward greater justice, Rawls now believes that citizens may call upon their full convictions at any time. The sole qualification is what he terms "the proviso": "in due course public reasons, given by a reasonable political conception, [must be] presented sufficient to support whatever the comprehensive doctrines are introduced to support" (*PL*, pp. li–lii). This "wide" view of public reason was introduced in "The Idea of Public Reason Revisited" (*CP*, pp. 584, 591–2), and in the introduction to the second edition of his book (1996) Rawls declares that it represents his considered position.

Nonetheless, I am not convinced that the change is for the better. To begin with, one cannot help but worry about the vagueness of the proviso. On whom does the obligation fall to satisfy it? And how is "in due course" to be defined? Rawls concedes the existence of these difficulties, adding that no hard-and-fast rules but only "good sense and understanding" (*CP*, p. 592) can serve to handle them. That is fine, if the proviso is truly necessary. But it is unclear what advantages favor this more permissive conception in the first place. What need would there be in a well-ordered society to abandon the constraints of public reason? Rawls's answer is that, in invoking their comprehensive views subject to the proviso, citizens make known how their ethical and religious convictions entail commitment to a common idea of justice. As a result, others feel more secure in their own commitment and stability is enhanced – a boon even in the best of circumstances (*CP*, pp. 592–3; *PL*, p. lii).

Certainly, mutual reassurance of this sort is important. But does it have a part to play in the process by which citizens arrive at legally binding decisions in a well-ordered society? When fair principles of justice are mutually acknowledged by all, why should citizens ever cast their nets more widely to establish how a remaining question of constitutional essentials should be authoritatively settled? It cannot be because they have found they disagree or that they feel uncertain about how the question is to be decided in accord with political reason. For, as we shall see, Rawls believes (and correctly so) that even then they should continue to heed the voice of public reason as each of them best understands it. The mutual reassurance which comes from citizens disclosing to one another the comprehensive roots of

their commitment to justice really has no place in the deliberations by which they decide what shall have the force of law. But it does have a point in the different sort of public debate I have called "open discussion." Indeed, Rawls's "wide" conception seems motivated by the wish to make room for the freewheeling arguments about political issues that belong to the public life of an energetic democracy. Such arguments are properly part of what he calls "the background culture." To permit them in the "public political culture" as Rawls now proposes (subject to the proviso) is to be misled by what that term suggests as opposed to the way that he himself defined it. In the forum where citizens officially decide the basic principles of their political association and where the canons of public reason therefore apply, appeals to comprehensive doctrines cannot but be out of place – at least in a well-ordered society. The earlier, "inclusive" conception, which allows departures from public reason only when its most elementary ingredients are in wide dispute, appears to be the better view.

So much for the question of how strictly the discipline of public reason applies. Another question concerns how much it should aim to accomplish within its sphere. Ideally, it should set its sights on settling all matters of fundamental justice, for they make up its domain. A political conception of justice, as Rawls says, should aim to be complete. Yet situations arise where citizens, reasoning as best they can from the common ground they endorse as free and equal persons, find that they cannot achieve reasonable agreement on an important issue of justice. The right move may be to put off its resolution, not so much to avoid conflict as to allow more time for reflection and experience to shape deliberation. Suspending judgment can promote democracy.¹⁷ Sometimes, however, a decision cannot be postponed. It would be wrong, Rawls argues (*PL*, pp. lv f., 240 f.), for citizens then to suppose that, public reason having failed to settle the issue, they may resort to considerations farther afield borrowed from those parts of their comprehensive views on which there is no overlapping consensus.¹⁸ Standoffs requiring a decision are indeed to be handled by a vote, but a vote carried out in the spirit of public reason. Citizens should follow their best sense of what public reason entails despite the disagreement about what that is and the uncertainty they may therefore feel in their own mind. Where possible, they should also seek to minimize their differences by giving extra weight to points of convergence (*PL*, p. 217).¹⁹

