

thinking with respect to the proper relationships between men and women.

Both Rawls and Rand are trying to capture what an ideally just society would look like. But the societies in which we live are not ideally just, and many of them are characterized by gross injustices. What should be our response to these injustices? Frantz Fanon argues in the selection from his *The Wretched of the Earth* (1963), that the oppressed of the world must meet violence with violence. During the French-Algerian War, Fanon, a psychiatrist, was assigned to a hospital in Algeria, and soon found his sympathies were with the lot of the rebels. So he joined the revolution and became one of its most articulate spokespersons.

According to Fanon, colonialism is violence in its natural state, and it will only yield when confronted with greater violence. Fanon claims that decolonization is putting into practice the biblical saying that the last shall be first and the first last.

As you can see, Fanon's approach to dealing with unjust and violent institutions is radically different from Gandhi's (see "Satyagraha"). Both Fanon and Gandhi were dealing with oppressive colonial systems, Fanon in Algeria and Gandhi primarily in India, but their approaches for dealing with these oppressive systems were radically opposed. Were the colonial systems they faced so different as to require their different approaches, or is one of their approaches simply better than the other? What do you think?

One possibility for reconciling Gandhi and Fanon is to view them as offering approaches for confronting injustice that must be tried in sequence. First, Gandhi's nonviolent approach must be given a chance to succeed, and only if that fails, would then the use of Fanon's violent approach be justified.

<sup>1</sup> John Hospers, *Libertarianism* (Los Angeles: Nash, 1971); Douglas Den Uyl and Douglas Rasmussen, *The Philosophic Thought of Ayn Rand* (Chicago: University of Illinois Press, 1984).

## A Theory of Justice

(1971)

JOHN RAWLS

### JUSTICE AS RATIONAL CHOICE BEHIND A VEIL OF IGNORANCE

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.<sup>1</sup> 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.<sup>2</sup> 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 favor 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 gen-

eral 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. I shall modify this concept to some extent . . . , but one must try to avoid introducing into it any controversial ethical elements. The initial situation must be characterized by stipulations that are widely accepted.

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.<sup>3</sup> 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.

The problem of the choice of principles, however, is extremely difficult. I do not expect the answer I shall suggest to be convincing to everyone. It is, therefore, 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 argued, would be agreed to. One may accept the first part of the theory (or some variant thereof), but not the other, and conversely. The concept of the initial contractual situation may seem reasonable although the particular principles proposed are rejected. To be sure, I want to maintain that the most appropriate conception of this situation does lead to principles of justice contrary to utilitarianism and perfectionism, and therefore that the contract doctrine provides an alternative to these views. Still, one may dispute this contention even though one grants that the contractarian method is a useful way of studying ethical theories and of setting forth their underlying assumptions.

Justice as fairness is an example of what I have called a contract theory. Now there may be an objection to the term "contract" and related expressions, but I think it will serve reasonably well. Many words have misleading connotations which at first are likely to confuse. The terms "utility" and "utilitarianism" are surely no exception. They too have unfortunate suggestions which hostile critics have been willing to exploit;

yet they are clear enough for those prepared to study utilitarian doctrine. The same should be true of the term "contract" applied to moral theories. As I have mentioned, to understand it one has to keep in mind that it implies a certain level of abstraction. In particular, the content of the relevant agreement is not to enter a given society or to adopt a given form of government, but to accept certain moral principles. Moreover, the undertakings referred to are purely hypothetical: a contract view holds that certain principles would be accepted in a well-defined initial situation.

The merit of the contract terminology is that it conveys the idea that principles of justice may be conceived as principles that would be chosen by rational persons, and that in this way conceptions of justice may be explained and justified. The theory of justice is a part, perhaps the most significant part, of the theory of rational choice. Furthermore, principles of justice deal with conflicting claims upon the advantages won by social cooperation; they apply to the relations among several persons or groups. The word "contract" suggests this plurality as well as the condition that the appropriate division of advantages must be in accordance with principles acceptable to all parties. The condition of publicity for principles of justice is also connoted by the contract phraseology. Thus, if these principles are the outcome of an agreement, citizens have a knowledge of the principles that others follow. It is characteristic of contract theories to stress the public nature of political principles. Finally there is the long tradition of the contract doctrine. Expressing the tie with this line of thought helps to define ideas and accords with natural piety. There are then several advantages in the use of the term "contract." With due precautions taken, it should not be misleading.

A final remark. Justice as fairness is not a complete contract theory. For it is clear that the contractarian idea can be extended to the choice of more or less an entire ethical system; that is, to a system including principles for all the virtues

and not only for justice. Now for the most part I shall consider only principles of justice and others closely related to them; I make no attempt to discuss the virtues in a systematic way. Obviously if justice as fairness succeeds reasonably well, a next step would be to study the more general view suggested by the name "rightness as fairness." But even this wider theory fails to embrace all moral relationships, since it would seem to include only our relations with other persons and to leave out of account how we are to conduct ourselves toward animals and the rest of nature. I do not contend that the contract notion offers a way to approach these questions, which are certainly of the first importance; and I shall have to put them aside. We must recognize the limited scope of justice as fairness and of the general type of view that it exemplifies. How far its conclusions must be revised once these other matters are understood cannot be decided in advance.

