

10 Rational Choice and Fairness: John Rawls, *A Theory of Justice**

Our next selection brings us down to the more recent past with an extract from the magnum opus of the influential Harvard philosopher John Rawls. Rawls's theory of justice has something in common with the Kantian approach to ethics (see extract 5, above), since it lays stress on asking what moral maxims could be rationally *chosen* or *willed* to operate in society. It also makes use of the idea of a hypothetical contract, familiar from traditional political theory (see below, Part X, extracts 3 and 4). The central idea of 'justice as fairness', as Rawls calls his

theory, is that we are to imagine having to decide what principles we would choose to see adopted in society, from a hypothetical 'original position' in which each of us is ignorant of certain crucial facts about our personal characteristics and our actual situation in society. Requiring the choice of principles to be made behind this 'veil of ignorance' is a way of ensuring that we make our decisions in an impartial and fair way. For example, if I already know I am a white male, I might favour arrangements which favour whites over blacks, or men over women; but from

¹ An argument based merely on defects in the opponent's position.

* J. Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1972), ch. 1, sections 3 and 4; with omissions.

behind the 'veil of ignorance' I will tend to avoid such unjust systems, since I might myself turn out to be a member of the disadvantaged group.

The problem of what *is* the rational choice under conditions of uncertainty is one that has much exercised both Rawls and his critics. Some have suggested that in Rawls's original position it would be rational to vote for whatever system produces the greatest good for all – in which case the theory turns out to generate yet another version of utilitarianism (see extracts 6 and 7, above). But Rawls argues that it would be most rational for the parties to select two principles of justice: first, that there be maximum equality in the assignment of basic liberties and duties; and second, that economic inequalities be permitted only if on balance they benefit the least advantaged members of society. The second principle would no doubt require some sacrifices on the part of the better off or more fortunate; but for the supporters of Rawls it is a crucial part of the appeal of his theory that it requires us to choose the principles of justice so as to counterbalance (at least in part) those accidents of natural endowment (intelligence, gender, race, etc.) that are, as Rawls puts it, 'arbitrary from the moral point of view'.

In the final part of our extract, important questions of methodology are addressed. Rawls

does not claim to arrive at his conception of the original position out of nowhere: it is partly designed to generate results that accord with our pre-reflective intuitions about morality and justice. As Rawls envisages it, however, developing a moral theory may require us on the one hand to discard some of our intuitions, where these cannot be fitted into the constraints of the theory, and on the other hand to modify the theory when it conflicts with too many of our central intuitions. Such a process of mutual adjustment continues until we reach the state of 'reflective equilibrium' – the (adjusted) theory is in balance with our (corrected) intuitions. Irrespective of the merits or otherwise of Rawls's account of justice, the notion of reflective equilibrium has stimulated considerable debate about the nature of moral theory: is it in the end just a descriptive endeavour, aimed merely at systematizing our existing intuitions, or can it aspire to lay down prescriptive rules about how human society should be regulated? The Rawlsian idea of pure rational choice under conditions of uncertainty also raises important questions about the extent to which moral theorizing can operate in abstraction from the particularities of our history and culture – a question that will be taken up in extract 11, below.

The main idea of the theory of justice



My aim is to present a conception of justice which generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau and Kant. In order to do this we are not to think of the original contract as one to enter a particular society or to set up a particular form of government. Rather, the guiding idea is that the principles of justice for the basic structure of society are the object of the original agreement. They are the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association. These principles are to regulate all further agreements; they specify the kinds of social cooperation that can be entered into and the forms of government that can be established. This way of regarding the principles of justice I shall call justice as fairness.

Thus we are to imagine that those who engage in social cooperation choose together, in one joint act, the principles which are to assign basic rights and duties and to determine the division of social benefits. Men are to decide in advance how they are to regulate their claims against one another and what is to be the foundation

charter of their society. Just as each person must decide by rational reflection what constitutes his good, that is, the system of ends which it is rational for him to pursue, so a group of persons must decide once and for all what is to count among them as just and unjust. The choice which rational men would make in this hypothetical situation of equal liberty, assuming for the present that this choice problem has a solution, determines the principles of justice.

