CHAPTER I. JUSTICE AS FAIRNESS

In this introductory chapter I sketch some of the main ideas of the theory of justice I wish to develop. The exposition is informal and intended to prepare the way for the more detailed arguments that follow. Unavoidably there is some overlap between this and later discussions. I begin by describing the role of justice in social cooperation and with a brief account of the primary subject of justice, the basic structure of society. I then present the main idea of justice as fairness, a theory of justice that generalizes and carries to a higher level of abstraction the traditional conception of the social contract. The compact of society is replaced by an initial situation that incorporates certain procedural constraints on arguments designed to lead to an original agreement on principles of justice. I also take up, for purposes of clarification and contrast, the classical utilitarian and intuitionist conceptions of justice and consider some of the differences between these views and justice as fairness. My guiding aim is to work out a theory of justice that is a viable alternative to these doctrines which have long dominated our philosophical tradition.

1. THE ROLE OF JUSTICE

Justice is the first virtue of social institutions, as truth is of systems of thought. A theory however elegant and economical must be rejected or revised if it is untrue; likewise laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust. Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. For this reason justice denies that the loss of freedom for some is made right by a greater good shared by others. It does not allow that the sacrifices imposed on a few are outweighed by the larger sum of advantages enjoyed by many. Therefore in a just society the liberties of equal citizenship are taken as settled;
the rights secured by justice are not subject to political bargaining or to the calculus of social interests. The only thing that permits us to acquiesce in an erroneous theory is the lack of a better one; analogously, an injustice is tolerable only when it is necessary to avoid an even greater injustice. Being first virtues of human activities, truth and justice are uncompromising.

These propositions seem to express our intuitive conviction of the primacy of justice. No doubt they are expressed too strongly. In any event I wish to inquire whether these contentions or others similar to them are sound, and if so how they can be accounted for. To this end it is necessary to work out a theory of justice in the light of which these assertions can be interpreted and assessed. I shall begin by considering the role of the principles of justice. Let us assume, to fix ideas, that a society is a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them. Suppose further that these rules specify a system of cooperation designed to advance the good of those taking part in it. Then, although a society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by 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 live solely by his own efforts. There is a conflict of interests since persons 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. A set of principles is required for choosing among the various social arrangements which determine this division of advantages and for underwriting an agreement on the proper distributive shares. These principles are the principles of social justice: they provide a way of assigning rights and duties in the basic institutions of society and they define the appropriate distribution of the benefits and burdens of social cooperation.

Now let us say that a society is well-ordered when it is not only designed to advance the good of its members but when it is also effectively regulated by a public conception of justice. That is, it is a society in which (1) everyone accepts and knows that the others accept the same principles of justice, and (2) the basic social institutions generally satisfy and are generally known to satisfy these principles. In this case while men may put forth excessive demands on one another, they nevertheless acknowledge a common point of view from which their claims may be adjudicated. If men’s inclination to self-interest makes their vigilance
against one another necessary, their public sense of justice makes their secure association together possible. Among individuals with disparate aims and purposes a shared conception of justice establishes the bonds of civic friendship; the general desire for justice limits the pursuit of other ends. One may think of a public conception of justice as constituting the fundamental charter of a well-ordered human association.

Existing societies are of course seldom well-ordered in this sense, for what is just and unjust is usually in dispute. Men disagree about which principles should define the basic terms of their association. Yet we may still say, despite this disagreement, that they each have a conception of justice. That is, they understand the need for, and they are prepared to affirm, a characteristic set of principles for assigning basic rights and duties and for determining what they take to be the proper distribution of the benefits and burdens of social cooperation. Thus it seems natural to think of the concept of justice as distinct from the various conceptions of justice and as being specified by the role which these different sets of principles, these different conceptions, have in common.¹ Those who hold different conceptions of justice can, then, still agree that institutions are just when no arbitrary distinctions are made between persons in the assigning of basic rights and duties and when the rules determine a proper balance between competing claims to the advantages of social life. Men can agree to this description of just institutions since the notions of an arbitrary distinction and of a proper balance, which are included in the concept of justice, are left open for each to interpret according to the principles of justice that he accepts. These principles single out which similarities and differences among persons are relevant in determining rights and duties and they specify which division of advantages is appropriate. Clearly this distinction between the concept and the various conceptions of justice settles no important questions. It simply helps to identify the role of the principles of social justice.

Some measure of agreement in conceptions of justice is, however, not the only prerequisite for a viable human community. There are other fundamental social problems, in particular those of coordination, efficiency, and stability. Thus the plans of individuals need to be fitted together so that their activities are compatible with one another and they can all be carried through without anyone’s legitimate expectations being severely disappointed. Moreover, the execution of these plans should lead to the

achievement of social ends in ways that are efficient and consistent with justice. And finally, the scheme of social cooperation must be stable: it must be more or less regularly complied with and its basic rules willingly acted upon; and when infractions occur, stabilizing forces should exist that prevent further violations and tend to restore the arrangement. Now it is evident that these three problems are connected with that of justice. In the absence of a certain measure of agreement on what is just and unjust, it is clearly more difficult for individuals to coordinate their plans efficiently in order to insure that mutually beneficial arrangements are maintained. Distrust and resentment corrode the ties of civility, and suspicion and hostility tempt men to act in ways they would otherwise avoid. So while the distinctive role of conceptions of justice is to specify basic rights and duties and to determine the appropriate distributive shares, the way in which a conception does this is bound to affect the problems of efficiency, coordination, and stability. We cannot, in general, assess a conception of justice by its distributive role alone, however useful this role may be in identifying the concept of justice. We must take into account its wider connections; for even though justice has a certain priority, being the most important virtue of institutions, it is still true that, other things equal, one conception of justice is preferable to another when its broader consequences are more desirable.

2. THE SUBJECT OF JUSTICE

Many different kinds of things are said to be just and unjust: not only laws, institutions, and social systems, but also particular actions of many kinds, including decisions, judgments, and imputations. We also call the attitudes and dispositions of persons, and persons themselves, just and unjust. Our topic, however, is that of social justice. For us the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation. By major institutions I understand the political constitution and the principal economic and social arrangements. Thus the legal protection of freedom of thought and liberty of conscience, competitive markets, private property in the means of production, and the monogamous family are examples of major social institutions. Taken together as one scheme, the major institutions define men’s rights and duties and influence their life prospects, what they can expect to be and how well they can hope to
do. The basic structure is the primary subject of justice because its effects are so profound and present from the start. The intuitive notion here is that this structure contains various social positions and that men born into different positions have different expectations of life determined, in part, by the political system as well as by economic and social circumstances. In this way the institutions of society favor certain starting places over others. These are especially deep inequalities. Not only are they pervasive, but they affect men’s initial chances in life; yet they cannot possibly be justified by an appeal to the notions of merit or desert. It is these inequalities, presumably inevitable in the basic structure of any society, to which the principles of social justice must in the first instance apply. These principles, then, regulate the choice of a political constitution and the main elements of the economic and social system. The justice of a social scheme depends essentially on how fundamental rights and duties are assigned and on the economic opportunities and social conditions in the various sectors of society.

The scope of our inquiry is limited in two ways. First of all, I am concerned with a special case of the problem of justice. I shall not consider the justice of institutions and social practices generally, nor except in passing the justice of the law of nations and of relations between states (§58). Therefore, if one supposes that the concept of justice applies whenever there is an allotment of something rationally regarded as advantageous or disadvantageous, then we are interested in only one instance of its application. There is no reason to suppose ahead of time that the principles satisfactory for the basic structure hold for all cases. These principles may not work for the rules and practices of private associations or for those of less comprehensive social groups. They may be irrelevant for the various informal conventions and customs of everyday life; they may not elucidate the justice, or perhaps better, the fairness of voluntary cooperative arrangements or procedures for making contractual agreements. The conditions for the law of nations may require different principles arrived at in a somewhat different way. I shall be satisfied if it is possible to formulate a reasonable conception of justice for the basic structure of society conceived for the time being as a closed system isolated from other societies. The significance of this special case is obvious and needs no explanation. It is natural to conjecture that once we have a sound theory for this case, the remaining problems of justice will prove more tractable in the light of it. With suitable modifications such a theory should provide the key for some of these other questions.

