

Kenneth Baynes

Constructivism and Practical Reason in Rawls

Reason must in all its undertakings subject itself to criticism; should it limit freedom of criticism by any prohibitions, it must harm itself, drawing upon itself a damaging suspicion. ... Reason depends on this freedom for its very existence. For reason has no dictatorial authority; its verdict is always simply the agreement of free citizens, of whom each one must be permitted to express, without let or hindrance, his objection or even his veto.

I. Kant, *Critique of Pure Reason* (A738/B766)

Abstract: This essay argues that Rawls's recent constructivist approach waivers between a relativist defense and a more Kantian account which grounds his conception of justice in the idea of an agreement between free and equal moral persons. It is suggested that this ambiguity lies at the center of his attempt to provide a "political not metaphysical" account which is also not "political in the wrong way".

In a series of essays since the publication of his Dewey Lectures, "Kantian Constructivism in Moral Theory", Rawls has sought to clarify certain ambiguities and misunderstandings associated with his conception of 'justice as fairness'. Yet, despite these efforts, interpreters remain deeply divided about the significance of the later essays, especially for the kind of justification he now claims to offer for his "political, not metaphysical" conception. On the one hand, there are those who emphasize the pragmatic and even relativist features of those essays. Some of these, such as Richard Rorty, regard this turn as a gain, while others, including 'moral realists', regard it as a weakness and unnecessary concession in his theory.¹ On the other hand, there are a number of commentators who dismiss this reading and see Rawls's more recent essays as offering a stronger (perhaps even mildly Kantian) defense of an egalitarian liberalism with reference to a model of practical reason and, more specifically, the (counterfactual and, hence, regulative) idea of a reasonable agreement between free and equal moral persons.²

In attempting to assess the merits of these competing interpretations, much will depend on how one understands Rawls's claim to be offering a 'constructivist' account of justice as fairness. It will also be necessary to clarify what it

¹ For a sympathetic relativist reading, see Richard Rorty 1991; for similar interpretations, but ones which criticize this turn, see David Brink 1989, Alan Gilbert 1990, Jean Hampton 1989 and Joseph Raz 1990.

² See, for example, Samuel Freeman 1990; 1991, Charles Larmore 1990, and Joshua Cohen forthcoming. Onora O'Neill in 1989 provides a useful overview of both interpretations of Rawls.

means for his account to be "political, not metaphysical" while still not being "political in the wrong way" - that is, a mere "consensus politics" in which principles of justice are accommodated to the dominant interests and power relations of the status quo (see Rawls 1989, 234). Finally, attempting to decide between these two positions will also require relating Rawls's remark that the task of justification is a practical task aimed at a 'reasonable agreement' that replaces the search for truth to other remarks he makes about the use of 'free public reason' and to the claim that the conception does not depend for its validity on a *de facto* consensus. In short, it will require an interpretation of the following summary of his justificatory strategy:

"What justifies a conception of justice is not its being true to an order antecedent to and given to us, but its congruence with our deeper understanding of ourselves and our aspirations, and our realization that, given our history and the traditions embedded in our public life, it is the most reasonable doctrine for us. We can find no better basic charter for our social world. Kantian constructivism holds that moral objectivity is to be understood in terms of a suitably constructed social point of view that all can accept." (Rawls 1975, 519)

Relativist readings emphasize the view that the justification of principles now depends on finding an 'overlapping consensus' concerning our common self-understandings, traditions, and public life; stronger readings, by contrast, draw attention to the conception of 'moral objectivity' as a 'reasonable' agreement that *all* could accept from a 'suitably constructed' point of view. My own view, which I will attempt to defend in the following essay, is that the tension that exists between these two interpretations does not simply result from a superficial juxtaposition of isolated texts, but in fact reflects a deep ambiguity in Rawls's own position that can in turn be traced back to a tension inherent in Kant's account of morality. Kant, it will be recalled, maintained that his formulation of the moral law did not contain anything that could not already be found in ordinary moral consciousness.³ At the same time, he held that the categorical imperative (in its various formulations) - as the 'supreme principle' of pure practical reason - could be derived from an analysis of the conditions of practical agency and was, as such, binding on all (finite) rational agents (see Hill 1989; Reath 1989; Allison 1986). Rawls, I want to suggest, plays on a similar ambivalence, at times claiming that his conception merely articulates or makes more perspicuous notions already present in ordinary moral experience and, at other times, offering a stronger claim that it can be grounded in a (normative) account of practical agency or the basic capacities for moral reasoning and deliberation. I will pursue this thesis by first identifying more precisely the important role performed by the model-conception of the person (as part of the model-conception of the well-ordered society) within Rawls's conception, especially as this is developed in "Kantian Constructivism". I will then consider the relation of this conception of

