Reconciliation Through the Public use of Reason: Remarks on John Rawls's Political Liberalism

Jurgen Habermas


Stable URL: http://links.jstor.org/sici?sici=0022-362X%28199503%2992%3A3%3C109%3ARTPUO%3E2.0.CO%3B2-P

*The Journal of Philosophy* is currently published by Journal of Philosophy, Inc..
RECONCILIATION THROUGH THE PUBLIC USE OF REASON: REMARKS ON JOHN RAWLS'S POLITICAL LIBERALISM*

John Rawls's *A Theory of Justice* marks a pivotal turning point in the most recent history of practical philosophy, for he restored long-suppressed moral questions to the status of serious objects of philosophical investigation. Immanuel Kant posed the fundamental question of morality in such a way that it admitted a rational answer: we ought to do what is equally good for all persons. Without espousing Kant's transcendental philosophical background assumptions, Rawls renewed this theoretical approach with particular reference to the issue of the organization of a just society. In opposition to utilitarianism and value skepticism he proposed an intersubjectivist version of Kant's principle of autonomy: we act autonomously when we obey those laws which could be accepted by all concerned on the basis of a public use of their reason. More recently, in *Political Liberalism*, in which Rawls has concluded a twenty-year process of extension and revision of his theory of justice, he exploits this moral concept of autonomy as the key to explaining the political autonomy of citizens of a democratic society: “Our exercise of political power is fully proper only when it is exercised in accordance with a constitution, the essentials of which all citizens as free and equal may be reasonably expected to endorse in the light of principles and ideals acceptable to their common human reason” (PL 137). Just as previously he took a stand against utilitarian positions, now he responds

* [This historic interchange between Jürgen Habermas and John Rawls is the result of an initiative by Michael Kelly, Managing Editor of this JOURNAL. The Trustees and Editors wish to acknowledge their appreciation for his enterprise in this matter and his dedication in carrying it through to fulfillment.]

1 Cambridge: Harvard, 1971; hereafter TJ.

2 New York: Columbia, 1993; hereafter PL.
primarily to contextualist positions that question the presuppositions of a reason common to all humans.

Because I admire this project, share its intentions, and regard its essential results as correct, the dissent I express here will remain within the bounds of a familial dispute. My doubts are limited to whether Rawls always brings to bear against his critics his important normative intuitions in their most compelling form. But first let me review briefly the outline of his project in its current state.

Rawls offers a justification of those principles on which a modern society must be constituted if it is to ensure the fair cooperation of its citizens as free and equal persons. His first step is to clarify the standpoint from which fictional representatives of the people could answer this question impartially. Rawls explains why the parties in the so-called original position would agree on two principles: first, on the liberal principle according to which everyone is entitled to an equal system of basic liberties, and, second, on a subordinate principle that establishes equal access to public offices and stipulates that social inequalities are acceptable only when they are also to the advantage of the least privileged. In a second step, Rawls shows that this conception of justice can expect to meet with agreement under those conditions of a pluralistic society which it itself promotes. Political liberalism, as a reasonable construction that does not raise a claim to truth, is neutral toward conflicting worldviews. In a third and final step, Rawls outlines the basic rights and principles of the constitutional state that can be derived from the two principles of justice. Taking these steps in sequence, I shall raise objections directed not so much against the project as such but against certain aspects of its execution. I fear that Rawls makes concessions to opposed philosophical positions which impair the cogency of his own project.

My critique is a constructive and immanent one. First, I doubt whether every aspect of the original position is designed to clarify and secure the standpoint of impartial judgment of deontological principles of justice (I). Further, I think that Rawls should make a sharper separation between questions of justification and questions of acceptance; he seems to want to purchase the neutrality of his conception of justice at the cost of forsaking its cognitive validity claim (II). These two theoretical decisions result in a construction of the constitutional state that accords liberal basic rights primacy over the democratic principle of legitimation. Rawls thereby fails to achieve his goal of bringing the liberties of the moderns into harmony with the liberties of the ancients (III). I conclude my remarks with a thesis on the self-understanding of political philosophy: under
conditions of postmetaphysical thought, this should be modest, but not in the wrong way.

My role in the present exchange with Rawls leads me to heighten tentative reservations into objections. This intensification is justified by my intention, at once friendly and provocative, of setting the not easily surveyable arguments of a highly complex and well thought-out theory in motion in such a way that the latter can reveal its strengths.5

I. THE DESIGN OF THE ORIGINAL POSITION
Rawls conceives of the original position as a situation in which rationally choosing representatives of the citizens are subject to the specific constraints that guarantee an impartial judgment of practical questions. The concept of full autonomy is reserved for the citizens who already live under the institutions of a well-ordered society. For the construction of the original position, Rawls splits this concept of political autonomy into two elements: the morally neutral characteristics of parties who seek their rational advantage, on the one hand, and the morally substantive situational constraints under which those parties choose principles for a system of fair cooperation, on the other. These normative constraints permit the parties to be endowed with a minimum of properties, in particular, "the capacity for a conception of the good (and thus to be rational)."6 Regardless of whether the parties entertain exclusively purposive-rational considerations or also address ethical questions of particular plans of life, they always reach their decisions in light of their value orientations (that is, from the perspective of the groups of citizens they represent). They need not regard matters from the moral point of view which would require them to take account of what is in the equal interest of all, for this impartiality is exacted by a situation that throws a veil of ignorance over the mutually disinterested, though free and equal parties. Because the latter do not know which positions they will occupy in the society that it is their task to order, they find themselves constrained already by their self-interest to reflect on what is equally good for all.

