

the reach of legal punishment or control has been accounted by whole nations, and by some of the best and wisest of men, not a crime but an act of exalted virtue; and that, right or wrong, it is not of the nature of assassination, but of civil war. As such, I hold that the instigation to it, in a specific case, may be a proper subject of punishment, but only if an overt act has followed, and at least a probable connection can be established between the act and the instigation. Even then it is not a foreign government but the very government assailed which alone, in the exercise of self-defense, can legitimately punish attacks directed against its own existence.

2. Thomas Pooley, Bodmin Assizes, July 31, 1857. In December following, he received a free pardon from the Crown.

3. George Jacob Holyoake, August 17, 1857; Edward Truelove, July 1857.

4. Baron de Gleichen, Marlborough-street Police Court, August 4, 1857.

5. Ample warning may be drawn from the large infusion of the passions of a persecutor, which mingled with the general display of the worst parts of our national character on the occasion of the Sepoy insurrection. The ravings of fanatics or charlatans from the pulpit may be unworthy of notice; but the heads of the Evangelical party have announced as their principle for the government of Hindus and Mohammedans that no schools be supported

by public money in which the Bible is not taught, and by necessary consequence that no public employment be given to any but real or pretended Christians. An Under-secretary of State, in a speech delivered to his constituents on the 12th of November, 1857, is reported to have said: "Toleration of their faith" (the faith of a hundred millions of British subjects), "the superstition which they called religion, by the British Government, had had the effect of retarding the ascendancy of the British name, and preventing the salutary growth of Christianity. Toleration was the great cornerstone of the religious liberties of this country; but do not let them abuse that precious word 'toleration.' As he understood it, it means the complete liberty to all, freedom of worship, *among Christians, who worshiped upon the same foundation*. It meant toleration of all sects and denominations of *Christians who believed in the one mediation*." I desire to call attention to the fact that a man who has been deemed fit to fill a high office in the government of this country under a liberal ministry maintains the doctrine that all who do not believe in the divinity of Christ are beyond the pale of toleration. Who, after this imbecile display, can indulge the illusion that religious persecution has passed away, never to return?

6. *The Spheres and Duties of Government*, from the German of Baron Wilhelm von Humboldt, pages 11 to 13.

JOHN RAWLS



INTRODUCTION

JOSHUA COHEN

John Rawls (1921–2003) was born in Baltimore, Maryland. He graduated from Princeton University, and then, after serving in the American army in the Pacific during World War II, returned to Princeton, where he received a Ph.D. in philosophy in 1950. An influential teacher of several generations of moral and political philosophers, Rawls taught at Cornell in the 1950s, moved to The Massachusetts Institute of Technology in 1960, and then to Harvard's Philosophy Department in 1962, where he remained until his retirement in 1991. From the time he received his Ph.D. until 1971, Rawls's principal intellectual project was a book about justice. That book appeared in 1971 under the title *A Theory of Justice*, and provided a comprehensive statement of a theory that Rawls called "justice as fairness." After the book's publication, Rawls spent nearly twenty

years rethinking the foundations of justice as fairness, with the aim of making the presentation of his outlook more consistent with the religious and philosophical pluralism characteristic of modern democracies. The fruits of those labors appeared in 1993 in his *Political Liberalism* (New York: Columbia University Press). In 1999, Rawls published *The Law of Peoples* (Cambridge, Mass.: Harvard University Press), which extended his ideas about justice to the international system.

For much of the past century, the idea of an egalitarian-liberal political philosophy seemed to many a contradiction in terms. Egalitarians troubled by vast differences between the lives of rich and poor commonly condemned liberalism for paying excessive attention to legal rights and liberties while exhibiting indifference to the real fate of ordinary people. Equality, they argued, could be found only in the rarified atmosphere of liberal legal and political discourse—with its claims about equality of persons before the law and of citizens in the state—in disturbing isolation from life on earth. Liberals concerned to ensure individual rights would condemn egalitarianism for being paternalistic and willing to sacrifice human freedom in the name of a bland sameness of circumstance or future utopia. Practically speaking, democratic welfare states tried, with more or less success, to respect liberal and egalitarian values: to ensure basic individual rights to personal and political liberties while protecting individuals from the contingencies of the market. But the philosophical options seemed starkly opposed. In between Friedrich Hayek's classical liberalism and Karl Marx's egalitarianism, everything was an unstable political compromise, or an ad hoc balancing of competing values.

Rawls's *Theory of Justice* reshaped this philosophical terrain. Rawls proposed a conception of justice committed to the individual rights associated with traditional liberalism, to an egalitarian ideal of fair distribution conventionally associated with socialist and radical democratic traditions, and to a reasonable faith in the practical possibility of a form of constitutional democracy ensuring both liberty and equality. In summarizing his view, he said that justice as fairness aims to effect a "reconciliation of liberty and equality."

