The Justification of Justice as Fairness: A Two Stage Process

TED VAGGALIS
University of Kansas

The tragic truth about philosophy is that misunderstanding occurs more frequently than understanding. Nowhere is this more evident than in the reception of Rawls' work. A common misinterpretation of his conception of justice as fairness is that it is an application of Kantian moral theory to the political structure of society. Viewed in this light, Rawls' theory seems to be open to a serious objection. One could argue that as a Kantian moral theory justice as fairness is too controversial to generate the consensus necessary for contractual agreement. It is controversial because it violates the most fundamental liberal requirement: that the state remain neutral in regard to competing conceptions of the good.[1] A critic who pressed this line of argument could then go on to argue that any attempt to weaken the moral claims of justice as fairness in order to mitigate the controversy, would only undermine the capacity of justice as fairness to generate support for itself. The supposed strength of this objection is that it reveals that what is problematic in Rawls' theory is only symptomatic of the problems faced by liberal political conceptions in general.

Much of Rawls' recent work has been an attempt at correcting this mistaken reading and answering this objection. In particular, the essay 'The Idea of an Overlapping Consensus' further clarifies his claim that justice as fairness is a political and not a metaphysical moral conception. Rawls defines an overlapping consensus as the articulation of that "shared basis of consensus on a political conception of justice" which is latent in the political culture of a constitutional democracy and which makes possible an orderly and stable community (IOC 25). As a political conception justice as fairness constitutes an overlapping consensus when it is found to be congruent with the variety of moral, philosophical and religious views that compose a society. This, in turn, establishes it as non-controversial.

But with the introduction of the overlapping consensus Rawls has created a problematic ambiguity in the interpretation of his theory. Since a political conception of justice is viable to the extent that it is capable of generating this overlapping consensus, what is the justificatory role of the original position? It no longer appears necessary as a device for deciding

---

[1] For an excellent discussion on the necessity of the neutrality of the state in liberal political conceptions see Larmore, chapter 3, pp. 40-48. (See the bibliography for a key to citations.)

on the principles of justice. Instead, the appeal can be made directly to the principles as ideas already latent in the public political culture. This move would no doubt appease many of Rawls's critics who regard the original position as the source of most of the controversial metaphysical claims of his theory.[2] But for Rawls to adopt this strategy, in my view, would rob justice as fairness of its critical justificatory feature. That the principles of justice are chosen in a fair decision procedure gives them a rational and principled superiority over other competing theories. To reduce the justification of a conception of justice to a matter of simply dredging up the ideas (or preferences) latent in the prevailing political culture, would leave justice as fairness open to the charge of merely representing the arbitrary and capricious preferences of a society. My paper, then, is concerned with what appear to be two opposed patterns of justifying the principles of justice in Rawls's recent work. I will argue that Rawls has not abandoned the original position, nor does he intend the overlapping consensus to be an alternative to it. Rather, they are component parts of a two-stage process of justification.

I. Overlapping Consensus: What is it?

Rawls starts with the basic presupposition that our historical and social situation requires us to conceive of justice in a particular way. Now traditionally justice has been conceived as a fragile consensus of self- and group-interests (IOC 2). But we have been part of a democratic tradition with its well-defined constitutional practices and this calls for something more principled. What Rawls has in mind is a "regulative political conception of justice that can articulate and order in a principled way the political ideas and values of a democratic regime, thereby specifying the aims the constitution is to achieve and the limits it must respect" (IOC 1). Accordingly the conception he has in mind has three defining features: i) it is a political conception, formulated for the purposes of addressing a specific subject, i.e., the basic structure of society (IOC 3); ii) as a political conception it is not part of any general and comprehensive moral conception (IOC 3); and iii) it is formulated in "terms of certain fundamental intuitive ideas viewed as latent in the public political culture of a democratic society" (IOC 6). Now, for Rawls, the most important of these intuitive ideas are the ideas of the person as free and equal and that society is a permanent cooperative venture for mutual advantage among such persons.[3]

---

[2] Sandel, chapter 1, pp. 15-65.