This construction of an original position that frames the freedom of choice of rational actors in a reasonable fashion is explained by Rawls's initial intention of representing the theory of justice as part of the general theory of choice. Rawls originally proceeded on the

5 In preparing the manuscript, the following works were especially helpful: Kenneth Baynes, *The Normative Grounds of Social Criticism* (Albany: SUNY, 1992); Rainer Forst, *Kontexte der Gerechtigkeit* (Frankfurt am Main: Suhrkamp, 1994).
assumption that the range of options open to rationally choosing parties only needed to be limited in an appropriate fashion in order to facilitate the derivation of principles of justice from their enlightened self-interest. But he soon realized that the reason of autonomous citizens cannot be reduced to rational choice conditioned by subjective preferences. Yet even after the revision of the initial goal that the original position was designed to achieve, he has held to the view that the meaning of the moral point of view can be operationalized in this way. This has some unfortunate consequences, three of which I would like to address in what follows: (1) Can the parties in the original position comprehend the highest-order interests of their clients solely on the basis of rational egoism? (2) Can basic rights be assimilated to primary goods? (3) Does the veil of ignorance guarantee the impartiality of judgment?

(1) Rawls cannot consistently stand by the decision that “fully” autonomous citizens are to be represented by parties who lack this autonomy. Citizens are assumed to be moral persons who possess a sense of justice and the capacity for their own conception of the good, as well as an interest in cultivating these dispositions in a rational manner. But in the case of the parties in the original position, these reasonable characteristics of moral persons are replaced by the constraints of a rational design. At the same time, however, the parties are supposed to be able to understand and take adequate account of the “highest-order interests” of the citizens that follow from these very characteristics. For example, they must take account of the fact that autonomous citizens respect the interests of others on the basis of just principles and not only from self-interest, that they can be obligated to loyalty, that they want to be convinced of the legitimacy of existing arrangements and policies through the public use of their reason, and so forth. Thus, the parties are supposed both to understand and to take seriously the implications and consequences of an autonomy that they themselves are denied. This may still be plausible for the advocacy of self-related interests and conceptions of the good that are not known in detail. But can the meaning of considerations of justice remain unaffected by the perspective of rational egoists? At any rate, the parties are incapable of achiev-

---


6 Thomas Scanlon also criticizes the traces of a decision-theoretical orientation, though from a different standpoint, in "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., Utilitarianism and Beyond (New York: Cambridge, 1982), pp. 123ff.
ing, within the bounds set by their rational egoism, the reciprocal perspective taking that the citizens they represent must undertake when they orient themselves in a just manner to what is equally good for all: "in their rational deliberations the parties...recognize no standpoint external to their own point of view as rational representatives" (PL 73). But if, despite this, the parties are to understand the meaning of the deontological principles they are seeking and to take sufficient account of their clients’ interests in justice, they must be equipped with cognitive competences that extend further than the capacities sufficient for rationally choosing actors who are blind to issues of justice.

Of course, it is open to Rawls to modify the design of the original position accordingly. Already in A Theory of Justice, he qualified the rationality of the contracting partners in various ways. On the one hand, they take no interest in one another, conducting themselves like players who “strive for as high an absolute score as possible” (TJ 144). On the other hand, they are equipped with a “purely formal” sense of justice, for they are supposed to know that they will conform to whatever principles are agreed upon in their future role as citizens living in a well-ordered society (TJ 145). This can be understood to mean that the parties in the original position are at least cognizant of the kind of binding mutuality that will characterize the life of their clients in the future, although they themselves must for the present conduct their negotiations under different premises. Such stipulations are perfectly admissible. My only question is whether, in being extended in this direction, the design loses its point by becoming too far removed from the original model. For as soon as the parties step outside the boundaries of their rational egoism and assume even a distant likeness to moral persons, the division of labor between the rationality of choice of subjects and appropriate objective constraints is destroyed, a division through which self-interested agents are nonetheless supposed to achieve morally sound decisions. This consequence may not have a great impact on the rest of the project; but it draws attention to the conceptual constraints imposed by the original (though in the meantime abandoned) intention to provide a decision-theoretical solution to Thomas Hobbes’s problem. For another consequence of the rational-choice format of the original position is the introduction of basic goods, and this determination is important for the further development of the theory.

(2) For rationally choosing actors bound to the first-person perspective, normative issues of whatever kind can be represented solely
in terms of interests or values that are satisfied by goods. Goods are what we strive for—indeed, what is good for us. Correspondingly, Rawls introduces “primary goods” as generalized means that people may need in order to realize their plans of life. Although the parties know that some of these primary goods assume the form of rights for citizens of a well-ordered society, in the original position they themselves can only describe rights as one category of “goods” among others. For them, the issue of principles of justice can only arise in the guise of the question of the just distribution of primary goods. Rawls thereby adopts a concept of justice that is proper to an ethics of the good, one that is more consistent with Aristotelian or utilitarian approaches than with a theory of rights, such as his own, that proceeds from the concept of autonomy. Precisely because Rawls adheres to a conception of justice on which the autonomy of citizens is constituted through rights, the paradigm of distribution generates difficulties for him. Rights can be “enjoyed” only by being exercised. They cannot be assimilated to distributive goods without forfeiting their deontological meaning. An equal distribution of rights results only if those who enjoy rights recognize one another as free and equal. Of course, there exist rights to a fair share of goods or opportunities, but rights in the first instance regulate relations between actors: they cannot be “possessed” like things. If I am correct, the conceptual constraints of the model of rational choice preclude Rawls from construing basic liberties from the outset as basic rights and compel him to interpret them as primary goods. This leads him to assimilate the deontological meaning of obligatory norms to the teleological meaning of preferred values. Rawls thereby blurs certain distinctions that I shall briefly mention in order to show how this limits his options in the further development of his project.