To appreciate the force of this reconciliation, consider the two principles of justice that Rawls explains and defends in *A Theory of Justice*. Rawls's first principle—a principle of *equal basic liberties*—says that each citizen has an equal right to the most extensive system of equal basic personal and political liberties compatible with a similar system of liberties for others. This principle does not assert a right to "liberty as such," that is, a right that would condemn restrictions on all manner of choices. Instead, the first principle requires stringent protections for certain *specific* liberties: liberty of thought and conscience; political liberty; freedom of association; liberty and integrity of the person; and the rights and liberties associated with the rule of law, and its requirements of generality and predictability. Rawls's first principle also includes a demanding norm of political equality, which holds that political liberty is to be assured a fair value—that the chance to hold office and to exercise influence on the political system ought to be independent of socioeconomic position. So citizens who have the motivation and ability to play an active political role should not be disadvantaged in their efforts by a lack of sufficient private wealth, or advantaged by their greater wealth.

Rawls's second principle of justice expresses egalitarian ideals of distributive justice. The second principle has two components, both of which set limits on acceptable socioeconomic inequalities. The first component states that when inequalities are attached to offices and positions—say, when different jobs are differently rewarded—those offices and posi-

tions must be open to everyone under conditions of *fair equality of opportunity*. In particular, people who are equally talented and motivated must have equal chances to attain desirable positions, regardless of their class background. Access to responsible and well-compensated work should not depend on the social circumstances in which people happen to have been born and raised.

But even a society that protects each person's basic personal and political liberties, and ensures fair equality of opportunity, might still have troubling inequalities. Thus, suppose some people, partly because of their native endowments, possess scarce talents that command high returns in the market, while others lack such skills. Assume people in both groups work hard, and contribute according to their abilities. Still, they will reap substantially different rewards, and those differences will have a large impact on their lives. But these inequalities of reward are founded on natural contingencies, bare undeserved luck in life's lottery, and "there is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune." So a second part of Rawls's second principle—the difference principle—requires an economic structure that mitigates inequalities in income and wealth owing to differences in natural talent. Instead of permitting differences of reward simply to reflect differences of native endowment, the difference principle requires that we maximize the lifetime expectations of those who are in the least advantaged social position. Thus someone might legitimately be paid more than someone else because the higher income compensates for expensive training and education that enable the person to take on socially desirable tasks; or inequalities might make sense as incentives that encourage people to take on tasks they would otherwise be unable or simply unwilling to take on. According to the difference principle, such inequalities are fully just only if they are to the greatest benefit of those who are least well-off.

While the requirement of fair equality of opportunity, then, condemns a society in which class background is a source of social or economic privilege, the difference principle condemns a society in which, as the sociologist Emile Durkheim put it, "social inequalities exactly express natural inequalities." In effect, then, Rawls urges us to reject the idea that our economic system is a race or talent contest, designed to reward the swift and gifted. Instead, it is one part of a fair scheme of cooperation, designed to ensure a reasonable life for all. "In justice as fairness," Rawls says, "men agree to share one another's fate. In designing institutions they undertake to avail themselves of the accidents of nature and social circumstance *only when* doing so is for the common benefit."

To see how justice as fairness reconciles liberty and equality, then, consider the joint operation of the two principles of justice. Assume that what matters to people is not only to have basic liberties that are legally protected from interferences by others, but that the legally protected liberties provide a meaningful or valuable liberty, that the liberties are worth something to us. Assume, too, that the worth of our liberty to us reflects the resources we have available for using the liberty. In particular, assume that the value of my liberties to me is an increasing function of the resources over which I exercise control: as my command of resources increases, I can do more with my liberties.

Now put the two principles together: the first principle ensures equal basic liberties, and the difference principle guarantees that the minimum level of resources is maximized. If, as I just suggested, the worth of a person's liberty—its value to the person—is an increasing

function of the level of that person's resources, then by maximizing the minimum level of resources, we also maximize the minimum worth of liberty. Thus the two principles together require that society "maximize the worth to the least advantaged of the complete scheme of equal liberty shared by all." Maximizing the minimum worth of liberty "defines," Rawls says, "the end of social justice"—what justice aims ultimately to achieve.

The central idea of justice as fairness, then, is that an egalitarian-liberal conception comprising the two principles of justice is the most reasonable conception of justice "for a democratic society." Abraham Lincoln said that the United States was conceived in liberty and dedicated to the proposition that all men are created equal. Justice as fairness argues that the two principles of justice are the most reasonable theory of justice for a society with that conception and dedication—more reasonable than, for example, utilitarianism, or libertarianism, or a less liberal egalitarianism, or a less egalitarian liberalism.

