

first-person dictatorship and the free-rider forms, since in each case a proper name, or pronoun, or a rigged definite description is needed, either to single out the dictator or to characterize the free-rider. Generality does not, however, exclude general egoism, for each person is allowed to do whatever, in his judgment, is most likely to further his own aims. The principle here can clearly be expressed in a perfectly general way. It is the ordering condition which renders general egoism inadmissible, for if everyone is authorized to advance his aims as he pleases, or if everyone ought to advance his own interests, competing claims are not ranked at all and the outcome is determined by force and cunning.

The several kinds of egoism, then, do not appear on the list presented to the parties. They are eliminated by the formal constraints. Of course, this is not a surprising conclusion, since it is obvious that by choosing one of the other conceptions the persons in the original position can do much better for themselves. Once they ask which principles all should agree to, no form of egoism is a serious candidate for consideration in any case. This only confirms what we knew already, namely, that although egoism is logically consistent and in this sense not irrational, it is incompatible with what we intuitively regard as the moral point of view. The significance of egoism philosophically is not as an alternative conception of right but as a challenge to any such conception. In justice as fairness this is reflected in the fact that we can interpret general egoism as the no-agreement point. It is what the parties would be stuck with if they were unable to reach an understanding.

## 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.<sup>11</sup>

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

11. The veil of ignorance is so natural a condition that something like it must have occurred to many. The closest explicit statement of it known to me 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). Harsanyi uses it to develop a utilitarian theory, as I discuss below in §§27-28.

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. It is an important feature of a conception of justice that it should generate its own support. That is, 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. Given the principles of moral learning, men 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. To conceive of the original position in either of these ways is to stretch fantasy too far; the conception would cease to be a natural guide to intuition. In any case, it is important that the original position 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. (Those based on the Kantian interpretation of the original position are given later, § 40.) 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 choice 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 I resolve the question of justice between generations in a different way by altering the motivation assumption. But with this adjustment no one is able to formulate principles especially designed to advance his own cause. Whatever his temporal position, each is forced to choose for everyone.<sup>12</sup>

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.

The notion of the veil of ignorance is implicit, I think, in Kant's

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

ethics (§40). Nevertheless the problem of defining the knowledge of the parties and of characterizing the alternatives open to them has often been passed over, even by contract theories. Sometimes the situation definitive of moral deliberation is presented in such an indeterminate way that one cannot ascertain how it will turn out. Thus Perry's doctrine is essentially contractarian: he holds that social and personal integration must proceed by entirely different principles, the latter by rational prudence, the former by the concurrence of persons of good will. He would appear to reject utilitarianism on much the same grounds suggested earlier: namely, that it improperly extends the principle of choice for one person to choices facing society. The right course of action is characterized as that which best advances social aims as these would be formulated by reflective agreement given that the parties have full knowledge of the circumstances and are moved by a benevolent concern for one another's interests. No effort is made, however, to specify in any precise way the possible outcomes of this sort of agreement. Indeed, without a far more elaborate account, no conclusions can be drawn.<sup>13</sup> I do not wish here to criticize others; rather, I want to explain the necessity for what may seem at times like so many irrelevant details.

Now the reasons for the veil of ignorance go beyond mere simplicity. We want to define the original position so that we get the desired solution. If a knowledge of particulars is allowed, then the outcome is biased by arbitrary contingencies. As already observed, to each according to his threat advantage is not a principle of justice. If the original position is to yield agreements that are just, the parties must be fairly situated and treated equally as moral persons. The arbitrariness of the world must be corrected for by adjusting the circumstances of the initial contractual situation. Moreover, if in choosing principles we required unanimity even when there is full information, only a few rather obvious cases could be decided. A conception of justice based on unanimity in these circumstances would indeed be weak and trivial. But once knowledge is excluded, the requirement of unanimity is not out of place and the fact that

13. See R. B. Perry, *The General Theory of Value* (New York, Longmans, Green and Company, 1926), pp. 674-682.

it can be satisfied is of great importance. It enables us to say of the preferred conception of justice that it represents a genuine reconciliation of interests.