The other limitation on our discussion is that for the most part I
examine the principles of justice that would regulate a well-ordered society. Everyone is presumed to act justly and to do his part in upholding just institutions. Though justice may be, as Hume remarked, the cautious, jealous virtue, we can still ask what a perfectly just society would be like. Thus I consider primarily what I call strict compliance as opposed to partial compliance theory (§§25, 39). The latter studies the principles that govern how we are to deal with injustice. It comprises such topics as the theory of punishment, the doctrine of just war, and the justification of the various ways of opposing unjust regimes, ranging from civil disobedience and conscientious objection to militant resistance and revolution. Also included here are questions of compensatory justice and of weighing one form of institutional injustice against another. Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things that we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems. The discussion of civil disobedience, for example, depends upon it (§§55–59). At least, I shall assume that a deeper understanding can be gained in no other way, and that the nature and aims of a perfectly just society is the fundamental part of the theory of justice.

Now admittedly the concept of the basic structure is somewhat vague. It is not always clear which institutions or features thereof should be included. But it would be premature to worry about this matter here. I shall proceed by discussing principles which do apply to what is certainly a part of the basic structure as intuitively understood; I shall then try to extend the application of these principles so that they cover what would appear to be the main elements of this structure. Perhaps these principles will turn out to be perfectly general, although this is unlikely. It is sufficient that they apply to the most important cases of social justice. The point to keep in mind is that a conception of justice for the basic structure is worth having for its own sake. It should not be dismissed because its principles are not everywhere satisfactory.

A conception of social justice, then, is to be regarded as providing in the first instance a standard whereby the distributive aspects of the basic structure of society are to be assessed. This standard, however, is not to be confused with the principles defining the other virtues, for the basic structure, and social arrangements generally, may be efficient or ineffi-

cient, liberal or illiberal, and many other things, as well as just or unjust. A complete conception defining principles for all the virtues of the basic structure, together with their respective weights when they conflict, is more than a conception of justice; it is a social ideal. The principles of justice are but a part, although perhaps the most important part, of such a conception. A social ideal in turn is connected with a conception of society, a vision of the way in which the aims and purposes of social cooperation are to be understood. The various conceptions of justice are the outgrowth of different notions of society against the background of opposing views of the natural necessities and opportunities of human life. Fully to understand a conception of justice we must make explicit the conception of social cooperation from which it derives. But in doing this we should not lose sight of the special role of the principles of justice or of the primary subject to which they apply.

In these preliminary remarks I have distinguished the concept of justice as meaning a proper balance between competing claims from a conception of justice as a set of related principles for identifying the relevant considerations which determine this balance. I have also characterized justice as but one part of a social ideal, although the theory I shall propose no doubt extends its everyday sense. This theory is not offered as a description of ordinary meanings but as an account of certain distributive principles for the basic structure of society. I assume that any reasonably complete ethical theory must include principles for this fundamental problem and that these principles, whatever they are, constitute its doctrine of justice. The concept of justice I take to be defined, then, by the role of its principles in assigning rights and duties and in defining the appropriate division of social advantages. A conception of justice is an interpretation of this role.

Now this approach may not seem to tally with tradition. I believe, though, that it does. The more specific sense that Aristotle gives to justice, and from which the most familiar formulations derive, is that of refraining from *pleonexia*, that is, from gaining some advantage for oneself by seizing what belongs to another, his property, his reward, his office, and the like, or by denying a person that which is due to him, the fulfillment of a promise, the repayment of a debt, the showing of proper respect, and so on.³ It is evident that this definition is framed to apply to actions, and

persons are thought to be just insofar as they have, as one of the permanent elements of their character, a steady and effective desire to act justly. Aristotle’s definition clearly presupposes, however, an account of what properly belongs to a person and of what is due to him. Now such entitlements are, I believe, very often derived from social institutions and the legitimate expectations to which they give rise. There is no reason to think that Aristotle would disagree with this, and certainly he has a conception of social justice to account for these claims. The definition I adopt is designed to apply directly to the most important case, the justice of the basic structure. There is no conflict with the traditional notion.

3. THE MAIN IDEA OF THE THEORY OF JUSTICE

My aim is to present a conception of justice which generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant. In order to do this we are not to think of the original contract as one to enter a particular society or to set up a particular form of government. Rather, the guiding idea is that the principles of justice for the basic structure of society are the object of the original agreement. They are the principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association. These principles are to regulate all further agreements; they specify the kinds of social cooperation that can be entered into and the forms of government that can be established. This way of regarding the principles of justice I shall call justice as fairness.

Thus we are to imagine that those who engage in social cooperation choose together, in one joint act, the principles which are to assign basic rights and duties and to determine the division of social benefits. Men are to decide in advance how they are to regulate their claims against one another and what is to be the foundation charter of their society. Just as each person must decide by rational reflection what constitutes his good,

that is, the system of ends which it is rational for him to pursue, so a
group of persons must decide once and for all what is to count among
them as just and unjust. The choice which rational men would make in
this hypothetical situation of equal liberty, assuming for the present that
this choice problem has a solution, determines the principles of justice.

In justice as fairness the original position of equality corresponds to
the state of nature in the traditional theory of the social contract. This
original position is not, of course, thought of as an actual historical state
of affairs, much less as a primitive condition of culture. It is understood
as a purely hypothetical situation characterized so as to lead to a certain
conception of justice. 5 Among the essential features of this situation is
that no one knows his place in society, his class position or social status,
nor does any one know his fortune in the distribution of natural assets and
abilities, his intelligence, strength, and the like. I shall even assume that
the parties do not know their conceptions of the good or their special
psychological propensities. The principles of justice are chosen behind a
veil of ignorance. This ensures that no one is advantaged or disadvan-
taged in the choice of principles by the outcome of natural chance or the
contingency of social circumstances. Since all are similarly situated and
no one is able to design principles to favor his particular condition, the
principles of justice are the result of a fair agreement or bargain. For
given the circumstances of the original position, the symmetry of every-
one’s relations to each other, this initial situation is fair between individu-
als as moral persons, that is, as rational beings with their own ends and
capable, I shall assume, of a sense of justice. The original position is, one
might say, the appropriate initial status quo, and thus the fundamental
agreements reached in it are fair. This explains the propriety of the name
“justice as fairness”: it conveys the idea that the principles of justice are
agreed to in an initial situation that is fair. The name does not mean that
the concepts of justice and fairness are the same, any more than the
phrase “poetry as metaphor” means that the concepts of poetry and meta-
phor are the same.

Justice as fairness begins, as I have said, with one of the most general
of all choices which persons might make together, namely, with the

5. Kant is clear that the original agreement is hypothetical. See The Metaphysics of Morals, pt. I
(Rechtslehre), especially §§47, 52; and pt. II of the essay “Concerning the Common Saying: This
May Be True in Theory but It Does Not Apply in Practice,” in Kant’s Political Writings, ed. Hans
136, for a further discussion.
choice of the first principles of a conception of justice which is to regulate all subsequent criticism and reform of institutions. Then, having chosen a conception of justice, we can suppose that they are to choose a constitution and a legislature to enact laws, and so on, all in accordance with the principles of justice initially agreed upon. Our social situation is just if it is such that by this sequence of hypothetical agreements we would have contracted into the general system of rules which defines it. Moreover, assuming that the original position does determine a set of principles (that is, that a particular conception of justice would be chosen), it will then be true that whenever social institutions satisfy these principles those engaged in them can say to one another that they are cooperating on terms to which they would agree if they were free and equal persons whose relations with respect to one another were fair. They could all view their arrangements as meeting the stipulations which they would acknowledge in an initial situation that embodies widely accepted and reasonable constraints on the choice of principles. The general recognition of this fact would provide the basis for a public acceptance of the corresponding principles of justice. No society can, of course, be a scheme of cooperation which men enter voluntarily in a literal sense; each person finds himself placed at birth in some particular position in some particular society, and the nature of this position materially affects his life prospects. Yet a society satisfying the principles of justice as fairness comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are autonomous and the obligations they recognize self-imposed.