To argue for the two principles of justice over alternatives, Rawls revives the social contract idea associated with Hobbes, Locke, Rousseau, and Kant. The social contract idea is that the most reasonable ordering of a society is the ordering that the members themselves would unanimously agree to as the basis for their own association. So Rawls asks us to imagine ourselves in a hypothetical situation—he calls it "the Original Position"—in which we are to choose the principles of justice that will be used in our own society. Whereas the Hobbesian social contract was based on certain fundamental truths about human nature, the construction of the Original Position is designed to reflect the moral idea that we are free and equal moral persons—with a capacity to cooperate with others on fair terms, to choose our ends and devotions, and to pursue the ends we set for ourselves—and that the principles of justice for our society should treat us as such. So certain of our characteristics—those that distinguish among free and equal persons—are irrelevant in deciding what we are entitled to as a matter of justice. Specifically, our social class position, natural talents, sex or race, and conception of the good are irrelevant from the standpoint of justice. When we imagine ourselves making a choice of principles of justice, then, we are to imagine choosing from behind a "veil of ignorance" in which we are assumed to lack knowledge of the irrelevant features. We know that we are free and equal moral persons, but we do not know our class background, native endowments, sex or race, and conception of the good. We do not know, in short, whether the natural and social contingencies have worked in our favor.

When we reason from behind the veil of ignorance, then, we focus only on interests we share as free and equal moral persons, and put aside what distinguishes us from one another. Thus the parties in the original position know only that they represent a person who has some conception of the good (though they do not know what that conception is); an interest in being able to choose and revise their ends; and an interest in forming and acting on a sense of justice. Moreover, advancing those shared interests requires certain goods, called "social primary goods," and so the parties to the social contract know that they need these goods—in particular, the basic liberties; freedom of movement and occupational choice; powers and prerogatives of office and positions of responsibility; income and wealth; and the social bases of self-respect.

Why would parties in the original position choose Rawls's principles? The argument is complicated, but the intuitive idea is reasonably straightforward. You are asked to choose principles for your society under conditions of ignorance about your talents, ideals, and social position, ignorance that models the irrelevance of these properties from the point of view of

justice. You do not know which person you will be, but have to live with the principles you choose, so you want to be sure—if this is possible—that your situation is acceptable whatever it turns out to be: you want to be sure that the society is acceptable from the point of view of each person, because you do not know which person you are. In particular, you want to be sure that it will be acceptable even if you land in the lowest social position, where acceptability is least likely. And, according to Rawls, this is precisely the “downside protection,” the insurance that the two principles provide. They ensure that social arrangements are acceptable to all members of a society of equals, in particular because they guarantee basic liberties to all, and ensure an acceptable worth of liberty, even at the minimum position.

Starting from the fundamental ideal of fair cooperation among free and equal moral persons, then, we are led to the social contract idea of finding principles that would be the object of an initial agreement. And that initial agreement, made under conditions of ignorance, would endorse an egalitarian-liberal political conception that promises to maximize the minimum worth of liberty and thus to provide terms that are acceptable to all members of a democratic society. It is an inspiring political vision, and in a better world than our own, it would guide the political judgment of democratic citizens.

Rawls's work is the topic of a vast secondary literature. The discussion of Rawls in Brian Barry's *Theories of Justice* (Berkeley and Los Angeles: University of California Press, 1989) is especially illuminating. *Reading Rawls*, edited by Norman Daniels (Stanford, Calif.: Stanford University Press, 1989), contains some of the classic early papers on *Theory of Justice*, including important articles by Thomas Nagel, Ronald Dworkin, and H. L. A. Hart. In chapter 7 of *Anarchy, State, and Utopia* (New York: Basic Books, 1974), Robert Nozick advances a forceful libertarian critique of justice as fairness. Susan Okin's *Justice, Gender, and the Family* (New York: Basic Books, 1989) presents a feminist critique of justice as fairness. In *Liberalism and the Limits of Justice*, 2d ed. (Cambridge: Cambridge University Press, 1998), Michael Sandel argues that there is a fundamental tension between the liberal and egalitarian commitments of justice as fairness—that Rawls's liberalism is founded on an individualistic outlook, whereas his egalitarianism suggests a communitarian philosophy. In a series of papers, G. A. Cohen has argued that Rawls's willingness, under the difference principle, to countenance incentives to people with scarce talents indicates an accommodation to unjust selfishness. See his “Incentives, Inequality, and Community,” in Grethe Peterson (ed.), *The Tanner Lectures on Human Values*, vol. 13 (Salt Lake City: University of Utah Press, 1992), pp. 263–329. The articles in a special issue of *Ethics* (October 1989) are helpful in understanding both the substance of justice as fairness and the movement from the presentation of justice as fairness in *Theory of Justice* to its reformulation in *Political Liberalism* as a political conception of justice.