In fact, disagreements are to be expected within public reason, Rawls declares (*PL*, pp. xlix, lii f., lvi), since its common point of view, rightly conceived, is not defined by any one political conception of justice – not even by the liberty and difference principles in his own theory of justice. This feature of his position has not been widely noted. It first appears in “The Idea of an Overlapping Consensus” (*CP*, p. 427), and after *Political Liberalism* Rawls returns to it in “The Idea of Public Reason Revisited” (*CP*, pp. 581, 583, 605 f.). Rawls has not himself given it the detailed treatment it deserves. But the thesis is exceptionally important. It is his way of dealing with the fact, often adduced against his political liberalism, that justice no less than the good life has been an enduring object of dispute even within liberal societies. It also represents a significant innovation in his thinking since, as we have seen, “publicity” and then “public reason” are notions first worked out as part of his account of the two principles of justice. They are now to be understood in less partisan a fashion.

Public reason, Rawls argues, must be able to welcome a family of liberal conceptions of justice, the essential conditions of a “liberal” conception being that it specify certain basic rights, liberties, and opportunities; that it assign a special priority to these elements of a constitutional regime; and that it aim to provide citizens with the means to make effective use of their freedoms (*PL*, pp. xlvi, 6, 223). These are broad conditions, and citizens may make use of their political traditions and theoretical imagination to flesh them out in various ways. Opposing views are likely to arise, and it is the sign of a vibrant democracy that controversy of this sort should go on and that individuals and social movements should be able to challenge the reigning interpretation of justice – not just in open debate but also in political decision making. Utilitarians, opposed though they must be to Rawls’s own principles of justice and wedded instead to the ideal of efficiency, can still conclude that the general happiness is maximized by a scheme of justice satisfying these three conditions.

In their exercise of public reason, citizens may therefore appeal to the different perspectives making up this family of liberal conceptions. Rawls does not explain how they can do so while still heeding the demands of public reason. But it is not difficult to figure out how they have to proceed. In deciding some disputed issue, they must invoke their own views about justice in a form that does not exceed

the bounds of the common point of view they share with their fellow citizens. This means that they must present them as ways of formulating more concretely the three conditions constitutive of a liberal outlook. Of course, their different conceptions of justice will also transcend this public rationale, which explains why they are likely to produce contrary interpretations of those basic, but very general, principles embraced by all. But such disputes revolve around the proper understanding of this common point of view and do not call into question the authority of public reason. As Rawls himself remarks in this context (*PL*, pp. xliv, xlix, li; *CP*, pp. 574, 581), citizens having different conceptions of justice have to share a commitment to reciprocity if public reason is to be possible: they must view one another as free and equal citizens and be prepared to offer one another terms of cooperation which all have good reason to affirm. This standard, of course, is tantamount to what Rawls means by fairness, and fairness, as we have seen, forms the core of the ideal of public reason.

Yet precisely when we see this defining feature of public reason for what it is, we may wonder whether Rawls's wish to accommodate a family of liberal conceptions can really be as generous as he supposes. On the one hand, his own two principles of justice are claimed to be one possible view among others that citizens may invoke as they settle basic questions about constitutional essentials and about social and economic inequalities. But on the other, the very exercise of public reason must embody a commitment to fairness. Does not public reason effectively exclude appeal to any idea of justice that does not, like Rawls's, view the distribution of rights and resources as a matter of arranging fair terms of social cooperation? Must not utilitarians, for example, find themselves debarred from speaking their minds? Believing that justice is to be achieved by institutions promoting the greatest net balance of satisfaction, they may well find reason to agree to the three broad principles characteristic of a liberal society. But if they must reason about how to give content to these principles in accord with the ideal of fairness, are they not being expected to switch philosophical allegiances and give up their distinctive way of thinking?