Norms inform decisions as to what one ought to do, values inform decisions as to what conduct is most desirable. Recognized norms impose equal and exceptionless obligations on their addressees, while values express the preferability of goods that are striven for by particular groups. Whereas norms are observed in the sense of a fulfillment of generalized behavioral expectations, values or goods can be realized or acquired only by purposive action. Furthermore, norms raise a binary validity claim in virtue of which they are said to be either valid or invalid: to ought statements, as to assertoric state-

---


8 This objection is not based on the thesis of the primacy of duties over rights, as in Onora O'Neill, *Constructions of Reason* (New York: Cambridge, 1989), ch. 12, pp. 206ff.
ments, we can respond only with 'yes' or 'no'—or refrain from judgment. Values, by contrast, fix relations of preference that signify that certain goods are more attractive than others: hence, we can assent to evaluative statements to a greater or lesser degree. The obligatory force of norms has the absolute meaning of an unconditional and universal duty: what one ought to do is what is equally good for all (that is, for all addressees). The attractiveness of values reflects an evaluation and a transitive ordering of goods that has become established in particular cultures or has been adopted by particular groups: important evaluative decisions or higher-order preferences express what is good for us (or for me), all things considered. Finally, different norms must not contradict each other when they claim validity for the same domain of addressees; they must stand in coherent relations to one another—in other words, they must constitute a system. Different values, by contrast, compete for priority; insofar as they meet with intersubjective recognition within a culture or group, they constitute shifting configurations fraught with tension. To sum up, norms differ from values, first, in their relation to rule-governed as opposed to purposive action; second, in a binary as opposed to a gradual coding of the respective validity claims; third, in their absolute as opposed to relative bindingness; and, last, in the criteria that systems of norms as opposed to systems of values must satisfy.

Nevertheless, Rawls wishes to do justice to the deontological intuition that finds expression in these distinctions; hence, he must compensate for the leveling of the deontological dimension which he—as a consequence of the design of the original position—initially accepts with the concept of primary goods. So he accords the first principle priority over the second. An absolute priority of equal liberties over the primary goods regulated by the second principle is, however, difficult to justify from the first-person perspective in which we orient ourselves to our own interests or values. H.L.A. Hart has developed this point clearly in his critique of Rawls. Interestingly, Rawls can meet this criticism only by building a subsequent qualification into the primary goods which secures them a relation to basic liberties as basic rights: he acknowledges as primary goods only those which are expedient for the life plans and the development of the moral faculties of citizens as free and equal persons. Furthermore, Rawls differentiates the primary goods that

---


are constitutive of the institutional framework of the well-ordered society in the moral sense from the remainder of the primary goods by incorporating the guarantee of the "fair value" of liberty into the first principle. 

This additional determination, however, tacitly presupposes a deontological distinction between rights and goods which contradicts the prima facie classification of rights as goods. Since the fair value of equal liberties requires the actual availability of equal opportunities to exercise these rights, only rights, not goods, can be qualified in this manner. Only in the case of rights can we distinguish between legal competence and the actual opportunities to choose and to act. Only between rights, on the one side, and actual chances to exercise rights, on the other, can there exist a chasm that is problematic from the perspective of justice; such a rupture cannot exist between the possession and enjoyment of goods. It would be either redundant or meaningless to speak of the "fair value" of equally distributed goods. The distinction between legal and factual equality has no application to "goods" for grammatical reasons, to put it in Wittgensteinian terms. But if the notion of primary goods is subject to correction in a second step, we may ask whether the first step—the design of the original position that necessitates this conception—was a wise one.

(3) The foregoing reflections show that, for the parties in the original position, the capacity to make rational decisions is not sufficient to comprehend the highest-order interests of their clients or to understand rights (in Ronald Dworkin's sense) as trumps that override collective goals. But why then are the parties deprived of practical reason in the first place and shrouded in an impenetrable veil of ignorance? Rawls's guiding intuition is clear: the role of the categorical imperative is taken over by an intersubjectively applied procedure which is embodied in participation conditions, such as the equality of parties, and in situational features, such as the veil of ignorance. In my view, however, the potential gains of this turn are dissipated precisely by the systematic deprivation of information. This third question reveals the perspective from which I also posed the two previous questions: I believe that Rawls could avoid the difficulties associated with the design of an original position if he operationalized the moral point of view in a different way, namely, if he kept the procedural conception of practical reason free of substantive connotations by developing it in a strictly procedural manner.

---


Kant's categorical imperative already goes beyond the egocentric character of the Golden Rule: “Do not do unto others what you would not have them do unto you.” Whereas this rule calls for a universalization test from the viewpoint of a given individual, the categorical imperative requires that all those possibly affected be able to will a just maxim as a general rule. But as long as we apply this more exacting test in a monological fashion, it still remains individually isolated perspectives from which each of us considers privately what all could will. This is inadequate. For only when the self-understanding of each individual reflects a transcendental consciousness, that is, a universally valid view of the world, would what from my point of view is equally good for all actually be in the equal interest of each individual. But this can no longer be assumed under conditions of social and ideological pluralism. If we wish to preserve the intuition underlying the Kantian universalization principle, we can respond to this fact of pluralism in different ways. Rawls imposes a common perspective on the parties in the original position through informational constraints and thereby neutralizes the multiplicity of particular interpretive perspectives from the outset. Discourse ethics, by contrast, views the moral point of view as embodied in an intersubjective practice of argumentation which enjoins those involved to an idealizing enlargement of their interpretive perspectives.