³ See Kant 1922, 403-4; for a critical discussion of this claim, see Paul Stern 1986.

the person to Rawls's more recent claim that his account is "political, not metaphysical". Finally, I will suggest that Rawls's constructivist account will only be attractive, that is, not "political in the wrong way", if it is developed in the direction of a reflexive or recursive justification of our basic capacities for moral reasoning or practical deliberation.⁴

In "Kantian Constructivism in Moral Theory" Rawls identifies three different points of view which are important for a proper understanding of his argument: "that of the parties in the original position; that of citizens in a well-ordered society; and, finally, that of ourselves - you and me who are examining justice as fairness as a basis for a conception of justice that may yield a suitable understanding of freedom and equality" (Rawls 1975, 533; see diagram). The first two points of view occur within the theory of justice itself: on the one hand, there is the ideal of the well-ordered society and, as part of this ideal, the model-conception of the person as a free and equal moral being; on the other hand, there is what Rawls calls the "mediating conception" of the original position: it is a "device of representation" that mediates between the ideals contained in the description of the well-ordered society and the definition of the principles of justice. Roughly stated, for its characterization we look primarily to the ideal of the well-ordered society; at the same time, it helps to make vivid to us why the ideals contained in the well-ordered society, when combined in a suitable way, yield the two principles of justice. Finally, there is the position of us who have the task of finding principles of justice for our society (Rawls 1975, 518 and 533). Rawls's differentiation among the three points of view is crucial for an understanding of his justificatory strategy. The description of the well-ordered society (together with its ideal of the person), and the characterization of the original position, as well as the specific principles of justice, must all cohere with "our considered judgments upon due reflection", that is, they must all agree with those judgments at which we would arrive as a result of carrying out the process of reflective equilibrium. The entire procedure is constructivist or non-foundationalist in that it does not accept any intuitions as indubitable and does not begin with the assumption that there are first principles that somehow exist prior to and independently of our conception of ourselves as free and equal moral persons. Rather, as David Richards has recently summarized it, "ethical principles do not track special properties in the world, but are the expression and acknowledgement of our common moral powers of rationality and reasonableness" (Richards 1988, 120). The principles of justice are the result of a process of construction in which, from 'our' point of view, various ideals or model-conceptions are articulated which are subsequently used in the construction or design of the original position. The parties in the original position, as agents of construction, then select the principles of justice from a list of alternative conceptions. Finally, each of these 'points of view' can serve as a corrective to considerations and features

⁴ I develop this idea of a reflexive or recursive justification at greater length in my 1992a.

introduced in the other; no one perspective is taken as fixed or absolute, although, at the end, the results must be acceptable to us as citizens faced with the task of finding principles of justice.

Kantian Constructivism and the Three Points of View

I. 'Ourselves'

'we who are faced with the task
of settling questions of justice'

II. The Well-ordered Society

Public Conception of Justice
Free and Equal Person with two Moral Powers:
- sense of justice
- conception of the good
'Circumstances of Justice'
Stability

III. The Original Position as a 'Device of Representation'

The Reasonable
Veil of Ignorance
Symmetry Conditions
Primary Social Goods
Formal Conditions
Basic Structure as Subject
The Rational
Parties with rational autonomy
'Maximin Rule'

This brief summary of the three points of view and its importance for Rawls's justificatory strategy underscores the central role assumed in his theory by the model-conception of the well-ordered society and, in particular, the model-conception of the person. In fact, as I shall argue, the original position, as a 'device

of representation', acquires whatever justificatory force it has because it adequately mirrors or models the ideals specified in the model-conceptions.

Rawls cites four distinct features that combine to make up the model-conception of the well-ordered society as a self-sufficient system of social cooperation for mutual advantage.⁵ First, a well-ordered society is one that is "effectively regulated by a public conception of justice" (Rawls 1971, 453; 1975, 537). This means (a) that each citizen accepts, and knows that others accept, the same conception of justice (whatever it turns out to be); (b) that the basic institutions of the society satisfy this conception of justice and are with good reason believed by everyone to satisfy it; and (c) that this public conception of justice is based upon reasonable beliefs established by widely-accepted methods of inquiry (i.e., it does not presuppose controversial metaphysical or religious doctrines), (Rawls 1971, 454; 1975, 537).