A Theory of Justice

3. THE MAIN IDEA OF THE THEORY OF JUSTICE

My aim is to present a conception of justice which generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant.¹ 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.² 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 general recognition of this fact would provide the basis for a public acceptance of the corresponding principles of justice. No society can, of course, be a scheme of cooperation which men enter voluntarily in a literal sense; each person finds himself placed at birth in some particular position in some particular society, and the nature of this position materially affects his life prospects. Yet a society satisfying the principles of justice as fairness comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are autonomous and the obligations they recognize self-imposed.

One feature of justice as fairness is to think of the parties in the initial situation as rational and mutually

disinterested. This does not mean that the parties are egoists, that is, individuals with only certain kinds of interests, say in wealth, prestige, and domination. But they are conceived as not taking an interest in one another's interests. They are to presume that even their spiritual aims may be opposed, in the way that the aims of those of different religions may be opposed. Moreover, the concept of rationality must be interpreted as far as possible in the narrow sense, standard in economic theory, of taking the most effective means to given ends. I shall modify this concept to some extent, as explained later, 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. These matters I shall take up in the immediately succeeding chapters. 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. Off-hand 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. The two principles mentioned seem to be a fair basis on 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.³ Once we decide to look for a conception of justice that prevents the use of the accidents of natural endowment and the contingencies of social circumstance as counters in a 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.

4. 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 be 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 insure 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 justice 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*.⁴ 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 principles, 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.⁵

11. TWO PRINCIPLES OF JUSTICE

I shall now state in a provisional form the two principles of justice that I believe would be agreed to in the original position. The first formulation of these principles is tentative. As we go on I shall consider 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 scheme of equal basic liberties compatible with a similar scheme of liberties for others.

Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all.

These principles primarily apply, as I have said, to the basic structure of society and govern the assignment of rights and duties and regulate the distribution of social and economic advantages. Their formulation presupposes that, for the purposes of a theory of justice, the social structure may be viewed as having two more or less distinct parts, the first principle applying to the one, the second principle to the other. Thus we distinguish between the aspects of the social system that define and secure the equal basic liberties and the aspects that specify and establish social and economic inequalities. Now it is essential to observe that the basic liberties are given by a list of such liberties. Important among these are political liberty (the right to vote and to hold public office) and freedom of speech and assembly; liberty of conscience and freedom of thought; freedom of the person, which includes freedom from psychological oppression and physical assault and dismemberment (integrity of the person); 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 to be equal by the first principle.

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. 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 responsibility 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 infringements of the basic equal liberties protected by the first principle cannot be justified, or compensated for, by greater social and economic advantages. These liberties have a central range of application within which they can be limited and compromised only when they conflict with other basic liberties. Since they may be limited when they clash with one another, none of these liberties is absolute; but however they are adjusted to form one system, this system is to be the same for all. It is difficult, and perhaps impossible, to give a complete specification of these liberties independently from the particular circumstances—social, economic, and technological—of a given society. The hypothesis is that the general form of such a list could be devised with sufficient exactness to sustain this conception of justice. Of course, liberties not on the list, for example, the right to own certain kinds of property (e.g., means of production) and freedom of contract as understood by the doctrine of *laissez-faire* are not basic; and so they are not protected by the priority of the first principle. Finally, in regard to the second principle, the distribution of wealth and income, and positions of authority and responsibility, are to be consistent with both the basic liberties and equality of opportunity.

The two principles are rather specific in their content, and their acceptance rests on certain assumptions that I must eventually try to explain and

justify. For the present, it should be observed that these principles 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 social 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, liberties, and opportunities, and income and wealth. (Later on in Part Three 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 differences in authority 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 people seem willing to

forego certain political rights when the economic returns are significant. It is this kind of exchange which the two principles rule out; being arranged in serial order they do not permit exchanges between basic liberties and economic and social gains except under extenuating circumstances.

The fact that the two principles apply to institutions has certain consequences. First of all, the rights and basic liberties referred to by these principles are those which 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 basic liberties and making them less extensive is that otherwise they would interfere with one another.

Further, 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 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. 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 regu-

late 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 in positions of authority and responsibility 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 counterbalanced in this way. It is obvious, however, that there are indefinitely many ways in which all may be advantaged when the initial arrangement of equality is taken as a benchmark. How then are we to choose among these possibilities? The principles must be specified so that they yield a determinate conclusion.

13. DEMOCRATIC EQUALITY AND THE DIFFERENCE PRINCIPLE

The democratic interpretation... is arrived at by combining the principle of fair equality of opportunity with the difference principle. Assuming the framework of institutions required by equal liberty and fair equality of opportunity, the higher expectations of those better situated are just if and only if they work as part of a scheme which improves the expectations of the least advantaged members of society. The intuitive idea is that the social order is not to establish and secure the more attractive prospects of those better off unless doing so is to the advantage of those less fortunate.

