



Rawls on Justice

Thomas Nagel

The Philosophical Review, Vol. 82, No. 2. (Apr., 1973), pp. 220-234.

Stable URL:

<http://links.jstor.org/sici?sici=0031-8108%28197304%2982%3A2%3C220%3AR0J%3E2.0.CO%3B2-9>

The Philosophical Review is currently published by Cornell University.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/sageschool.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

RAWLS ON JUSTICE

A *THEORY OF JUSTICE*¹ is a rich, complicated, and fundamental work. It offers an elaborate set of arguments and provides many issues for discussion. This review will focus on its contribution to the more abstract portions of ethical theory.

The book contains three elements. One is a vision of men and society as they should be. Another is a conception of moral theory. The third is a construction that attempts to derive principles expressive of the vision, in accordance with methods that reflect the conception of moral theory. In that construction Rawls has pursued the contractarian tradition in moral and political philosophy. His version of the social contract, a hypothetical choice situation called the original position, was first presented in 1958 and is here developed in great and explicit detail. The aim is to provide a way of treating the basic problems of social choice, for which no generally recognized methods of precise solution exist, through the proxy of a specially constructed parallel problem of individual choice, which can be solved by the more reliable intuitions and decision procedures of rational prudence.

If this enterprise is to succeed, and the solution to the clearer prudential problem is to be accepted as a solution to the more obscure moral one, then the alleged correspondence between the two problems must bear a great deal of weight. Critics of the theory have tended to take issue with Rawls over what principles would be chosen in the original position, but it is also necessary to examine those features of the position that are thought to support the most controversial choices and to ask why the results of a decision taken under these highly specific and rather peculiar conditions should confirm the justice of the principles chosen. This doctrine of correspondence is both fundamental and obscure, and its defense is not easy to extract from the book. A proper treatment of the subject will have to cover considerable ground, and it is probably best to begin with Rawls's moral epistemology.

Rawls believes that it will be more profitable to investigate the foundations of ethics when there are more substantive ethical results

¹ *A Theory of Justice*. By John Rawls. (Cambridge, Mass., Harvard University Press, 1971. Pp. xv, 607.)

to seek the foundations of. Nevertheless, in Section 9 he expounds a general position that helps to explain his method of proceeding. Ethics, he says, cannot be derived from self-evident axioms nor from definitions, but must be developed, like any other scientific subject, through the constant interaction between theoretical construction and particular observation. In this case, the particular observations are not experiments but substantive moral judgments. It is a bit like linguistics: ethics explores our moral sense as grammar explores our linguistic competence.²

Intuitionism attempts to capture the moral sense by summarizing our particular moral intuitions in principles of maximum generality, relying on further intuitions to settle conflicts among those principles. This is not what Rawls means. He intends rather that the underlying principles should possess intuitive moral plausibility of their own, and that the total theory should not merely summarize but illuminate and make plausible the particular judgments that it explains. Moreover, its intrinsic plausibility may persuade us to modify or extend our intuitions, thereby achieving greater theoretical coherence. Our knowledge of contingent facts about human nature and society will play a substantial part in the process.

When this interplay between general and particular has produced a relatively stable outcome, and no immediate improvements on either level suggest themselves, then our judgments are said to be in a state of *reflective equilibrium*. Its name implies that the state is always subject to change, and that our current best approximation to the truth will eventually be superseded. The indefinite article in Rawls's title is significant: he believes that all present moral theories "are primitive and have grave defects" (p. 52). His own results are provisional. "I doubt," he says (p. 581), "that the principles of justice (as I have defined them) will be the preferred conception on anything resembling a complete list."

If the principles and judgments of a theory are controversial and do not command immediate intuitive assent, then the support they receive from the underlying moral conception assumes special impor-

² This seems to me a false analogy, because the intuitions of native speakers are decisive as regards grammar. Whatever native speakers agree on is English, but whatever ordinary men agree in condemning is not necessarily wrong. Therefore the intrinsic plausibility of an ethical theory can impel a change in our moral intuitions. Nothing corresponds to this in linguistics (*pace* Rawls's suggestion on p. 49), where the final test of a theory is its ability to explain the data.

tance. To a certain extent that conception may reveal itself directly in the basic principles of the theory, but it is more clearly visible when the theory contains a model or construction that accounts for the principles and for their relation to one another. Alternative theories of justice are intuitively represented by different models (utilitarianism, for example, by the impartial sympathetic observer). Rawls's model is the original position, and the principles it is used to support are controversial. To enhance their appeal, the construction must express an intuitive idea that has independent plausibility. Before turning to the model itself, it will be useful to review briefly the substantive conclusions of the theory, identifying their controversial elements and thus the respects in which they are most in need of independent support.