[3] Rawls, in his *Dewey Lectures*, regards the idea of the person, the idea of social cooperation and the original position as the 'model conceptions' presupposed by an overlapping consensus. However, for the purposes of this paper I assume that the reader is aware of this. My main concern here

Historically conceptions of justice have been formulated in one of two ways. Either it is a balancing of opposed interests or it is the extension of a general moral principle to the political realm (IOC 2-5). Rawls argues that both approaches are inadequate for the task at hand. A conception of justice is supposed to generate a consensus out of the various religious, philosophical and moral ideals which compete with each other. On the one hand the balancing of interests is always a tenuous arrangement at best. It is conceded that the various parties are ready to pursue their interests at the expense of each other should the occasion arise. On the other hand, extending a moral doctrine to cover the political realm is problematic because it espouses ideas and values that are not widely shared in a community (IOC 6). At a certain point the right of conscience must be violated in order to bring about harmony within the community. In both cases, then, their respective conceptions of justice cause conflict and inevitably rely upon an oppressive use of force in order to establish them (IOC 14).

Rawls believes that in the case of his political conception of justice such outcomes can be avoided. This is because it is designed to accommodate a very important fact of contemporary democratic life--pluralism (IOC 4). In order to achieve this goal, a political conception must remain free of so-called 'longstanding controversies' (IOC 13). It must allow for a wide variety of general and comprehensive doctrines as well as a 'plurality of conflicting and incommensurable' conceptions of the good (IOC 4). For Rawls, then, the emphasis will be on the agreement necessary for social cooperation and not the truth about the nature of the political. If it proves necessary for individuals to assert aspects of their comprehensive views, then they are to be given in a minimalist form. "The question is: what is the least that must be asserted; and if it must be asserted, what is its least controversial form?" (IOC 8). But this brings us to a serious problem: how can such an agreement maintain its efficacy given the lack of any unifying principle or interest? In other words, how can a constitutional or basic political consensus generate the kind of allegiance that will outweigh self- or group-interests on the one hand, and moral/religious doctrines on the other?

Rawls gives two related answers to this question. First, he claims that allegiance to a political conception is not necessarily determined by one's interests or moral/religious views. In fact, most moral/religious doctrines are not fully general and comprehensive views to begin with. This means that in many cases one's political conception has nothing (or very little) to do with one's other interests and comprehensive views (IOC 18-19). Rawls claims that this fact creates a 'slippage,' which allows a political conception to loosely cohere with other beliefs (IOC 19). In the event that one does

---

is in the relationship between the original position and the overlapping consensus in the pattern of justification for 'justice as fairness.'

discover a discrepancy between one's political conception and one's other immediate views, one is just as likely to revise these immediate views rather than the political conception. The reason for this is that one has come to value the capacity of the political conception to achieve the public good (IOC 19).

This last point leads to the other reason that Rawls gives for the ability of an overlapping consensus to generate allegiance to itself. He claims that a political conception, when it effectively regulates basic political institutions, meets three requirements essential for a stable constitutional regime. First it sets forth the content of basic rights and liberties and gives them priority over all other values (IOC 19). In doing this, it places these items beyond the debate and calculus of social interests (IOC 19-20). This insures that social cooperation will take place on terms of mutual respect (IOC 20). Next it meets the requirements of the idea of free public reason (IOC 20). In addition to detaching itself from controversial debates, the overlapping consensus recognizes certain guidelines of public enquiry and rules for assessing evidence (IOC 20). This will include the forms of reasoning available to common sense, the non-controversial aspects of scientific inquiry, and guarantees of freedom of speech and thought (IOC 2). Finally the overlapping consensus encourages the 'cooperative virtues of political life.' These are the virtues of reasonableness, fairness and a spirit of compromise (IOC 21). Social cooperation in terms of mutual respect engenders a tendency on the part of others to put aside self-interested aims in favor of broader social aims. The primary task of a political conception of justice, then, is to order the political institutions and specify basic rights and liberties. When it has met the three requirements of a stable constitutional regime, it attains a fixed quality that enables it to affect the political character of its citizens (IOC 21). Allegiance to the political conception occurs because the citizens see it as consistent with their own mutually opposed views and, consequently, one that will order their conduct without sacrificing their self- or moral-interests. A political conception of justice becomes an overlapping consensus that operates only at the level of the basic structure of society. It cannot assert any moral, religious, philosophical or metaphysical doctrines (IOC 7). Nor can it take part in any dispute about these matters. Instead, it unites incommensurable views of the good into a consensus by locating the point at which these views overlap each other (IOC 6).

> The conjecture, then, is that as citizens come to appreciate what a liberal conception does, they acquire an allegiance to it, an allegiance that becomes stronger over time. They come to think it both reasonable and wise for them to confirm their allegiance to its principles of justice as expressing values that, under the reasonably favorable conditions make democracy possible, normally

> counter-balance whatever values oppose them. With this an overlapping consensus is achieved (IOC 22).