In reality, Rawls's latitudinarian vision of public reason is not the sham that it might seem at first glance. Recall the distinction I mentioned earlier (Section I) between a scheme of distributive justice and the basis – for Rawls, the indispensably public basis – on which

the individuals to whom it applies ought to adhere to its principles. The actual import of Rawls's contractarianism, I noted then, lies in aiming to handle both these matters by means of a theory of justice centered on the single notion of fairness. And thus among the liberal conceptions consonant with public reason justice as fairness has a special standing. In this conception, as Rawls himself observes (*PL*, p. 225), "the guidelines of inquiry of public reason . . . have the same basis as the substantial principles of justice." But nothing prevents other members of the group from treating these two topics by separate means. It is not therefore incoherent to consider questions of distributive justice as simply questions of efficiency while admitting that principles of justice, to have the force of law, must satisfy the criteria of public reason. Still, utilitarians who hold such a position will have to give their philosophical doctrine an "indirect" form: they will have to support as the public basis for affirming principles of justice a viewpoint (fairness) different from the one they themselves occupy when judging the ultimate reasons for any moral principles. No doubt Rawls continues to think (as in *A Theory of Justice*) that indirect utilitarianism is therefore inferior to his own liberty and difference principles of justice. But he admits that it can belong to the overlapping consensus of a liberal society and thus take part in the discourse of public reason (*CP*, pp. 433–4). Indeed, he welcomes this fact, given the prominent place of utilitarian thought in the democratic tradition.

V. CONCLUSION

The loosening of the link between the ideal of public reason and his two principles of justice, the recognition that controversy is an inescapable part of public reason, comprises one of the most interesting developments in Rawls's recent writings. Clearly it involves many complexities, perhaps difficulties too, which have yet to be explored. Yet we should not lose sight of the way in which fairness continues to shape the heart of that ideal. Even though citizens may understand differently the core principles of a liberal society, and even though some may base their views on the idea that maximizing average utility, not guaranteeing the least well off their best prospects (Rawls's difference principle), ought to govern the distribution of resources, they exercise public reason by determining the fundamentals of their political life within a shared and broadly liberal framework. Their

adherence to this common point of view is itself a commitment to fairness. And thus the disagreements that mark their deliberations embody at the same time the fundamental sort of respect for one another which fairness involves.

As I showed earlier in Section I, fairness and respect are notions which shape Rawls's thought at the deepest level. Nowhere does he subject them to sustained analysis in their own right. Rather, they are deployed in a variety of ways, sometimes (as with fairness) in the metaphor of the social contract, sometimes (as with respect) more implicitly than otherwise, and sometimes in the guise of cognate notions such as reasonableness and reciprocity. But we will not understand his thought unless we trace their ramifications and perceive the overall conception they define. Fairness and respect inspire the social ideal to which his philosophical work has sought to give systematic expression, an image of society that the early essay "Justice as Fairness" evoked as the "mutual acknowledgement of principles by free persons" (*CP*, p. 59). The essential question is not so much the total good achieved as the relations in which people stand to one another as members of a collective undertaking. To borrow a phrase from the German Idealist tradition, we may say that for Rawls the just society is, first and foremost, a matter of "mutual recognition."

Public reason is the practice in which citizens make this vision a reality. Though an implicit theme already in *A Theory of Justice*, the idea of public reason has assumed its true dimensions only in the "political liberalism" Rawls has fashioned in recent years. The prominence it now enjoys is the fitting culmination of a philosophy devoted to exploring the meaning of fairness for political life.

ENDNOTES

- 1 References to Rawls's writings are generally given in the text and in accord with the following abbreviations: *A Theory of Justice* (=TJ) (Cambridge, MA: Harvard University Press, 1971; revised edition, 1999); *Political Liberalism* (=PL) (New York: Columbia University Press, 2nd paperback edition, 1996); *Collected Papers* (=CP) (Cambridge, MA: Harvard University Press, 1999).
- 2 For a theory of deliberative democracy building on Rawls's work, see Joshua Cohen, "Democracy and Liberty," in Jon Elster (ed.), *Deliberative Democracy* (Cambridge, UK: Cambridge University Press, 1998), pp. 185–231.