Second, citizens in a well-ordered society are, and recognize themselves as being, free and equal moral persons (Rawls 1971, 505; 1975, 525). Rawls calls this feature the "model-conception of the person". Although it is present in *A Theory of Justice*, there is no doubt that it has become more prominent in his later writings, especially to counter those interpretations that view the earlier work as an attempt to ground principles of justice in a neutral (game-theoretical) conception of rational choice. Since this model-conception is central to my own interpretation, I will outline its three most important features more fully.

(a) Citizens in the well-ordered society are characterized as having *two basic moral powers*: the capacity for a sense of justice, that is, "the capacity to understand, to apply and to act from (and not merely in accordance with) the principles of justice"; and "the capacity to form, to revise, and rationally to pursue a conception of the good" (Rawls 1975, 525; 1971, 505). In ascribing the first moral power to citizens in the well-ordered society, it is assumed that whatever the appropriate conception of justice turns out to be citizens are capable of being effectively motivated by it. The second moral power suggests that citizens have their own conceptions of the good which motivate them and give them a sense of purpose or worth in life. However, citizens are not viewed as being permanently attached to any one conception of the good; rather, they are considered capable of changing their ends and ideals at various times in their life, and the recognition of this capacity to revise their conception of the good is an important feature of their characterization as moral persons.

(b) Citizens are *equal* in that "they each have, and view themselves as having, a right to equal respect and consideration in determining the principles by which the basic arrangements of their society are to be regulated".⁶ This notion of the

⁵ For Rawls's discussion of the notion of a well-ordered society, see Rawls 1971, 453-62; Rawls 1975, 521-22; and 1974 esp. pp. 633-37; in order to emphasize the continuity between Rawls 1971 and 1975, I will cite references to both of these in the text.

⁶ Rawls 1975, 94; see also, Rawls 1975, 521 and 546; Rawls's notion of equality is thus quite similar to Dworkin's notion of equal concern and respect.

equality of citizens in determining the principles of justice is more fundamental than (and the basis for) other ideals of equality that are realized within the basic structure of society (e.g., formal equality before the law, equality of opportunity with respect to powers and offices defined by the social structure, etc.). It is based on their common status as moral beings.

(c) Citizens in the well-ordered society are *free* and recognize each others' freedom, in several respects. First, each is viewed as having the capacity to form, revise and pursue a conception of the good (Rawls 1971, 505; 1975, 544). This repeats the second moral power noted above. It also reflects Rawls's liberal commitment to a plurality of conceptions of the good and his belief that principles of justice should not rely upon any particular conception of the good, but upon a conception of the person (as a free and equal moral being). Thus, it parallels Kant's attempt to distinguish between the moral agent and the ends to which she might be attached at any given time. Second, citizens are free in that they are, and recognize one another as being, "self-originating sources of valid claims" (Rawls 1975, 543). This follows directly from the view that there is a plurality of conceptions of the good, all of which are acceptable so long as they are pursued within the limits of justice. Finally, citizens are free in that they recognize one another as being responsible for their ends or conceptions of the good. Persons are not merely the subjective bearers of preferences that, so to speak, assail them from without. This implies, among other things, that a utility function defined solely according to the strength or intensity of preferences or desires does not provide an acceptable criterion for the interpersonal comparison of levels of well-being, since it does not reflect the way in which persons have come to have those preferences.⁷

Finally, the third and fourth features of the model-conception of the well-ordered society primarily refer to the background conditions and arrangement of its institutions. Rawls assumes that even in a well-ordered society the "circumstances of justice" will obtain (Rawls 1971, 126 and 1975, 525).⁸ Even in a well-ordered society it is assumed that moderate scarcity will persist since it is unlikely that there will be enough natural and social resources to satisfy every legitimate demand. This is due in part to the "subjective" circumstance that in a well-ordered society there will still be a plurality of conflicting conceptions of the good within the limits of justice. Finally, a well-ordered society is said to be stable with respect to its conception of justice (Rawls 1971, 454; 1975, 522). "This means that, viewing the society as an ongoing concern, its members ac-

⁷ Rawls discusses this aspect of the person as free in his 1982, 169; for a careful discussion of this question, see Scanlon 1975 and Buchanan 1975.