In justice as fairness the original position of equality corresponds to the state of nature in the traditional theory of the social contract. This original position is not, of course, thought of as an actual historical state of affairs, much less as a primitive condition of culture. It is understood as a purely hypothetical situation characterized so as to lead to a certain conception of justice. Among the essential features of this situation is that no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength and the like. I shall even assume that the parties do not know their conceptions of the good or their special psychological propensities. The principles of justice are chosen behind a veil of ignorance. This ensures that no one is advantaged or disadvantaged in the choice of principles by the outcome of natural chance or the contingency of social circumstances. Since all are similarly situated and no one is able to design principles to favour his particular condition, the principles of justice are the result of a fair agreement or bargain. For given the circumstances of the original position, the symmetry of everyone's relations to each other, this initial situation is fair between individuals as moral persons, that is, as rational beings with their own ends and capable, I shall assume, of a sense of justice. The original position is, one might say, the appropriate initial status quo, and thus the fundamental agreements reached in it are fair. This explains the propriety of the name 'justice as fairness' – it conveys the idea that the principles of justice are agreed to in an initial situation that is fair. The name does not mean that the concepts of justice and fairness are the same, any more than the phrase 'poetry as metaphor' means that the concepts of poetry and metaphor are the same.

Justice as fairness begins, as I have said, with one of the most general of all choices which persons might make together, namely, with the choice of the first principles of a conception of justice which is to regulate all subsequent criticism and reform of institutions. Then, having chosen a conception of justice, we can suppose that they are to choose a constitution and a legislature to enact laws, and so on, all in accordance with the principles of justice initially agreed upon. Our social situation is just if it is such that by this sequence of hypothetical agreements we would have contracted into the general system of rules which defines it. Moreover, assuming that the original position does determine a set of principles (that is, that a particular conception of justice would be chosen), it will then be true that whenever social institutions satisfy these principles those engaged in them can say to one another that they are cooperating on terms to which they would agree if they were free and equal persons whose relations with respect to one another were fair. They could all view their arrangements as meeting the stipulations which they would acknowledge in an initial situation that embodies widely accepted and reasonable constraints on the choice of principles. The general recognition of this fact would provide the basis for a public acceptance of the corresponding principles of justice. No society can, of course, be a scheme of cooperation which men enter voluntarily in a literal sense; each person finds himself

placed at birth in some particular position in some particular society, and the nature of this position materially affects his life prospects. Yet a society satisfying the principles of justice as fairness comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are autonomous and the obligations they recognize self-imposed.

One feature of justice as fairness is to think of the parties in the initial situation as rational and mutually disinterested. This does not mean that the parties are egoists, that is, individuals with only certain kinds of interests, say in wealth, prestige and domination. But they are conceived as not taking an interest in one another's interests. They are to presume that even their spiritual aims may be opposed in the way that the aims of those of different religions may be opposed. Moreover, the concept of rationality must be interpreted as far as possible in the narrow sense, standard in economic theory, of taking the most effective means to given ends...

In working out the conception of justice as fairness one main task clearly is to determine which principles of justice would be chosen in the original position. To do this we must describe this situation in some detail and formulate with care the problem of choice which it presents... It may be observed, however, that once the principles of justice are thought of as arising from an original agreement in a situation of equality, it is an open question whether the principle of utility would be acknowledged. Offhand it hardly seems likely that persons who view themselves as equals, entitled to press their claims upon one another, would agree to a principle which may require lesser life prospects for some simply for the sake of a greater sum of advantages enjoyed by others. Since each desires to protect his interests, his capacity to advance his conception of the good, no one has a reason to acquiesce in an enduring loss for himself in order to bring about a greater net balance of satisfaction. In the absence of strong and lasting benevolent impulses, a rational man would not accept a basic structure merely because it maximized the algebraic sum of advantages irrespective of its permanent effects on his own basic rights and interests. Thus it seems that the principle of utility is incompatible with the conception of social cooperation among equals for mutual advantage. It appears to be inconsistent with the idea of reciprocity implicit in the notion of a well-ordered society. Or, at any rate, so I shall argue.

I shall maintain instead that the persons in the initial situation would choose two rather different principles: the first requires equality in the assignment of basic rights and duties, while the second holds that social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society. These principles rule out justifying institutions on the grounds that the hardships of some are offset by a greater good in the aggregate. It may be expedient but it is not just that some should have less in order that others may prosper. But there is no injustice in the greater benefits earned by a few provided that the situation of persons not so fortunate is thereby improved. The intuitive idea is that since everyone's well-being depends upon a scheme of cooperation without which no one could have a satisfactory life, the division of advantages should be such as to draw forth the willing cooperation of everyone taking part in it, including those less well situated. Yet this can be expected only if reasonable terms are proposed. The two principles mentioned

seem to be a fair agreement on the basis of which those better endowed, or more fortunate in their social position, neither of which we can be said to deserve, could expect the willing cooperation of others when some workable scheme is a necessary condition of the welfare of all. Once we decide to look for a conception of justice that nullifies the accidents of natural endowment and the contingencies of social circumstance as counters in quest for political and economic advantage, we are led to these principles. They express the result of leaving aside those aspects of the social world that seem arbitrary from a moral point of view...