Rawls's substantive doctrine is a rather pure form of egalitarian liberalism, whose controversial elements are its egalitarianism, its anti-perfectionism and anti-meritocracy, the primacy it gives to liberty, and the fact that it is more egalitarian about liberty than about other goods. The justice of social institutions is measured not by their tendency to maximize the sum or average of certain advantages, but by their tendency to counteract the natural inequalities deriving from birth, talent, and circumstance, pooling those resources in the service of the common good. The common good is measured in terms of a very restricted, basic set of benefits to individuals: personal and political liberty, economic and social advantages, and self-respect.

The justice of institutions depends on their conformity to two principles. The first requires the greatest equal liberty compatible with a like liberty for all. The second (the difference principle) permits only those inequalities in the distribution of primary economic and social advantages that benefit everyone, in particular the worst off. Liberty is prior in the sense that it cannot be sacrificed for economic and social advantages, unless they are so scarce or unequal as to prevent the meaningful exercise of equal liberty until material conditions have improved.

The view is firmly opposed to mere equality of opportunity, which allows too much influence to the morally irrelevant contingencies of birth and talent; it is also opposed to counting a society's advanced cultural or intellectual achievements among the gains which can make sacrifice of the more primary goods just. What matters is that everyone be provided with the basic conditions for the realization of his own aims, regardless of the absolute level of achievement that may represent.

When the social and political implications of this view are worked out in detail, as is done in Part Two of the book, it is extremely appealing, but far from self-evident. In considering its theoretical basis, one should therefore ask whether the contractarian approach, realized in terms of the original position, depends on assumptions any less controversial than the substantive conclusions it is adduced to support.

The notion that a contract is the appropriate model for a theory of social justice depends on the view that it is fair to require people to submit to procedures and institutions only if, given the opportunity, they could in some sense have agreed in advance on the principles to which they must submit. That is why Rawls calls the theory "justice as fairness." (Indeed, he believes that a similar contractual basis can be found for the principles of individual morality, yielding a theory of rightness as fairness.) The fundamental attitude toward persons on which justice as fairness depends is a respect for their autonomy or freedom.³ Since social institutions are simply there and people are born into them, submission cannot be literally voluntary, but (p. 13) "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."

Before considering whether the original position embodies these conditions, we must ask why respect for the freedom of others, and the desire to make society as near to voluntary as possible, should be taken as the mainspring of the sense of justice. That gives liberty a position of great importance from the very beginning, an importance

³ Expanding on this point, Rawls submits that his view is susceptible to a Kantian interpretation, but the details of the analogy are not always convincing. See, e.g., the claim on p. 253 that the principles of justice are categorical imperatives, because the argument for them does not assume that the parties to the agreement have particular ends, but only that they desire those primary goods that it is rational to want whatever else one wants. First of all, the desire for those primary goods is not itself the motive for obeying the principles of justice in real life, but only for choosing them in the original position. Secondly, imperatives deriving from such a desire would be hypothetical and assertoric in Kant's system, not categorical. But since our adherence to the two principles is supposed to be motivated by a sense of justice growing out of gratitude for the benefits received from just institutions, the imperatives of justice as fairness would in fact appear to be hypothetical and problematic (*Foundation of the Metaphysics of Morals*, pp. 415-416 of the Prussian Academy Edition).

that it retains in the resulting substantive theory. But we must ask how the respect for autonomy by itself can be expected to yield further results as well.

