CHAPTER III. THE ORIGINAL POSITION

In this chapter I discuss the favored philosophical interpretation of the initial situation. I refer to this interpretation as the original position. I begin by sketching the nature of the argument for conceptions of justice and explaining how the alternatives are presented so that the parties are to choose from a definite list of traditional conceptions. Then I describe the conditions which characterize the initial situation under several headings: the circumstances of justice, the formal constraints of the concept of right, the veil of ignorance, and the rationality of the contracting parties. In each case I try to indicate why the features adopted for the favored interpretation are reasonable from a philosophical point of view. Next the natural lines of reasoning leading to the two principles of justice and to the principle of average utility are examined prior to a consideration of the relative advantages of these conceptions of justice. I argue that the two principles would be acknowledged and set out some of the main grounds to support this contention. In order to clarify the differences between the various conceptions of justice, the chapter concludes with another look at the classical principle of utility.

20. THE NATURE OF THE ARGUMENT FOR CONCEPTIONS OF JUSTICE

The intuitive idea of justice as fairness is to think of the first principles of justice as themselves the object of an original agreement in a suitably defined initial situation. These principles are those which rational persons concerned to advance their interests would accept in this position of equality to settle the basic terms of their association. It must be shown, then, that the two principles of justice are the solution for the problem of choice presented by the original position. In order to do this, one must establish that, given the circumstances of the parties, and their knowl-
edge, beliefs, and interests, an agreement on these principles is the best way for each person to secure his ends in view of the alternatives available.

Now obviously no one can obtain everything he wants; the mere existence of other persons prevents this. The absolutely best for any man is that everyone else should join with him in furthering his conception of the good whatever it turns out to be. Or failing this, that all others are required to act justly but that he is authorized to exempt himself as he pleases. Since other persons will never agree to such terms of association these forms of egoism would be rejected. The two principles of justice, however, seem to be a reasonable proposal. In fact, I should like to show that these principles are everyone’s best reply, so to speak, to the corresponding demands of the others. In this sense, the choice of this conception of justice is the unique solution to the problem set by the original position.

By arguing in this way one follows a procedure familiar in social theory. That is, a simplified situation is described in which rational individuals with certain ends and related to each other in certain ways are to choose among various courses of action in view of their knowledge of the circumstances. What these individuals will do is then derived by strictly deductive reasoning from these assumptions about their beliefs and interests, their situation and the options open to them. Their conduct is, in the phrase of Pareto, the resultant of tastes and obstacles. 1 In the theory of price, for example, the equilibrium of competitive markets is thought of as arising when many individuals each advancing his own interests give way to each other what they can best part with in return for what they most desire. Equilibrium is the result of agreements freely struck between willing traders. For each person it is the best situation that he can reach by free exchange consistent with the right and freedom of others to further their interests in the same way. It is for this reason that this state of affairs is an equilibrium, one that will persist in the absence of further changes in the circumstances. No one has any incentive to alter it. If a departure from this situation sets in motion tendencies which restore it, the equilibrium is stable.

Of course, the fact that a situation is one of equilibrium, even a stable one, does not entail that it is right or just. It only means that given men’s estimate of their position, they act effectively to preserve it. Clearly a

balance of hatred and hostility may be a stable equilibrium; each may think that any feasible change will be worse. The best that each can do for himself may be a condition of lesser injustice rather than of greater good. The moral assessment of equilibrium situations depends upon the background circumstances which determine them. It is at this point that the conception of the original position embodies features peculiar to moral theory. For while the theory of price, say, tries to account for the movements of the market by assumptions about the actual tendencies at work, the philosophically favored interpretation of the initial situation incorporates conditions which it is thought reasonable to impose on the choice of principles. By contrast with social theory, the aim is to characterize this situation so that the principles that would be chosen, whatever they turn out to be, are acceptable from a moral point of view. The original position is defined in such a way that it is a status quo in which any agreements reached are fair. It is a state of affairs in which the parties are equally represented as moral persons and the outcome is not conditioned by arbitrary contingencies or the relative balance of social forces. Thus justice as fairness is able to use the idea of pure procedural justice from the beginning.

It is clear, then, that the original position is a purely hypothetical situation. Nothing resembling it need ever take place, although we can by deliberately following the constraints it expresses simulate the reflections of the parties. The conception of the original position is not intended to explain human conduct except insofar as it tries to account for our moral judgments and helps to explain our having a sense of justice. Justice as fairness is a theory of our moral sentiments as manifested by our considered judgments in reflective equilibrium. These sentiments presumably affect our thought and action to some degree. So while the conception of the original position is part of the theory of conduct, it does not follow at all that there are actual situations that resemble it. What is necessary is that the principles that would be accepted play the requisite part in our moral reasoning and conduct.

One should note also that the acceptance of these principles is not conjectured as a psychological law or probability. Ideally anyway, I should like to show that their acknowledgment is the only choice consistent with the full description of the original position. The argument aims eventually to be strictly deductive. To be sure, the persons in the original position have a certain psychology, since various assumptions are made about their beliefs and interests. These assumptions appear along with other premises in the description of this initial situation. But clearly arguments
from such premises can be fully deductive, as theories in politics and economics attest. We should strive for a kind of moral geometry with all the rigor which this name connotes. Unhappily the reasoning I shall give will fall far short of this, since it is highly intuitive throughout. Yet it is essential to have in mind the ideal one would like to achieve.