⁸ In response to several Marxist objections (notably, MacPherson) Rawls no longer insists upon the objective circumstance of justice (Rawls 1975, 539); however, even with the advances of technology and science (and the freeing of production for the satisfaction of generalized interests) there is no guarantee that moderate scarcity can be overcome given the plurality of conflicting conceptions of the good.

quire as they grow up a sufficiently strong and effective sense of justice, one that usually overcomes the temptations and stresses of social life." (Rawls 1974, 634)

In "Kantian Constructivism" the description and design of the original position is then developed with reference to these features of the well-ordered society. Each element of construction in the original position is defended by showing how it appropriately models or represents the model-conceptions. Thus, the veil of ignorance, the various formal constraints on choice, and the symmetry and rationality of the parties are defended with reference to the four features cited. Similarly, the list of primary social goods and the restriction of deliberation to the basic social structure are defended with reference to the model-conception of the person and the two basic moral powers (see especially Rawls 1982). For our own purposes, the point worth stressing again is that the original position does not reflect an allegedly neutral, 'game-theoretical' model of rational choice, but reflects more explicitly normative assumptions as a modeling of the ideals specified by the model-conceptions.

In "Kantian Constructivism" Rawls also suggests that the four features of the well-ordered society can be united under "the Reasonable" and "the Rational" terms of social cooperation. (These two notions correspond closely to Kant's distinction between *Vernunft* and *Verstand*.) Social cooperation is Reasonable since it incorporates the ideals of mutuality and reciprocity (Rawls 1975, 528): All who cooperate reasonably share in the benefits and burdens as established by principles agreed to from an appropriate perspective, namely, the perspective of citizens regarded as free and equal moral persons. The notion of the Reasonable is also connected to the notion of a citizen as fully autonomous, that is, as having the capacity to act from a sense of justice and his right to equal consideration in determining principles for the regulation of the basic structure. The terms of social cooperation are "Rational" since they allow for "each participant's rational advantage, what, as individuals, they are trying to advance" (Rawls 1975, 528). This is reflected in the person's second basic moral power, the capacity to form, revise, and pursue a conception of the good, as well as the commitment to a plurality of such conceptions within the limits of justice. In a rather elegant passage, Rawls clarifies the relationship between the Reasonable and the Rational in a well-ordered society:

"The Reasonable presupposes and subordinates the Rational. It defines the fair terms of cooperation acceptable to all within some group of separately identifiable persons, each of whom possesses and can exercise the two moral powers. All have a conception of their good which defines their rational advantage, and everyone has a normally effective sense of justice; a capacity to honor the fair terms of cooperation. The Reasonable presupposes the Rational, because, without conceptions of the good that move members of the group, there is no point to social cooperation nor to notions of right and justice, even though such cooperation realizes values that go beyond what conceptions of the good specify taken alone. The

Reasonable subordinates the Rational because its principles limit, and in a Kantian doctrine limit absolutely, the final ends that can be pursued." (Rawls 1975, 530)

The description and design of the original position can also be seen as expressing the ideals of the Reasonable and the Rational (see diagram). The Rational is exhibited primarily in the rational autonomy of the parties and the adoption of the maximin rule. The Reasonable, by contrast, is expressed in the formal constraints, the veil of ignorance, the symmetry conditions, the list of primary goods, and the designation of the basic structure as the subject of justice. Thus, just as the design of the original position models the ideals of the model-conception, it can also be seen as expressing the notions of the Reasonable and the Rational as complementary aspects of our capacity for moral deliberation. The reason why the principles hypothetically chosen by the parties in the original position are binding on us is because as a 'device of representation' it adequately mirrors our basic moral powers and thus models ideals it would be unreasonable for us to reject. It now remains to be seen how this account of the original position (and thus the selection of his two principles) which emphasizes the model-conception of the person and our basic moral powers is related to Rawls's more recent emphasis on his account as a *political* conception.

In "Justice as Fairness: Political, Not Metaphysical" Rawls suggests that the account of justice as fairness developed in *A Theory of Justice* and subsequent essays is misunderstood if it is seen as part of a more general moral theory or philosophical approach (in the way, for example, one might regard utilitarianism or Kantian moral theory). Though some passages in the earlier writings may encourage such a reading, in this essay Rawls argues that his account is offered as a *political* conception. This means that the conception is limited in the scope of its application to what Rawls calls the "basic structure" of society. More importantly, however, it means that its justification does not appeal to "general and comprehensive" moral, religious or philosophical doctrines that are bound to be controversial in a society characterized by the "fact of pluralism," but relies instead on "fundamental intuitive ideas" (including the model-conception of the person) implicit in the public culture of a democratic society. Rawls describes this strategy as the "method of avoidance" and considers it to be simply the extension of the principle of (religious) toleration to the aims of political philosophy itself: we prescind from controversial metaphysical or philosophical doctrines in order to find a practicable solution to the long-standing disputes concerning social justice that still beset liberal democratic regimes. The idea is to remove deeply controversial moral and religious issues from the political agenda so that a common basis can be gained for finding principles of justice and defining constitutional essentials. As Rawls puts it, "the question is: what is the least

that must be asserted; and if it must be asserted, what is its least controversial form?"⁹