A final comment. For the most part I shall suppose that the parties possess all general information. No general facts are closed to them. I do this mainly to avoid complications. Nevertheless a conception of justice is to be the public basis of the terms of social cooperation. Since common understanding necessitates certain bounds on the complexity of principles, there may likewise be limits on the use of theoretical knowledge in the original position. Now clearly it would be very difficult to classify and to grade for complexity the various sorts of general facts. I shall make no attempt to do this. We do however recognize an intricate theoretical construction when we meet one. Thus it seems reasonable to say that other things equal one conception of justice is to be preferred to another when it is founded upon markedly simpler general facts, and its choice does not depend upon elaborate calculations in the light of a vast array of theoretically defined possibilities. It is desirable that the grounds for a public conception of justice should be evident to everyone when circumstances permit. This consideration favors, I believe, the two principles of justice over the criterion of utility.

## 25. THE RATIONALITY OF THE PARTIES

I have assumed throughout that the persons in the original position are rational. In choosing between principles each tries as best he can to advance his interests. But I have also assumed that the parties do not know their conception of the good. This means that while they know that they have some rational plan of life, they do not know the details of this plan, the particular ends and interests which it is calculated to promote. How, then, can they decide which conceptions of justice are most to their advantage? Or must we suppose that they are reduced to mere guessing? To meet this difficulty, I postulate that they accept the account of the good touched upon in the preceding chapter: they assume that they would prefer more primary social goods rather than less. Of course, it may turn out, once the veil of ignorance is removed, that some of them for religious or other reasons may not, in fact, want more of these goods.

But from the standpoint of the original position, it is rational for the parties to suppose that they do want a larger share, since in any case they are not compelled to accept more if they do not wish to, nor does a person suffer from a greater liberty. Thus even though the parties are deprived of information about their particular ends, they have enough knowledge to rank the alternatives. They know that in general they must try to protect their liberties, widen their opportunities, and enlarge their means for promoting their aims whatever these are. Guided by the theory of the good and the general facts of moral psychology, their deliberations are no longer guesswork. They can make a rational decision in the ordinary sense.

The concept of rationality invoked here, with the exception of one essential feature, is the standard one familiar in social theory.<sup>14</sup> Thus in the usual way, a rational person is thought to have a coherent set of preferences between the options open to him. He ranks these options according to how well they further his purposes; he follows the plan which will satisfy more of his desires rather than less, and which has the greater chance of being successfully executed. The special assumption I make is that a rational individual does not suffer from envy. He is not ready to accept a loss for himself if only others have less as well. He is not downcast by the knowledge or perception that others have a larger index of primary social goods. Or at least this is true as long as the differences between himself and others do not exceed certain limits, and he does not believe that the existing inequalities are founded on injustice or are the result of letting chance work itself out for no compensating social purpose (§ 80).

14. For this notion of rationality, see the references to Sen and Arrow above, §23, note 9. The discussion in I. M. D. Little, *The Critique of Welfare Economics*, 2nd ed. (Oxford, Clarendon Press, 1957), ch. II, is also relevant here. For rational choice under uncertainty, see below §26, note 18. H. A. Simon discusses the limitations of the classical conceptions of rationality and the need for a more realistic theory in "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, vol. 69 (1955). See also his essay in *Surveys of Economic Theory*, vol. 3 (London, Macmillan, 1967). For philosophical discussions see Donald Davidson, "Actions, Reasons, and Causes," *Journal of Philosophy*, vol. 60 (1963); C. G. Hempel, *Aspects of Scientific Explanation* (New York, The Free Press, 1965), pp. 463-486; Jonathan Bennett, *Rationality* (London, Routledge and Kegan Paul, 1964), and J. D. Mabbott, "Reason and Desire," *Philosophy*, vol. 28 (1953).

revisions:  
Main lines by  
Davidson

The assumption that the parties are not moved by envy raises certain questions. Perhaps we should also assume that they are not liable to various other feelings such as shame and humiliation (§ 67). Now a satisfactory account of justice will eventually have to deal with these matters too, but for the present I shall leave these complications aside. Another objection to our procedure is that it is too unrealistic. Certainly men are afflicted with these feelings. How can a conception of justice ignore this fact? I shall meet this problem by dividing the argument for the principles of justice into two parts. In the first part, the principles are derived on the supposition that envy does not exist; while in the second, we consider whether the conception arrived at is feasible in view of the circumstances of human life.