A final remark. There are, as I have said, many possible interpretations of the initial situation. This conception varies depending upon how the contracting parties are conceived, upon what their beliefs and interests are said to be, upon which alternatives are available to them, and so on. In this sense, there are many different contract theories. Justice as fairness is but one of these. But the question of justification is settled, as far as it can be, by showing that there is one interpretation of the initial situation which best expresses the conditions that are widely thought reasonable to impose on the choice of principles yet which, at the same time, leads to a conception that characterizes our considered judgments in reflective equilibrium. This most favored, or standard, interpretation I shall refer to as the original position. We may conjecture that for each traditional conception of justice there exists an interpretation of the initial situation in which its principles are the preferred solution. Thus, for example, there are interpretations that lead to the classical as well as the average principle of utility. These variations of the initial situation will be mentioned as we go along. The procedure of contract theories provides, then, a general analytic method for the comparative study of conceptions of justice. One tries to set out the different conditions embodied in the contractual situation in which their principles would be chosen. In this way one formulates the various underlying assumptions on which these conceptions seem to depend. But if one interpretation is philosophically most favored, and if its principles characterize our considered judgments, we have a procedure for justification as well. We cannot know at first whether such an interpretation exists, but at least we know what to look for.

21. THE PRESENTATION OF ALTERNATIVES

Let us now turn from these remarks on method to the description of the original position. I shall begin with the question of the alternatives open to the persons in this situation. Ideally of course one would like to say that they are to choose among all possible conceptions of justice. One obvious difficulty is how these conceptions are to be characterized so that those in the original position can be presented with them. Yet granting
that these conceptions could be defined, there is no assurance that the parties could make out the best option; the principles that would be most preferred might be overlooked. Indeed, there may exist no best alternative: conceivably for each conception of justice there is another that is better. Even if there is a best alternative, it seems difficult to describe the parties’ intellectual powers so that this optimum, or even the more plausible conceptions, are sure to occur to them. Some solutions to the choice problem may be clear enough on careful reflection; it is another matter to describe the parties so that their deliberations generate these alternatives. Thus although the two principles of justice may be superior to those conceptions known to us, perhaps some hitherto unformulated set of principles is still better.

In order to handle this problem I shall resort to the following device. I shall simply take as given a short list of traditional conceptions of justice, for example those discussed in the first chapter, together with a few other possibilities suggested by the two principles of justice. I then assume that the parties are presented with this list and required to agree unanimously that one conception is the best among those enumerated. We may suppose that this decision is arrived at by making a series of comparisons in pairs. Thus the two principles would be shown to be preferable once all agree that they are to be chosen over each of the other alternatives. In this chapter I shall consider for the most part the choice between the two principles of justice and two forms of the principle of utility (the classical and the average principle). Later on, the comparisons with perfectionism and mixed theories are discussed. In this manner I try to show that the two principles would be chosen from the list.

Now admittedly this is an unsatisfactory way to proceed. It would be better if we could define necessary and sufficient conditions for a uniquely best conception of justice and then exhibit a conception that fulfilled these conditions. Eventually one may be able to do this. For the time being, however, I do not see how to avoid rough and ready methods. Moreover, using such procedures may point to a general solution of our problem. Thus it may turn out that, as we run through these comparisons, the reasoning of the parties singles out certain features of the basic structure as desirable, and that these features have natural maximum and minimum properties. Suppose, for example, that it is rational for the persons in the original position to prefer a society with the greatest equal liberty. And suppose further that while they prefer social and economic advantages to work for the common good they insist that they mitigate the ways in which men are advantaged or disadvantaged by natural and
social contingencies. If these two features are the only relevant ones, and if the principle of equal liberty is the natural maximum of the first feature, and the difference principle (constrained by fair equality of opportunity) the natural maximum of the second, then, leaving aside the problem of priority, the two principles are the optimum solution. The fact that one cannot constructively characterize or enumerate all possible conceptions of justice, or describe the parties so that they are bound to think of them, is no obstacle to this conclusion.

It would not be profitable to pursue these speculations any further. For the present, no attempt is made to deal with the general problem of the best solution. I limit the argument throughout to the weaker contention that the two principles would be chosen from the conceptions of justice on the following list.

A. The Two Principles of Justice (in serial order)
   1. The principle of greatest equal liberty
   2. (a) The principle of (fair) equality of opportunity
      (b) The difference principle

B. Mixed Conceptions. Substitute one for A2 above
   1. The principle of average utility; or
   2. The principle of average utility, subject to a constraint, either:
      (a) That a certain social minimum be maintained, or
      (b) That the overall distribution not be too wide; or
   3. The principle of average utility subject to either constraint in B2 plus that of equality of fair opportunity

C. Classical Teleological Conceptions
   1. The classical principle of utility
   2. The average principle of utility
   3. The principle of perfection

D. Intuitionistic Conceptions
   1. To balance total utility against the principle of equal distribution
   2. To balance average utility against the principle of redress
   3. To balance a list of prima facie principles (as appropriate)

E. Egoistic Conceptions (See §23 where it is explained why strictly speaking the egoistic conceptions are not alternatives.)
   1. First-person dictatorship: Everyone is to serve my interests
   2. Free-rider: Everyone is to act justly except for myself, if I choose not to
   3. General: Everyone is permitted to advance his interests as he pleases
The merits of these traditional theories surely suffice to justify the effort to rank them. And in any case, the study of this ranking is a useful way of feeling one’s way into the larger question. Now each of these conceptions presumably has its assets and liabilities; there are reasons for and against any alternative one selects. The fact that a conception is open to criticism is not necessarily decisive against it, nor are certain desirable features always conclusive in its favor. The decision of the persons in the original position hinges, as we shall see, on a balance of various considerations. In this sense, there is an appeal to intuition at the basis of the theory of justice. Yet when everything is tallied up, it may be perfectly clear where the balance of reason lies. The relevant reasons may have been so factored and analyzed by the description of the original position that one conception of justice is distinctly preferable to the others. The argument for it is not strictly speaking a proof, not yet anyway; but, in Mill’s phrase, it may present considerations capable of determining the intellect.2