To illustrate the difference principle, consider the distribution of income among social classes. Let us suppose that the various income groups correlate with representative individuals by reference to

whose expectations we can judge the distribution. Now those starting out as members of the entrepreneurial class in property-owning democracy, say, have a better prospect than those who begin in the class of unskilled laborers. It seems likely that this will be true even when the social injustices which now exist are removed. What, then, can possibly justify this kind of initial inequality in life prospects? According to the difference principle, it is justifiable only if the difference in expectation is to the advantage of the representative man who is worse off, in this case the representative unskilled worker. The inequality in expectation is permissible only if lowering it would make the working class even more worse off. Supposedly, given the rider in the second principle concerning open positions, and the principle of liberty generally, the greater expectations allowed to entrepreneurs encourages them to do things which raise the prospects of laboring class. Their better prospects act as incentives so that the economic process is more efficient, innovation proceeds at a faster pace, and so on. I shall not consider how far these things are true. The point is that something of this kind must be argued if these inequalities are to satisfy by the difference principle.

17. THE TENDENCY TO EQUALITY

I wish to conclude this discussion of the two principles by explaining the sense in which they express an egalitarian conception of justice. Also I should like to forestall the objection to the principle of fair opportunity that it leads to a meritocratic society. In order to prepare the way for doing this, I note several aspects of the conception of justice that I have set out.

First we may observe that the difference principle gives some weight to the considerations singled out by the principle of redress. This is the principle that undeserved inequalities call for redress; and since inequalities of birth and natural endowment are undeserved, these inequalities are to be somehow compensated for.⁶ Thus the principle holds that in order to treat all persons equally, to provide genuine

equality of opportunity, society must give more attention to those with fewer native assets and to those born into the less favorable social positions. The idea is to redress the bias of contingencies in the direction of equality. In pursuit of this principle greater resources might be spent on the education of the less rather than the more intelligent, at least over a certain time of life, say the earlier years of school.

Now the principle of redress has not to my knowledge been proposed as the sole criterion of justice, as the single aim of the social order. It is plausible as most such principles are only as a *prima facie* principle, one that is to be weighed in the balance with others. For example, we are to weigh it against the principle to improve the average standard of life, or to advance the common good.⁷ But whatever other principles we hold, the claims of redress are to be taken into account. It is thought to represent one of the elements in our conception of justice. Now the difference principle is not of course the principle of redress. It does not require society to try to even out handicaps as if all were expected to compete on a fair basis in the same race. But the difference principle would allocate resources in education, say, so as to improve the long-term expectation of the least favored. If this end is attained by giving more attention to the better endowed, it is permissible; otherwise not. And in making this decision, the value of education should not be assessed solely in terms of economic efficiency and social welfare. Equally if not more important is the role of education in enabling a person to enjoy the culture of his society and to take part in its affairs, and in this way to provide for each individual a secure sense of his own worth.

Thus although the difference principle is not the same as that of redress, it does achieve some of the intent of the latter principle. It transforms the aims of the basic structure so that the total scheme of institutions no longer emphasizes social efficiency and technocratic values. The difference principle represents, in effect, an agreement to regard the distribution of natural talents as in some respects a common asset and to share in the greater social and economic benefits made possible by the complementarities of

this distribution. Those who have been favored by nature, whoever they are, may gain from their good fortune only on terms that improve the situation of those who have lost out. The naturally advantaged are not to gain merely because they are more gifted, but only to cover the costs of training and education and for using their endowments in ways that help the less fortunate as well. No one deserves his greater natural capacity nor merits a more favorable starting place in society. But, of course, this is no reason to ignore, much less to eliminate these distinctions. Instead, the basic structure can be arranged so that these contingencies work for the good of the least fortunate. Thus we are led to the difference principle if we wish to set up the social system so that no one gains or loses from his arbitrary place in the distribution of natural assets or his initial position in society without giving or receiving compensating advantages in return.

In view of these remarks we may reject the contention that the ordering of institutions is always defective because the distribution of natural talents and the contingencies of social circumstance are unjust, and this injustice must inevitably carry over to human arrangements. Occasionally this reflection is offered as an excuse for ignoring injustice, as if the refusal to acquiesce in injustice is on a par with being unable to accept death. The natural distribution is neither just nor unjust; nor is it unjust that persons are born into society at some particular position. These are simply natural facts. What is just and unjust is the way that institutions deal with these facts. Aristocratic and caste societies are unjust because they make these contingencies the ascriptive basis for belonging to more or less enclosed and privileged social classes. The basic structure of these societies incorporates the arbitrariness found in nature. But there is no necessity for men to resign themselves to these contingencies. The social system is not an unchangeable order beyond human control but a pattern of human action. In justice as fairness men agree to avail themselves of the accidents of nature and social circumstance only when doing so is for the common benefit. The two principles are a fair way of meeting the arbitrariness of fortune; and while no doubt imperfect in other

ways, the institutions which satisfy these principles are just.