Discourse ethics rests on the intuition that the application of the principle of universalization, properly understood, calls for a joint process of “ideal role taking.” It interprets this idea of G. H. Mead in terms of a pragmatic theory of argumentation. Under the pragmatic presuppositions of an inclusive and noncoercive rational discourse among free and equal participants, everyone is required to take the perspective of everyone else, and thus project herself into the understandings of self and world of all others; from this interlocking of perspectives there emerges an ideally extended we-perspective from which all can test in common whether they wish to make a controversial norm the basis of their shared practice; and this should include mutual criticism of the appropriateness of the languages in terms of which situations and needs are interpreted. In the course of

---

successively undertaken abstractions, the core of generalizable interests can then emerge step by step.\textsuperscript{14}

Things are different when the veil of ignorance constrains \textit{from the beginning} the field of vision of parties in the original position to the basic principles on which presumptively free and equal citizens would agree, notwithstanding their divergent understandings of self and world. It is important to see that with this initial abstraction Rawls accepts a \textit{double} burden of proof. The veil of ignorance must extend to all particular viewpoints and interests that could impair an impartial judgment; at the same time, it may extend \textit{only} to such normative matters as can be disqualified without further ado as candidates for the common good to be accepted by free and equal citizens. This second condition makes a demand on the theory that is difficult to meet, as is shown by a brief reflection. Following the justification of the principles of justice, the veil of ignorance is gradually raised at the successive steps of framing the constitution, of legislation, and of applying law. Since the new information that thereby flows in must harmonize with the basic principles already selected under conditions of informational constraint, unpleasant surprises must be avoided. If we are to ensure that no discrepancies arise, we must construct the original position already with knowledge, and even foresight, of all the normative contents that could potentially nourish the shared self-understanding of free and equal citizens in the future. In other words, the theoretician himself would have to shoulder the burden of anticipating at least parts of the information of which he previously relieved the parties in the original position! The impartiality of judgment would be guaranteed in the original position only if the basic normative concepts employed in its construction—those of the politically autonomous citizen, of fair cooperation, and of a well-ordered society, in the specific sense Rawls attaches to these terms—could withstand revision in light of morally significant future experiences and learning processes.

If such a heavy burden of proof is generated by the deprivation of information imposed on the parties in the original position by the veil of ignorance, a convenient response would be to lighten this burden by operationalizing the moral point of view in a different way. I have in mind the more open procedure of an argumentative practice that proceeds under the demanding presuppositions of the "public use of reason" and does not bracket the pluralism of convic-

\textsuperscript{14} Cf. William Rehg, \textit{Insight and Solidarity: The Discourse Ethics of Jürgen Habermas} (Berkeley: California UP, 1993).
tions and worldviews from the outset. This procedure can be ex-
plained without having recourse to the substantive concepts that Rawls
employs in the construction of the original position.

II. THE FACT OF PLURALISM & THE IDEA OF AN OVERLAPPING CONSENSUS

Since his Dewey Lectures, "Kantian Constructivism in Moral Theory," Rawls has stressed the political character of justice as fairness. This
shift is motivated by disquiet concerning the fact of social and, above all, ideological pluralism. In discussing the "veil of ignorance," I have
already clarified the burden of proof that the theory of justice takes
upon itself with its initial theoretical decisions. The decisive issue in
the justification of the two highest principles of justice is less the
deliberations in the original position than the intuitions and basic
concepts that guide the design of the original position itself. Rawls
introduces normative contents into the very procedure of justifica-
tion, above all those ideas he associates with the concept of the
moral person: the sense of fairness and the capacity for one's own
conception of the good. Thus, the concept of the citizen as a moral
person, which also underlies the concept of the fair cooperation of
politically autonomous citizens, stands in need of a prior justifi-
cation. Further, it needs to be shown that this conception is neutral toward
conflicting worldviews and remains uncontroversial after the veil of
ignorance has been lifted. This explains Rawls's interest in a political,
as opposed to a metaphysical, conception of justice. I suspect that this
terminology indicates a certain unclarity about the precise character
of what is in need of justification; from this, in turn, there results an
indecisiveness as to how the validity claim of the theory itself should
be understood. I shall examine whether the overlapping consensus,
on which the theory of justice depends, plays a cognitive or merely
instrumental role: whether it primarily contributes to the further jus-
tification of the theory or whether it serves, in light of the prior justi-
fication of the theory, to explicate a necessary condition of social
stability (1). Connected with this is the question of the sense in
which Rawls uses the predicate 'reasonable': as a predicate for the
validity of moral judgments or for the reflective attitude of enlight-
ened tolerance (2).

(1) In order to pin down the underlying normative ideas, Rawls
has recourse to the so-called method of reflective equilibrium. The
philosopher arrives at the basic concept of the moral person and the
adjunct concepts of the politically autonomous citizen, of fair coop-
eration, of the well-ordered society, and so forth, via a rational

15This JOURNAL, LXXVII, 9 (September 1980): 515–72.
reconstruction of proven intuitions, that is, intuitions actually found in the practices and traditions of a democratic society. Reflective equilibrium is achieved at the moment when the philosopher has attained the assurance that those involved can no longer reject with good reasons intuitions reconstructed and clarified in this manner. The procedure of rational reconstruction already fulfills Thomas Scanlon's criterion of what it is "not reasonable to reject" (op. cit.). Of course, Rawls does not wish to limit himself solely to the fundamental normative convictions of a particular political culture: even the present-day Rawls, pace Richard Rorty, has not become a contextualist. His aim, as before, is to reconstruct a substratum of intuitive ideas latent in the political culture of his society and its democratic traditions. But if experiences associated with an incipiently successful institutionalization of principles of justice have already become sedimented in the existing political culture—in the American, for example—such a reconstructive appropriation can accomplish more than merely the hermeneutic clarification of a contingent tradition. The concept of justice worked out on this basis must nonetheless be examined once again as to whether it can expect to meet with acceptance in a pluralistic society. How is this second step related to the first stage of justification of the two highest principles? Is it even properly a second step of justification?