One feature of justice as fairness is to think of the parties in the initial situation as rational and mutually disinterested. This does not mean that the parties are egoists, that is, individuals with only certain kinds of interests, say in wealth, prestige, and domination. But they are conceived as not taking an interest in one another’s interests. They are to presume that even their spiritual aims may be opposed, in the way that the aims of those of different religions may be opposed. Moreover, the concept of rationality must be interpreted as far as possible in the narrow sense, standard in economic theory, of taking the most effective means to given ends. I shall modify this concept to some extent, as explained later (§25), but one must try to avoid introducing into it any controversial ethical elements. The initial situation must be characterized by stipulations that are widely accepted.

In working out the conception of justice as fairness one main task
clearly is to determine which principles of justice would be chosen in the original position. To do this we must describe this situation in some detail and formulate with care the problem of choice which it presents. These matters I shall take up in the immediately succeeding chapters. It may be observed, however, that once the principles of justice are thought of as arising from an original agreement in a situation of equality, it is an open question whether the principle of utility would be acknowledged. Off-hand it hardly seems likely that persons who view themselves as equals, entitled to press their claims upon one another, would agree to a principle which may require lesser life prospects for some simply for the sake of a greater sum of advantages enjoyed by others. Since each desires to protect his interests, his capacity to advance his conception of the good, no one has a reason to acquiesce in an enduring loss for himself in order to bring about a greater net balance of satisfaction. In the absence of strong and lasting benevolent impulses, a rational man would not accept a basic structure merely because it maximized the algebraic sum of advantages irrespective of its permanent effects on his own basic rights and interests. Thus it seems that the principle of utility is incompatible with the conception of social cooperation among equals for mutual advantage. It appears to be inconsistent with the idea of reciprocity implicit in the notion of a well-ordered society. Or, at any rate, so I shall argue.

I shall maintain instead that the persons in the initial situation would choose two rather different principles: the first requires equality in the assignment of basic rights and duties, while the second holds that social and economic inequalities, for example inequalities of wealth and authority, are just only if they result in compensating benefits for everyone, and in particular for the least advantaged members of society. These principles rule out justifying institutions on the grounds that the hardships of some are offset by a greater good in the aggregate. It may be expedient but it is not just that some should have less in order that others may prosper. But there is no injustice in the greater benefits earned by a few provided that the situation of persons not so fortunate is thereby improved. The intuitive idea is that since everyone’s well-being depends upon a scheme of cooperation without which no one could have a satisfactory life, the division of advantages should be such as to draw forth the willing cooperation of everyone taking part in it, including those less well situated. The two principles mentioned seem to be a fair basis on which those better endowed, or more fortunate in their social position, neither of which we can be said to deserve, could expect the willing cooperation of others when some workable scheme is a necessary condition of the wel-
fare of all.\textsuperscript{6} Once we decide to look for a conception of justice that prevents the use of the accidents of natural endowment and the contingencies of social circumstance as counters in a quest for political and economic advantage, we are led to these principles. They express the result of leaving aside those aspects of the social world that seem arbitrary from a moral point of view.

The problem of the choice of principles, however, is extremely difficult. I do not expect the answer I shall suggest to be convincing to everyone. It is, therefore, worth noting from the outset that justice as fairness, like other contract views, consists of two parts: (1) an interpretation of the initial situation and of the problem of choice posed there, and (2) a set of principles which, it is argued, would be agreed to. One may accept the first part of the theory (or some variant thereof), but not the other, and conversely. The concept of the initial contractual situation may seem reasonable although the particular principles proposed are rejected. To be sure, I want to maintain that the most appropriate conception of this situation does lead to principles of justice contrary to utilitarianism and perfectionism, and therefore that the contract doctrine provides an alternative to these views. Still, one may dispute this contention even though one grants that the contractarian method is a useful way of studying ethical theories and of setting forth their underlying assumptions.

Justice as fairness is an example of what I have called a contract theory. Now there may be an objection to the term “contract” and related expressions, but I think it will serve reasonably well. Many words have misleading connotations which at first are likely to confuse. The terms “utility” and “utilitarianism” are surely no exception. They too have unfortunate suggestions which hostile critics have been willing to exploit; yet they are clear enough for those prepared to study utilitarian doctrine. The same should be true of the term “contract” applied to moral theories.

As I have mentioned, to understand it one has to keep in mind that it implies a certain level of abstraction. In particular, the content of the relevant agreement is not to enter a given society or to adopt a given form of government, but to accept certain moral principles. Moreover, the undertakings referred to are purely hypothetical: a contract view holds that certain principles would be accepted in a well-defined initial situation.

The merit of the contract terminology is that it conveys the idea that principles of justice may be conceived as principles that would be chosen by rational persons, and that in this way conceptions of justice may be

\textsuperscript{6.} For the formulation of this intuitive idea I am indebted to Allan Gibbard.
explained and justified. The theory of justice is a part, perhaps the most significant part, of the theory of rational choice. Furthermore, principles of justice deal with conflicting claims upon the advantages won by social cooperation; they apply to the relations among several persons or groups. The word “contract” suggests this plurality as well as the condition that the appropriate division of advantages must be in accordance with principles acceptable to all parties. The condition of publicity for principles of justice is also connotated by the contract phraseology. Thus, if these principles are the outcome of an agreement, citizens have a knowledge of the principles that others follow. It is characteristic of contract theories to stress the public nature of political principles. Finally there is the long tradition of the contract doctrine. Expressing the tie with this line of thought helps to define ideas and accords with natural piety. There are then several advantages in the use of the term “contract.” With due precautions taken, it should not be misleading.

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

4. THE ORIGINAL POSITION AND JUSTIFICATION

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

If this view of the problem of justification is to succeed, we must, of course, describe in some detail the nature of this choice problem. A problem of rational decision has a definite answer only if we know the beliefs and interests of the parties, their relations with respect to one another, the alternatives between which they are to choose, the procedure whereby they make up their minds, and so on. As the circumstances are presented in different ways, correspondingly different principles are accepted. The concept of the original position, as I shall refer to it, is that of the most philosophically favored interpretation of this initial choice situation for the purposes of a theory of justice.

But how are we to decide what is the most favored interpretation? I assume, for one thing, that there is a broad measure of agreement that principles of justice should be chosen under certain conditions. To justify a particular description of the initial situation one shows that it incorporates these commonly shared presumptions. One argues from widely accepted but weak premises to more specific conclusions. Each of the presumptions should by itself be natural and plausible; some of them may seem innocuous or even trivial. The aim of the contract approach is to establish that taken together they impose significant bounds on acceptable principles of justice. The ideal outcome would be that these conditions determine a unique set of principles; but I shall be satisfied if they suffice to rank the main traditional conceptions of social justice.

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

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

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

In searching for the most favored description of this situation we work from both ends. We begin by describing it so that it represents generally shared and preferably weak conditions. We then see if these conditions are strong enough to yield a significant set of principles. If not, we look for further premises equally reasonable. But if so, and these principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted. This state of affairs I refer to as reflective equilibrium. It is an equilibrium because at last our principles and judgments coincide; and it is reflective since we know to what principles our judgments conform and the premises of their derivation. At the moment everything is in order. But this equilibrium is not necessarily stable. It is liable to be upset by further examination of the conditions which should be imposed on the contractual situation and by particular cases which may lead us to revise our judgments. Yet for the time being we have done what we can to render coherent and to justify our convictions of social justice. We have reached a conception of the original position.

I shall not, of course, actually work through this process. Still, we may think of the interpretation of the original position that I shall present as the result of such a hypothetical course of reflection. It represents the

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

A final comment. We shall want to say that certain principles of justice are justified because they would be agreed to in an initial situation of equality. I have emphasized that this original position is purely hypothetical. It is natural to ask why, if this agreement is never actually entered into, we should take any interest in these principles, moral or otherwise. The answer is that the conditions embodied in the description of the original position are ones that we do in fact accept. Or if we do not, then perhaps we can be persuaded to do so by philosophical reflection. Each aspect of the contractual situation can be given supporting grounds. Thus what we shall do is to collect together into one conception a number of conditions on principles that we are ready upon due consideration to recognize as reasonable. These constraints express what we are prepared to regard as limits on fair terms of social cooperation. One way to look at the idea of the original position, therefore, is to see it as an expository device which sums up the meaning of these conditions and helps us to extract their consequences. On the other hand, this conception is also an intuitive notion that suggests its own elaboration, so that led on by it we are drawn to define more clearly the standpoint from which we can best interpret moral relationships. We need a conception that enables us to envision our objective from afar: the intuitive notion of the original position is to do this for us.8

5. CLASSICAL UTILITARIANISM

There are many forms of utilitarianism, and the development of the theory has continued in recent years. I shall not survey these forms here, nor

take account of the numerous refinements found in contemporary discussions. My aim is to work out a theory of justice that represents an alternative to utilitarian thought generally and so to all of these different versions of it. I believe that the contrast between the contract view and utilitarianism remains essentially the same in all these cases. Therefore I shall compare justice as fairness with familiar variants of intuitionism, perfectionism, and utilitarianism in order to bring out the underlying differences in the simplest way. With this end in mind, the kind of utilitarianism I shall describe here is the strict classical doctrine which receives perhaps its clearest and most accessible formulation in Sidgwick. The main idea is that society is rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it.  