On one interpretation of his work, referred to in my opening remarks, this construal of the "method of avoidance" is central to Rawls's understanding of the political. What he now refers to as the "domain of the political" is characterized, from a socio-cultural perspective, in terms of certain basic or fundamental ideas and self-understandings latent in the public culture.¹⁰ The idea of society as a fair system of cooperation between free and equal citizens is the most important of these, but other self-understandings are drawn from this idea of the political as well. For example, the assumption that citizens are able to distinguish between values they hold in connection with comprehensive religious and metaphysical (or philosophical) doctrines and those 'very great political values' associated with the idea of society as a fair system of cooperation and that, when a conflict arises, citizens give priority to the latter helps to delineate the domain of the political (see Rawls 1988, 275; 1989, 243 and Galston 1989, 717). So too does the assumption that citizens are able to distinguish between their "public" and "non-public" (or "moral") identities.¹¹ Rawls's socio-cultural characterization of the political also includes a fairly specific understanding of the use of "free public reason" and idea of a reasonable agreement between free and equal citizens: The principles of justice which are to regulate the basic structure must be justified not only with reference to modes of inquiry and rules for assessing evidence that are publicly recognized, but also on grounds that it is reasonable for all to accept in view of their "common human reason" and "the criteria and procedures of common sense knowledge" (Rawls 1989, 236 and 244).

It is important to observe, however, that on the first interpretation I have proposed even these further self-understandings are drawn from and - in a crucial sense - constrained by the "fundamental intuitive ideas" contained in the "overlapping consensus". According to Rawls, it would be contrary to the aims of a *political* conception of justice - that is, one that applied the principle of toleration to itself - if the justification of these ideas were pursued otherwise. For example, the characterization of the scope and constraints implicit in the idea of a reasona-

⁹ Rawls 1987, 8; for a more detailed account and criticism of Rawls's "method of avoidance" see my 1992b.

¹⁰ See Rawls 1985, 225; 1987, 6; and 1989, 240. This *socio-cultural* characterization of the domain of the political should be distinguished from Rawls's related *institutional* characterization of the political as a non-voluntary association vested with the authority for making collectively binding and coercively enforced decisions. In this second sense it refers to a (functionally-specific) subset of the "basic structure" of society and is distinguished from the associational, the familial, and the personal.

¹¹ See Rawls 1985, 240-41 and 1987, 16. An ambiguity arises here, however, since the public or political conception of the citizen is also a moral conception. Thus, if the distinction between public and non-public identity is not to be reduced solely to the question of whether it is tied to the institutions of the "basic structure" or not, Rawls must offer some further means for distinguishing between these two "moral" aspects of our identity.

ble agreement between free and equal persons is not itself clarified reflexively or recursively in relation to a more general analysis of the requirements of practical reason, but is made to depend upon the fundamental ideas latent in our public culture.¹² Whereas the former strategy (pursued by Kant and, more recently, Habermas) is "metaphysical" or philosophical in a manner Rawls wishes to avoid, the latter contributes to a "political, not metaphysical" characterization of the domain of the political.

However, this first interpretation, which is closest to relativist readings such as Rorty's, faces serious challenges from within Rawls's own texts. It does not provide a way for distinguishing (as Rawls wishes to do) between principles that rest simply on a *de facto* consensus concerning fundamental intuitive ideas and the correct account (Rawls 1989, 250). Similarly, insofar as it suggests that the best account of the conception of justice simply reflects the shared understandings of a political culture, it leaves unexplained Rawls's concern that a political conception not be "political in the wrong way". It also neglects Rawls's recent remark that the idea of an overlapping consensus comes into play only at the second stage in the exposition of his conception of justice, that is, as part of the response to the question of the stability of his conception *after* it has already been articulated as a "free-standing political conception" (Rawls 1989, 234). Finally, it would seem to play down the significance of Rawls's constructivist account of objectivity (in Rawls 1975) and the related notion of justification as a reasonable agreement between free and equal persons.