When one justifies a policy on the ground that the affected parties would have (or even have) agreed to it, much depends on the reasons for their agreement. If it is motivated by ignorance or fear or helplessness or a defective sense of what is reasonable, then actual or possible prior agreement does not sanction anything. In other cases, prior agreement for the right reasons can be obtained or presumed, but it is not the agreement that justifies what has been agreed to, but rather whatever justifies the agreement itself. If, for example, certain principles would be agreed to because they are just, that cannot be what makes them just. In many cases the appeal to hypothetical prior agreement is actually of this character. It is not a final justification, not a mark of respect for autonomy, but merely a way of recalling someone to the kind of *moral* judgment he would make in the absence of distorting influences derived from his special situation.

Actual or presumable consent can be the *source* of a justification only if it is already accepted that the affected parties are to be treated as certain reasons would incline each of them to want to be treated. The circumstances of consent are designed to bring those reasons into operation, suppressing irrelevant considerations, and the fact that the choice would have been made becomes a further reason for adhering to the result.

When the interests of the parties do not naturally coincide, a version of consent may still be preserved if they are able to agree in advance on a procedure for settling conflicts. They may agree unanimously that the procedure treats them equally in relevant respects, though they would not be able to agree in advance to any of the particular distributions of advantages that it might yield. (An example would be a lottery to determine the recipient of some indivisible benefit.)

For the result of such a choice to be morally acceptable, two things must be true: (a) the choice must be unanimous; (b) the circumstances that make unanimity possible must not undermine the equality of the parties in other respects. Presumably they must be deprived of some knowledge (for example, of who will win the lottery) in order to reach agreement, but it is essential that they not be unequally deprived (as would be the case, for example, if they agreed to submit a dispute to an arbitrator who, unknown to any of them, was extremely biased).

RAWLS ON JUSTICE

The more disparate the conflicting interests to be balanced, however, the more information the parties must be deprived of to insure unanimity, and doubts begin to arise whether any procedure can be relied on to treat everyone equally in respect of the relevant interests. There is then a real question whether hypothetical choice under conditions of ignorance, as a representation of consent, can by itself provide a moral justification for outcomes that could not be unanimously agreed to if they were known in advance.

Can such a procedure be used to justify principles for evaluating the basic structure of social institutions? Clearly the preferences of individuals are so divergent that they would not voluntarily agree on a common set of principles if all were given an equal voice. According to the theory of the original position, the appeal to prior agreement can be utilized nevertheless, by requiring the hypothetical choice to be made on the basis of reasons that all men have in common, omitting those which would lead them to select different principles and institutions. By restricting the basis of the hypothetical agreement in this way, however, one may lose some of its justifying power. We must therefore look carefully at the conditions imposed on a choice in the original position. Since Rawls does not, in any case, offer an abstract argument for the contractarian approach, its defense must be found in its application.

The original position is supposed to be the most philosophically favored interpretation of a hypothetical initial status quo in which fundamental agreements would be fair. The agreements can then be appealed to in disputes over the justice of institutions. The parties have an equal voice and they choose freely: in fact, they can all arrive independently at the same conclusions. Each of us, moreover, can enter the original position at any time simply by observing its rather special restrictions on arguments, and choosing principles from that point of view.

All this is possible because the grounds of choice are severely restricted as follows. The parties are mutually disinterested—that is, neither altruistic nor envious. About their own desires they know only what is true of everyone: that they have some life plan or conception of the good and a personal commitment to certain other individuals. Whatever the details, they know these interests can be advanced by the employment of very basic primary goods under conditions of liberty. They also possess general knowledge about economics, politics, and sociology and they know that the circumstances of justice, conflicting interests and moderate scarcity, obtain. Finally,

they believe that they have a sense of justice which will help them to adhere to the principles selected, but they know enough about moral psychology to realize that their choices must take into account the strains of commitment which will be felt when the principles are actually adopted, and the importance of choosing principles that will, when put into application, evoke their own support and thereby acquire psychological stability. Everything else—their talents, their social position, even the general nature or stage of development of their particular society—is covered over with a thick veil of ignorance on the ground that it is morally irrelevant. The choice should not be influenced by social and natural contingencies that would lead some parties to press for special advantages, or give some of them special bargaining power.

Rawls contends (p. 21) that these restrictions “collect together into one conception a number of conditions on principles that we are ready upon due consideration to recognize as reasonable. . . . One argues,” he says (p. 18), “from widely accepted but weak premises to more specific conclusions. Each of the presumptions should by itself be natural and plausible; some of them may seem innocuous or even trivial. The aim of the contract approach is to establish that taken together they impose significant bounds on acceptable principles of justice.”