One reason for this procedure is that envy tends to make everyone worse off. In this sense it is collectively disadvantageous. Presuming its absence amounts to supposing that in the choice of principles men should think of themselves as having their own plan of life which is sufficient for itself. They have a secure sense of their own worth so that they have no desire to abandon any of their aims provided others have less means to further theirs. I shall work out a conception of justice on this stipulation to see what happens. Later I shall try to show that when the principles adopted are put into practice, they lead to social arrangements in which envy and other destructive feelings are not likely to be strong. The conception of justice eliminates the conditions that give rise to disruptive attitudes. It is, therefore, inherently stable (§§ 80–81).

The assumption of mutually disinterested rationality, then, comes to this: the persons in the original position try to acknowledge principles which advance their system of ends as far as possible. They do this by attempting to win for themselves the highest index of primary social goods, since this enables them to promote their conception of the good most effectively whatever it turns out to be. The parties do not seek to confer benefits or to impose injuries on one another; they are not moved by affection or rancor. Nor do they try to gain relative to each other; they are not envious or vain. Put in terms of a game, we might say: they strive for as high an absolute score as possible. They do not wish a high or a low score for their opponents, nor do they seek to maximize or minimize the differ-

ence between their successes and those of others. The idea of a game does not really apply, since the parties are not concerned to win but to get as many points as possible judged by their own system of ends.

There is one further assumption to guarantee strict compliance. The parties are presumed to be capable of a sense of justice and this fact is public knowledge among them. This condition is to insure the integrity of the agreement made in the original position. It does not mean that in their deliberations the parties apply some particular conception of justice, for this would defeat the point of the motivation assumption. Rather, it means that the parties can rely on each other to understand and to act in accordance with whatever principles are finally agreed to. Once principles are acknowledged the parties can depend on one another to conform to them. In reaching an agreement, then, they know that their undertaking is not in vain: their capacity for a sense of justice insures that the principles chosen will be respected. It is essential to observe, however, that this assumption still permits the consideration of men's capacity to act on the various conceptions of justice. The general facts of human psychology and the principles of moral learning are relevant matters for the parties to examine. If a conception of justice is unlikely to generate its own support, or lacks stability, this fact must not be overlooked. For then a different conception of justice might be preferred. The assumption only says that the parties have a capacity for justice in a purely formal sense: taking everything relevant into account, including the general facts of moral psychology, the parties will adhere to the principles eventually chosen. They are rational in that they will not enter into agreements they know they cannot keep, or can do so only with great difficulty. Along with other considerations, they count the strains of commitment (§ 29). Thus in assessing conceptions of justice the persons in the original position are to assume that the one they adopt will be strictly complied with. The consequences of their agreement are to be worked out on this basis.

With the preceding remarks about rationality and motivation of the parties the description of the original position is for the most part complete. We can summarize this description with the following list of elements of the initial situation and their varia-

tions. (The asterisks mark the interpretations that constitute the original position.)

1. The Nature of the Parties (§ 22)
  - \*a. continuing persons (family heads, or genetic lines)
  - b. single individuals
  - c. associations (states, churches, or other corporate bodies)
2. Subject of Justice (§ 2)
  - \*a. basic structure of society
  - b. rules of corporate associations
  - c. law of nations
3. Presentation of Alternatives (§ 21)
  - \*a. shorter (or longer) list
  - b. general characterization of the possibilities
4. Time of Entry (§ 24)
  - \*a. any time (during age of reason) for living persons
  - b. all actual persons (those alive at some time) simultaneously
  - c. all possible persons simultaneously
5. Circumstances of Justice (§ 22)
  - \*a. Hume's conditions of moderate scarcity
  - b. the above plus further extremes
6. Formal Conditions on Principles (§ 23)
  - \*a. generality, universality, publicity, ordering, and finality
  - b. the above less publicity, say
7. Knowledge and Beliefs (§ 24)
  - \*a. veil of ignorance
  - b. full information
  - c. partial knowledge
8. Motivation of the Parties (§ 25)
  - \*a. mutual disinterestedness (limited altruism)
  - b. elements of social solidarity and good will
  - c. perfect altruism
9. Rationality (§§ 25, 28)
  - \*a. taking effective means to ends with unified expectations and objective interpretation of probability
  - b. as above but without unified expectations and using the principle of insufficient reason