The list of conceptions is largely self-explanatory. A few brief comments, however, may be useful. Each conception is expressed in a reasonably simple way, and each holds unconditionally, that is, whatever the circumstances or state of society. None of the principles is contingent upon certain social or other conditions. Now one reason for this is to keep things simple. It would be easy to formulate a family of conceptions each designed to apply only if special circumstances obtain, these various conditions being exhaustive and mutually exclusive. For example one conception might hold at one stage of culture, a different conception at another. Such a family could be counted as itself a conception of justice; it would consist of a set of ordered pairs, each pair being a conception of justice matched with the circumstances in which it applies. But if conceptions of this kind were added to the list, our problem would become very complicated if not unmanageable. Moreover, there is a reason for excluding alternatives of this kind, for it is natural to ask what underlying principle determines the ordered pairs. Here I assume that some recognizably ethical conception specifies the appropriate principles given each of the conditions. It is really this unconditional principle that defines the conception expressed by the set of ordered pairs. Thus to allow such families on the list is to include alternatives that conceal their proper basis. So for this reason as well I shall exclude them. It also turns out to be desirable to characterize the original position so that the parties are to choose princi-

2. Utilitarianism, ch. I, par. 5.
ples that hold unconditionally whatever the circumstances. This fact is connected with the Kantian interpretation of justice as fairness. But I leave this matter aside until later (§40).

Finally, an obvious point. An argument for the two principles, or indeed for any conception, is always relative to some list of alternatives. If we change the list, the argument will, in general, have to be different. A similar sort of remark applies to all features of the original position. There are indefinitely many variations of the initial situation and therefore no doubt indefinitely many theorems of moral geometry. Only a few of these are of any philosophical interest, since most variations are irrelevant from a moral point of view. We must try to steer clear of side issues while at the same time not losing sight of the special assumptions of the argument.

22. THE CIRCUMSTANCES OF JUSTICE

The circumstances of justice may be described as the normal conditions under which human cooperation is both possible and necessary. Thus, as I noted at the outset, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as an identity of interests. There is an identity of interests since social cooperation makes possible a better life for all than any would have if each were to try to live solely by his own efforts. There is a conflict of interests since men are not indifferent as to how the greater benefits produced by their collaboration are distributed, for in order to pursue their ends they each prefer a larger to a lesser share. Thus principles are needed for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These requirements define the role of justice. The background conditions that give rise to these necessities are the circumstances of justice.

These conditions may be divided into two kinds. First, there are the objective circumstances which make human cooperation both possible and necessary. Thus, many individuals coexist together at the same time on a definite geographical territory. These individuals are roughly similar

---

in physical and mental powers; or at any rate, their capacities are comparable in that no one among them can dominate the rest. They are vulnerable to attack, and all are subject to having their plans blocked by the united force of others. Finally, there is the condition of moderate scarcity understood to cover a wide range of situations. Natural and other resources are not so abundant that schemes of cooperation become superfluous, nor are conditions so harsh that fruitful ventures must inevitably break down. While mutually advantageous arrangements are feasible, the benefits they yield fall short of the demands men put forward.

The subjective circumstances are the relevant aspects of the subjects of cooperation, that is, of the persons working together. Thus while the parties have roughly similar needs and interests, or needs and interests in various ways complementary, so that mutually advantageous cooperation among them is possible, they nevertheless have their own plans of life. These plans, or conceptions of the good, lead them to have different ends and purposes, and to make conflicting claims on the natural and social resources available. Moreover, although the interests advanced by these plans are not assumed to be interests in the self, they are the interests of a self that regards its conception of the good as worthy of recognition and that advances claims in its behalf as deserving satisfaction. I also suppose that men suffer from various shortcomings of knowledge, thought, and judgment. Their knowledge is necessarily incomplete, their powers of reasoning, memory, and attention are always limited, and their judgment is likely to be distorted by anxiety, bias, and a preoccupation with their own affairs. Some of these defects spring from moral faults, from selfishness and negligence; but to a large degree, they are simply part of men’s natural situation. As a consequence individuals not only have different plans of life but there exists a diversity of philosophical and religious belief, and of political and social doctrines.

Now this constellation of conditions I shall refer to as the circumstances of justice. Hume’s account of them is especially perspicuous and the preceding summary adds nothing essential to his much fuller discussion. For simplicity I often stress the condition of moderate scarcity (among the objective circumstances), and that of conflict of interests (among the subjective circumstances). Thus, one can say, in brief, that the circumstances of justice obtain whenever persons put forward conflicting claims to the division of social advantages under conditions of moderate scarcity. Unless these circumstances existed there would be no occasion for the virtue of justice, just as in the absence of threats of injury to life and limb there would be no occasion for physical courage.
Several clarifications should be noted. First of all, I shall, of course, assume that the persons in the original position know that these circumstances of justice obtain. This much they take for granted about the conditions of their society. A further assumption is that the parties try to advance their conception of the good as best they can, and that in attempting to do this they are not bound by prior moral ties to each other.