### *The Original Position and Justification*

I have said that the original position is the appropriate initial status quo which insures that the fundamental agreements reached in it are fair. This fact yields the name "justice as fairness." It is clear, then, that I want to say that one conception of justice is more reasonable than another, or justifiable with respect to it, if rational persons in the initial situation would choose its principles over those of the other for the role of justice. Conceptions of justice are to be ranked by their acceptability to persons so circumstanced. Understood in this way the question of justification is settled by working out a problem of deliberation: we have to ascertain which principles it would be rational to adopt given the contractual situation. This connects the theory of justice with the theory of rational choice.

If this view of the problem of justification is to succeed, we must, of course, describe in some detail the nature of this choice problem. A problem of rational decision has a definite answer only if we know the beliefs and interests of the parties,

their relations with respect to one another, the alternatives between which they are to choose, the procedure whereby they make up their minds, and so on. As the circumstances are presented in different ways, correspondingly different principles are accepted. The concept of the original position, as I shall refer to it, is that of the most philosophically favored interpretation of this initial choice situation for the purposes of a theory of justice.

But how are we to decide what is the most favored interpretation? I assume, for one thing, that there is a broad measure of agreement that principles of justice should be chosen under certain conditions. To justify a particular description of the initial situation one shows that it incorporates these commonly shared presumptions. One argues 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. The ideal outcome would be that these conditions determine a unique set of principles; but I shall be satisfied if they suffice to rank the main traditional conceptions of social justice.

One should not be misled, then, by the somewhat unusual conditions which characterize the original position. The idea here is simply to make vivid to ourselves the restrictions that it seems reasonable to impose on arguments for principles of justice, and therefore on these principles themselves. Thus it seems reasonable and generally acceptable that no one should be advantaged or disadvantaged by natural fortune or social circumstances in the choice of principles. It also seems widely agreed that it should be impossible to tailor principles to the circumstances of one's own case. We should ensure further that particular inclinations and aspirations, and persons' conceptions of their good, do not affect the principles adopted. The aim 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. For example, if a man knew that he was wealthy, he might find it rational to advance the principle that various taxes for welfare measures be counted unjust; if he knew that he was poor, he would most likely propose the contrary principle. To represent the desired restrictions one imagines a situation in which everyone is deprived of this sort of information. One excludes the knowledge of those contingencies which sets men at odds and allows them to be guided by their prejudices. In this manner the veil of ignorance is arrived at in a natural way. This concept should cause no difficulty if we keep in mind the constraints on arguments that it is meant to express. At any time we can enter the original position, so to speak, simply by following a certain procedure; namely, by arguing for principles of justice in accordance with these restrictions.

It seems reasonable to suppose that the parties in the original position are equal. That is, all have the same rights in the procedure for choosing principles; each can make proposals, submit reasons for their acceptance, and so on. Obviously the purpose of these conditions is to represent equality between human beings as moral persons, as creatures having a conception of their good and capable of a sense of justice. The basis of equality is taken to be similarity in these two respects. Systems of ends are not ranked in value, and each man is presumed to have the requisite ability to understand and to act upon whatever principles are adopted. Together with the veil of ignorance, these conditions define the principles of justice as those which rational persons concerned to advance their interests would consent to as equals when none are known to be advantaged or disadvantaged by social and natural contingencies.

There is, however, another side to justifying a particular description of the original position. This is to see if the principles which would be chosen match our considered convictions of jus-

tice or extend them in an acceptable way. We can note whether applying these principles would lead us to make the same judgments about the basic structure of society which we now make intuitively and in which we have the greatest confidence; or whether, in cases where our present judgments are in doubt and given with hesitation, these principles offer a resolution which we can affirm on reflection. There are questions which we feel sure must be answered in a certain way. For example, we are confident that religious intolerance and racial discrimination are unjust. We think that we have examined these things with care and have reached what we believe is an impartial judgment not likely to be distorted by an excessive attention to our own interests. These convictions are provisional fixed points which we presume any conception of justice must fit. But we have much less assurance as to what is the correct distribution of wealth and authority. Here we may be looking for a way to remove our doubts. We can check an interpretation of the initial situation, then, by the capacity of its principles to accommodate our firmest convictions and to provide guidance where guidance is needed.

In searching for the most favored description of this situation we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial

situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium.<sup>4</sup> It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation. At the moment everything is in order. But this equilibrium is not necessarily stable. It is liable to be upset by further examination of the conditions which should be imposed on the contractual situation and by particular cases which may lead us to revise our judgments. Yet for the time being we have done what we can to render coherent and to justify our convictions of social justice. We have reached a conception of the original position.

I shall not, of course, actually work through this process. Still, we may think of the interpretation of the original position that I shall present as the result of such a hypothetical course of reflection. It represents the attempt to accommodate within one scheme both reasonable philosophical conditions on principles as well as our considered judgments of justice. In arriving at the favored interpretation of the initial situation there is no point at which an appeal is made to self-evidence in the traditional sense either of general conceptions or particular convictions. I do not claim for the principles of justice proposed that they are necessary truths or derivable from such truths. A conception of justice cannot be deduced from self-evident premises or conditions on principles; instead, its justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view.