10. Agreement Condition (§ 24)
  - \*a. unanimity in perpetuity
  - b. majority acceptance, or whatever, for limited period
11. Compliance Condition (§ 25)
  - \*a. strict compliance
  - b. partial compliance in various degrees
12. No Agreement Point (§ 23)
  - \*a. general egoism
  - b. the state of nature

We can turn now to the choice of principles. But first I shall mention a few misunderstandings to be avoided. First of all, we must keep in mind that the parties in the original position are theoretically defined individuals. The grounds for their consent are set out by the description of the contractual situation and their preference for primary goods. Thus to say that the principles of justice would be adopted is to say how these persons would decide being moved in ways our account describes. Of course, when we try to simulate the original position in everyday life, that is, when we try to conduct ourselves in moral argument as its constraints require, we will presumably find that our deliberations and judgments are influenced by our special inclinations and attitudes. Surely it will prove difficult to correct for our various propensities and aversions in striving to adhere to the conditions of this idealized situation. But none of this affects the contention that in the original position rational persons so characterized would make a certain decision. This proposition belongs to the theory of justice. It is another question how well human beings can assume this role in regulating their practical reasoning.

Since the persons in the original position are assumed to take no interest in one another's interests (although they may have a concern for third parties), it may be thought that justice as fairness is itself an egoistic theory. It is not, of course, one of the three forms of egoism mentioned earlier, but some may think, as Schopenhauer thought of Kant's doctrine, that it is egoistic nevertheless.<sup>15</sup> Now this is a misconception. For the fact that in the original

15. See *On the Basis of Ethics* (1840), trans. E. F. J. Payne (New York, The Liberal Arts Press, Inc., 1965), pp. 89-92.

position the parties are characterized as not interested in one another's concerns does not entail that persons in ordinary life who hold the principles that would be agreed to are similarly disinterested in one another. Clearly the two principles of justice and the principles of obligation and natural duty require us to consider the rights and claims of others. And the sense of justice is a normally effective desire to comply with these restrictions. The motivation of the persons in the original position must not be confused with the motivation of persons in everyday life who accept the principles that would be chosen and who have the corresponding sense of justice. In practical affairs an individual does have a knowledge of his situation and he can, if he wishes, exploit contingencies to his advantage. Should his sense of justice move him to act on the principles of right that would be adopted in the original position, his desires and aims are surely not egoistic. He voluntarily takes on the limitations expressed by this interpretation of the moral point of view.

This conclusion is supported by a further reflection. Once we consider the idea of a contract theory it is tempting to think that it will not yield the principles we want unless the parties are to some degree at least moved by benevolence, or an interest in one another's interests. Perry, as I mentioned before, thinks of the right standards and decisions as those promoting the ends reached by reflective agreement under circumstances making for impartiality and good will. Now the combination of mutual disinterest and the veil of ignorance achieves the same purpose as benevolence. For this combination of conditions forces each person in the original position to take the good of others into account. In justice as fairness, then, the effects of good will are brought about by several conditions working jointly. The feeling that this conception of justice is egoistic is an illusion fostered by looking at but one of the elements of the original position. Furthermore, this pair of assumptions has enormous advantages over that of benevolence plus knowledge. As I have noted, the latter is so complex that no definite theory at all can be worked out. Not only are the complications caused by so much information insurmountable, but the motivational assumption requires clarification. For example, what is the relative strength of benevolent desires? In brief, the combination of mutual disin-

terestedness plus the veil of ignorance has the merits of simplicity and clarity while at the same time insuring the effects of what are at first sight morally more attractive assumptions. And if it is asked why one should not postulate benevolence with the veil of ignorance, the answer is that there is no need for so strong a condition. Moreover, it would defeat the purpose of grounding the theory of justice on weak stipulations, as well as being incongruous with the circumstances of justice.