The question arises, however, whether the persons in the original position have obligations and duties to third parties, for example, to their immediate descendants. To say that they do would be one way of handling questions of justice between generations. However, the aim of justice as fairness is to try to derive all duties and obligations of justice from other reasonable conditions. So, if possible, this way out should be avoided. There are several other courses open to us. We can adopt a motivation assumption and think of the parties as representing a continuing line of claims. For example, we can assume that they are heads of families and therefore have a desire to further the well-being of at least their more immediate descendants. Or we can require the parties to agree to principles subject to the constraint that they wish all preceding generations to have followed the very same principles. By an appropriate combination of such stipulations, I believe that the whole chain of generations can be tied together and principles agreed to that suitably take into account the interests of each (§§24, 44). If this is right, we will have succeeded in deriving duties to other generations from reasonable conditions.

It should be noted that I make no restrictive assumptions about the parties’ conceptions of the good except that they are rational long-term plans. While these plans determine the aims and interests of a self, the aims and interests are not presumed to be egoistic or selfish. Whether this is the case depends upon the kinds of ends which a person pursues. If wealth, position, and influence, and the accolades of social prestige, are a person’s final purposes, then surely his conception of the good is egoistic. His dominant interests are in himself, not merely, as they must always be, interests of a self.4 There is no inconsistency, then, in supposing that once the veil of ignorance is removed, the parties find that they have ties of sentiment and affection, and want to advance the interests of others and to see their ends attained. But the postulate of mutual disinterest in the original position is made to insure that the principles of justice do not depend upon strong assumptions. Recall that the original position is meant to incorporate widely shared and yet weak conditions. A conception of

justice should not presuppose, then, extensive ties of natural sentiment. At
the basis of the theory, one tries to assume as little as possible.

Finally, I shall assume that the parties in the original position are
mutually disinterested: they are not willing to have their interests sac-
rificed to the others. The intention is to model men’s conduct and motives
in cases where questions of justice arise. The spiritual ideals of saints and
heroes can be as irreconcilably opposed as any other interests. Conflicts
in pursuit of these ideals are the most tragic of all. Thus justice is the
virtue of practices where there are competing interests and where persons
feel entitled to press their rights on each other. In an association of saints
agreeing on a common ideal, if such a community could exist, disputes
about justice would not occur. Each would work selflessly for one end as
determined by their common religion, and reference to this end (assum-
ing it to be clearly defined) would settle every question of right. But a
human society is characterized by the circumstances of justice. The ac-
count of these conditions involves no particular theory of human motiva-
tion. Rather, its aim is to reflect in the description of the original position
the relations of individuals to one another which set the stage for ques-
tions of justice.

23. THE FORMAL CONSTRAINTS OF
THE CONCEPT OF RIGHT

The situation of the persons in the original position reflects certain con-
straints. The alternatives open to them and their knowledge of their cir-
cumstances are limited in various ways. These restrictions I refer to as the
constraints of the concept of right since they hold for the choice of all
ethical principles and not only for those of justice. If the parties were to
acknowledge principles for the other virtues as well, these constraints
would also apply.

I shall consider first the constraints on the alternatives. There are
certain formal conditions that it seems reasonable to impose on the con-
ceptions of justice that are to be allowed on the list presented to the
parties. I do not claim that these conditions follow from the concept of
right, much less from the meaning of morality. I avoid an appeal to the
analysis of concepts at crucial points of this kind. There are many con-
straints that can reasonably be associated with the concept of right, and
different selections can be made from these and counted as definitive
within a particular theory. The merit of any definition depends upon the
soundness of the theory that results; by itself, a definition cannot settle any fundamental question.\textsuperscript{5}

The propriety of these formal conditions is derived from the task of principles of right in adjusting the claims that persons make on their institutions and one another. If the principles of justice are to play their role, that of assigning basic rights and duties and determining the division of advantages, these requirements are natural enough. Each of them is suitably weak and I assume that they are satisfied by the traditional conceptions of justice. These conditions do, however, exclude the various forms of egoism, as I note below, which shows that they are not without moral force. This makes it all the more necessary that the conditions not be justified by definition or the analysis of concepts, but only by the reasonableness of the theory of which they are a part. I arrange them under five familiar headings.

First of all, principles should be general. That is, it must be possible to formulate them without the use of what would be intuitively recognized as proper names, or rigged definite descriptions. Thus the predicates used in their statement should express general properties and relations. Unfortunately deep philosophical difficulties seem to bar the way to a satisfactory account of these matters.\textsuperscript{6} I shall not try to deal with them here. In presenting a theory of justice one is entitled to avoid the problem of defining general properties and relations and to be guided by what seems reasonable. Further, since the parties have no specific information about themselves or their situation, they cannot identify themselves anyway. Even if a person could get others to agree, he does not know how to tailor principles to his advantage. The parties are effectively forced to stick to general principles, understanding the notion here in an intuitive fashion.

The naturalness of this condition lies in part in the fact that first


\textsuperscript{6} See, for example, W. V. Quine, \textit{Ontological Relativity and Other Essays} (New York, Columbia University Press, 1969), ch. 5 entitled “Natural Kinds.”
principles must be capable of serving as a public charter of a well-ordered society in perpetuity. Being unconditional, they always hold (under the circumstances of justice), and the knowledge of them must be open to individuals in any generation. Thus, to understand these principles should not require a knowledge of contingent particulars, and surely not a reference to individuals or associations. Traditionally the most obvious test of this condition is the idea that what is right is that which accords with God’s will. But in fact this doctrine is normally supported by an argument from general principles. For example, Locke held that the fundamental principle of morals is the following: if one person is created by another (in the theological sense), then that person has a duty to comply with the precepts set to him by his creator. This principle is perfectly general and given the nature of the world on Locke’s view, it singles out God as the legitimate moral authority. The generality condition is not violated, although it may appear so at first sight.