Already in the final chapters of A Theory of Justice, Rawls addressed the issue of whether a society constituted in accordance with the principles of justice could stabilize itself: for example, whether it could generate the functionally necessary motivations from its own resources through the requisite political socialization of its citizens (TJ 496ff.). In view of the fact of social and ideological pluralism which he subsequently took more seriously, Rawls now wants to examine in a similar way whether the theoretical conception of justice falls under the "art of the possible" and hence is "practicable." First of all, the central concept of the person, on which the theory ultimately rests, must be sufficiently neutral to be acceptable from the interpretive perspectives of different worldviews. Hence, it must be shown that justice as fairness can form the basis of an "overlapping consensus." So far, so good. What bothers me is Rawls's working assumption that such a test of acceptability is of the same kind as the test of consistency he previously undertook with reference to the well-ordered society's potential for self-stabilization.

This methodological parallel is problematic because the test cannot be undertaken in an immanent manner in the case of acceptability; it is no longer a move within the theory. The test of the neutrality of the basic normative concepts with respect to conflicting worldviews now rests on different premises: it is different from a hypothetical examination of the capacity of a society already organized in accordance with principles of justice to reproduce itself. Rawls himself distinguishes in his present work between "two stages" of theory formation. The principles justified at the first stage must be exposed to public discussion at the second stage. Only when the theoretical design is completed can the fact of pluralism be brought into play and the abstractions of the original position revoked. The theory as a whole must be subjected to criticism by citizens in the forum of public reason. But this now refers not to the fictional citizens of a just society about whom statements are made within the theory but to real citizens of flesh and blood. The theory, therefore, must leave the outcome of such a test of acceptability undetermined. For Rawls has in mind real discourses whose outcome is open: "What if it turns out that the principles of justice as fairness cannot gain the support of reasonable doctrines, so that the case for stability fails?... We should have to see whether acceptable changes in the principles of justice would achieve stability" (PL 65–66). Clearly, the philosopher can at most attempt to anticipate in reflection the direction of real discourses as they would probably unfold under conditions of a pluralistic society. But such a more or less realistic simulation of real discourses cannot be incorporated into the theory in the same way as the derivation of possibilities of self-stabilization from the underlying premises of a just society. For now the citizens themselves debate about the premises developed by the parties in the original position.

The misleading parallel would be of no further consequence if it did not cast in the wrong light the "overlapping consensus" with which the principles of justice are supposed to converge. Because Rawls situates the "question of stability" in the foreground, the overlapping consensus merely expresses the functional contribution that the theory of justice can make to the peaceful institutionalization of social cooperation; but in this the intrinsic value of a justified theory must already be presupposed. From this functionalist perspective, the question of whether the theory can meet with public agreement—that is, from the perspective of different worldviews in the forum of the public use of reason—would lose an epistemic meaning essential to the theory itself. The overlapping consensus would then be merely an index of the utility, and no longer a confirmation of
the correctness of the theory; it would no longer be of interest from the point of view of acceptability, and hence of validity, but only from that of acceptance, that is, of securing social stability. If I understand Rawls correctly, however, he does not wish to distinguish in this way between questions of justification and questions of stability. When he calls his conception of justice political, his intention appears to be rather to collapse the distinction between its justified acceptability and its actual acceptance: "the aim of justice as fairness as a political conception is practical, and not metaphysical or epistemological. That is, it presents itself not as a conception of justice that is true, but one that can serve as a basis of informed and willing political agreement between citizens viewed as free and equal persons."17

In my view, Rawls must make a sharper distinction between acceptability and acceptance. A purely instrumental understanding of the theory is already invalidated by the fact that the citizens must first be convinced by the proposed conception of justice before such a consensus can come about. The conception of justice must not be political in the wrong sense and should not merely lead to a modus vivendi. The theory itself must furnish the premises "that we and others recognize as true, or as reasonable for the purpose of reaching a working agreement on the fundamentals of political justice."18 But if Rawls rules out a functionalist interpretation of justice as fairness, he must allow some epistemic relation between the validity of his theory and the prospect of its neutrality toward competing worldviews being confirmed in public discourses. The stabilizing effect of an overlapping consensus would then be explained in cognitive terms, that is, in terms of the confirmation of the assumption that justice as fairness is neutral toward "comprehensive doctrines." I do not mean to say that Rawls accepts premises that would prevent him from drawing this consequence; I mean only that he hesitates to assert it because he associates with the characterization 'political' the proviso that the theory of justice should not be burdened with an epistemic claim and that its anticipated practical effect should not be made contingent on the rational acceptability of its assertions. Thus, we have reason to ask why Rawls does not think his theory admits of truth and in what sense he here uses the predicate 'reasonable' in place of the predicate 'true'.

(2) On a weak interpretation, the claim that a theory of justice cannot be true or false has merely the unproblematic sense that normative statements do not describe an independent order of moral facts. On a strong interpretation, this thesis has the value-skeptical sense that behind the validity claim of normative statements there lurks something purely subjective: feelings, wishes, or decisions expressed in a grammatically misleading fashion. But for Rawls, both moral realism and value skepticism are equally unacceptable. He wants to secure for normative statements—and for the theory of justice as a whole—a form of rational obligatoriness founded on justified intersubjective recognition, but without according them an epistemic meaning. For this reason, he introduces the predicate ‘reasonable’ as a complementary concept to ‘true’. The difficulty here is in specifying in what sense the one is a “complementary concept” to the other. Two alternative interpretations suggest themselves. Either we understand ‘reasonable’ in the sense of practical reason as synonymous with ‘morally true’, that is, as a validity concept analogous to truth and on the same plane as propositional truth; this reading is supported by at least one line of argumentation (a). Or we understand ‘reasonable’ in more or less the same sense as ‘thoughtfulness’ in dealing with debatable views whose truth is for the present undecided; then ‘reasonable’ is employed as a higher-level predicate concerned more with “reasonable disagreements,” and hence with the fallibilistic consciousness and civil demeanor of persons, than with the validity of their assertions. In general, Rawls seems to favor this latter reading (b).