Finally, if the parties are conceived as themselves making proposals, they have no incentive to suggest pointless or arbitrary principles. For example, none would urge that special privileges be given to those exactly six feet tall or born on a sunny day. Nor would anyone put forward the principle that basic rights should depend on the color of one's skin or the texture of one's hair. No one can tell whether such principles would be to his advantage. Furthermore, each such principle is a limitation of one's liberty of action, and such restrictions are not to be accepted without a reason. Certainly we might imagine peculiar circumstances in which these characteristics are relevant. Those born on a sunny day might be blessed with a happy temperament, and for some positions of authority this might be a qualifying attribute. But such distinctions would never be proposed in first principles, for these must have some rational connection with the advancement of human interests broadly defined. The rationality of the parties and their situation in the original position guarantees that ethical principles and conceptions of justice have this general content.<sup>16</sup> Inevitably, then, racial and sexual discrimination presupposes that some hold a favored place in the social system which they are willing to exploit to their advantage. From the standpoint of persons similarly situated in an initial situation which is fair, the principles of explicit racist doctrines are not only unjust. They are irrational. For this reason we could say that they are not moral conceptions at all, but simply means of suppression. They have no place on a reason-

16. For a different way of reaching this conclusion, see Philippa Foot, "Moral Arguments," *Mind*, vol. 67 (1958), and "Moral Beliefs," *Proceedings of the Aristotelian Society*, vol. 59 (1958-1959); and R. W. Beardsmore, *Moral Reasoning* (New York, Schocken Books, 1969), especially ch. IV. The problem of content is discussed briefly in G. F. Warnock, *Contemporary Moral Philosophy* (London, Macmillan, 1967), pp. 55-61.

able list of traditional conceptions of justice.<sup>17</sup> Of course, this contention is not at all a matter of definition. It is rather a consequence of the conditions characterizing the original position, especially the conditions of the rationality of the parties and the veil of ignorance. That conceptions of right have a certain content and exclude arbitrary and pointless principles is, therefore, an inference from the theory.

## 26. THE REASONING LEADING TO THE TWO PRINCIPLES OF JUSTICE

In this and the next two sections I take up the choice between the two principles of justice and the principle of average utility. Determining the rational preference between these two options is perhaps the central problem in developing the conception of justice as fairness as a viable alternative to the utilitarian tradition. I shall begin in this section by presenting some intuitive remarks favoring the two principles. I shall also discuss briefly the qualitative structure of the argument that needs to be made if the case for these principles is to be conclusive.

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.

<sup>17</sup> For a similar view, see B. A. O. Williams, "The Idea of Equality," *Philosophy, Politics, and Society*, Second Series, ed. Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1962), p. 113.

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. From time to time in later chapters this problem will be considered (§§ 39, 82). But roughly, the idea underlying this ordering is that if the parties assume that their basic liberties

↪ inevitable possible?

↪ existence

↪ which also ...?

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. Part II is devoted to this. 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>18</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 pos-

18. 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, including the diagram used in §13 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.

sible 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 principles. 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		
	$c_1$	$c_2$	$c_3$
$d_1$	-7	8	12
$d_2$	-8	7	14
$d_3$	5	6	8

Sacrificing  
people now,  
though.

A. S.  
intention

playing  
against  
nature.

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>19</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 combina-

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

tion. 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 of 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 enu-

merate 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, especially in Part Two, 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. In parts of Chapters IV and IX the case for this ordering is discussed.