A final comment. We shall want to say that certain principles of justice are justified because they would be agreed to in an initial situation of equality. I have emphasized that this original position is purely hypothetical. It is natural to ask why, if this agreement is never actually entered into, we should take any interest in these prin-

ciples, moral or otherwise. The answer is that the conditions embodied in the description of the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded to do so by philosophical reflection. Each aspect of the contractual situation can be given supporting grounds. Thus what we shall do is to collect together into one conception a number of conditions on principles that we are ready upon due consideration to recognize as reasonable. These constraints express what we are prepared to regard as limits on fair terms of social cooperation. One way to look at the idea of the original position, therefore, is to see it as an expository device which sums up the meaning of these conditions and helps us to extract their consequences. On the other hand, this conception is also an intuitive notion that suggests its own elaboration, so that led on by it we are drawn to define more clearly the standpoint from which we can best interpret moral relationships. We need a conception that enables us to envision our objective from afar: the intuitive notion of the original position is to do this for us. . . .

### *Two Principles of Justice*

I shall now state in a provisional form the two principles of justice that I believe would be chosen in the original position. In this section I wish to make only the most general comments, and therefore the first formulation of these principles is tentative. As we go on I shall run through several formulations and approximate step by step the final statement to be given much later. I believe that doing this allows the exposition to proceed in a natural way.

The first statement of the two principles reads as follows:

First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty for others.

Second: social and economic inequalities are to be arranged so that they are both (a) reason-

ably expected to be to everyone's advantage, and  
(b) attached to positions and offices open to all.

There are two ambiguous phrases in the second principle, namely "everyone's advantage" and "open to all." Determining their sense more exactly will lead to a second formulation of the principle. . . .

By way of general comment, these principles primarily apply, as I have said, to the basic structure of society. They are to govern the assignment of rights and duties and to regulate the distribution of social and economic advantages. As their formulation suggests, these principles presuppose that the social structure can be divided into two more or less distinct parts, the first principle applying to the one, the second to the other. They distinguish between those aspects of the social system that define and secure the equal liberties of citizenship and those that specify and establish social and economic inequalities. The basic liberties of citizens are, roughly speaking, political liberty (the right to vote and to be eligible for public office) together with freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person along with the right to hold (personal) property; and freedom from arbitrary arrest and seizure as defined by the concept of the rule of law. These liberties are all required to be equal by the first principle, since citizens of a just society are to have the same basic rights.

The second principle applies, in the first approximation, to the distribution of income and wealth and to the design of organizations that make use of differences in authority and responsibility, or chains of command. While the distribution of wealth and income need not be equal, it must be to everyone's advantage, and at the same time, positions of authority and offices of command must be accessible to all. One applies the second principle by holding positions open, and then, subject to this constraint, arranges social and economic inequalities so that everyone benefits.

These principles are to be arranged in a serial order with the first principle prior to the second. This ordering means that a departure from the institutions of equal liberty required by the first principle cannot be justified by, or compensated for, by greater social and economic advantages. The distribution of wealth and income, and the hierarchies of authority, must be consistent with both the liberties of equal citizenship and equality of opportunity.

It is clear that these principles are rather specific in their content, and their acceptance rests on certain assumptions that I must eventually try to explain and justify. A theory of justice depends upon a theory of society in ways that will become evident as we proceed. For the present, it should be observed that the two principles (and this holds for all formulations) are a special case of a more general conception of justice that can be expressed as follows:

All social values—liberty and opportunity, income and wealth, and the bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage.

Injustice, then, is simply inequalities that are not to the benefit of all. Of course, this conception is extremely vague and requires interpretation.

As a first step, suppose that the basic structure of society distributes certain primary goods, that is, things that every rational man is presumed to want. These goods normally have a use whatever a person's rational plan of life. For simplicity, assume that the chief primary goods at the disposition of society are rights and liberties, powers and opportunities, income and wealth. (Later on . . . the primary good of self-respect has a central place.) These are the social primary goods. Other primary goods such as health and vigor, intelligence and imagination, are natural goods; although their possession is influenced by the basic structure, they are not so directly under its control. Imagine, then, a hypothetical initial arrangement in which all the social primary goods



are equally distributed: everyone has similar rights and duties, and income and wealth are evenly shared. This state of affairs provides a benchmark for judging improvements. If certain inequalities of wealth and organizational powers would make everyone better off than in this hypothetical starting situation, then they accord with the general conception.

Now it is possible, at least theoretically, that by giving up some of their fundamental liberties men are sufficiently compensated by the resulting social and economic gains. The general conception of justice imposes no restrictions on what sort of inequalities are permissible; it only requires that everyone's position be improved. We need not suppose anything so drastic as consenting to a condition of slavery. Imagine instead that men forgo certain political rights when the economic returns are significant and their capacity to influence the course of policy by the exercise of these rights would be marginal in any case. It is this kind of exchange which the two principles as stated rule out; being arranged in serial order they do not permit exchanges between basic liberties and economic and social gains. The serial ordering of principles expresses an underlying preference among primary social goods. When this preference is rational so likewise is the choice of these principles in this order.