(a) Rawls first introduces the ‘reasonable’ as a property of moral persons. People count as reasonable who possess a sense of justice and thus are both willing and able to take account of fair conditions of cooperation, but who are also aware of the fallibility of knowledge and—in recognition of these “burdens of reason”—are willing to justify their conception of political justice publicly. By contrast, persons act merely “rationally” as long as they are prudently guided by their conception of the good. What it means to be reasonable can indeed be explicated in terms of such qualities of moral persons. But the concept of a person itself already presupposes the concept of practical reason.

Ultimately, Rawls explains the meaning of practical reason by reference to two dimensions: on the one hand, the deontological dimension of normative validity (which I here leave to one side as
unproblematic) and, on the other, the pragmatic dimension of a public sphere and the process of public reasoning (which is of particular interest in the present context). The public use is in a sense inscribed in reason. *Publicity* is the common perspective from which the citizens *mutually* convince one another of what is just and unjust by the force of the better argument. This perspective of the public use of reason, in which all participate, first lends moral convictions their objectivity. Rawls calls valid normative statements *objective* and he explains "objectivity" in a procedural manner with reference to a public use of reason that satisfies certain counterfactual conditions: "Political convictions (which are also, of course, moral convictions) are objective—actually founded on an order of reasons—if reasonable and rational persons, who are sufficiently intelligent and conscientious in exercising their powers of practical reason...would eventually endorse those convictions...provided that these persons know the relevant facts and have sufficiently surveyed the grounds that bear on the matter under conditions favorable to due reflection" (PL 119). Rawls does add in this passage that grounds are only specified as good grounds in the light of a recognized concept of justice; but this concept must in turn meet with agreement under the same ideal conditions (PL 137). Hence, Rawls must be understood to mean that, on his view as well, the procedure of the public use of reason remains the final court of appeal for normative statements.

In the light of this reflection, it could be said that the predicate 'reasonable' points to the discursive redemption of a validity claim. By analogy with a nonsemantic concept of truth purified of all connotations of correspondence, one could understand 'reasonable'—correctly, in my view—as a predicate for the validity of normative statements.\(^9\) This is clearly not how Rawls himself uses this term. Otherwise, he would not say that worldviews need not be true even when they are reasonable—and vice versa. For me, the problem is not Rawls's rejection of moral realism or the consequent rejection of a semantic truth predicate for normative statements but the fact that he does attach such a truth predicate to worldviews (comprehensive doctrines). He thereby precludes the possibility of exploiting the epistemic connotations of the term 'reasonable', connotations which he must nevertheless attribute to his own conception of justice if this is to lay claim to some sort of normative binding force.

(b) On Rawls's conception, metaphysical doctrines and religious world-interpretations admit of truth and falsity. As a consequence, a

\(^9\)Cf. my reflections in *Justification and Application*, pp. 25ff.
political conception of justice could only be true if it were not merely compatible with such doctrines but also derivable from a true doctrine. Yet from the point of view of a political philosophy that is neutral toward worldviews, we cannot determine whether and when this is the case. From this secular viewpoint, the truth claims of all reasonable worldviews have equal weight, where those worldviews count as reasonable which compete with one another in a reflexive attitude, that is, on the assumption that one's own truth claim could prevail in public discourse in the long run only through the force of better reasons. "Reasonable comprehensive doctrines" are ultimately distinguished by their recognition of the burdens of proof, which enables groups with competing ideologies to accept—for the time being—a "reasonable disagreement" as the basis of their peaceful coexistence.

Since disputes concerning metaphysical and religious truths remain unresolved under conditions of enduring pluralism, only the reasonableness of this kind of reflexive consciousness can be transferred as a validity predicate to a political conception of justice compatible with all reasonable doctrines. By way of this transference, a reasonable conception of justice preserves an oblique relation to a truth claim projected into the future. But it cannot be certain that one of the reasonable doctrines from which it is derivable is also the true one. A political conception of justice is reasonable in the sense that it can afford a kind of tolerance toward worldviews that are not unreasonable, in the sense advocated by Gottfried Lessing. What remains is an act of faith in reason: "reasonable faith in the real possibility of a just constitutional regime."\textsuperscript{21} This view may appeal to some of our better intuitions, but how can it be harmonized with Rawls's reasons for accepting the priority of the right over the good in the first place?

Questions of justice or moral questions admit of justifiable answers—justifiable in the sense of rational acceptability—because they are concerned with what, from an ideally expanded perspective, is in the equal interest of all. Ethical questions, by contrast, do not admit of such impartial treatment because they refer to what, from the first-person perspective, is in the long run good for me or for us—even if this is not equally good for all. Now, metaphysical or religious worldviews are at the very least permeated with answers to basic ethical questions; they articulate in an exemplary fashion collective identities and guide individual plans of life. Hence, world-

\textsuperscript{21} "The Idea of an Overlapping Consensus," p. 25.
views are measured more by the authenticity of the life styles they shape than by the truth of the statements they admit. Because such doctrines are "comprehensive" in precisely the sense that they offer interpretations of the world as a whole, they cannot merely be understood as an ordered set of statements of fact; their contents cannot be expressed completely in sentences that admit of truth and they do not form a symbolic system that can be true or false as such. So, at least, it appears under the conditions of postmetaphysical thinking in which justice as fairness is to be justified.