I do not believe that the assumptions of the original position are either weak or innocuous or uncontroversial. In fact, the situation thus constructed may not be fair. Rawls says that the aim of the veil of ignorance is “to rule out those principles that it would be rational to propose for acceptance, however little the chance of success, only if one knew certain things that are irrelevant from the standpoint of justice” (p. 18). Let us grant that the parties should be equal and should not be in possession of information which would lead them to seek advantages on morally irrelevant grounds like race, sex, parentage, or natural endowments. But they are deprived also of knowledge of their particular conception of the good. It seems odd to regard that as morally irrelevant from the standpoint of justice. If someone favors certain principles because of his conception of the good, he will not be seeking special advantages for himself so long as he does not know who in the society he is. Rather he will be opting for principles that advance the good for everyone, as defined by that conception. (I assume a conception of the good is just that, and not simply a system of tastes or preferences.) Yet Rawls appears to believe that it would be as unfair to permit people to press for the realization

RAWLS ON JUSTICE

of their conception of the good as to permit them to press for the advantage of their social class.

It is true that men's different conceptions of the good divide them and produce conflict, so allowing this knowledge to the parties in the original position would prevent unanimity. Rawls concludes that the information must be suppressed and a common idea substituted which will permit agreement without selecting any particular conception of the good. This is achieved by means of the class of primary goods that it is supposedly rational to want whatever else one wants. Another possible conclusion, however, is that the model of the original position will not work because in order to secure spontaneous unanimity and avoid the necessity of bargaining one must suppress information that is morally relevant, and moreover suppress it in a way that does not treat the parties equally.

What Rawls wishes to do, by using the notion of primary goods, is to provide an Archimedean point, as he calls it, from which choice is possible without unfairness to any of the fuller conceptions of the good that lead people to differ. A *theory* of the good is presupposed, but it is ostensibly neutral between divergent particular conceptions, and supplies a least common denominator on which a choice in the original position can be based without unfairness to any of the parties. Only later, when the principles of justice have been reached on this basis, will it be possible to rule out certain particular interests or aims as illegitimate because they are unjust. It is a fundamental feature of Rawls's conception of the fairness of the original position that it should not permit the choice of principles of justice to depend on a particular conception of the good over which the parties may differ.

The construction does not, I think, accomplish this, and there are reasons to believe that it cannot be successfully carried out. Any hypothetical choice situation which requires agreement among the parties will have to impose strong restrictions on the grounds of choice, and these restrictions can be justified only in terms of a conception of the good. It is one of those cases in which there is no neutrality to be had, because neutrality needs as much justification as any other position.

Rawls's minimal conception of the good does not amount to a weak assumption: it depends on a strong assumption of the sufficiency of that reduced conception for the purposes of justice. The refusal to rank particular conceptions of the good implies a very marked tolerance for individual inclinations. Rawls is opposed not only to

teleological conceptions according to which justice requires adherence to the principles that will maximize the good. He is also opposed to the natural position that even in a nonteleological theory what is just must depend on what is good, at least to the extent that a correct conception of the good must be used in determining what counts as an advantage and what as a disadvantage, and how much, for purposes of distribution and compensation. I interpret him as saying that the principles of justice are objective and interpersonally recognizable in a way that conceptions of the good are not. The refusal to rank individual conceptions and the reliance on primary goods are intended to insure this objectivity.

Objectivity may not be so easily achieved.⁴ The suppression of knowledge required to achieve unanimity is not equally fair to all the parties, because the primary goods are not equally valuable in pursuit of all conceptions of the good. They will serve to advance many different individual life plans (some more efficiently than others), but they are less useful in implementing views that hold a good life to be readily achievable only in certain well-defined types of social structure, or only in a society that works concertedly for the realization of certain higher human capacities and the suppression of baser ones, or only given certain types of economic relations among men. The model contains a strong individualistic bias, which is further strengthened by the motivational assumptions of mutual disinterest and absence of envy. These assumptions have the effect of discounting the claims of conceptions of the good that depend heavily on the relation between one's own position and that of others (though Rawls is prepared to allow such considerations to enter in so far as they affect self-esteem). The original position seems to presuppose not just a neutral theory of the good, but a liberal, individualistic conception according to which the best that can be wished for someone is the unimpeded pursuit of his own path, provided it does not interfere with the rights of others. The view is persuasively developed in the later portions of the book, but without a sense of its controversial character.