Next, principles are to be universal in application. They must hold for everyone in virtue of their being moral persons. Thus I assume that each can understand these principles and use them in his deliberations. This imposes an upper bound of sorts on how complex they can be, and on the kinds and number of distinctions they draw. Moreover, a principle is ruled out if it would be self-contradictory, or self-defeating, for everyone to act upon it. Similarly, should a principle be reasonable to follow only when others conform to a different one, it is also inadmissible. Principles are to be chosen in view of the consequences of everyone’s complying with them.

As defined, generality and universality are distinct conditions. For example, egoism in the form of first-person dictatorship (Everyone is to serve my—or Pericles’—interests) satisfies universality but not generality. While all could act in accordance with this principle, and the results might in some cases not be at all bad, depending on the interests of the dictator, the personal pronoun (or the name) violates the first condition. Again, general principles may not be universal. They may be framed to hold for a restricted class of individuals, for instance those singled out by special biological or social characteristics, such as hair color or class situation, or whatever. To be sure, in the course of their lives individuals acquire obligations and assume duties that are peculiar to them. Never-

theless these various duties and obligations are the consequence of first principles that hold for all as moral persons; the derivation of these requirements has a common basis.

A third condition is that of publicity, which arises naturally from a contractarian standpoint. The parties assume that they are choosing principles for a public conception of justice. They suppose that everyone will know about these principles all that he would know if their acceptance were the result of an agreement. Thus the general awareness of their universal acceptance should have desirable effects and support the stability of social cooperation. The difference between this condition and that of universality is that the latter leads one to assess principles on the basis of their being intelligently and regularly followed by everyone. But it is possible that all should understand and follow a principle and yet this fact not be widely known or explicitly recognized. The point of the publicity condition is to have the parties evaluate conceptions of justice as publicly acknowledged and fully effective moral constitutions of social life. The publicity condition is clearly implicit in Kant’s doctrine of the categorical imperative insofar as it requires us to act in accordance with principles that one would be willing as a rational being to enact as law for a kingdom of ends. He thought of this kingdom as an ethical commonwealth, as it were, which has such moral principles for its public charter.

A further condition is that a conception of right must impose an ordering on conflicting claims. This requirement springs directly from the role of its principles in adjusting competing demands. There is a difficulty, however, in deciding what counts as an ordering. It is clearly desirable that a conception of justice be complete, that is, able to order all the claims that can arise (or that are likely to in practice). And the ordering

8. Publicity is clearly implied in Kant’s notion of the moral law, but the only place I know of where he discusses it expressly is in Perpetual Peace, appendix II; see Political Writings, ed. Hans Reiss and trans. H. B. Nisbet (Cambridge, The University Press, 1970), pp. 125–130. There are of course brief statements elsewhere. For example, in The Metaphysics of Morals, pt. I (Rechtslehre), §43, he says: “Public Right is the sum total of those laws which require to be made universally public in order to produce a state of right.” In “Theory and Practice” he remarks in a footnote: “No right in a state can be tacitly and treacherously included by a secret reservation, and least of all a right which the people claim to be a part of the constitution, for all laws within it must be thought of as arising out of a public will. Thus if a constitution allowed rebellion, it would have to declare this right publicly and make clear how it might be implemented.” Political Writings, pp. 136, 84n, respectively. I believe Kant intends this condition to apply to a society’s conception of justice. See also note 4, §51, below; and Baier, cited in note 5 above. There is a discussion of common knowledge and its relation to agreement in D. K. Lewis, Convention (Cambridge, Mass., Harvard University Press, 1969), esp. pp. 52–60, 83–88.
should in general be transitive: if, say, a first arrangement of the basic structure is ranked more just than a second, and the second more just than a third, then the first should be more just than the third. These formal conditions are natural enough, though not always easy to satisfy. But is trial by combat a form of adjudication? After all, physical conflict and resort to arms result in an ordering; certain claims do win out over others. The main objection to this ordering is not that it may be intransitive. Rather, it is to avoid the appeal to force and cunning that the principles of right and justice are accepted. Thus I assume that to each according to his threat advantage is not a conception of justice. It fails to establish an ordering in the required sense, an ordering based on certain relevant aspects of persons and their situation which are independent from their social position, or their capacity to intimidate and coerce.