But then it is impossible to make the validity of a conception of justice contingent on the truth of a worldview, however reasonable it may be. Rather, under these premises it makes sense to analyze the different validity claims that we associate, respectively, with descriptive, evaluative, and normative statements (of various kinds) independently of the characteristic complex of validity claims that are obscurely fused together in religious and metaphysical interpretations of reality.23

Why does Rawls nevertheless think that identity-stabilizing worldviews admit of truth? A possible motive might be the conviction that a profane, freestanding morality is untenable, that moral convictions must be embedded in metaphysical or religious doctrines. That, at any rate, would cohere with Rawls's way of posing the problem of an overlapping consensus: he takes as his model that political institutionalization of freedom of belief and conscience which brought the religious civil wars of the modern period to an end. But could the religious conflicts have been brought to an end if the principle of tolerance and freedom of belief and conscience had not been able to appeal, with good reasons, to a moral validity independent of religion and metaphysics?

III. PRIVATE AND PUBLIC AUTONOMY

The objections I raised in the first part against the design of the original position and in the second against the assimilation of questions of validity to those of acceptance point in the same direction. By subjecting rationally choosing parties to reasonable procedural constraints, Rawls remains dependent on substantive normative assumptions; at the same time, by tailoring a universalistic theory of justice to questions of political stability through an overlapping consensus, he compromises its epistemic status. Both strategies are pursued at the cost of a strict proceduralist program. By contrast with this

21 See Forst, "Justice as Fairness: Ethical, Political, or Moral" (manuscript, 1992).
approach, Rawls could satisfy more elegantly the burdens of proof he incurs with his strong and presumptively neutral concept of the moral person if he developed his substantive concepts and assumptions out of the procedure of the public use of reason.

In my view, the moral point of view is already implicit in the socio-ontological constitution of the public practice of argumentation, comprising the complex relations of mutual recognition that participants in rational discourse "must" accept (in the sense of weak transcendental necessity). Rawls believes that a theory of justice developed in such exclusively procedural terms could not be "sufficiently structured." Since I subscribe to a division of labor between moral theory and the theory of action, I do not regard this as a serious reservation: the conceptual structuring of the contexts of interaction to which questions of political justice refer is not within the province of moral theory. Together with the content of action conflicts in need of resolution, a whole conceptual frame for normatively regulated interaction is forced upon us—a network of concepts in which persons and interpersonal relations, actors and actions, norm-conforming and deviant behavior, responsibility and autonomy, and even intersubjectively structured moral feelings all find their place. Each of these concepts deserves a prior analysis. If we then take the concept of practical reason in the procedural sense that Rawls himself intimates with his notion of the public use of reason, we could say that precisely those principles are valid which meet with uncoerced intersubjective recognition under conditions of rational discourse. It remains as a further, and primarily empirical, question whether and when such valid principles ensure political stability under conditions of pluralism. In what follows, I am interested in the consistent execution of the proceduralist program only with reference to one implication it has for the explanation of the constitutional state.

Liberals have stressed the "liberties of the moderns": liberty of belief and conscience, the protection of life, personal liberty, and property—in sum, the core of subjective private rights. Republicanism, by contrast, has defended the "liberties of the ancients": the political rights of participation and communication that make possible the citizens' exercise of self-determination. Jean-Jacques Rousseau and Kant shared the aspiration of deriving both elements from the same root, namely, from moral and political autonomy; the liberal rights may neither be merely foisted on the practice of self-determination as extrinsic constraints nor be made merely instrumental to its exercise. Rawls, too, subscribes to this intuition;
nevertheless, the two-stage character of his theory generates a priority of liberal rights which demotes the democratic process to an inferior status.

Rawls certainly proceeds from the idea of political autonomy and models it at the level of the original position: it is represented by the interplay between the rationally choosing parties and the framework conditions that guarantee impartiality of judgment. But this idea is brought to bear only selectively at the institutional level of the democratic procedure for the political will formation of free and equal citizens from which it is nonetheless borrowed. The form of political autonomy granted virtual existence in the original position, and thus on the first level of theory formation, does not fully unfold in the heart of the justly constituted society. For the higher the veil of ignorance is raised and the more Rawls's citizens themselves take on real flesh and blood, the more deeply they find themselves subject to principles and norms that have been anticipated in theory and have already become institutionalized beyond their control. In this way, the theory deprives the citizens of too many of the insights that they would have to assimilate anew in each generation. From the perspective of the theory of justice, the act of founding the democratic constitution cannot be repeated under the institutional conditions of an already constituted just society, and the process of realizing the system of basic rights cannot be assured on an ongoing basis. It is not possible for the citizens to experience this process as open and incomplete, as the shifting historical circumstances nonetheless demand. They cannot reignite the radical democratic embers of the original position in the civic life of their society, for from their perspective all of the essential discourses of legitimation have already taken place within the theory; and they find the results of the theory already sedimented in the constitution. Because the citizens cannot conceive of the constitution as a project, the public use of reason does not actually have the significance of a present exercise of political autonomy but merely promotes the nonviolent preservation of political stability. Granted, this reading does not reflect Rawls's intention in formulating his theory, but if I am correct it uncovers one of its undesired consequences. This is shown, for example, by the rigid boundary between the political and the nonpublic identities of the citizens. According to Rawls, this boundary is set by basic liberal

\[\text{ Cf. the Tanner Lectures, where he writes at the end of section VII: "The idea is to incorporate into the basic structure of society an effective political procedure which mirrors in that structure the fair representation of persons achieved by the original position" (p. 45).}\]
rights that constrain democratic self-legislation, and with it the sphere of the political, *from the beginning*, that is, prior to all political will formation.