II. The Relationship between the Original Position and the Overlapping Consensus.

From the foregoing it is clear that in articulating the role of the overlapping consensus Rawls has also carried out a justification of the principles of justice. But it is a much different kind of justification than the one given in *A Theory of Justice*. The most significant change is in Rawls' direct appeal to interests that were excluded from consideration in the model of the original position. Also, the parties have a rather informed view of their place in society, as well as their talents, beliefs, etc. Some have speculated that Rawls has given up on the original position in order to accommodate criticisms of his theory that: I) the original position, given its Kantian formulation, made metaphysical claims that were too controversial for agreement; ii) that even if one granted the results of the original position, it was not clear why anyone would adhere to the agreement once his position in society and interests were known; and finally, iii) any attempt to enforce the agreement would be to impose a particular conception of the good upon a citizen, thus violating her/his right to conscience. Given Rawls' detailed emphasis on the necessity of a non-controversial political conception and the very constrained nature of the social debate in his analysis of the overlapping consensus, along with the disappearance of any sustained discussion of the original position in his most recent works, such an interpretation carries a high degree of plausibility. This is enhanced when one also considers the standard criticism that because the original position stacks the deck in favor of justice as fairness by including certain moral conceptions in the background, it is not a decision procedure but an expository device. Thus, the notion of an overlapping consensus as the justification of the principles of justice would go a long way towards accommodating many criticisms and appeasing many critics.

But such an interpretation also rests on a fundamental misunderstanding of Rawls' theory. It holds that justice as fairness is an application of Kantian moral theory to the basic structure of society. However, Rawls is quite explicit that in referring to his theory as 'Kantian,' he intends this claim to be understood as an analogy and not an identity. This is a crucial distinction. Its significance is best appreciated only when considering it in connection with the original position. Only after having done this will one be in a position to see that the idea of an overlapping consensus is insufficient to justify the principles of justice and requires a decision procedure like the original position.

The original position is a device for deciding which scheme of principles is most appropriate "for realizing liberty and equality once

society is viewed as a system of cooperation between free and equal persons" (JP 235). In order to adequately understand this decision procedure, we need to be clear about two aspects of the original position: its background assumptions and the conditions that must prevail in order to render a fair decision. There are two ideas that constitute the background assumptions of the original position. They are the idea of the person and the idea of society as a fair system of cooperation (or mutual advantage). Cooperation is to be understood in the following way. First it is a system guided by publicly recognized rules and procedures. It is not a system where orders are issued by a central authority. Second each participant may reasonably accept the terms of cooperation, provided that the others do the same. Also, it means that those who cooperate must benefit in an appropriate way. Finally, each individual is to be guided by their own conception of the good (JP 232).

By person we are to understand the following. First, a person is someone who can be a fully cooperating member of society over a complete life (or, in other words, a citizen). Second, insofar as these persons have the powers of reason, thought and judgement they are free. Finally, to the extent that these persons have these powers to the requisite degree they are equal (JP 233).

Now in regard to the background assumptions, Rawls indicates that they are *nothing more than weak, normative assumptions*. They do not imply any deep theoretical commitments that would prejudice the decision procedure of the original position. In fact, they are simply minimal ideas formulated with a view to establishing consensus, ideas that all could accept given their various, and in many cases conflicting, personal views of the world. This is especially true in regard to the idea of the person. Here one is not committed to any deep metaphysical conception of the person, Kantian or otherwise. In placing persons behind a veil of ignorance one does not come across any doctrine of the person as a 'self shorn of all its contingently given attributes' or a self that 'assumes a supra-empirical status, given prior to its ends, a pure subject, ultimately thin,' as Michael Sandel has suggested (JP 239; see also Sandel 93-95). Instead one is simply stating that minimal common ground that must be presupposed about the person by all the competing views of the good if agreement about a conception of justice is going to be reached in a reasonable and principled way.

This method of minimally stating the background assumptions insures that the conditions that prevail in the original position will be such that they are not prejudicial to any of the competing schemes of principles or to the bargaining position of any of the parties. As we said above, the background ideas, although normative (or moral) ones, are to be weak minimal statements. The veil of ignorance is employed in order to filter out any irrelevant information that arises from the contingencies of the social world and would prejudice the decision process (JP 236-237). It acts

to constrain the arguments that will be presented so that whichever arguments are successful will be universalizable (within a given society) and have undoubted authority in regard to settling which scheme of principles is to be chosen. Thus, the participants will have a reflective knowledge about themselves, which means they will be abstract or representative types and not metaphysical selves. All of this enables the original position to exclude threats of force, coercion, deception and fraud (JP 235).