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. Thus if the list of traditional views (§21)

represents the possible decisions, these principles would be selected by the rule. The original position clearly exhibits these special features to a very high degree in view of the fundamental character of the choice of a conception of justice. These remarks about the maximin rule are intended only to clarify the structure of the choice problem in the original position. They depict its qualitative anatomy. The arguments for the two principles will be presented more fully as we proceed. I want to conclude this section by taking up an objection which is likely to be made against the difference principle and which leads into an important question. The objection is that since we are to maximize (subject to the usual constraints) the long-term prospects of the least advantaged, it seems that the justice of large increases or decreases in the expectations of the more advantaged may depend upon small changes in the prospects of those worst off. To illustrate: the most extreme disparities in wealth and income are allowed provided that the expectations of the least fortunate are raised in the slightest degree. But at the same time similar inequalities favoring the more advantaged are forbidden when those in the worst position lose by the least amount. Yet it seems extraordinary that the justice of increasing the expectations of the better placed by a billion dollars, say, should turn on whether the prospects of the least favored increase or decrease by a penny. This objection is analogous to the following difficulty with the maximin rule. Consider the sequence of gain-and-loss tables:

0	n
1/n	1

for all natural numbers  $n$ . Even if for some smallish number it is reasonable to select the second row, surely there is another point later in the sequence when it is irrational not to choose the first row contrary to the rule.

Part of the answer is that the difference principle is not intended to apply to such abstract possibilities. As I have said, the problem of social justice is not that of allocating *ad libitum* various amounts of something, whether it be money, or property, or whatever, among given individuals. Nor is there some substance of which

expectations are made that can be shuffled from one representative man to another in all possible combinations. The possibilities which the objection envisages cannot arise in real cases; the feasible set is so restricted that they are excluded.<sup>20</sup> The reason for this is that the two principles are tied together as one conception of justice which applies to the basic structure of society as a whole. The operation of the principles of equal liberty and open positions prevents these contingencies from occurring. For as we raise the expectations of the more advantaged the situation of the worst off is continuously improved. Each such increase is in the latter's interest, up to a certain point anyway. For the greater expectations of the more favored presumably cover the costs of training and encourage better performance thereby contributing to the general advantage. While nothing guarantees that inequalities will not be significant, there is a persistent tendency for them to be leveled down by the increasing availability of educated talent and ever widening opportunities. The conditions established by the other principles insure that the disparities likely to result will be much less than the differences that men have often tolerated in the past.

We should also observe that the difference principle not only assumes the operation of other principles, but it presupposes as well a certain theory of social institutions. In particular, as I shall discuss in some detail in Chapter V, it relies on the idea that in a competitive economy (with or without private ownership) with an open class system excessive inequalities will not be the rule. Given the distribution of natural assets and the laws of motivation, great disparities will not long persist. Now the point to stress here is that there is no objection to resting the choice of first principles upon the general facts of economics and psychology. As we have seen, the parties in the original position are assumed to know the general facts about human society. Since this knowledge enters into the premises of their deliberations, their choice of principles is relative to these facts. What is essential, of course, is that these premises be true and sufficiently general. It is often objected, for example, that utilitarianism may allow for slavery and serfdom, and for other infractions of liberty. Whether these institutions are justified is

20. I am indebted to S. A. Marglin for this point.

made to depend upon whether actuarial calculations show that they yield a higher balance of happiness. To this the utilitarian replies that the nature of society is such that these calculations are normally against such denials of liberty. Utilitarians seek to account for the claims of liberty and equality by making certain standard assumptions, as I shall refer to them. Thus they suppose that persons have similar utility functions which satisfy the condition of diminishing marginal utility. It follows from these stipulations that, given a fixed amount of income say, the distribution should be equal, once we leave aside effects on future production. For so long as some have more than others, total utility can be increased by transfers to those who have less. The assignment of rights and liberties can be regarded in much the same way. There is nothing wrong with this procedure provided the assumptions are sound.