In developing justice as fairness I shall, for the most part, leave aside the general conception of justice and examine instead the special case of the two principles in serial order. The advantage of this procedure is that from the first the matter of priorities is recognized and an effort made to find principles to deal with it. One is led to attend throughout to the conditions under which the acknowledgment of the absolute weight of liberty with respect to social and economic advantages, as defined by the lexical order of the two principles, would be reasonable. Offhand, this ranking appears extreme and too special a case to be of much interest; but there is more justification for it than would appear at first sight. Or at any rate, so I shall maintain.

... Furthermore, the distinction between fundamental rights and liberties and economic and social benefits marks a difference among primary social goods that one should try to exploit. It suggests an important division in the social system. Of course, the distinctions drawn and the ordering proposed are bound to be at best only approximations. There are surely circumstances in which they fail. But it is essential to depict clearly the main lines of a reasonable conception of justice; and under many conditions, anyway, the two principles in serial order may serve well enough. When necessary we can fall back on the more general conception.

The fact that the two principles apply to institutions has certain consequences. Several points illustrate this. First of all, the rights and liberties referred to by these principles are those that are defined by the public rules of the basic structure. Whether men are free is determined by the rights and duties established by the major institutions of society. Liberty is a certain pattern of social forms. The first principle simply requires that certain sorts of rules, those defining basic liberties, apply to everyone equally and that they allow the most extensive liberty compatible with a like liberty for all. The only reason for circumscribing the rights defining liberty and making men's freedom less extensive than it might otherwise be is that these equal rights as institutionally defined would interfere with one another.

Another thing to bear in mind is that when principles mention persons, or require that everyone gain from an inequality, the reference is to representative persons holding the various social positions, or offices, or whatever, established by the basic structure. Thus in applying the second principle I assume that it is possible to assign an expectation of well-being to representative individuals holding these positions. This expectation indicates their life prospects as viewed from their social station. In general, the expectations of representative persons depend upon the distribution of rights and duties throughout the basic

structure. When this changes, expectations change. I assume, then, that expectations are connected: by raising the prospects of the representative man in one position we presumably increase or decrease the prospects of representative men in other positions. Since it applies to institutional forms, the second principle (or rather the first part of it) refers to the expectations of representative individuals. As I shall discuss below, neither principle applies to distributions of particular goods to particular individuals who may be identified by their proper names. The situation where someone is considering how to allocate certain commodities to needy persons who are known to him is not within the scope of the principles. They are meant to regulate basic institutional arrangements. We must not assume that there is much similarity from the standpoint of justice between an administrative allotment of goods to specific persons and the appropriate design of society. Our common sense intuitions for the former may be a poor guide to the latter.

Now the second principle insists that each person benefit from permissible inequalities in the basic structure. This means that it must be reasonable for each relevant representative man defined by this structure, when he views it as a going concern, to prefer his prospects with the inequality, to his prospects without it. One is not allowed to justify differences in income or organizational powers on the ground that the disadvantages of those in one position are outweighed by the greater advantages of those in another. Much less can infringements of liberty be counter-balanced in this way. Applied to the basic structure, the principle of utility would have us maximize the sum of expectations of representative men (weighted by the number of persons they represent, on the classical view); and this would permit us to compensate for the losses of some by the gains of others. Instead the two principles require that everyone benefit from economic and social inequalities.

### *The Reasoning Leading to the Two Principles of Justice*

It will be recalled that the general conception of justice as fairness requires that all primary social goods be distributed equally unless an unequal distribution would be to everyone's advantage. No restrictions are placed on exchanges of these goods and therefore a lesser liberty can be compensated for by greater social and economic benefits. Now looking at the situation from the standpoint of one person selected arbitrarily, there is no way for him to win special advantages for himself. Nor, on the other hand, are there grounds for his acquiescing in special disadvantages. Since it is not reasonable for him to expect more than an equal share in the division of social goods, and since it is not rational for him to agree to less, the sensible thing for him to do is to acknowledge as the first principle of justice one requiring an equal distribution. Indeed, this principle is so obvious that we would expect it to occur to anyone immediately.

Thus, the parties start with a principle establishing equal liberty for all, including equality of opportunity, as well as an equal distribution of income and wealth. But there is no reason why this acknowledgment should be final. If there are inequalities in the basic structure that work to make everyone better off in comparison with the benchmark of initial equality, why not permit them? The immediate gain which a greater equality might allow can be regarded as intelligently invested in view of its future return. If, for example, these inequalities set up various incentives which succeed in eliciting more productive efforts, a person in the original position may look upon them as necessary to cover the costs of training and to encourage effective performance. One might think that ideally individuals should want to serve one another. But since the parties are assumed not to take an interest in one another's interests, their acceptance of these inequalities is only the acceptance of the relations

in which men stand in the circumstances of justice. They have no grounds for complaining of one another's motives. A person in the original position would, therefore, concede the justice of these inequalities. Indeed, it would be shortsighted of him not to do so. He would hesitate to agree to these regularities only if he would be dejected by the bare knowledge or perception that others were better situated; and I have assumed that the parties decide as if they are not moved by envy. In order to make the principle regulating inequalities determinate, one looks at the system from the standpoint of the least advantaged representative man. Inequalities are permissible when they maximize, or at least all contribute to, the long-term expectations of the least fortunate group in society.