A further point is that the difference principle expresses a conception of reciprocity. It is a principle of mutual benefit. At first sight, however, it may appear unfairly biased towards the least favored. To consider this question in an intuitive way, suppose for simplicity that there are only two groups in society, one noticeably more fortunate than the other. Subject to the usual constraints (defined by the priority of the first principle and fair equality of opportunity), society could maximize the expectations of either group but not both, since we can maximize with respect to only one aim at a time. It seems clear that society should not do the best it can for those initially more advantaged; so if we reject the difference principle, we must prefer maximizing some weighted mean of the two expectations. But if we give any weight to the more fortunate, we are valuing for their own sake the gains to those already more favored by natural and social contingencies. No one had an antecedent claim to be benefited in this way, and so to maximize a weighted mean is, so to speak, to favor the more fortunate twice over. Thus the more advantaged, when they view the matter from a general perspective, recognize that the well-being of each depends on a scheme of social cooperation without which no one could have a satisfactory life; they recognize also that they can expect the willing cooperation of all only if the terms of the scheme are reasonable. So they regard themselves as already compensated, as it were, by the advantages to which no one (including themselves) had a prior claim. They forego the idea of maximizing a weighted mean and regard the difference principle as a fair basis for regulating the basic structure.

One may object that those better situated deserve the greater advantages they could acquire for themselves under other schemes of cooperation whether or not these advantages are gained in ways that benefit others. Now it is true that given a just system of cooperation as a framework of public rules, and the expectations set up by it, those who, with the prospect of improving their condition, have done what the system announces it will reward are entitled to have their

expectations met. In this sense the more fortunate have title to their better situation; their claims are legitimate expectations established by social institutions and the community is obligated to fulfill them. But this sense of desert is that of entitlement. It presupposes the existence of an ongoing cooperative scheme and is irrelevant to the question whether this scheme itself is to be designed in accordance with the difference principle or some other criterion.

Thus it is incorrect that individuals with greater natural endowments and the superior character that has made their development possible have a right to a cooperative scheme that enables them to obtain even further benefits in ways that do not contribute to the advantages of others. We do not deserve our place in the distribution of native endowments, any more than we deserve our initial starting place in society. That we deserve the superior character that enables us to make the effort to cultivate our abilities is also problematic; for such character depends in good part upon fortunate family and social circumstances in early life for which we can claim no credit. The notion of desert does not apply here. To be sure, the more advantaged have a right to their natural assets, as does everyone else; this right is covered by the first principle under the basic liberty protecting the integrity of the person. And so the more advantaged are entitled to whatever they can acquire in accordance with the rules of a fair system of social cooperation. Our problem is how this scheme, the basic structure of society, is to be designed. From a suitably general standpoint, the difference principle appears acceptable to both the more advantaged and the less advantaged individual. Of course, none of this is strictly speaking an argument for the principle, since in a contract theory arguments are made from the point of view of the original position. But these intuitive considerations help to clarify the principle and the sense in which it is egalitarian.

24. THE VEIL OF IGNORANCE

The idea of the original position is to set up a fair procedure so that any principles agreed to will be

just. The aim is to use the notion of pure procedural justice as a basis of theory. Somehow we must nullify the effects of specific contingencies which put men at odds and tempt them to exploit social and natural circumstances to their own advantage. Now in order to do this I assume that the parties are situated behind a veil of ignorance. They do not know how the various alternatives will affect their own particular case and they are obliged to evaluate principles solely on the basis of general considerations.⁸

It is assumed, then, that the parties do not know certain kinds of particular facts. First of all, no one knows his place in society, his class position or social status; nor does he know his fortune in the distribution of natural assets and abilities, his intelligence and strength, and the like. Nor, again, does anyone know his conception of the good, the particulars of his rational plan of life, or even the special features of his psychology such as his aversion to risk or liability to optimism or pessimism. More than this, I assume that the parties do not know the particular circumstances of their own society. That is, they do not know its economic or political situation, or the level of civilization and culture it has been able to achieve. The persons in the original position have no information as to which generation they belong. These broader restrictions on knowledge are appropriate in part because questions of social justice arise between generations as well as within them, for example, the question of the appropriate rate of capital saving and of the conservation of natural resources and the environment of nature. There is also, theoretically anyway, the question of a reasonable genetic policy. In these cases too, in order to carry through the idea of the original position, the parties must not know the contingencies that set them in opposition. They must choose principles the consequences of which they are prepared to live with whatever generation they turn out to belong to.