We may note first that there is, indeed, a way of thinking of society which makes it easy to suppose that the most rational conception of jus-

9. I shall take Henry Sidgwick’s The Methods of Ethics, 7th ed. (London, 1907), as summarizing the development of utilitarian moral theory. Book III of his Principles of Political Economy (London, 1883) applies this doctrine to questions of economic and social justice, and is a precursor of A. C. Pigou, The Economics of Welfare (London, Macmillan, 1920). Sidgwick’s Outlines of the History of Ethics, 5th ed. (London, 1902), contains a brief history of the utilitarian tradition. We may follow him in assuming, somewhat arbitrarily, that it begins with Shaftesbury’s An Inquiry Concerning Virtue and Merit (1711) and Hutcheson’s An Inquiry Concerning Moral Good and Evil (1725). Hutcheson seems to have been the first to state clearly the principle of utility. He says in Inquiry, sec. 111, §8, that “that action is best, which procures the greatest happiness for the greatest numbers; and that, worst, which, in like manner, occasions misery.” Other major eighteenth century works are Hume’s A Treatise of Human Nature (1739), and An Enquiry Concerning the Principles of Morals (1751); Adam Smith’s A Theory of the Moral Sentiments (1759); and Bentham’s The Principles of Morals and Legislation (1789). To these we must add the writings of J. S. Mill represented by Utilitarianism (1863) and F. Y. Edgeworth’s Mathematical Psychics (London, 1888).


tice is utilitarian. For consider: each man in realizing his own interests is certainly free to balance his own losses against his own gains. We may impose a sacrifice on ourselves now for the sake of a greater advantage later. A person quite properly acts, at least when others are not affected, to achieve his own greatest good, to advance his rational ends as far as possible. Now why should not a society act on precisely the same principle applied to the group and therefore regard that which is rational for one man as right for an association of men? Just as the well-being of a person is constructed from the series of satisfactions that are experienced at different moments in the course of his life, so in very much the same way the well-being of society is to be constructed from the fulfillment of the systems of desires of the many individuals who belong to it. Since the principle for an individual is to advance as far as possible his own welfare, his own system of desires, the principle for society is to advance as far as possible the welfare of the group, to realize to the greatest extent the comprehensive system of desire arrived at from the desires of its members. Just as an individual balances present and future gains against present and future losses, so a society may balance satisfactions and dissatisfactions between different individuals. And so by these reflections one reaches the principle of utility in a natural way: a society is properly arranged when its institutions maximize the net balance of satisfaction.

The principle of choice for an association of men is interpreted as an extension of the principle of choice for one man. Social justice is the principle of rational prudence applied to an aggregative conception of the welfare of the group (§30).10

This idea is made all the more attractive by a further consideration. The two main concepts of ethics are those of the right and the good; the concept of a morally worthy person is, I believe, derived from them. The structure of an ethical theory is, then, largely determined by how it defines and connects these two basic notions. Now it seems that the simplest way of relating them is taken by teleological theories: the good is defined

independently from the right, and then the right is defined as that which maximizes the good. More precisely, those institutions and acts are right which of the available alternatives produce the most good, or at least as much good as any of the other institutions and acts open as real possibilities (a rider needed when the maximal class is not a singleton). Teleological theories have a deep intuitive appeal since they seem to embody the idea of rationality. It is natural to think that rationality is maximizing something and that in morals it must be maximizing the good. Indeed, it is tempting to suppose that it is self-evident that things should be arranged so as to lead to the most good.

It is essential to keep in mind that in a teleological theory the good is defined independently from the right. This means two things. First, the theory accounts for our considered judgments as to which things are good (our judgments of value) as a separate class of judgments intuitively distinguishable by common sense, and then proposes the hypothesis that the right is maximizing the good as already specified. Second, the theory enables one to judge the goodness of things without referring to what is right. For example, if pleasure is said to be the sole good, then presumably pleasures can be recognized and ranked in value by criteria that do not presuppose any standards of right, or what we would normally think of as such. Whereas if the distribution of goods is also counted as a good, perhaps a higher order one, and the theory directs us to produce the most good (including the good of distribution among others), we no longer have a teleological view in the classical sense. The problem of distribution falls under the concept of right as one intuitively understands it, and so the theory lacks an independent definition of the good. The clarity and simplicity of classical teleological theories derives largely from the fact that they factor our moral judgments into two classes, the one being characterized separately while the other is then connected with it by a maximizing principle.

Teleological doctrines differ, pretty clearly, according to how the conception of the good is specified. If it is taken as the realization of human excellence in the various forms of culture, we have what may be called perfectionism. This notion is found in Aristotle and Nietzsche, among others. If the good is defined as pleasure, we have hedonism; if as happiness, eudaimonism, and so on. I shall understand the principle of utility in its classical form as defining the good as the satisfaction of desire, or

perhaps better, as the satisfaction of rational desire. This accords with the view in all essentials and provides, I believe, a fair interpretation of it. The appropriate terms of social cooperation are settled by whatever in the circumstances will achieve the greatest sum of satisfaction of the rational desires of individuals. It is impossible to deny the initial plausibility and attractiveness of this conception.

The striking feature of the utilitarian view of justice is that it does not matter, except indirectly, how this sum of satisfactions is distributed among individuals any more than it matters, except indirectly, how one man distributes his satisfactions over time. The correct distribution in either case is that which yields the maximum fulfillment. Society must allocate its means of satisfaction whatever these are, rights and duties, opportunities and privileges, and various forms of wealth, so as to achieve this maximum if it can. But in itself no distribution of satisfaction is better than another except that the more equal distribution is to be preferred to break ties.\textsuperscript{12} It is true that certain common sense precepts of justice, particularly those which concern the protection of liberties and rights, or which express the claims of desert, seem to contradict this contention. But from a utilitarian standpoint the explanation of these precepts and of their seemingly stringent character is that they are those precepts which experience shows should be strictly respected and departed from only under exceptional circumstances if the sum of advantages is to be maximized.\textsuperscript{13} Yet, as with all other precepts, those of justice are derivative from the one end of attaining the greatest balance of satisfaction. Thus there is no reason in principle why the greater gains of some should not compensate for the lesser losses of others; or more importantly, why the violation of the liberty of a few might not be made right by the greater good shared by many. It simply happens that under most conditions, at least in a reasonably advanced stage of civilization, the greatest sum of advantages is not attained in this way. No doubt the strictness of common sense precepts of justice has a certain usefulness in limiting men’s propensities to injustice and to socially injurious actions, but the utilitarian believes that to affirm this strictness as a first principle of morals is a mistake. For just as it is rational for one man to maximize the fulfillment of his system of desires, it is right for a society to maximize the net balance of satisfaction taken over all of its members.