The fifth and last condition is that of finality. The parties are to assess the system of principles as the final court of appeal in practical reasoning. There are no higher standards to which arguments in support of claims can be addressed; reasoning successfully from these principles is conclusive. If we think in terms of the fully general theory which has principles for all the virtues, then such a theory specifies the totality of relevant considerations and their appropriate weights, and its requirements are decisive. They override the demands of law and custom, and of social rules generally. We are to arrange and respect social institutions as the princi-


10. Theory of Games as a Tool for the Moral Philosopher (Cambridge, The University Press, 1955). On the analysis he presents, it turns out that the fair division of playing time between Matthew and Luke depends on their preferences, and these in turn are connected with the instruments they wish to play. Since Matthew has a threat advantage over Luke, arising from the fact that Matthew, the trumpeter, prefers both of them playing at once to neither of them playing, whereas Luke, the pianist, prefers silence to cacophony, Matthew is allotted twenty-six evenings of play to Luke’s seventeen. If the situation were reversed, the threat advantage would be with Luke. See pp. 36f. But we have only to suppose that Matthew is a jazz enthusiast who plays the drums, and Luke a violinist who plays sonatas, in which case it will be fair on this analysis for Matthew to play whenever and as often as he likes, assuming as it is plausible to assume that he does not care whether Luke plays or not. Clearly something has gone wrong. What is lacking is a suitable definition of a status quo that is acceptable from a moral point of view. We cannot take various contingencies as known and individual preferences as given and expect to elucidate the concept of justice (or fairness) by theories of bargaining. The conception of the original position is designed to meet the problem of the appropriate status quo. A similar objection to Braithwaite’s analysis is found in J. R. Lucas, “Moralists and Gamesmen,” Philosophy, vol. 34 (1959), pp. 9f. For another discussion, consult Sen, Collective Choice and Social Welfare, pp. 118–123, who argues that the solution of J. F. Nash in “The Bargaining Problem,” Econometrica, vol. 18 (1950), is similarly defective from an ethical point of view.
ples of right and justice direct. Conclusions from these principles also
override considerations of prudence and self-interest. This does not mean
that these principles insist upon self-sacrifice; for in drawing up the
conception of right the parties take their interests into account as best
they can. The claims of personal prudence are already given an appropri-
ate weight within the full system of principles. The complete scheme is
final in that when the course of practical reasoning it defines has reached
its conclusion, the question is settled. The claims of existing social ar-
rangements and of self-interest have been duly allowed for. We cannot at
the end count them a second time because we do not like the result.

Taken together, then, these conditions on conceptions of right come to
this: a conception of right is a set of principles, general in form and
universal in application, that is to be publicly recognized as a final court
of appeal for ordering the conflicting claims of moral persons. Principles
of justice are identified by their special role and the subject to which they
apply. Now by themselves the five conditions exclude none of the tradi-
tional conceptions of justice. It should be noted, however, that they do
drive out the listed variants of egoism. The generality condition eliminates
both 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. Genera-
liaty 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 sig-
nificance 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.11

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 re-

11. 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. See the last paragraph of §27.
sources and the environment of nature. There is also, theoretically any-
way, 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 what-
ever 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 organiza-
tion 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 gen-
eral 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 princi-
pies. 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 restric-
tions. In arguing for a conception of justice we must be sure that it is
among the permitted alternatives and satisfies the stipulated formal con-
straints. 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 applica-
tion, 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. (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 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 situ-
ation 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 (§22). 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 (§44). Whatever a person’s temporal position, each is forced to choose for all.¹²

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 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. 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 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 the complexity of 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. But I have also assumed that they 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 normally 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. 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. 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).

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 con-
ception 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
difference 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 accord-
ance 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 cho-
sen will be respected. It is essential to observe, however, that this assump-
tion 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 exam-
ine. 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 variations. (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 individu-
als. 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 Schopenhauers thought of Kant’s doctrine,
that it is egoistic nevertheless.15 Now this is a misconception. For the fact
that in the original position the parties are characterized as mutually
disinterested does not entail that persons in ordinary life, or in a well-or-
dered society, who hold the principles that would be agreed to are simi-
larly 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 effec-
tive desire to comply with these restrictions. The motivation of the per-
sons in the original position must not be confused with the motivation of
persons in everyday life who accept the principles of justice 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 posi-
tion, his desires and aims are surely not egoistic. He voluntarily takes on
the limitations expressed by this interpretation of the moral point of view.
Thus, more generally, the motivation of the parties in the original position
does not determine directly the motivation of people in a just society. For
in the latter case, we assume that its members grow up and live under a
just basic structure, as the two principles require; and then we try to work
out what kind of conception of the good and moral sentiments people
would acquire (Ch. VIII). Therefore the mutual disinterestedness of the
parties determines other motivations only indirectly, that is, via its effects
on the agreement on principles. It is these principles, together with the
laws of psychology (as these work under the conditions of just institu-
tions), which shape the aims and moral sentiments of citizens of a well-
ordered society.

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
much the same purpose as benevolence. For this combination of condi-
tions 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 disinterestedness 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.

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.16

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 reasonable list of traditional conceptions of justice.\(^\text{17}\) 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.

Now consider the point of view of anyone in the original position. 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 primary goods, and since it is not rational for him to agree to less, the sensible thing is to acknowledge as the first step a principle of justice requiring an equal distribution. Indeed, this principle is so obvious given the symmetry of the parties that it would occur to everyone immediately. Thus the parties start with a principle requiring equal basic liberties for all, as well as fair equality of opportunity and equal division of income and wealth.

But even holding firm to the priority of the basic liberties and fair equality of opportunity, there is no reason why this initial acknowledgment should be final. Society should take into account economic efficiency and the requirements of organization and technology. If there are inequalities in income and wealth, and differences in authority and de-

degrees of responsibility, that work to make everyone better off in comparison with the benchmark of equality, why not permit them? One might think that ideally individuals should want to serve one another. But since the parties are assumed to be mutually disinterested, their acceptance of these economic and institutional inequalities is only the recognition of the relations of opposition in which men stand in the circumstances of justice. They have no grounds for complaining of one another’s motives. Thus the parties would object to these differences only if they would be dejected by the bare knowledge or perception that others are better situated; but I suppose that they decide as if they are not moved by envy. Thus the basic structure should allow these inequalities so long as these improve everyone’s situation, including that of the least advantaged, provided that they are consistent with equal liberty and fair opportunity. Because the parties start from an equal division of all social primary goods, those who benefit least have, so to speak, a veto. Thus we arrive at the difference principle. Taking equality as the basis of comparison, those who have gained more must do so on terms that are justifiable to those who have gained the least.