Now this general conception imposes no constraints on what sorts of inequalities are allowed, whereas the special conception, by putting the two principles in serial order (with the necessary adjustments in meaning), forbids exchanges between basic liberties and economic and social benefits. I shall not try to justify this ordering here. . . . But roughly, the idea underlying this ordering is that if the parties assume that their basic liberties can be effectively exercised, they will not exchange a lesser liberty for an improvement in economic well-being. It is only when social conditions do not allow the effective establishment of these rights that one can concede their limitation; and these restrictions can be granted only to the extent that they are necessary to prepare the way for a free society. The denial of equal liberty can be defended only if it is necessary to raise the level of civilization so that in due course these freedoms can be enjoyed. Thus in adopting a serial order we are in effect making a special assumption in the original position, namely, that the parties know that the conditions of their society, whatever they are, admit the effective realization of the equal liberties. The serial ordering of the two principles of justice eventually comes to be reasonable if the general

conception is consistently followed. This lexical ranking is the long-run tendency of the general view. For the most part I shall assume that the requisite circumstances for the serial order obtain.

It seems clear from these remarks that the two principles are at least a plausible conception of justice. The question, though, is how one is to argue for them more systematically. Now there are several things to do. One can work out their consequences for institutions and note their implications for fundamental social policy. In this way they are tested by a comparison with our considered judgments of justice. . . . But one can also try to find arguments in their favor that are decisive from the standpoint of the original position. In order to see how this might be done, it is useful as a heuristic device to think of the two principles as the maximin solution to the problem of social justice. There is an analogy between the two principles and the maximin rule for choice under uncertainty.<sup>5</sup> This is evident from the fact that the two principles are those a person would choose for the design of a society in which his enemy is to assign him his place. The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative the worst outcome of which is superior to the worst outcomes of the others. The persons in the original position do not, of course, assume that their initial place in society is decided by a malevolent opponent. As I note below, they should not reason from false premises. The veil of ignorance does not violate this idea, since an absence of information is not misinformation. But that the two principles of justice would be chosen if the parties were forced to protect themselves against such a contingency explains the sense in which this conception is the maximin solution. And this analogy suggests that if the original position has been described so that it is rational for the parties to adopt the conservative attitude expressed by this rule, a conclusive argument can indeed be constructed for these prin-

ciples. Clearly the maximin rule is not, in general, a suitable guide for choices under uncertainty. But it is attractive in situations marked by certain special features. My aim, then, is to show that a good case can be made for the two principles based on the fact that the original position manifests these features to the fullest possible degree, carrying them to the limit, so to speak.

Consider the gain-and-loss table below. It represents the gains and losses for a situation which is not a game of strategy. There is no one playing against the person making the decision; instead he is faced with several possible circumstances which may or may not obtain. Which circumstances happen to exist does not depend upon what the person choosing decides or whether he announces his moves in advance. The numbers in the table are monetary values (in hundreds of dollars) in comparison with some initial situation. The gain ( $g$ ) depends upon the individual's decision ( $d$ ) and the circumstances ( $c$ ). Thus  $g = f(d, c)$ . Assuming that there are three possible decisions and three possible circumstances, we might have this gain-and-loss table.

Decisions	Circumstances		
	C1	C2	C3
d1	-7	8	12
d2	-8	7	14
d3	5	6	8

The maximin rule requires that we make the third decision. For in this case the worst that can happen is that one gains five hundred dollars, which is better than the worst for the other actions. If we adopt one of these we may lose either eight or seven hundred dollars. Thus, the choice of  $d_3$  maximizes  $f(d, c)$  for that value of  $c$  which for a given  $d$ , minimizes  $f$ . The term "maximin" means the *maximum minimorum*; and the rule directs our attention to the worst that can happen under any proposed course of action, and to decide in the light of that.

Now there appear to be three chief features of situations that give plausibility to this unusual rule.<sup>6</sup> First, since the rule takes no account of the likelihoods of the possible circumstances, there must be some reason for sharply discounting estimates of these probabilities. Offhand, the most natural rule of choice would seem to be to compute the expectation of monetary gain for each decision and then to adopt the course of action with the highest prospect. (This expectation is defined as follows: let us suppose that  $g_{ij}$  represent the numbers in the gain-and-loss table, where  $i$  is the row index and  $j$  is the column index; and let  $p_j$ ,  $j = 1, 2, 3$ , be the likelihoods of the circumstances, with  $\sum p_j = 1$ . Then the expectation for the  $i$ th decision is equal to  $\sum p_j g_{ij}$ .) Thus it must be, for example, that the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure. In this case it is unreasonable not to be skeptical of probabilistic calculations unless there is no other way out, particularly if the decision is a fundamental one that needs to be justified to others.