Among different life plans of this general type the construction is neutral. But given that many conceptions of the good do not fit into the individualistic pattern, how can this be described as a fair choice situation for principles of justice? Why should parties in the original position be prepared to commit themselves to principles that

⁴ For the ideas in this paragraph I am indebted to Mary Gibson.

RAWLS ON JUSTICE

may frustrate or contravene their deepest convictions, just because they are deprived of the knowledge of those convictions?

There does not seem to be any way of redesigning the original position to do away with a restrictive assumption of this kind. One might think it would be an improvement to allow the parties full information about everyone's preferences and conception of the good, merely depriving them of the knowledge of who they were. But this, as Rawls points out (pp. 173-174), would yield no result at all. For either the parties would retain their conceptions of the good and, choosing from different points of view, would not reach unanimity, or else they would possess no aims of their own and would be asked to choose in terms of the aims of all the people they might be—an unintelligible request which provides no basis for a unified choice, in the absence of a dominant conception. The reduction to a common ground of choice is therefore essential for the model to operate at all, and the selection of that ground inevitably represents a strong assumption.

Let us now turn to the argument leading to the choice of the two principles in the original position as constructed. The core of this argument appears in Sections 26-29, intertwined with an argument against the choice of the principle of average utility. Rawls has gone to some lengths to defend his controversial claim that in the original position it is rational to adopt the maximin rule which leads one to choose principles that favor the bottom of the social hierarchy, instead of accepting a greater risk at the bottom in return for the possibility of greater benefits at the top (as might be prudentially rational if one had an equal chance of being anyone in the society).

Rawls states (p. 154) that three conditions which make maximin plausible hold in the original position to a high degree. (1) "There must be some reason for sharply discounting estimates of . . . probabilities." (2) "The person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule." (3) "The rejected alternatives have outcomes that one can hardly accept." Let us consider these in turn.

The first condition is very important, and the claim that it holds in the original position is not based simply on a general rejection of the principle of insufficient reason (that is, the principle that where probabilities are unknown they should be regarded as equal). For one could characterize the original position in such a way that the parties would be prudentially rational to choose as if they had an

equal chance of being anyone in the society, and the problem is to see why this would be an inappropriate representation of the grounds for a choice of principles.

One factor mentioned by Rawls is that the subject matter of the choice is extremely serious, since it involves institutions that will determine the total life prospects for the parties and those close to them. It is not just a choice of alternatives for a single occasion. Now this would be a reason for a conservative choice even if one knew the relative probabilities of different outcomes. It would be irresponsible to accept even a small risk of dreadful life prospects for oneself and one's descendants in exchange for a good chance of wealth or power. But what is needed is an account of why probabilities should be totally discounted, and not just with regard to the most unacceptable outcomes. The difference principle, for example, is supposed to apply at all levels of social development, so it is not justified merely by the desire to avoid grave risks. The fact that total life prospects are involved does not seem an adequate explanation. There must be some reason against allowing probabilities (proportional, for instance, to the number of persons in each social position) to enter into the choice of distributions above an acceptable minimum. Let me stress that I am posing a question not about decision theory but about the design of the original position and the comprehensiveness of the veil of ignorance. Why should it be thought that a just solution will be reached only if these considerations are suppressed?

Their suppression is justified, I think, only on the assumption that the proportions of people in various social positions are regarded as morally irrelevant, and this must be because it is not thought acceptable to sum advantages and disadvantages over persons, so that a loss for some is compensated by a gain for others. This aspect of the design of the original position appears, therefore, to be motivated by the wish to avoid extending to society as a whole the principle of rational choice for one man. Now this is supposed to be one of the *conclusions* of the contract approach, not one of its presuppositions. Yet the constraints on choice in Rawls's version of the original position are designed to rule out the possibility of such an extension,⁵ by requiring that probabilities be discounted. I can see no way to avoid presupposing some definite view on this matter in the design of a contract situation. If that is true, then a contract approach cannot give any particular view very much support.