By some such reasoning, then, the parties might arrive at the two principles of justice in serial order. I shall not try to justify this ordering here, but the following remarks may convey the intuitive idea. I assume that the parties view themselves as free persons who have fundamental aims and interests in the name of which they think it legitimate for them to make claims on one another concerning the design of the basic structure of society. The religious interest is a familiar historical example; the interest in the integrity of the person is another. In the original position the parties do not know what particular forms these interests take; but they do assume that they have such interests and that the basic liberties necessary for their protection are guaranteed by the first principle. Since they must secure these interests, they rank the first principle prior to the second. The case for the two principles can be strengthened by spelling out in more detail the notion of a free person. Very roughly the parties regard themselves as having a highest-order interest in how all their other interests, including even their fundamental ones, are shaped and regulated by social institutions. They do not think of themselves as inevitably bound to, or as identical with, the pursuit of any particular complex of fundamental interests that they may have at any given time, although they want the right to advance such interests (provided they are admissible). Rather, free persons conceive of themselves as beings who can revise and alter their final ends and who give first priority to preserving their liberty.
in these matters. Hence, they not only have final ends that they are in principle free to pursue or to reject, but their original allegiance and continued devotion to these ends are to be formed and affirmed under conditions that are free. Since the two principles secure a social form that maintains these conditions, they would be agreed to rather than the principle of utility. Only by this agreement can the parties be sure that their highest-order interest as free persons is guaranteed.

The priority of liberty means that whenever the basic liberties can be effectively established, a lesser or an unequal liberty cannot be exchanged for an improvement in economic well-being. It is only when social circumstances do not allow the effective establishment of these basic rights that one can concede their limitation; and even then these restrictions can be granted only to the extent that they are necessary to prepare the way for the time when they are no longer justified. The denial of the equal liberties can be defended only when it is essential to change the conditions of civilization so that in due course these liberties can be enjoyed. Thus in adopting the serial order of the two principles, the parties are assuming that the conditions of their society, whatever they are, admit the effective realization of the equal liberties. Or that if they do not, circumstances are nevertheless sufficiently favorable so that the priority of the first principle points out the most urgent changes and identifies the preferred path to the social state in which all the basic liberties can be fully instituted. The complete realization of the two principles in serial order is the long-run tendency of this ordering, at least under reasonably fortunate conditions.

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. 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 a relation between the two principles and the maximin rule for choice under uncertainty. ¹⁸ This is evident from the fact that the two principles

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-
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.\footnote{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.}

\[
\begin{array}{c|ccc}
\text{Circumstances} & c_1 & c_2 & c_3 \\
\hline
\text{Decisions} & d_1 & d_2 & d_3 \\
\hline
\text{d}_1 & -7 & 8 & 12 \\
\text{d}_2 & -8 & 7 & 14 \\
\text{d}_3 & 5 & 6 & 8 \\
\end{array}
\]

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 \textit{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. 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 $\Sigma p_j = 1$. Then the expectation for the $i$th decision is equal to $\Sigma p_j g_{ij}$.) Thus it must be, for example, that the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure. In this case it is unreasonable not to be skeptical of probabilistic calculations unless there is no other way out, particularly if the decision is a fundamental one that needs to be justified to others.

The second feature that suggests the maximin rule is the following: the person choosing has a conception of the good such that he cares very little, if anything, for what he might gain above the minimum stipend that he can, in fact, be sure of by following the maximin rule. It is not worthwhile for him to take a chance for the sake of a further advantage, especially when it may turn out that he loses much that is important to him. This last provision brings in the third feature, namely, that the rejected alternatives have outcomes that one can hardly accept. The situation involves grave risks. Of course these features work most effectively in combination. The paradigm situation for following the maximin rule is when all three features are realized to the highest degree.

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

20. Here I borrow from William Fellner, *Probability and Profit* (Homewood, Ill., R. D. Irwin, Inc., 1965), pp. 140–142, where these features are noted.
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, especially in Part Two, 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 (§§33–35).

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.

So much, then, for a brief sketch of the features of situations in which the maximin rule is a useful maxim 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 exhibits these special features to a sufficiently 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. I shall 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 con-
straints) the 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 they are necessary to raise the expectations of the least
fortunate 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 fol-
lowing familiar difficulty with the maximin rule. Consider the sequence
of gain-and-loss tables:

\[
\begin{array}{cc}
0 & n \\
\frac{1}{n} & 1 \\
\end{array}
\]

for all natural numbers \( n \). Even if for some smallish number it is reason-
able 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 \textit{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 combina-
tions. The possibilities which the objection envisages cannot arise in real
cases; the feasible set is so restricted that they are excluded.\textsuperscript{21} 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 fair equality of opportu-
nity prevents these contingencies from occurring. For we raise the expec-
tations of the more advantaged only in ways required to improve the
situation of the worst off. For the greater expectations of the more favored
presumably cover the costs of training or answer to organizational re-
quirements, thereby contributing to the general advantage. While nothing

\textsuperscript{21}. I am indebted to S. A. Marglin for this point,
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 estab-
lished 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 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 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 calcula-
tions are normally against such denials of liberty.