The second feature that suggests the maximin rule is the following: 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. It is not worthwhile for him to take a chance for the sake of a further advantage, especially when it may turn out that he loses much that is important to him. This last provision brings in the third feature; namely, that the rejected alternatives have outcomes that one can hardly accept. The situation involves grave risks. Of course these features work most effectively in combination. The paradigm situation for following the maximin rule is when all three features are realized to the highest degree. This rule does not, then, generally apply, nor of course is it self-evident. Rather, it is a maxim, a rule of thumb, that comes into its own in special circumstances. Its application depends upon the qualitative structure of the possible gains and losses in relation to one's conception of

the good, all this against a background in which it is reasonable to discount conjectural estimates of likelihoods.

It should be noted, as the comments on the gain-and-loss table say, that the entries in the table represent monetary values and not utilities. This difference is significant since for one thing computing expectations on the basis of such objective values is not the same thing as computing expected utility and may lead to different results. The essential point, though, is that in justice as fairness the parties do not know their conception of the good and cannot estimate their utility in the ordinary sense. In any case, we want to go behind *de facto* preferences generated by given conditions. Therefore, expectations are based upon an index or primary goods and the parties make their choice accordingly. The entries in the example are in terms of money and not utility to indicate this aspect of the contract doctrine.

Now, as I have suggested, the original position has been defined so that it is a situation in which the maximin rule applies. In order to see this, let us review briefly the nature of this situation with these three special features in mind. To begin with, the veil of ignorance excludes all but the vaguest knowledge of likelihoods. The parties have no basis for determining the probable nature of their society, or their place in it. Thus they have strong reasons for being wary of probability calculations if any other course is open to them. They must also take into account the fact that their choice of principles should seem reasonable to others, in particular their descendants, whose rights will be deeply affected by it. There are further grounds for discounting that I shall mention as we go along. For the present it suffices to note that these considerations are strengthened by the fact that the parties know very little about the gain-and-loss table. Not only are they unable to conjecture the likelihoods of the various possible circumstances, they cannot say much about what the possible circumstances are, much less enumerate them and foresee the outcome of each

alternative available. Those deciding are much more in the dark than the illustration by a numerical table suggests. It is for this reason that I have spoken of an analogy with the maximin rule.

Several kinds of arguments for the two principles of justice illustrate the second feature. Thus, if we can maintain that these principles provide a workable theory of social justice, and that they are compatible with reasonable demands of efficiency, then this conception guarantees a satisfactory minimum. There may be, on reflection, little reason for trying to do better. Thus much of the argument . . . is to show, by their application to the main questions of social justice, that the two principles are a satisfactory conception. These details have a philosophical purpose. Moreover, this line of thought is practically decisive if we can establish the priority of liberty, the lexical ordering of the two principles. For this priority implies that the persons in the original position have no desire to try for greater gains at the expense of the equal liberties. The minimum assured by the two principles in lexical order is not one that the parties wish to jeopardize for the sake of greater economic and social advantages. . . .

Finally, the third feature holds if we can assume that other conceptions of justice may lead to institutions that the parties would find intolerable. For example, it has sometimes been held that under some conditions the utility principle (in either form) justifies, if not slavery or serfdom, at any rate serious infractions of liberty for the sake of greater social benefits. We need not consider here the truth of this claim, or the likelihood that the requisite conditions obtain. For the moment, this contention is only to illustrate the way in which conceptions of justice may allow for outcomes which the parties may not be able to accept. And having the ready alternative of the two principles of justice which secure a satisfactory minimum, it seems unwise, if not irrational, for them to take a chance that these outcomes are not realized.



So much, then, for a brief sketch of the features of situations in which the maximin rule comes into its own and of the way in which the arguments for the two principles of justice can be subsumed under them. . . .

### *The Final Formulation of the Principles of Justice*

. . . I now wish to give the final statement of the two principles of justice for institutions. For the sake of completeness, I shall give a full statement including earlier formulations.

#### *FIRST PRINCIPLE*

Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all.

#### *SECOND PRINCIPLE*

Social and economic inequalities are to be arranged so that they are both:

- a. to the greatest benefit of the least advantaged, consistent with the just savings principle, and
- b. attached to offices and positions open to all under conditions of fair equality of opportunity.

#### *FIRST PRIORITY RULE (THE PRIORITY OF LIBERTY)*

The principles of justice are to be ranked in lexical order and therefore liberty can be restricted only for the sake of liberty. There are two cases:

- a. a less extensive liberty must strengthen the total system of liberty shared by all;
- b. a less than equal liberty must be acceptable to those with the lesser liberty.

#### *SECOND PRIORITY RULE (THE PRIORITY OF JUSTICE OVER EFFICIENCY AND WELFARE)*

The second principle of justice is lexically prior to the principle of efficiency and to that of maximiz-

ing the sum of advantages; and fair opportunity is prior to the difference principle. There are two cases:

- a. an inequality of opportunity must enhance the opportunities of those with the lesser opportunity;
- b. an excessive rate of saving must on balance mitigate the burden of those bearing this hardship.