The most natural way, then, of arriving at utilitarianism (although not,

\textsuperscript{12} On this point see Sidgwick, \textit{The Methods of Ethics}, pp. 416f.
\textsuperscript{13} See J. S. Mill, \textit{Utilitarianism}, ch. V, last two pars.
of course, the only way of doing so) is to adopt for society as a whole the principle of rational choice for one man. Once this is recognized, the place of the impartial spectator and the emphasis on sympathy in the history of utilitarian thought is readily understood. For it is by the conception of the impartial spectator and the use of sympathetic identification in guiding our imagination that the principle for one man is applied to society. It is this spectator who is conceived as carrying out the required organization of the desires of all persons into one coherent system of desire; it is by this construction that many persons are fused into one. Endowed with ideal powers of sympathy and imagination, the impartial spectator is the perfectly rational individual who identifies with and experiences the desires of others as if these desires were his own. In this way he ascertains the intensity of these desires and assigns them their appropriate weight in the one system of desire the satisfaction of which the ideal legislator then tries to maximize by adjusting the rules of the social system. On this conception of society separate individuals are thought of as so many different lines along which rights and duties are to be assigned and scarce means of satisfaction allocated in accordance with rules so as to give the greatest fulfillment of wants. The nature of the decision made by the ideal legislator is not, therefore, materially different from that of an entrepreneur deciding how to maximize his profit by producing this or that commodity, or that of a consumer deciding how to maximize his satisfaction by the purchase of this or that collection of goods. In each case there is a single person whose system of desires determines the best allocation of limited means. The correct decision is essentially a question of efficient administration. This view of social cooperation is the consequence of extending to society the principle of choice for one man, and then, to make this extension work, conflating all persons into one through the imaginative acts of the impartial sympathetic spectator. Utilitarianism does not take seriously the distinction between persons.

6. SOME RELATED CONTRASTS

It has seemed to many philosophers, and it appears to be supported by the convictions of common sense, that we distinguish as a matter of principle between the claims of liberty and right on the one hand and the desirability of increasing aggregate social welfare on the other; and that we give a certain priority, if not absolute weight, to the former. Each member of society is thought to have an inviolability founded on justice or, as some
say, on natural right, which even the welfare of every one else cannot override. Justice denies that the loss of freedom for some is made right by a greater good shared by others. The reasoning which balances the gains and losses of different persons as if they were one person is excluded. Therefore in a just society the basic liberties are taken for granted and the rights secured by justice are not subject to political bargaining or to the calculus of social interests.

Justice as fairness attempts to account for these common sense convictions concerning the priority of justice by showing that they are the consequence of principles which would be chosen in the original position. These judgments reflect the rational preferences and the initial equality of the contracting parties. Although the utilitarian recognizes that, strictly speaking, his doctrine conflicts with these sentiments of justice, he maintains that common sense precepts of justice and notions of natural right have but a subordinate validity as secondary rules; they arise from the fact that under the conditions of civilized society there is great social utility in following them for the most part and in permitting violations only under exceptional circumstances. Even the excessive zeal with which we are apt to affirm these precepts and to appeal to these rights is itself granted a certain usefulness, since it counterbalances a natural human tendency to violate them in ways not sanctioned by utility. Once we understand this, the apparent disparity between the utilitarian principle and the strength of these persuasions of justice is no longer a philosophical difficulty. Thus while the contract doctrine accepts our convictions about the priority of justice as on the whole sound, utilitarianism seeks to account for them as a socially useful illusion.

A second contrast is that whereas the utilitarian extends to society the principle of choice for one man, justice as fairness, being a contract view, assumes that the principles of social choice, and so the principles of justice, are themselves the object of an original agreement. There is no reason to suppose that the principles which should regulate an association of men is simply an extension of the principle of choice for one man. On the contrary: if we assume that the correct regulative principle for anything depends on the nature of that thing, and that the plurality of distinct persons with separate systems of ends is an essential feature of human societies, we should not expect the principles of social choice to be utilitarian. To be sure, it has not been shown by anything said so far that the parties in the original position would not choose the principle of utility to define the terms of social cooperation. This is a difficult question which I shall examine later on. It is perfectly possible, from all that
one knows at this point, that some form of the principle of utility would be adopted, and therefore that contract theory leads eventually to a deeper and more roundabout justification of utilitarianism. In fact a derivation of this kind is sometimes suggested by Bentham and Edgeworth, although it is not developed by them in any systematic way and to my knowledge it is not found in Sidgwick. 14 For the present I shall simply assume that the persons in the original position would reject the utility principle and that they would adopt instead, for the kinds of reasons previously sketched, the two principles of justice already mentioned. In any case, from the standpoint of contract theory one cannot arrive at a principle of social choice merely by extending the principle of rational prudence to the system of desires constructed by the impartial spectator. To do this is not to take seriously the plurality and distinctness of individuals, nor to recognize as the basis of justice that to which men would consent. Here we may note a curious anomaly. It is customary to think of utilitarianism as individualistic, and certainly there are good reasons for this. The utilitarians were strong defenders of liberty and freedom of thought, and they held that the good of society is constituted by the advantages enjoyed by individuals. Yet utilitarianism is not individualistic, at least when arrived at by the more natural course of reflection, in that, by conflating all systems of desires, it applies to society the principle of choice for one man. And thus we see that the second contrast is related to the first, since it is this conflation, and the principle based upon it, which subjects the rights secured by justice to the calculus of social interests.

The last contrast that I shall mention now is that utilitarianism is a teleological theory whereas justice as fairness is not. By definition, then, the latter is a deontological theory, one that either does not specify the good independently from the right, or does not interpret the right as maximizing the good. (It should be noted that deontological theories are defined as non-teleological ones, not as views that characterize the rightness of institutions and acts independently from their consequences. All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy.) Justice as fairness is a deontological theory in the second way. For if it is assumed that the persons in the original position would choose a principle of equal liberty and restrict economic and social inequalities to those in

everyone’s interests, there is no reason to think that just institutions will maximize the good. (Here I suppose with utilitarianism that the good is defined as the satisfaction of rational desire.) Of course, it is not impossible that the most good is produced but it would be a coincidence. The question of attaining the greatest net balance of satisfaction never arises in justice as fairness; this maximum principle is not used at all.

There is a further point in this connection. In utilitarianism the satisfaction of any desire has some value in itself which must be taken into account in deciding what is right. In calculating the greatest balance of satisfaction it does not matter, except indirectly, what the desires are for.15 We are to arrange institutions so as to obtain the greatest sum of satisfactions; we ask no questions about their source or quality but only how their satisfaction would affect the total of well-being. Social welfare depends directly and solely upon the levels of satisfaction or dissatisfaction of individuals. Thus if men take a certain pleasure in discriminating against one another, in subjecting others to a lesser liberty as a means of enhancing their self-respect, then the satisfaction of these desires must be weighed in our deliberations according to their intensity, or whatever, along with other desires. If society decides to deny them fulfillment, or to suppress them, it is because they tend to be socially destructive and a greater welfare can be achieved in other ways.

In justice as fairness, on the other hand, persons accept in advance a principle of equal liberty and they do this without a knowledge of their more particular ends. They implicitly agree, therefore, to conform their conceptions of their good to what the principles of justice require, or at least not to press claims which directly violate them. An individual who finds that he enjoys seeing others in positions of lesser liberty understands that he has no claim whatever to this enjoyment. The pleasure he takes in others’ deprivations is wrong in itself: it is a satisfaction which requires the violation of a principle to which he would agree in the original position. The principles of right, and so of justice, put limits on which satisfactions have value; they impose restrictions on what are reasonable conceptions of one’s good. In drawing up plans and in deciding on aspirations men are to take these constraints into account. Hence in justice as fairness one does not take men’s propensities and inclinations as given, whatever they are, and then seek the best way to fulfill them. Rather, their desires and aspirations are restricted from the outset by the principles of justice which specify the boundaries that men’s systems of

ends must respect. We can express this by saying that in justice as fairness the concept of right is prior to that of the good. A just social system defines the scope within which individuals must develop their aims, and it provides a framework of rights and opportunities and the means of satisfaction within and by the use of which these ends may be equitably pursued. The priority of justice is accounted for, in part, by holding that the interests requiring the violation of justice have no value. Having no merit in the first place, they cannot override its claims.16

This priority of the right over the good in justice as fairness turns out to be a central feature of the conception. It imposes certain criteria on the design of the basic structure as a whole; these arrangements must not tend to generate propensities and attitudes contrary to the two principles of justice (that is, to certain principles which are given from the first a definite content) and they must insure that just institutions are stable. Thus certain initial bounds are placed upon what is good and what forms of character are morally worthy, and so upon what kinds of persons men should be. Now any theory of justice will set up some limits of this kind, namely, those that are required if its first principles are to be satisfied given the circumstances. Utilitarianism excludes those desires and propensities which if encouraged or permitted would, in view of the situation, lead to a lesser net balance of satisfaction. But this restriction is largely formal, and in the absence of fairly detailed knowledge of the circumstances it does not give much indication of what these desires and propensities are. This is not, by itself, an objection to utilitarianism. It is simply a feature of utilitarian doctrine that it relies very heavily upon the natural facts and contingencies of human life in determining what forms of moral character are to be encouraged in a just society. The moral ideal of justice as fairness is more deeply embedded in the first principles of the ethical theory. This is characteristic of natural rights views (the contractarian tradition) in comparison with the theory of utility.