As far as possible, then, the only particular facts which the parties know is that their society is subject to the circumstances of justice and whatever this implies. It is taken for granted, however, that they know the general facts about human society. They understand political affairs and the principles

of economic theory; they know the basis of social organization and the laws of human psychology. Indeed, the parties are presumed to know whatever general facts affect the choice of the principles of justice. There are no limitations on general information, that is, on general laws and theories, since conceptions of justice must be adjusted to the characteristics of the systems of social cooperation which they are to regulate, and there is no reason to rule out these facts. It is, for example, a consideration against a conception of justice that, in view of the laws of moral psychology, men would not acquire a desire to act upon it even when the institutions of their society satisfied it. For in this case there would be difficulty in securing the stability of social cooperation. An important feature of a conception of justice is that it should generate its own support. Its principles should be such that when they are embodied in the basic structure of society men tend to acquire the corresponding sense of justice and develop a desire to act in accordance with its principles. In this case a conception of justice is stable. This kind of general information is admissible in the original position.

The notion of the veil of ignorance raises several difficulties. Some may object that the exclusion of nearly all particular information makes it difficult to grasp what is meant by the original position. Thus it may be helpful to observe that one or more persons can at any time enter this position, or perhaps better, simulate the deliberations of this hypothetical situation, simply by reasoning in accordance with the appropriate restrictions. In arguing for a conception of justice we must be sure that it is among the permitted alternatives and satisfies the stipulated formal constraints. No considerations can be advanced in its favor unless they would be rational ones for us to urge were we to lack the kind of knowledge that is excluded. The evaluation of principles must proceed in terms of the general consequences of their public recognition and universal application, it being assumed that they will be complied with by everyone. To say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and

restrictions would reach a certain conclusion. If necessary, the argument to this result could be set out more formally. I shall, however, speak throughout in terms of the notion of the original position. It is more economical and suggestive, and brings out certain essential features that otherwise one might easily overlook.

These remarks show that the original position is not to be thought of as a general assembly which includes at one moment everyone who will live at some time or, much less, as an assembly of everyone who could live at some time. It is not a gathering of all actual or possible persons. If we conceived of the original position in either of these ways, the conception would cease to be a natural guide to intuition and would lack a clear sense. In any case, the original position must be interpreted so that one can at any time adopt its perspective. It must make no difference when one takes up this viewpoint, or who does so: the restrictions must be such that the same principles are always chosen. The veil of ignorance is a key condition in meeting this requirement. It insures not only that the information available is relevant, but that it is at all times the same.

It may be protested that the condition of the veil of ignorance is irrational. Surely, some may object, principles should be chosen in the light of all the knowledge available. There are various replies to this contention. Here I shall sketch those which emphasize the simplifications that need to be made if one is to have any theory at all. . . . To begin with, it is clear that since the differences among the parties are unknown to them, and everyone is equally rational and similarly situated, each is convinced by the same arguments. Therefore, we can view the agreement in the original position from the standpoint of one person selected at random. If anyone after due reflection prefers a conception of justice to another, then they all do, and a unanimous agreement can be reached. We can, to make the circumstances more vivid, imagine that the parties are required to communicate with each other through a referee as intermediary, and that he is to announce which alternatives have been suggested and the reasons offered in their support. He forbids the attempt

to form coalitions, and he informs the parties when they have come to an understanding. But such a referee is actually superfluous, assuming that the deliberations of the parties must be similar.

Thus there follows the very important consequence that the parties have no basis for bargaining in the usual sense. No one knows his situation in society nor his natural assets, and therefore no one is in a position to tailor principles to his advantage. We might imagine that one of the contractees threatens to hold out unless the others agree to principles favorable to him. But how does he know which principles are especially in his interests? The same holds for the formation of coalitions: if a group were to decide to band together to the disadvantage of the others, they would not know how to favor themselves in the choice of principles. Even if they could get everyone to agree to their proposal, they would have no assurance that it was to their advantage, since they cannot identify themselves either by name or description. The one case where this conclusion fails is that of saving. Since the persons in the original position know that they are contemporaries (taking the present time of entry interpretation), they can favor their generation by refusing to make any sacrifices at all for their successors; they simply acknowledge the principle that no one has a duty to save for posterity. Previous generations have saved or they have not; there is nothing the parties can now do to affect that. So in this instance the veil of ignorance fails to secure the desired result. Therefore, to handle the question of justice between generations, I modify the motivation assumption and add a further constraint. . . . With these adjustments, no generation is able to formulate principles especially designed to advance its own cause and some significant limits on savings principles can be derived. . . . Whatever a person's temporal position, each is forced to choose for all.⁹

The restrictions on particular information in the original position are, then, of fundamental importance. Without them we would not be able to work out any definite theory of justice at all. We would have to be content with a vague formula stating that justice is what would be agreed to without being able to say

much, if anything, about the substance of the agreement itself. The formal constraints of the concept of right, those applying to principles directly, are not sufficient for our purpose. The veil of ignorance makes possible a unanimous choice of a particular conception of justice. Without these limitations on knowledge the bargaining problem of the original position would be hopelessly complicated. Even if theoretically a solution were to exist, we would not, at present anyway, be able to determine it. . . .