In the second interpretation, also referred to in my opening remarks, the model-conception of the person is not *simply* a fundamental intuitive idea latent in our public culture. More importantly, in this account, the model-conception of the person and the basic moral powers are taken to be the best (most correct) account of our capacities for practical reason. The focus is not on the shared *content* of a political culture, but on the citizen's moral powers or capacities for practical deliberation. On this interpretation, the relation between the 'method of avoidance' and the account of practical reasoning is thereby reversed. That is, rather than limiting the ideal of the person and the basic moral powers to what is contained within an overlapping consensus, what it is reasonable for us to assume and where one can expect reasonable disagreement to continue must also depend, among other things, on the best account that can be given of our basic moral powers as well as on the real possibilities for their effective exercise in light of the idea of 'free public reason' (see below). This reading has the advantage of providing a means for distinguishing between a *de facto* and a normative consensus. It would also enable Rawls to explain how his conception might not be "political in the wrong way".

¹² For a critique of Rawls's strategy along these lines, see O'Neill 1989a, 210-11; for an alternative 'Kantian' account of public reason see her 1989b, and my own 1992a, 76, note 68. For a similar criticism of Rawls, see Nagel 1991, 159.

In the third part of "Kantian Constructivism" Rawls considers his constructivist account of moral objectivity. Principles of justice are correct not because they are true (in the sense that they correspond to an independent moral order) but because they are ones that would be agreed to by all if viewed from a "suitably constructed social point of view" (Rawls 1975, 554). This constructed social point of view is, of course, the original position, and what makes it "suitable" is that it appropriately mirrors or represents the model conception of the person as a free and equal citizen with two basic moral powers. Or, to state the same point differently, it is suitable because it appropriately models both the "rational" and the "reasonable" features of our basic moral capacities. At times, Rawls describes this constructivist account of objectivity, in a manner consistent with his "method of avoidance," as indifferent to realist and anti-realist debates. More often, however, he contrasts constructivism to "rational intuitionism" and thereby suggests that he is not simply sidestepping this controversy, but proposing his own alternative account of moral objectivity (see Brinks discussion in his 1989, 303ff.). A strong reason for preferring the second position is that it fits better with his account of "free public reason" and responds more appropriately to the need for a *public* basis of justification for political principles.

The idea of free public reason, Rawls suggests, is an "essential companion" of a political conception of justice (Rawls 1987, 8). It specifies not only that specific "guidelines of inquiry and publicly recognized rules of assessing evidence be used" and that the virtues of reasonableness and fair-mindedness be adhered to, but also, and perhaps most importantly, that any acceptable principles of justice conform to the "idea of publicity" (Rawls 1987, 8). This last condition, which echoes Kant's "transcendental principle of publicity," reads as follows: "Publicity ensures, so far as the feasible design of institutions can allow, that free and equal persons are in a position to know and to accept the background social influences that shape their conception of themselves as persons, as well as their character and conception of their good. Being in this position is a precondition of freedom; it means that nothing is or need be hidden."¹³

The idea of free public reason requires that the justification of political principles appeal only to norms, values, and ideals that citizens, in view of their conception of themselves as free and equal persons with two basic moral powers, could reasonably be expected to accept. Any *de facto* agreement may fail to meet this requirement since it makes reference to certain ideal conditions which may not have been satisfied. It presents, as a test of legitimacy, the counterfactual ideal of a reasonable agreement between free and equal persons.

This idea of publicity and the related requirement that acceptable principles of justice must be capable of public justification is not a free-floating constraint

¹³ Rawls 1975, 539; in his 1987 Rawls also refers approvingly to an essay by Jeremy Waldron offering an analogous account of publicity as central to liberalism: "Liberals demand that the social order should in principle be capable of explaining itself at the tribunal of each person's understanding." (Waldron 1987, 149)

arbitrarily introduced by Rawls. It is, he claims, already implicit in the first moral power outlined above - the capacity for a sense of justice - and is thus connected with our basic capacities for practical deliberation. This moral power (together with the "highest-order interest" in its realization) "... implies a regulative desire to conform the pursuit of one's good, as well as the demands one makes on others, to public principles of justice which all can reasonably be expected to accept."¹⁴ In a careful and instructive analysis, Samuel Freeman describes this feature of our moral powers as "answerability": "Answerability implies more than being held accountable (responsible) for one's acts. It means that we (or someone in our stead) be able publicly to justify our conduct, aims, and expectations on terms others could freely accept."¹⁵ According to the second interpretation of Rawls I am proposing, "answerability" is a fundamental component of our moral personality or capacity for practical deliberation and, as such, assumes priority over other ends and aspirations we may have.