#### *GENERAL CONCEPTION*

All social primary goods—liberty and opportunity, income and wealth, and the bases of self-respect—are to be distributed equally unless an unequal distribution of any or all of these goods is to the advantage of the least favored.

By way of comment, these principles and priority rules are no doubt incomplete. Other modifications will surely have to be made, but I shall not further complicate the statement of the principles. It suffices to observe that when we come to nonideal theory, we do not fall back straightway upon the general conception of justice. The lexical ordering of the two principles, and the valuations that this ordering implies, suggest priority rules which seem to be reasonable enough in many cases. By various examples I have tried to illustrate how these rules can be used and to indicate their plausibility. Thus the ranking of the principles of justice in ideal theory reflects back and guides the application of these principles to nonideal situations. It identifies which limitations need to be dealt with first. The drawback of the general conception of justice is that it lacks the definite structure of the two principles in serial order. In more extreme and tangled instances of nonideal theory there may be no alternative to it. At some point the priority of rules for nonideal cases will fail; and indeed, we may be able to find no satisfactory answer at all. But we must try to postpone the day of reckoning as long as possible, and try to arrange society so that it never comes. . . .

### *The Kantian Interpretation*

Kant held, I believe, that a person is acting autonomously when the principles of his action are chosen by him as the most adequate possible expression of his nature as a free and equal rational being. The principles he acts upon are not adopted because of his social position or natural endowments, or in view of the particular kind of society in which he lives or the specific things that he happens to want. To act on such principles is to act heteronomously. Now the veil of ignorance deprives the persons in the original position of the knowledge that would enable them to choose heteronomous principles. The parties arrive at their choice together as free and equal rational persons knowing only that those circumstances obtain which give rise to the need for principles of justice.

To be sure, the argument for these principles does add in various ways to Kant's conception. For example, it adds the feature that the principles chosen are to apply to the basic structure of society; and premises characterizing this structure are used in deriving the principles of justice. But I believe that this and other additions are natural enough and remain fairly close to Kant's doctrine, at least when all of his ethical writings are viewed together. Assuming, then, that the reasoning in favor of the principles of justice is correct, we can say that when persons act on these principles they are acting in accordance with principles that they would choose as rational and independent persons in an original position of equality. The principles of their actions do not depend upon social or natural contingencies, nor do they reflect the bias of the particulars of their plan of life or the aspirations that motivate them. By acting from these principles persons express their nature as free and equal rational beings subject to the general conditions of human life. For to express one's nature as a being of a particular kind is to act on the principles that would be chosen if this nature were the decisive determining element. Of

course, the choice of the parties in the original position is subject to the restrictions of that situation. But when we knowingly act on the principles of justice in the ordinary course of events, we deliberately assume the limitations of the original position. One reason for doing this, for persons who can do so and want to, is to give expression to one's nature.

The principles of justice are also categorical imperatives in Kant's sense. For by a categorical imperative Kant understands a principle of conduct that applies to a person in virtue of his nature as a free and equal rational being. The validity of the principle does not presuppose that one has a particular desire or aim. Whereas a hypothetical imperative by contrast does assume this: it directs us to take certain steps as effective means to achieve a specific end. Whether the desire is for a particular thing, or whether it is for something more general, such as certain kinds of agreeable feelings or pleasures, the corresponding imperative is hypothetical. Its applicability depends upon one's having an aim which one need not have as a condition of being a rational human individual. The argument for the two principles of justice does not assume that the parties have particular ends, but only that they desire certain primary goods. These are things that it is rational to want whatever else one wants. Thus given human nature, wanting them is part of being rational; and while each is presumed to have some conception of the good, nothing is known about his final ends. The preference for primary goods is derived, then, from only the most general assumptions about rationality and the conditions of human life. To act from the principles of justice is to act from categorical imperatives in the sense that they apply to us whatever in particular our aims are. This simply reflects the fact that no such contingencies appear as premises in their derivation.

We may note also that the motivational assumption of mutual disinterest accords with Kant's notion of autonomy, and gives another

reason for this condition. So far this assumption has been used to characterize the circumstances of justice and to provide a clear conception to guide the reasoning of the parties. We have also seen that the concept of benevolence, being a second-order notion, would not work out well. Now we can add that the assumption of mutual disinterest is to allow for freedom in the choice of a system of final ends.<sup>7</sup> Liberty in adopting a conception of the good is limited only by principles that are deduced from a doctrine which imposes no prior constraints on these conceptions. Presuming mutual disinterest in the original position carries out this idea. We postulate that the parties have opposing claims in a suitably general sense. If their ends were restricted in some specific way, this would appear at the outset as an arbitrary restriction on freedom. Moreover, if the parties were conceived as altruists, or as pursuing certain kinds of pleasures, then the principles chosen would apply, as far as the argument would have shown, only to persons whose freedom was restricted to choices compatible with altruism or hedonism. As the argument now runs, the principles of justice cover all persons with rational plans of life, whatever their content, and these principles represent the appropriate restrictions on freedom. Thus it is possible to say that the constraints on conceptions of the good are the result of an interpretation of the contractual situation that puts no prior limitations on what men may desire. There are a variety of reasons, then, for the motivational premise of mutual disinterest. This premise is not only a matter of realism about the circumstances of justice or a way to make the theory manageable. It also connects up with the Kantian idea of autonomy. . . .