26. THE REASONING LEADING TO THE TWO PRINCIPLES OF JUSTICE

It seems 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 a relation between the two principles and the maximin rule for choice under uncertainty.¹⁰ 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.¹¹ 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 principles. Clearly the maximin rule is not, in general, a suitable guide for choices under uncertainty. But it holds only 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 has these features to a very high degree.

Now there appear to be three chief features of situations that give plausibility to this unusual rule.¹² 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. . . . 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.

Let us review briefly the nature of the original

position with these three special features in mind. To begin with, the veil of ignorance excludes all knowledge of likelihoods. The parties have no basis for determining the probable nature of their society, or their place in it. Thus they have no basis for probability calculations. 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. These considerations are strengthened by the fact that the parties know very little about the possible states of society. 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 illustrations by numerical tables suggest. It is for this reason that I have spoken only of a relation to 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 some 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. For this priority implies that the persons in the original position have no desire to try for greater gains at the expense of the basic 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. 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 conditions are not realized.

Notes

1 As the text suggests, I shall regard Locke's *Second Treatise of Government*, Rousseau's *The 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. Henri Poincaré remarks: "Il nous faut une faculté

qui nous fasse voir le but de loin, et, cette faculté, c'est l'intuition." *La Valeur de la science* (Paris, Flammarion, 1909), p. 27.

6. See Herbert Spiegelberg, "A Defense of Human Equality," *Philosophical Review*, vol. 53 (1944), pp. 101, 113–123; and D. D. Raphael, "Justice and Liberty," *Proceedings of the Aristotelian Society*, vol. 51 (1950–1951), pp. 187f.

7. See, for example, Spiegelberg, pp. 120f.

8. The veil of ignorance is so natural a condition that something like it must have occurred to many. The formulation in the text is implicit, I believe, in Kant's doctrine of the categorical imperative, both in the way this procedural criterion is defined and the use Kant makes of it. Thus when Kant tells us to test our maxim by considering what would be the case were it a universal law of nature, he must suppose that we do not know our place within this imagined system of nature. See, for example, his discussion of the topic of practical judgment in *The Critique of Practical Reason*, Academy Edition, vol. 5, pp. 68–72. A similar restriction on information is found in J. C. Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking," *Journal of Political Economy*, vol. 61 (1953). However, other aspects of Harsanyi's view are quite different, and he uses the restriction to develop a utilitarian theory.

9. Rousseau, *The Social Contract*, bk. II, ch. IV, par. 5.

10. 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 Inc., 1965), ch. 24. Baumol gives a geometric interpretation of these rules 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.

11. 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
d ₁	-7	8	12
d ₂	-8	7	14
d ₃	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 d₃ 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

12. Here I borrow from William Fellner, *Probability and Profit* (Homewood, Ill., R. D. Irwin, Inc., 1965), pp. 140–142, where these features are noted

ROBERT NOZICK



INTRODUCTION

THOMAS CHRISTIANO

Robert Nozick (1938–2003) was born in Brooklyn, New York, was graduated from Columbia College, and received his Ph.D. from Princeton University. He was Pellegrino University Professor at Harvard University, a past president of the American Philosophical Association (Eastern Division), a Fellow of the American Academy of Arts and Sciences, a Corresponding Fellow of the British Academy, and a Senior Fellow in the Society of Fellows at Harvard University.

While Nozick wrote path-breaking works in an unusually wide variety of areas in philosophy, his most famous book, winner of the 1975 National Book Award, was *Anarchy, State, and Utopia*. It revived classical liberalism as a serious option in political philosophy, offered the first full-length challenge to John Rawls’s *A Theory of Justice*, and defended a version of libertarianism. Nozick’s book *Philosophical Explanations* (1981) made lasting contributions to epistemology and our understanding of personal identity, free will, and the foundations of ethics. *The Examined Life* (1989) explored the nature of the good life and the meaning of life. *The Nature of Rationality* (1993) culminated a lifelong study of theories of rational decision and rational belief. Nozick also published *Socratic Puzzles* (1997), a collection of essays, and *Invariances* (2001), a book on the notion of an objective world and the role that invariance plays in it, including topics in philosophy of science, philosophy of mind, metaphysics, and ethics.

In *Anarchy, State, and Utopia*, Nozick’s arguments proceed in the Lockean tradition of natural rights to liberty and property. For Nozick, justice entails absolute constraints on the behavior of people toward others. No one may abridge the liberty of another, harm the other in life or limb, or take property from another, without the other’s consent. Persons have the right to act in self-defense, and rectification is appropriate if the rights to property and liberty have been violated. In short, whatever comes about by the voluntary consent of people