Contract theory agrees, then, with utilitarianism in holding that the fundamental principles of justice quite properly depend upon the natural facts about men in society. This dependence is made explicit by the description of the original position: the decision of the parties is taken in the light of general knowledge. Moreover, the various elements of the original position presuppose many things about the circumstances of human life. Some philosophers have thought that ethical first principles should be independent of all contingent assumptions, that they should take for granted no truths except those of logic and others that follow from these by an analysis of concepts. Moral conceptions should hold for all possible worlds. Now this view makes moral philosophy the study of the ethics of creation: an examination of the reflections an omnipotent deity might entertain in determining which is the best of all possible worlds. Even the general facts of nature are to be chosen. Certainly we have a natural religious interest in the ethics of creation. But it would appear to outrun human comprehension. From the point of view of contract theory it amounts to supposing that the persons in the original position know nothing at all about themselves or their world. How, then, can they possibly make a decision? A problem of choice is well defined only if the alternatives are suitably restricted by natural laws and other constraints, and those deciding already have certain inclinations to choose among them. Without a definite structure of this kind the question posed

is indeterminate. For this reason we need have no hesitation in making the choice of the principles of justice presuppose a certain theory of social institutions. Indeed, one cannot avoid assumptions about general facts any more than one can do without a conception of the good on the basis of which the parties rank alternatives. If these assumptions are true and suitably general, everything is in order, for without these elements the whole scheme would be pointless and empty.

It is evident from these remarks that both general facts as well as moral conditions are needed even in the argument for the first principles of justice. (Of course, it has always been obvious that secondary moral rules and particular ethical judgments depend upon factual premises as well as normative principles.) In a contract theory, these moral conditions take the form of a description of the initial contractual situation. It is also clear that there is a division of labor between general facts and moral conditions in arriving at conceptions of justice, and this division can be different from one theory to another. As I have noted before, principles differ in the extent to which they incorporate the desired moral ideal. It is characteristic of utilitarianism that it leaves so much to arguments from general facts. The utilitarian tends to meet objections by holding that the laws of society and of human nature rule out the cases offensive to our considered judgments. Justice as fairness, by contrast, embeds the ideals of justice, as ordinarily understood, more directly into its first principles. This conception relies less on general facts in reaching a match with our judgments of justice. It insures this fit over a wider range of possible cases.

There are two reasons that justify this embedding of ideals into first principles. First of all, and most obviously, the utilitarian's standard assumptions that lead to the wanted consequences may be only probably true, or even doubtfully so. Moreover, their full meaning and application may be highly conjectural. And the same may hold for all the requisite general suppositions that support the principle of utility. From the standpoint of the original position it may be unreasonable to rely upon these hypotheses and therefore far more sensible to embody the ideal more expressly in the principles chosen. Thus it seems that the parties would prefer to secure their liberties straightway rather than have them depend upon what

may be uncertain and speculative actuarial calculations. These remarks are further confirmed by the desirability of avoiding complicated theoretical arguments in arriving at a public conception of justice (§24). In comparison with the reasoning for the two principles, the grounds for the utility criterion trespass upon this constraint. But secondly, there is a real advantage in persons' announcing to one another once and for all that even though theoretical computations of utility always happen to favor the equal liberties (assuming that this is indeed the case here), they do not wish that things had been different. Since in justice as fairness moral conceptions are public, the choice of the two principles is, in effect, such an announcement. And the benefits of this collective profession favor these principles even though the standard utilitarian assumptions should be true. These matters I shall consider in more detail in connection with publicity and stability (§29). The relevant point here is that while, in general, an ethical theory can certainly invoke natural facts, there may nevertheless be good reasons for embedding convictions of justice more directly into first principles than a theoretically complete grasp of the contingencies of the world may actually require.

## 27. THE REASONING LEADING TO THE PRINCIPLE OF AVERAGE UTILITY

I now wish to examine the reasoning that favors the principle of average utility. The classical principle is discussed later (§30). One of the merits of contract theory is that it reveals these principles to be markedly distinct conceptions, however much their practical consequences may coincide. Their underlying analytic assumptions are far apart in the sense that they are associated with contrasting interpretations of the initial situation. Or so I shall try to show.

Applied to the basic structure, the classical principle requires that institutions be arranged to maximize the absolute weighted sum of the expectations of the relevant representative men. This sum is arrived at by weighting each expectation by the number of persons in the corresponding position and then adding. Thus, other