- 3 Ronald Dworkin, *Taking Rights Seriously* (Cambridge, MA: Harvard University Press, 1978), Chapter 6 ("Justice and Rights," originally published 1973). For Rawls's reply, see *CP*, pp. 400–401n.
- 4 See Rawls, "Justice as Fairness," *CP*, pp. 59 and 70, and contrast the less perspicuous, contractarian presentation of the idea of fairness in *TJ*, p. 11.
- 5 This material is taken up again in *PL*, pp. 66 ff.
- 6 The distinction goes back to *A Theory of Justice* (*TJ*, pp. 580–1/508 rev.). It also reappears in "The Idea of Public Reason Revisited" (1997), *CP*, p. 594.
- 7 See "Justice as Fairness: Political not Metaphysical" (1985), *CP*, p. 390 and the two subsequent essays which explore the notion in detail, "The Idea of an Overlapping Consensus" (1987; *CP*, pp. 421–48) and "The Domain of the Political and Overlapping Consensus" (1989; *CP*, pp. 473–96) reworked to form Chapter IV of *Political Liberalism*. The term with a somewhat different sense appears in *TJ*, p. 388/340 rev., whereas the concept itself is at work at *TJ*, pp. 220–1/193–4 rev.
- 8 "Pluralism" can be a misleading term in this context if it suggests the sort of ethical conception made famous by Isaiah Berlin. Berlin's pluralism is a positive doctrine according to which there are many ultimate, irreducible, and sometimes incompatible ends of life. The pluralism Rawls has in mind might be better described as the existence of reasonable disagreement about the nature of the human good (Berlin's value-pluralism being one of the views in dispute). For more on this distinction, see my book *The Morals of Modernity* (Cambridge, UK: Cambridge University Press, 1996), Chapter 7 ("Pluralism and Reasonable Disagreement").
- 9 "Justice as Fairness: Political not Metaphysical" (1985), *CP*, p. 388. See also *PL*, p. 10.
- 10 To this extent, Rawls's thinking moves in the direction of the all-embracing view of deliberative democracy defended in Amy Gutmann and Dennis Thompson, *Democracy and Disagreement* (Cambridge, MA: Harvard University Press, 1996), pp. 34–49.
- 11 Quebec is the model for the less "procedural," more "communitarian" form of liberalism favored by Charles Taylor, as in "The Politics of Recognition," pp. 242–8 in *Philosophical Arguments* (Cambridge, MA: Harvard University Press, 1995). It is not obviously at odds with Rawls's doctrine of public reason.
- 12 Jürgen Habermas, *Strukturwandel der Öffentlichkeit* (Darmstadt: Luchterhand, 1962). Rawls notes this terminological difference with Habermas at *PL*, pp. 1, 382.

- 13 Two examples are Michael Sandel in his review of *Political Liberalism* (*Harvard Law Review* 107 (7) (May 1994): 1765–94, particularly at 1789 ff.) and Nicholas Wolterstorff in “Why We Should Reject What Liberalism Tells Us about Speaking and Acting in Public for Religious Reasons,” in Paul Weithman (ed.), *Religion and Contemporary Liberalism* (South Bend, IN: University of Notre Dame Press, 1997), pp. 162–81.
- 14 For reflections along these lines, see Jeremy Waldron, “Religious Contributions in Public Deliberation,” *San Diego Law Review* 30 (4) (fall 1993): 817–48.
- 15 I pursue this subject in “The Moral Basis of Political Liberalism,” *Journal of Philosophy* 96 (12) (December 1999): 1–27.
- 16 Notably, Michael Sandel, in his review of *Political Liberalism*, *op. cit.*, pp. 1779–82; see also his *Democracy’s Discontent* (Cambridge, MA: Harvard University Press, 1996), pp. 21–3. Rawls points out the unfoundedness of this charge in “The Idea of Public Reason Revisited” (*CP*, pp. 609–10).
- 17 This is one of the themes in the “judicial minimalism” of Cass Sunstein, *One Case at a Time* (Cambridge, MA: Harvard University Press, 1999). For Rawls’s sympathy with Sunstein’s approach, see *CP*, p. 618.
- 18 For the contrary view, see Kent Greenawalt, *Religious Convictions and Political Choice* (New York: Oxford University Press, 1988) and *Private Consciences and Public Reasons* (New York: Oxford University Press, 1995).
- 19 Here Rawls alludes to the “principles of accommodation” advocated by Gutmann and Thompson. See their *Democracy and Disagreement*, pp. 79–91.