Further, in addition to the idea of publicity and account of answerability, there is yet another "requirement of reason" which can be inferred from Rawls's model-conception of the person. In describing various "precepts of reasonable discussion", Rawls says that we should also be prepared to recognize the "burdens of reason" and to accept that there will be many issues on which even reasonable people may disagree (Rawls 1989, 238). In view of the "obvious sources of reasonable disagreement", such as conflicting evidence, the indeterminacy of many of our concepts, alternative modes of assessment, and "limited social space", Rawls argues that it is unreasonable to expect that public discussion will produce a wide-ranging agreement even among reasonable people. It is important to note, however, that this constraint is not simply a concession to the status quo, but is itself a requirement of reason. Furthermore, at least in certain limiting cases, whether or not disagreement on a particular issue is reasonable cannot be decided solely by appeal to intuitive ideas and common sense (as might be supposed on the first interpretation above), but can only be determined in a public discussion (or, more precisely, in a series of overlapping and institutionally structured public discussions) in which all those affected have an opportunity to participate and all relevant considerations and perspectives are taken into view.¹⁶

In view of these "requirements of reason" inferred from our basic moral powers, it is clear that Rawls's model-conception of the person entails a fairly rich (and controversial) account of practical agency. It is, for example, at odds

¹⁴ Rawls 1982, 165; compare also Scanlon's contractualist account of morality in his 1982.

¹⁵ Freeman 1991, 285; much of Freeman's essay is also a commentary on Scanlon's contractualist (and constructivist) account of morality.

¹⁶ This would, of course, suggest that Rawls's model-conception must be complemented by an account of the democratic institutions in which such discourses might be realized; see my 1992b.

with Humean or "desire-based" accounts of practical reason since it assumes that the agent has fundamental interests which are not directly tied to (contingent) objects of desire.¹⁷ Rather, we are assumed to possess, as part of our deeper motivational structure, two "highest-order" interests - namely, interests in realizing the two basic moral powers (Rawls 1975, 525). Like primary desires, these interests can also be "action-guiding" in that they are taken to be "supremely regulative" for other aims we may have. However, since they presuppose the idea of a public justification they may be called "principle-dependent" (Freeman 1991, 292; see also Scanlon 1988, 173): They reflect our capacity and willingness to act on norms or principles that could not reasonably be rejected by others as a basis for general agreement. They also represent what Kant called an "interest of reason" since they refer not to the objects of desire, but to our basic capacities for practical deliberation.

If we inquire about the justification of this model of practical agency, however, Rawls faces a difficulty similar to that encountered by Kant: He must not only show that the model-conception and its implicit "requirements of reason" reflect a particular model of practical agency, but also that this model of agency is somehow appropriately binding on us. Within the framework of his constructivist account of moral objectivity, the direction his response must take to this difficulty seems clear: The model-conception of the person is not *simply* a "fundamental intuitive idea" latent in our public culture, but is, so to speak, reflexively or recursively grounded in the ongoing practice of public justification. That is, the model-conception is not simply an account that we would recognize after due reflection in a condition of reflective equilibrium, it is also an account of the basic moral capacities that we - those of us who occupy the "third point of view" - must already be assumed to possess as free and equal persons engaged in a process of justification. It is, in other words, not just one possible account of our self-conception, but the best account that can be given of the basic capacities presupposed by the practice of justification in which we are involved as co-deliberators. On this account, what Kant described as the necessity of acting "under the idea of freedom" may be seen as having been replaced with the necessity of acting "under the idea of providing justification" (Kant 1922, 448). Under this idea, the account of practical agency becomes intelligible; outside this idea, the very notion of acting *for a reason* seems threatened.

These two competing interpretations of the significance of Rawls's political conception of justice are, I hope to have shown, not simply the result of an obvious or willful misreading by one side or the other. Rather, they reflect a deeper tension in Rawls's own project itself reminiscent of a profound ambiguity in Kant's moral theory. I believe that the project outlined in the second interpretation will finally prove to be the more promising. But, in view of the controversy that persists over the interpretation of Rawls's later essays, it is clear that

¹⁷ Unfortunately, I cannot pursue this important debate further in this essay; see, however, Korsgaard 1986, and Wallace 1990.

neither Rawls nor his interpreters have as yet adequately articulated that project, let alone provided it with a satisfactory defense.