The original position may be viewed, then, as a procedural interpretation of Kant's conception of autonomy and the categorical imperative. The principles regulative of the kingdom of ends are those that would be chosen in this position, and the description of this situation enables us to explain the sense in which acting from these

principles expresses our nature as free and equal rational persons. No longer are these notions purely transcendent and lacking explicable connections with human conduct, for the procedural conception of the original position allows us to make these ties. . . .

## NOTES

1. As the text suggests, I shall regard Locke's *The Second Treatise of Government*, Rousseau's *Social Contract*, and Kant's ethical works beginning with *The Foundations of the Metaphysics of Morals* as definitive of the contract tradition. For all of its greatness, Hobbes's *Leviathan* raises special problems. A general historical survey is provided by J. W. Gough, *The Social Contract*, 2nd ed. (Oxford, The Clarendon Press, 1957), and Otto Gierke, *Natural Law and the Theory of Society*, trans. with an introduction by Ernest Barker (Cambridge, The University Press, 1934). A presentation of the contract view as primarily an ethical theory is to be found in G. R. Grice, *The Grounds of Moral Judgment* (Cambridge, The University Press, 1967). . . .
2. Kant is clear that the original agreement is hypothetical. See *The Metaphysics of Morals*, pt. I (*Rechtslehre*), especially §§ 47, 52; and pt. II of the essay "Concerning the Common Saying: This May Be True in Theory But It Does Not Apply in Practice," in *Kant's Political Writings*, ed. Hans Reiss and trans. by H. B. Nisbet (Cambridge, The University Press, 1970), pp. 73-87. See Georges Vlachos, *La Pensée politique de Kant* (Paris, Presses Universitaires de France, 1962), pp. 326-335; and J. G. Murphy, *Kant: The Philosophy of Right* (London, Macmillan, 1970), pp. 109-112, 133-136, for a further discussion.
3. For the formulation of this intuitive idea I am indebted to Allan Gibbard.
4. The process of mutual adjustment of principles and considered judgments is not peculiar to moral philosophy. See Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, Mass., Harvard University Press, 1955), pp. 65-68, for parallel remarks concerning the justification of the principles of deductive and inductive inference.
5. An accessible discussion of this and other rules of choice under uncertainty can be found in W. J.

Baumol, *Economic Theory and Operations Analysis*, 2nd ed. (Englewood Cliffs, N.J., Prentice Hall, 1965), ch. 24. Baumol gives a geometric interpretation of these rules, including the diagram used . . . to illustrate the difference principle. See pp. 558–562. See also R. D. Luce and Howard Raiffa, *Games and Decisions* (New York,

John Wiley and Sons, Inc., 1957), ch. XIII, for a fuller account.

6. Here I borrow from William Fellner, *Probability and Profit* (Homewood, Ill., Richard D. Irwin, 1965), pp. 140–142, where these features are noted.
7. For this point I am indebted to Charles Fried.

## The Domain of the Political and Overlapping Consensus

JOHN RAWLS

### *IV: Features of a Political Conception of Justice*

. . . [T]HE THIRD GENERAL FACT [is] that an enduring and stable democratic regime is one that at least a substantial majority of its politically active citizens freely support. Given this fact, what are the more general features of a political doctrine underlying a regime able to gain such allegiance? Plainly, it must be a doctrine that a diversity of comprehensive religious, philosophical, and moral doctrines can endorse, each from its own point of view.<sup>1</sup> This follows not only from the third general fact but also from the first, the fact of pluralism: for a democratic regime will eventually, if not from the outset, lead to a pluralism of comprehensive doctrines.

Let us say that a political conception of justice (in contrast to a political regime) is stable if it meets the following condition: those who grow up in a society well-ordered by it—a society whose institutions are publicly recognized to be just, as specified by that conception itself—develop a sufficient allegiance to those institutions, that is, a sufficiently strong sense of justice guided by appropriate principles and ideals, so that they normally act as justice requires, pro-

vided they are assured that others will act likewise.<sup>2</sup>

Now what more general features of a political conception of justice does this definition of stability suggest? The idea of a political conception of justice includes three such features.<sup>3</sup>

First, while a political conception of justice is, of course, a moral conception, it is worked out for a specific subject, namely, the basic structure of a constitutional democratic regime. This structure consists in society's main political, social, and economic institutions, and how they fit together into one unified system of social cooperation.

Second, accepting a political conception of justice does not presuppose accepting any particular comprehensive doctrine. The conception presents itself as a reasonable conception for the basic structure alone.<sup>4</sup>

Third, a political conception of justice is formulated so far as possible solely in terms of certain fundamental intuitive ideas viewed as implicit in the public political culture of a democratic society. Two examples are the idea of society as a fair system of social cooperation over time from one generation to the next, and the idea of citizens as free and equal persons fully capable of engaging