Rawls uses the term ‘political’ in a threefold sense. Thus far, we have become acquainted with the theoretical meaning: a conception of justice is political and not metaphysical when it is neutral toward conflicting worldviews. Further, Rawls uses the term ‘political’ in the usual sense to classify matters of public interest, so that political philosophy limits itself to the justification of the institutional framework and the basic structure of society. Both meanings are ultimately combined in an interesting way in Rawls’s treatment of “political values.” The ‘political’ in this third sense constitutes a fund both of shared convictions of citizens and of perspectives for delimiting an object domain. Rawls treats the political value sphere, which is distinguished in modern societies from other cultural value spheres, as something given, almost in the manner of a neo-Kantian like Max Weber. For only with reference to political values, whatever they may be, can he split the moral person into the public identity of a citizen and the nonpublic identity of a private person shaped by her individual conception of the good. These two identifications then constitute the reference points for two domains, the one constituted by rights of political participation and communication, the other protected by basic liberal rights. The constitutional protection of the private sphere in this way enjoys priority while “the role of the political liberties is...largely instrumental in preserving the other liberties.” Thus, with reference to the political value sphere, a prepolitical domain of liberties is delimited which is withdrawn from the reach of democratic self-legislation.

But such an a priori boundary between private and public autonomy not only contradicts the republican intuition that popular sovereignty and human rights are nourished by the same root. It also conflicts with historical experience, above all with the fact that the historically shifting boundary between the private and public spheres has always been problematic from a normative point of view. In addition, the development of the welfare state shows that the boundaries between the private and public autonomy of citizens are in flux and that such differentiations must be subjected to the political will formation of the citizens if the latter are to have the opportunity to press a legal claim to the “fair value” of their liberties.

A theory of justice can take better account of this circumstance if it differentiates the "political" in accordance with the criterion of "legal regulation" (mentioned only in passing by Rawls). It is ultimately by means of positive and coercive law that the life of a political community is legitimately regulated (PL 215). The basic question then is: Which rights must free and equal persons mutually accord one another if they wish to regulate their coexistence by the legitimate means of positive and coercive law?

According to Kant's conception of legality, *coercive law* extends only to the external relations among persons and addresses the freedom of choice of subjects who are allowed to follow their own conception of the good. Hence modern law, on the one hand, constitutes the status of legal subjects in terms of actionable subjective liberties that may be exercised by each according to her own preferences. Since it must also be possible to obey a legal order for moral reasons, the status of private legal subjects is legitimately determined by the right to *equal subjective liberties*. As positive or *codified law*, on the other hand, this medium calls for a political legislator, where the legitimacy of legislation is accounted for by a democratic procedure that secures the autonomy of the citizens. Citizens are politically autonomous only if they can view themselves jointly as authors of the laws to which they are subject as individual addressees.

The dialectical relation between private and public autonomy becomes clear in light of the fact that the status of such democratic citizens equipped with law-making competences can be institutionalized in turn only by means of coercive law. But because this law is directed to persons who could not even assume the status of legal subjects without subjective private rights, the private and public autonomy of citizens mutually presuppose each other. As we have seen, both elements are already interwoven in the concept of positive and coercive law: there can be no law at all without actionable subjective liberties that guarantee the private autonomy of individual legal subjects; and no legitimate law without democratic law making by citizens in common who, as free and equal, are entitled to participate in this process. Once the concept of law has been clarified in this way, it becomes clear that the normative substance of basic liberal rights is already contained in the indispensable medium for the legal institutionalization of the public use of reason of sovereign citizens. The main objects of further analysis are then the communicative presuppositions and the procedures of a discursive process of opinion and will formation in which the public use of reason is man-

---

This principle of Kantian legal theory is taken up in Rawls's first principle.
ifested. I cannot discuss this alternative in greater detail in the present context.28

Such a procedural moral and legal theory is at the same time more and less modest than Rawls’s theory. It is more modest because it focuses exclusively on the procedural aspects of the public use of reason and derives the system of rights from the idea of its legal institutionalization. It can leave more questions open because it entrusts more to the process of rational opinion and will formation. Philosophy shoulders different theoretical burdens when, as on Rawls’s conception, it claims to elaborate the idea of a just society, while the citizens then use this idea as a platform from which to judge existing arrangements and policies. By contrast, I propose that philosophy limit itself to the clarification of the moral point of view and the procedure of democratic legitimation, to the analysis of the conditions of rational discourses and negotiations. In this more modest role, philosophy need not proceed in a constructive, but only in a reconstructive fashion. It leaves substantial questions that must be answered here and now to the more or less enlightened engagement of participants, which does not mean that philosophers may not also participate in the public debate, though in the role of intellectuals, not of experts.

Rawls insists on a modesty of a different kind. He wants to extend the “method of avoidance,” which is intended to lead to an overlapping consensus on questions of political justice, to the philosophical enterprise. He hopes to develop political philosophy into a sharply focused discipline and thereby to avoid most of the controversial questions of a more general nature. This avoidance strategy can lead to an impressively self-contained theory, as we can see from the wonderful example before us. But even Rawls cannot develop his theory in as freestanding a fashion as he would like. As we have seen, his “political constructivism” draws him willy-nilly into the dispute concerning concepts of rationality and truth. His concept of the person also oversteps the boundaries of political philosophy. These and other preliminary theoretical decisions involve him in as many long-running and still unresolved debates. The subject matter itself, it seems to me, makes a presumptuous encroachment on neighboring fields often unavoidable and at times even fruitful.

JÜRGEN HABERMAS

Johann Wolfgang Goethe-Universität
translated by Ciaran Cronin


NOTE: The reference numbering from the original has been maintained in this citation list.