Bibliography

- Allison, Henry (1986), Morality and Freedom: Kant's Reciprocity Thesis, in: *The Philosophical Review* 95, 393-425
- Baynes, Kenneth (1992a), *The Normative Grounds of Social Criticism: Kant, Rawls, and Habermas*, Albany
- (1992b), Liberal Neutrality, Pluralism, and Deliberative Politics, in: *Praxis International* 12, 50-69
- Brink, David (1989), *Moral Realism and the Foundation of Ethics*, New York
- Buchanan, Allen (1975), Revisability and Rational Choice, in: *Canadian Journal of Philosophy* 5, 395-408.
- Cohen, Joshua (forthcoming), Moral Pluralism and Political Consensus, in: *The Idea of Democracy*, New York
- Freeman, Samuel (1990), Reason and Agreement in Social Contract Views, in: *Philosophy and Public Affairs* 19, 122-57
- (1991), Contractualism, Moral Motivation, and Practical Reason, in: *Journal of Philosophy* 88, 281-303
- Galston, William (1989), Pluralism and Social Unity, in: *Ethics* 99, 700-722
- Gilbert, Alan (1990), *Democratic Individuality*, New York
- Hampton, Jean (1989), Should Political Philosophy Be Done Without Metaphysics?, in: *Ethics* 99, 791-814
- Hill, Thomas, (1989), Kant's Theory of Practical Reason, in: *The Monist* 72, 363-83
- Kant, Immanuel (1922), *Grundlegung zur Metaphysik der Sitten*, (Cassirer ed.) Berlin
- Korsgaard, Christine (1986), Skepticism about Practical Reason, in: *Journal of Philosophy* 83, 5-25
- Larmore, Charles (1990), Political Liberalism, in: *Political Theory* 18, 339-60
- Nagel, Thomas (1991), *Equality and Partiality*, New York-Oxford
- O'Neill, Onora (1989a), Constructivisms in Ethics, in: *Constructions of Reason*, New York, 201-218
- (1989b), The Public Use of Reason, in: *Constructions of Reason*, New York, 28-50
- Rawls, John (1971), *A Theory of Justice*, Cambridge
- (1974), Reply to Alexander and Musgrave, in: *Quarterly Journal of Economic* 88, 633-55
- (1975), A Kantian Conception of Equality, in: *Cambridge Review* 96, 94-99
- (1980), Kantian Constructivism in Moral Theory, in: *Journal of Philosophy* 77.9, 515-72
- (1982), Social Unity and Primary Goods, in: Amartya Sen/Bernard Williams (eds.), *Utilitarianism and Beyond*, Cambridge, 159-185
- (1985), Justice as Fairness: Political Not Metaphysical, in: *Philosophy and Public Affairs* 14, 223-51
- (1987), The Idea of an Overlapping Consensus, in: *Oxford Journal of Legal Studies* 7, 1-25
- (1988), The Priority of Right and Ideas of the Good, in: *Philosophy and Public Affairs* 17, 251-76

- (1989), The Domain of the Political and Overlapping Consensus, in: *New York University Law Review* 64.2, 233-55
- Raz, Joseph (1990), Facing Diversity: The Case of Epistemic Abstinence, in: *Philosophy and Public Affairs* 19, 3-46
- Reath, Andrews (1989), The Categorical Imperative and Kant's Conception of Practical Reason, in: *The Monist* 72, 384-410
- Richards, David (1988), Prescriptivism, Constructivism, and Rights, in: Douglas Seanor/Nicholas Fotion (eds.), *Hare and His Critics*, Oxford, 135-152
- Rorty, Richard (1991), The Priority of Democracy to Philosophy, in: *Objectivism, Relativism, and Truth*, New York, 175-196
- Scanlon, Thomas M. (1975), Preference and Urgency, in: *Journal of Philosophy* 72, 655-69
- (1982), Contractualism and Utilitarianism, in: Amartya Sen/Bernard Williams (eds.), *Utilitarianism and Beyond*, Cambridge, 103-128
- (1988), The Significance of Choice, *The Tanner Lectures on Human Values VIII*, Cambridge
- Stern, Paul, (1986), The Problem of History and Temporality in Kant's Ethics, in: *Review of Metaphysics* 39, 505-545
- Waldron, Jeremy (1987), Theoretical Foundations of Liberalism, in: *The Philosophical Quarterly* 37, 127-150
- Wallace, R. Jay (1990), How To Argue About Practical Reason, in: *Mind* 99, 355-385