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 utilitarian’s 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. But first a word about the meaning of utility. It is understood in the traditional sense as the satisfaction of desire; and it admits of interpersonal comparisons that can at least be summed at the margin. I assume also that utility is measured by some procedure that is independent of choices involving risk, say by postulating an ability to rank differences between levels of satisfaction. These are the traditional assumptions; and while they are very strong, they will not be criticized here. As far as possible, I want to examine the historical doctrine on its own terms.

Applied to the basic structure, the classical principle requires that institutions be arranged to maximize the absolute weighted sum of the ex-
pectations 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 things equal, when the number of persons in society doubles, total utility is twice as great. (Of course, on the utilitarian view expectations are to measure total satisfactions enjoyed and foreseen. They are not, as in justice as fairness, merely indexes of primary goods.) By contrast, the principle of average utility directs society to maximize not the total but the average utility (per capita). This seems to be a more modern view: it was held by Mill and Wicksell, and recently others have given it a new foundation. To apply this conception to the basic structure, institutions are set up so as to maximize the percentage weighted sum of the expectations of representative individuals. To compute this sum we multiply expectations by the fraction of society at the corresponding position. Thus it is no longer true that, other things equal, when a community doubles its population the utility is twice as great. To the contrary, as long as the percentages in the various positions are unchanged, the utility remains the same.

Which of these principles of utility would be preferred in the original position? To answer this question, one should note that both variations come to the same thing if population size is constant. But when population is subject to change, there is a difference. The classical principle requires that so far as institutions affect the size of families, the age of marriage, and the like, they should be arranged so that the maximum of total utility is achieved. This entails that so long as the average utility per person falls slowly enough when the number of individuals increases, the population should be encouraged to grow indefinitely no matter how low the average has fallen. In this case the sum of utilities added by the greater number of persons is sufficiently great to make up for the decline in the share per capita. As a matter of justice and not of preference, a very low average of well-being may be required. (See the following figure.)

INDEFINITE INCREASE OF POPULATION

Formally the condition for increasing population size indefinitely is that the curve $y = F(x)$, where $y$ is average per capita utility and $x$ is population size, should be flatter than the rectangular hyperbola $xy = c$. For $xy$ equals the total utility, and the area of the rectangle representing this total increases as $x$ increases whenever the curve $y = F(x)$ is flatter than $xy = c$.

Now this consequence of the classical principle seems to show that it would be rejected by the parties in favor of the average principle. The two principles would be equivalent only if it is supposed that average well-being always falls sufficiently fast (beyond a certain point anyway) so that there is no serious conflict between them. But this assumption seems questionable. From the standpoint of the persons in the original position, it would appear more rational to agree to some sort of floor to hold up average welfare. Since the parties aim to advance their own interests, they have no desire in any event to maximize the sum total of satisfaction. I assume, therefore, that the more plausible utilitarian alternative to the two principles of justice is the average and not the classical principle.

I now wish to consider how the parties might arrive at the average principle. The reasoning I shall sketch is perfectly general and if it were sound it would sidestep entirely the problem of how to present the alternatives. The average principle would be recognized as the only reasonable candidate. Imagine a situation in which a single rational individual can choose which of several societies to enter. To fix ideas, assume first that the members of these societies all have the same preferences. And assume also that these preferences satisfy conditions that enable one to

---

define a cardinal utility. Further, each society has the same resources and
the same distribution of natural talents. Nevertheless, individuals with
different talents have different incomes; and each society has a redistribu-
tion policy which if pushed beyond a certain point weakens incentives
and thereby lowers production. Supposing that different policies are fol-
lowed in these societies, how will a single individual decide which soci-
ety to join? If he knows his own abilities and interests precisely, and if he
has detailed information about these societies, he may be able to foresee
the well-being that he will almost certainly enjoy in each one. He can
then decide on this basis. There is no need for him to make any prob-
abilistic calculations.

But this case is rather special. Let us alter it step by step so that it
increasingly resembles that of someone in the original position. Thus,
suppose first that the hypothetical joiner is unsure about the role his
talents will enable him to fill in these various societies. If he assumes that
his preferences are the same as everyone else, he may decide by trying to
maximize his expected well-being. He computes his prospect for a given
society by taking as the alternative utilities those of the representative
members of that society and as the likelihoods for each position his
estimates of his chances of attaining it. His expectation is defined, then,
by a weighted sum of utilities of representative individuals, that is, by the
expression \[ \sum p_i u_i \], where \( p_i \) is the likelihood of his achieving the ith
position, and \( u_i \) the utility of the corresponding representative man. He
then chooses the society offering the highest prospect.

Several further modifications bring the situation closer to that of the
original position. Assume that the hypothetical joiner knows nothing
about either his abilities or the place he is likely to hold in each society. It
is still assumed, though, that his preferences are the same as the people in
these societies. Now suppose that he continues to reason along prob-
abilistic lines by holding that he has an equal chance of being any individ-
ual (that is, that his chance of falling under any representative man is the
fraction of society that this man represents). In this case his prospects are
still identical with the average utility for each society. These modifica-
tions have at last brought his expected gains for each society in line with
its average welfare.

So far we have assumed that all individuals have similar preferences
whether or not they belong to the same society. Their conceptions of the
good are roughly the same. Once this highly restrictive assumption is
dropped, we take the final step and arrive at a variation of the initial
situation. Nothing is known, let us say, about the particular preferences of