Rawls claims that given these restrictions we have a conception of a decision procedure which is fair and principled. While it does favor one scheme of principles over the others, it does so on the basis that the overall balance of reasons is in its favor. Without this device the principles of justice would merely reflect the blind and arbitrary preferences of a society, which are subject to change by the calculus of social interests. It is in this sense that we are to understand that Rawls' theory is 'Kantian by analogy.' Clearly he rejects Kant's theoretical claims, but he retains the importance of determining a conception of right that can unify and generalize our considered convictions in a principled way and that will achieve greater mutual agreement and self-understanding (JP 238-239).

This last point uncovers what lies at the heart of this misinterpretation of justice as fairness. Rawls' critics regard the original position as a device of abstraction whose function is to articulate that vision of the good which all rational and autonomous persons must accept. The conclusions produced are binding on the contractual parties. But as we have seen, Rawls makes no such claim. Since the original position is only a device of representation and the persons in it are not actual persons, the resulting agreement is merely hypothetical and nonbinding (JP 238). It is only preparatory for underwriting in a principled way, the consensus based on real interests that constitutes an agreement on a political conception of justice (JP 246-247). The original position represents that neutral ground to which we can retreat when our various and conflicting views of the good make agreement impossible. It specifies those terms of social cooperation which will be acceptable to the competing conceptions of the good. As a result contractual parties can be justified in their conviction that:

> within the scope allowed by the basic liberties and the other provisions of a just constitution, all citizens can pursue their way of life on fair terms and properly respect its (non-public) values. So long as those constitutional guarantees are secure, they think no conflict of values is likely to arise that would justify their opposing the political conception as a whole, or on such fundamental matters as liberty of conscience, or equal political liberties, or basic civil rights, and the like (IOC 16).

The original position, then, is a necessary, but not a sufficient condition for justifying a political conception of justice. This is the reason why Rawls develops an account of the overlapping consensus in his theory. A political conception can never be fully accepted unless it proves itself congruent with a variety of conflicting conceptions of the good, both in theory and in practice. The overlapping consensus is itself only a necessary condition. Its purpose is to show how the principles of justice are consistent with the various social interests that one encounters in one's actual life. Here is where the allegiance to a conception of justice is generated. But without a device like the original position that allegiance would lack any principled justification for itself. It would reflect the blind and arbitrary preferences of the citizens which would be subject to change given the variability of the calculus of social interests. Thus the original position and the overlapping consensus constitutes the jointly sufficient conditions for laying the foundations for a society which is neither a 'modus vivendi' nor a general, comprehensive moral order, both of which have the tendency of employing coercion in unconscionable ways. Justice as fairness offers a principled means of ordering the values of a democratic regime, specifying the aims of the constitution and the limits it must respect (IOC I).

In conclusion it is now clear that the original position and the overlapping consensus are not alternative patterns of justification. Nor has Rawls abandoned any essential features of the justification of justice as fairness. We must read him as claiming that justification is a complex, two-stage process (much like that of John Locke's *Second Treatise*).[4] This version of Rawls' theory is much more powerful and formidable. The principles of justice are now seen to mediate the demands of two very different levels of consideration. It not only provides the theoretic grounds for choosing the principles of justice. It also can claim allegiance from the wide and diverse range of interests that constitutes a democratic society. Maintaining the analogy to Kant, justice as fairness is not only acceptable in theory, but in practice as well.

---

[4] I am grateful to Rex Martin for pointing this out to me. One could then view the original position as the stage of social contract. This is the point in Locke where the parties in the state of nature come together and specify the terms of the contract. The overlapping consensus would then be viewed as the stage where a constitution is developed and ratified by the various parties.

REFERENCES

Larmore, Charles E. *Patterns of Moral Complexity*. Cambridge: Cambridge University Press, 1987. (Cited as Larmore.)

Rawls, John. "Justice as Fairness: Political not Metaphysical." *Philosophy and Public Affairs* 14, 3 (Summer 1985): 223-51. (Cited as JP.)

Rawls, John. "The Idea of an Overlapping Consensus." *Oxford Journal of Legal Studies* vol. 7, no. 1. Oxford: Oxford University Press, 1987, pp. 1-25. (Cited as IOC.)

Sandel, Michael. *Liberalism and the Limits of Justice*. Cambridge: Cambridge University Press, 1982. (Cited as Sandel.)