In setting forth these contrasts between justice as fairness and utilitarianism, I have had in mind only the classical doctrine. This is the view of Bentham and Sidgwick and of the utilitarian economists Edgeworth and Pigou. The kind of utilitarianism espoused by Hume would not serve my purpose; indeed, it is not strictly speaking utilitarian. In his well-known arguments against Locke’s contract theory, for example, Hume maintains

16. The priority of right is a central feature of Kant’s ethics. See, for example, The Critique of Practical Reason, ch. II, bk. I of pt. I, esp. pp. 62–65 of vol. 5 of Kants Gesammelte Schriften, Preussische Akademie der Wissenschaften (Berlin, 1913). A clear statement is to be found in “Theory and Practice” (to abbreviate the title), Political Writings, pp. 67f.
that the principles of fidelity and allegiance both have the same foundation in utility, and therefore that nothing is gained from basing political obligation on an original contract. Locke’s doctrine represents, for Hume, an unnecessary shuffle: one might as well appeal directly to utility. But all Hume seems to mean by utility is the general interests and necessities of society. The principles of fidelity and allegiance derive from utility in the sense that the maintenance of the social order is impossible unless these principles are generally respected. But then Hume assumes that each man stands to gain, as judged by his long-term advantage, when law and government conform to the precepts founded on utility. No mention is made of the gains of some outweighing the disadvantages of others. For Hume, then, utility seems to be identical with some form of the common good; institutions satisfy its demands when they are to everyone’s interests, at least in the long run. Now if this interpretation of Hume is correct, there is offhand no conflict with the priority of justice and no incompatibility with Locke’s contract doctrine. For the role of equal rights in Locke is precisely to ensure that the only permissible departures from the state of nature are those which respect these rights and serve the common interest. It is clear that all the transformations from the state of nature which Locke approves of satisfy this condition and are such that rational men concerned to advance their ends could consent to them in a state of equality. Hume nowhere disputes the propriety of these constraints. His critique of Locke’s contract doctrine never denies, or even seems to recognize, its fundamental contention.

The merit of the classical view as formulated by Bentham, Edgeworth, and Sidgwick is that it clearly recognizes what is at stake, namely, the relative priority of the principles of justice and of the rights derived from these principles. The question is whether the imposition of disadvantages on a few can be outweighed by a greater sum of advantages enjoyed by others; or whether the weight of justice requires an equal liberty for all and permits only those economic and social inequalities which are to each person’s interests. Implicit in the contrasts between classical utilitarianism and justice as fairness is a difference in the underlying conceptions of society. In the one we think of a well-ordered society as a scheme of cooperation for reciprocal advantage regulated by principles which persons would choose in an initial situation that is fair, in the other as the efficient administration of social resources to maximize the satisfaction

of the system of desire constructed by the impartial spectator from the many individual systems of desires accepted as given. The comparison with classical utilitarianism in its more natural derivation brings out this contrast.

7. INTUITIONISM

I shall think of intuitionism in a more general way than is customary: namely, as the doctrine that there is an irreducible family of first principles which have to be weighed against one another by asking ourselves which balance, in our considered judgment, is the most just. Once we reach a certain level of generality, the intuitionist maintains that there exist no higher-order constructive criteria for determining the proper emphasis for the competing principles of justice. While the complexity of the moral facts requires a number of distinct principles, there is no single standard that accounts for them or assigns them their weights. Intuitionist theories, then, have two features: first, they consist of a plurality of first principles which may conflict to give contrary directives in particular types of cases; and second, they include no explicit method, no priority rules, for weighing these principles against one another: we are simply to strike a balance by intuition, by what seems to us most nearly right. Or if there are priority rules, these are thought to be more or less trivial and of no substantial assistance in reaching a judgment.18

Various other contentions are commonly associated with intuitionism, for example, that the concepts of the right and the good are unanalyz-


able, that moral principles when suitably formulated express self-evident propositions about legitimate moral claims, and so on. But I shall leave these matters aside. These characteristic epistemological doctrines are not a necessary part of intuitionism as I understand it. Perhaps it would be better if we were to speak of intuitionism in this broad sense as pluralism. Still, a conception of justice can be pluralistic without requiring us to weigh its principles by intuition. It may contain the requisite priority rules. To emphasize the direct appeal to our considered judgment in the balancing of principles, it seems appropriate to think of intuitionism in this more general fashion. How far such a view is committed to certain epistemological theories is a separate question.

Now so understood, there are many kinds of intuitionism. Not only are our everyday notions of this type but so perhaps are most philosophical doctrines. One way of distinguishing between intuitionist views is by the level of generality of their principles. Common sense intuitionism takes the form of groups of rather specific precepts, each group applying to a particular problem of justice. There is a group of precepts which applies to the question of fair wages, another to that of taxation, still another to punishment, and so on. In arriving at the notion of a fair wage, say, we are to balance somehow various competing criteria, for example, the claims of skill, training, effort, responsibility, and the hazards of the job, as well as to make some allowance for need. No one presumably would decide by any one of these precepts alone, and some compromise between them must be struck. The determination of wages by existing institutions also represents, in effect, a particular weighting of these claims. This weighting, however, is normally influenced by the demands of different social interests and so by relative positions of power and influence. It may not, therefore, conform to any one’s conception of a fair wage. This is particularly likely to be true since persons with different interests are likely to stress the criteria which advance their ends. Those with more ability and education are prone to emphasize the claims of skill and training, whereas those lacking these advantages urge the claim of need. But not only are our everyday ideas of justice influenced by our own situation, they are also strongly colored by custom and current expectations. And by what criteria are we to judge the justice of custom itself and the legitimacy of these expectations? To reach some measure of understanding and agreement which goes beyond a mere de facto resolution of competing interests and a reliance on existing conventions and established expectations, it is necessary to move to a more general scheme for determining the balance of precepts, or at least for confining it within narrower limits.
Thus we can consider the problems of justice by reference to certain ends of social policy. Yet this approach also is likely to rely on intuition, since it normally takes the form of balancing various economic and social objectives. For example, suppose that allocative efficiency, full employment, a larger national income, and its more equal distribution are accepted as social ends. Then, given the desired weighting of these aims, and the existing institutional setup, the precepts of fair wages, just taxation, and so on will receive their due emphasis. In order to achieve greater efficiency and equity, one may follow a policy which has the effect of stressing skill and effort in the payment of wages, leaving the precept of need to be handled in some other fashion, perhaps by welfare transfers. An intuitionism of social ends provides a basis for deciding whether the determination of fair wages makes sense in view of the taxes to be imposed. How we weigh the precepts in one group is adjusted to how we weigh them in another. In this way we have managed to introduce a certain coherence into our judgments of justice; we have moved beyond the narrow de facto compromise of interests to a wider view. Of course we are still left with an appeal to intuition in the balancing of the higher-order ends of policy themselves. Different weightings for these are not by any means trivial variations but often correspond to profoundly opposed political convictions.

The principles of philosophical conceptions are of the most general kind. Not only are they intended to account for the ends of social policy, but the emphasis assigned to these principles should correspondingly determine the balance of these ends. For purposes of illustration, let us discuss a rather simple yet familiar conception based on the aggregative-distributive dichotomy. It has two principles: the basic structure of society is to be designed first to produce the most good in the sense of the greatest net balance of satisfaction, and second to distribute satisfactions equally. Both principles have, of course, ceteris paribus clauses. The first principle, the principle of utility, acts in this case as a standard of efficiency, urging us to produce as large a total as we can, other things equal; whereas the second principle serves as a standard of justice constraining the pursuit of aggregate well-being and evening out of the distribution of advantages.