⁵ I.e., they do not just refuse to assume that the extension is acceptable: they assume that it is unacceptable.

Consider next the second condition. Keeping in mind that the parties in the original position do not know the stage of development of their society, and therefore do not know what minimum will be guaranteed by a maximin strategy, it is difficult to understand how an individual can know that he "cares very little, if anything, for what he might gain above the minimum." The explanation Rawls offers (p. 156) seems weak. Even if parties in the original position accept the priority of liberty, and even if the veil of ignorance leaves them with a skeletal conception of the good, it seems impossible that they should care very little for increases in primary economic and social goods above what the difference principle guarantees at any given stage of social development.

Finally, the third condition, that one should rule out certain possibilities as unacceptable, is certainly a ground for requiring a social minimum and the priority of basic personal liberties, but it is not a ground for adopting the maximin rule in that general form needed to justify the choice of the difference principle. That must rely on stronger egalitarian premises.⁶

Some of these premises reveal themselves in other parts of the argument. For example, the strongly egalitarian idea that sacrifice at the bottom is always worse than sacrifice at the top plays a central role in the appeal to strains of commitment and psychological stability. It is urged against the utilitarian alternative to the difference principle, for example, that the sacrifices utilitarianism might require would be psychologically unacceptable.

⁶ A factor not considered in Rawls's argument, which suggests that the difference principle may be too weak, is the following. If differential social and economic benefits are allowed to provide incentives, then the people at the top will tend to be those with certain talents and abilities, and the people at the bottom, even though they are better off than they would be otherwise, will tend to lack those qualities. Such a consistent schedule of rewards inevitably affects people's sense of their intrinsic worth, and any society operating on the difference principle will have a meritocratic flavor. This is very different from the case where an unequal distribution that benefits the worst off is not visibly correlated with any independent qualities. Rawls does suggest (p. 546) that "excusable envy" may be given its due in the operation of the difference principle by including self-esteem among the primary goods. But he does not stress the *bases* of income inequality. The phenomenon I have described is not *envy*. Rawls is too willing to rely on equal liberty as the support of self-esteem; this leads him to underrate the effect of differential rewards on people's conception of themselves. A reward that is consistently attached to a certain quality stops being perceived as mere good luck.

The principles of justice apply to the basic structure of the social system and to the determination of life prospects. What the principle of utility asks is precisely a sacrifice of these prospects. We are to accept the greater advantages of others as a sufficient reason for lower expectations over the whole course of our life. This is surely an extreme demand. In fact, when society is conceived as a system of cooperation designed to advance the good of its members, it seems quite incredible that some citizens should be expected, on the basis of political principles, to accept lower prospects of life for the sake of others [p. 178].

Notice that if we substitute the words "difference principle" for "principle of utility," we get an argument that might be offered against the difference principle by someone concentrating on the sacrifices it requires of those at the top of the social order. They must live under institutions that limit their life prospects unless an advantage to them also benefits those beneath them. The only difference between the two arguments is in the relative position of the parties and of their sacrifices.⁷ It is of course a vital difference, but that depends on a moral judgment—namely, that sacrifices which lessen social inequality are acceptable while sacrifices which increase inequality are not.

This appeal to psychological stability and the strains of commitment therefore adds to the grounds of choice in the original position a moral view that belongs to the substantive theory. The argument may receive some support from Rawls's idea about the natural development of moral sentiments, but they in turn are not independent of his ethical theory. If a hypothetical choice in the original position must be based on what one can expect to find morally acceptable in real life, then that choice is not the true ground of acceptability.⁸

⁷ Exactly the same sacrifice could, after all, be either at the bottom or at the top, depending on the stage of advancement of the society.