This conception is intuitionist because no priority rule is provided for determining how these two principles are to be balanced against each other. Widely different weights are consistent with accepting these principles. No doubt it is natural to make certain assumptions about how most people would in fact balance them. For one thing, at different combina-
tions of total satisfaction and degrees of equality, we presumably would give these principles different weights. For example, if there is a large total satisfaction but it is unequally distributed, we would probably think it more urgent to increase equality than if the large aggregate well-being were already rather evenly shared. This can be put more formally by using the economist’s device of indifference curves. Assume that we can measure the extent to which particular arrangements of the basic structure satisfy these principles; and represent total satisfaction on the positive X-axis and equality on the positive Y-axis. (The latter may be supposed to have an upper bound at perfect equality.) The extent to which an arrangement of the basic structure fulfills these principles can now be represented by a point in the plane.

Now clearly a point which is northeast of another is a better arrangement: it is superior on both counts. For example, the point B is better than the point A in figure 1. Indifference curves are formed by connecting points judged equally just. Thus curve I in figure 1 consists of the points rated equally with point A which lies on that curve; curve II consists of the points ranked along with point B, and so on. We may assume that these curves slope downward to the right; and also that they do not intersect, otherwise the judgments they represent would be inconsistent. The slope of the curve at any point expresses the relative weights of equality and total satisfaction at the combination the point represents; the changing slope along an indifference curve shows how the relative urgency of the principles shifts as they are more or less satisfied. Thus, moving along either of the indifference curves in figure 1, we see that as equality de-

creases a larger and larger increase in the sum of satisfactions is required to compensate for a further decrease in equality.

Moreover, very different weightings are consistent with these principles. Let figure 2 represent the judgments of two different persons. The solid lines depict the judgments of the one who gives a relatively strong weight to equality, while the dashed lines depict the judgments of the other who gives a relatively strong weight to total welfare. Thus while the first person ranks arrangement D equal with C, the second judges D superior. This conception of justice imposes no limitations on what are the correct weightings; and therefore it allows different persons to arrive at a different balance of principles. Nevertheless such an intuitionist conception, if it were to fit our considered judgments on reflection, would be by no means without importance. At least it would single out the criteria which are significant, the apparent axes, so to speak, of our considered judgments of social justice. The intuitionist hopes that once these axes, or principles, are identified, men will in fact balance them more or less similarly, at least when they are impartial and not moved by an excessive attention to their own interests. Or if this is not so, then at least they can agree to some scheme whereby their assignment of weights can be compromised.

It is essential to observe that the intuitionist does not deny that we can describe how we balance competing principles, or how any one man does so, supposing that we weigh them differently. The intuitionist grants the possibility that these weights can be depicted by indifference curves. Knowing the description of these weights, the judgments which will be made can be foreseen. In this sense these judgments have a consistent and definite structure. Of course, it may be claimed that in the assignment of weights we are guided, without being aware of it, by certain further standards or by how best to realize a certain end. Perhaps the weights we assign are those which would result if we were to apply these standards or to pursue this end. Admittedly any given balancing of principles is subject to interpretation in this way. But the intuitionist claims that, in fact, there is no such interpretation. He contends that there exists no expressible ethical conception which underlies these weights. A geometrical figure or a mathematical function may describe them, but there are no constructive moral criteria that establish their reasonableness. Intuitionism holds that in our judgments of social justice we must eventually reach a plurality of first principles in regard to which we can only say that it seems to us more correct to balance them this way rather than that.

Now there is nothing intrinsically irrational about this intuitionist doc-
trine. Indeed, it may be true. We cannot take for granted that there must be a complete derivation of our judgments of social justice from recognizably ethical principles. The intuitionist believes to the contrary that the complexity of the moral facts defies our efforts to give a full account of our judgments and necessitates a plurality of competing principles. He contends that attempts to go beyond these principles either reduce to triviality, as when it is said that social justice is to give every man his due, or else lead to falsehood and oversimplification, as when one settles everything by the principle of utility. The only way therefore to dispute intuitionism is to set forth the recognizably ethical criteria that account for the weights which, in our considered judgments, we think appropriate to give to the plurality of principles. A refutation of intuitionism consists in presenting the sort of constructive criteria that are said not to exist. To be sure, the notion of a recognizably ethical principle is vague, although it is easy to give many examples drawn from tradition and common sense. But it is pointless to discuss this matter in the abstract. The intuitionist and his critic will have to settle this question once the latter has put forward his more systematic account.

It may be asked whether intuitionistic theories are teleological or deontological. They may be of either kind, and any ethical view is bound to rely on intuition to some degree at many points. For example, one could maintain, as Moore did, that personal affection and human understanding, the creation and the contemplation of beauty, and the gaining and appreciation of knowledge are the chief good things, along with pleasure.\textsuperscript{20} And one might also maintain (as Moore did not) that these are the sole intrinsic goods. Since these values are specified independently from the right, we have a teleological theory of a perfectionist type if the right is defined as maximizing the good. Yet in estimating what yields the most good, the theory may hold that these values have to be balanced against each other by intuition: it may say that there are no substantive criteria for guidance here. Often, however, intuitionist theories are deontological. In the definitive presentation of Ross, the distribution of good things according to moral worth (distributive justice) is included among the goods to be advanced; and while the principle to produce the most good ranks as a first principle, it is but one such principle which must be balanced by intuition against the claims of the other prima facie principles.\textsuperscript{21} The dis-

\textsuperscript{20} See \textit{Principia Ethica}, ch. VI. The intuitionist nature of Moore’s doctrine is assured by his principle of organic unity, pp. 27–31.

\textsuperscript{21} See W. D. Ross, \textit{The Right and the Good}, pp. 21–27.
tinctive feature, then, of intuitionistic views is not their being teleological or deontological, but the especially prominent place that they give to the appeal to our intuitive capacities unguided by constructive and recognizably ethical criteria. Intuitionism denies that there exists any useful and explicit solution to the priority problem. I now turn to a brief discussion of this topic.

8. THE PRIORITY PROBLEM

We have seen that intuitionism raises the question of the extent to which it is possible to give a systematic account of our considered judgments of the just and the unjust. In particular, it holds that no constructive answer can be given to the problem of assigning weights to competing principles of justice. Here at least we must rely on our intuitive capacities. Classical utilitarianism tries, of course, to avoid the appeal to intuition altogether. It is a single-principle conception with one ultimate standard; the adjustment of weights is, in theory anyway, settled by reference to the principle of utility. Mill thought that there must be but one such standard, otherwise there would be no umpire between competing criteria, and Sidgwick argues at length that the utilitarian principle is the only one which can assume this role. They maintain that our moral judgments are implicitly utilitarian in the sense that when confronted with a clash of precepts, or with notions which are vague and imprecise, we have no alternative except to adopt utilitarianism. Mill and Sidgwick believe that at some point we must have a single principle to straighten out and to systematize our judgments.\(^\text{22}\) Undeniably one of the great attractions of the classical doctrine is the way it faces the priority problem and tries to avoid relying on intuition.

As I have already remarked, there is nothing necessarily irrational in the appeal to intuition to settle questions of priority. We must recognize the possibility that there is no way to get beyond a plurality of principles. No doubt any conception of justice will have to rely on intuition to some degree. Nevertheless, we should do what we can to reduce the direct appeal to our considered judgments. For if men balance final principles

\(^{22}\) For Mill, see *A System of Logic*, bk. VI, ch. XII, §7; and *Utilitarianism*, ch. V, pars. 26–31, where this argument is made in connection with common sense precepts of justice. For Sidgwick, see *The Methods of Ethics*, for example, bk. IV, chs. II and III, which summarize much of the argument of bk. III.
differently, as presumably they often do, then their conceptions of justice are different. The assignment of weights is an essential and not a minor part of a conception of justice. If we cannot explain how these weights are to be determined by reasonable ethical criteria, the means of rational discussion have come to an end. An intuitionist conception of justice is, one might say, but half a conception. We should do what we can to formulate explicit principles for the priority problem, even though the dependence on intuition cannot be eliminated entirely.