⁸ A similar objection could be made to Rawls's claim that the difference principle provides a condition of reciprocal advantage that allows everyone to co-operate willingly in the social order. Obviously, those at the bottom could not prefer any other arrangement, but what about those at the top? Rawls says the following:

"To begin with, it is clear that the well-being of each depends on a scheme of social cooperation without which no one could have a satisfactory life. Secondly, we can ask for the willing cooperation of everyone only if the terms of the scheme are reasonable. The difference principle, then, seems to be a fair basis on which those better endowed, or more fortunate in their social circumstances, could expect others to collaborate with them when some workable arrangement is a necessary condition of the good of all" [p. 103]. But if some scheme of social cooperation is necessary for *anyone* to have a satisfactory life, everyone will benefit from a wide range of schemes. To assume that the worst off need further benefits to co-operate willingly while the best off do not is simply to repeat the egalitarian principle.

Another strong conclusion of the theory is the priority of equal liberty, expressed in the lexical ordering of the two principles. The argument for *equal* liberty as a natural goal is straightforward. No analogue of the difference principle can apply permanently to liberty because it cannot be indefinitely increased. There will come a point at which increases in the liberty of the worst off can be achieved not by further increasing the liberty of the best off, but only by closing the gap. If one tries to maximize for everyone what really has a maximin, the result is equality.

The priority of liberty over other goods, however, is chosen in the original position on the basis of a judgment that the fundamental interest in determining one's plan of life assumes priority once the most basic material needs have been met, and that further increases in other goods depend for their value primarily on the ability to employ them under conditions of maximum liberty. "Thus the desire for liberty is the chief regulative interest that the parties must suppose they all will have in common in due course. The veil of ignorance forces them to abstract from the particulars of their plans of life thereby leading to this conclusion. The serial ordering of the two principles then follows" (p. 543). The parties also reflect that equal liberty guarantees them all a basic self-esteem against the background of which some differences in social position and wealth will be acceptable. Here again an explicitly liberal conception of individual good is used to defend a choice in the original position.

I have attempted to argue that the presumptions of the contract method Rawls employs are rather strong, and that the original position therefore offers less independent support to his conclusions than at first appears. The egalitarian liberalism which he develops and the conception of the good on which it depends are extremely persuasive, but the original position serves to model rather than to justify them. The contract approach allied with a non-liberal conception of the good would yield different results, and some conceptions of the good are incompatible with a contract approach to justice altogether. I believe that Rawls's conclusions can be more persuasively defended by direct moral arguments for liberty and equality, some of which he provides and some of which are indirectly represented in his present account through the grounds and conditions of choice in the original position. He remarks that it is worth

noting from the outset that justice as fairness, like other contract views, consists of two parts: (1) an interpretation of the initial situation and of the problem of choice posed there, and (2) a set of principles which, it is

THOMAS NAGEL

argued, would be agreed to. One may accept the first part of the theory (or some variant thereof), but not the other, and conversely [p. 15].

He suggests that the principles are more likely to be rejected than their contractual basis, but I suspect the reverse. It seems to me likely that over the long term this book will achieve its permanent place in the literature of political theory because of the substantive doctrine that it develops so eloquently and persuasively. The plausibility of the results will no doubt be taken to confirm the validity of the method, but such inferences are not always correct. It is possible that the solution to the combinatorial problems of social choice can be reached by means of a self-interested individual choice under carefully specified conditions of uncertainty, but the basis of such a solution has yet to be discovered.

This is already a famous and influential book, and inevitably for a certain time it will engage the attention of students of philosophy, politics, law, and economics. The longer life of a work and its broader impact on the habits of thought of reflective persons can never be predicted with certainty, but it is an interesting question. Although *A Theory of Justice* is for the most part very readable, it does not possess the literary distinction that has helped to make other important political works—those of Hobbes or Mill, for example—part of the common intellectual property of mankind. It does, however, possess another feature of great importance. Reading it is a powerful experience, because one is in direct contact at every point with a striking temperament and cast of mind. It is in that sense a very personal work, and the perceptions and attitudes one finds in it are vivid, intelligent, and appealing. The outlook expressed by this book is not characteristic of its age, for it is neither pessimistic nor alienated nor angry nor sentimental nor utopian. Instead it conveys something that today may seem incredible: a hopeful affirmation of human possibilities. Yet the hope has a basis, for Rawls possesses a deep sense of the multiple connections between social institutions and individual psychology. Without illusion he describes a pluralistic social order that will call forth the support of free men and evoke what is best in them. To have made such a vision precise, alive, and convincing is a memorable achievement.

THOMAS NAGEL

Princeton University