In justice as fairness the role of intuition is limited in several ways. Since the whole question is rather difficult, I shall only make a few comments here the full sense of which will not be clear until later on. The first point is connected with the fact that the principles of justice are those which would be chosen in the original position. They are the outcome of a certain choice situation. Now being rational, the persons in the original position recognize that they should consider the priority of these principles. For if they wish to establish agreed standards for adjudicating their claims on one another, they will need principles for assigning weights. They cannot assume that their intuitive judgments of priority will in general be the same; given their different positions in society they surely will not. Thus I suppose that in the original position the parties try to reach some agreement as to how the principles of justice are to be balanced. Now part of the value of the notion of choosing principles is that the reasons which underlie their adoption in the first place may also support giving them certain weights. Since in justice as fairness the principles of justice are not thought of as self-evident, but have their justification in the fact that they would be chosen, we may find in the grounds for their acceptance some guidance or limitation as to how they are to be balanced. Given the situation of the original position, it may be clear that certain priority rules are preferable to others for much the same reasons that principles are initially assented to. By emphasizing the role of justice and the special features of the initial choice situation, the priority problem may prove more tractable.

A second possibility is that we may be able to find principles which can be put in what I shall call a serial or lexical order.23 (The correct term

23. The term “lexicographical” derives from the fact that the most familiar example of such an ordering is that of words in a dictionary. To see this, substitute numerals for letters, putting “1” for “a” “2” for “b” and so on, and then rank the resulting strings of numerals from left to right, moving to the right only when necessary to break ties. In general, a lexical ordering cannot be represented by a continuous real-valued utility function; such a ranking violates the assumption of continuity. See I. F.
is “lexicographical,” but it is too cumbersome.) This is an order which requires us to satisfy the first principle in the ordering before we can move on to the second, the second before we consider the third, and so on. A principle does not come into play until those previous to it are either fully met or do not apply. A serial ordering avoids, then, having to balance principles at all; those earlier in the ordering have an absolute weight, so to speak, with respect to later ones, and hold without exception. We can regard such a ranking as analogous to a sequence of constrained maximum principles. For we can suppose that any principle in the order is to be maximized subject to the condition that the preceding principles are fully satisfied. As an important special case I shall, in fact, propose an ordering of this kind by ranking the principle of equal liberty prior to the principle regulating economic and social inequalities. This means, in effect, that the basic structure of society is to arrange the inequalities of wealth and authority in ways consistent with the equal liberties required by the preceding principle. Certainly the concept of a lexical, or serial, order does not offh hand seem very promising. Indeed, it appears to offend our sense of moderation and good judgment. Moreover, it presupposes that the principles in the order be of a rather special kind. For example, unless the earlier principles have but a limited application and establish definite requirements which can be fulfilled, later principles will never come into play. Thus the principle of equal liberty can assume a prior position since it may, let us suppose, be satisfied. Whereas if the


In the history of moral philosophy the conception of a lexical order occasionally appears though it is not explicitly discussed. A clear example may be found in Hutcheson, *A System of Moral Philosophy* (1755). He proposes that in comparing pleasures of the same kind, we use their intensity and duration; in comparing pleasures of different kinds, we must consider their duration and dignity jointly. Pleasures of higher kinds may have a worth greater than those of lower kinds however great the latter’s intensity and duration. See L. A. Selby-Bigge, *British Moralists*, vol. I (Oxford, 1897), pp. 421–423. J. S. Mill’s well-known view in *Utilitarianism*, ch. II, pars. 6–8, is similar to Hutcheson’s. It also is natural to rank moral worth as lexically prior to non-moral values. See for example Ross, *The Right and the Good*, pp. 149–154. And of course the primacy of justice noted in §1, as well as the priority of right as found in Kant, are further cases of such an ordering.

principle of utility were first, it would render otiose all subsequent criteria. I shall try to show that at least in certain social circumstances a serial ordering of the principles of justice offers an approximate solution to the priority problem.

Finally, the dependence on intuition can be reduced by posing more limited questions and by substituting prudential for moral judgment. Thus someone faced with the principles of an intuitionist conception may reply that without some guidelines for deliberation he does not know what to say. He might maintain, for example, that he could not balance total utility against equality in the distribution of satisfaction. Not only are the notions involved here too abstract and comprehensive for him to have any confidence in his judgment, but there are enormous complications in interpreting what they mean. The aggregative-distributive dichotomy is no doubt an attractive idea, but in this instance it seems unmanageable. It does not factor the problem of social justice into small enough parts. In justice as fairness the appeal to intuition is focused in two ways. First we single out a certain position in the social system from which the system is to be judged, and then we ask whether, from the standpoint of a representative man in this position, it would be rational to prefer this arrangement of the basic structure rather than that. Given certain assumptions, economic and social inequalities are to be judged in terms of the long-run expectations of the least advantaged social group. Of course, the specification of this group is not very exact, and certainly our prudential judgments likewise give considerable scope to intuition, since we may not be able to formulate the principle which determines them. Nevertheless, we have asked a much more limited question and have substituted for an ethical judgment a judgment of rational prudence. Often it is quite clear how we should decide. The reliance on intuition is of a different nature and much less than in the aggregative-distributive dichotomy of the intuitionist conception.

In addressing the priority problem the task is that of reducing and not of eliminating entirely the reliance on intuitive judgments. There is no reason to suppose that we can avoid all appeals to intuition, of whatever kind, or that we should try to. The practical aim is to reach a reasonably reliable agreement in judgment in order to provide a common conception of justice. If men’s intuitive priority judgments are similar, it does not matter, practically speaking, that they cannot formulate the principles which account for these convictions, or even whether such principles exist. Contrary judgments, however, raise a difficulty, since the basis for adjudicating claims is to that extent obscure. Thus our object should be to
formulate a conception of justice which, however much it may call upon intuition, ethical or prudential, tends to make our considered judgments of justice converge. If such a conception does exist, then, from the standpoint of the original position, there would be strong reasons for accepting it, since it is rational to introduce further coherence into our common convictions of justice. Indeed, once we look at things from the standpoint of the initial situation, the priority problem is not that of how to cope with the complexity of already given moral facts which cannot be altered. Instead, it is the problem of formulating reasonable and generally acceptable proposals for bringing about the desired agreement in judgments. On a contract doctrine the moral facts are determined by the principles which would be chosen in the original position. These principles specify which considerations are relevant from the standpoint of social justice. Since it is up to the persons in the original position to choose these principles, it is for them to decide how simple or complex they want the moral facts to be. The original agreement settles how far they are prepared to compromise and to simplify in order to establish the priority rules necessary for a common conception of justice.

I have reviewed two obvious and simple ways of dealing constructively with the priority problem: namely, either by a single overall principle, or by a plurality of principles in lexical order. Other ways no doubt exist, but I shall not consider what they might be. The traditional moral theories are for the most part single-principled or intuitionistic, so that the working out of a serial ordering is novelty enough for a first step. While it seems clear that, in general, a lexical order cannot be strictly correct, it may be an illuminating approximation under certain special though significant conditions (§82). In this way it may indicate the larger structure of conceptions of justice and suggest the directions along which a closer fit can be found.

9. SOME REMARKS ABOUT MORAL THEORY

It seems desirable at this point, in order to prevent misunderstanding, to discuss briefly the nature of moral theory. I shall do this by explaining in more detail the concept of a considered judgment in reflective equilibrium and the reasons for introducing it.24

Let us assume that each person beyond a certain age and possessed of the requisite intellectual capacity develops a sense of justice under normal social circumstances. We acquire a skill in judging things to be just and unjust, and in supporting these judgments by reasons. Moreover, we ordinarily have some desire to act in accord with these pronouncements and expect a similar desire on the part of others. Clearly this moral capacity is extraordinarily complex. To see this it suffices to note the potentially infinite number and variety of judgments that we are prepared to make. The fact that we often do not know what to say, and sometimes find our minds unsettled, does not detract from the complexity of the capacity we have.

Now one may think of moral theory at first (and I stress the provisional nature of this view) as the attempt to describe our moral capacity; or, in the present case, one may regard a theory of justice as describing our sense of justice. By such a description is not meant simply a list of the judgments on institutions and actions that we are prepared to render, accompanied with supporting reasons when these are offered. Rather, what is required is a formulation of a set of principles which, when conjoined to our beliefs and knowledge of the circumstances, would lead us to make these judgments with their supporting reasons were we to apply these principles conscientiously and intelligently. A conception of justice characterizes our moral sensibility when the everyday judgments we do make are in accordance with its principles. These principles can serve as part of the premises of an argument which arrives at the matching judgments. We do not understand our sense of justice until we know in some systematic way covering a wide range of cases what these principles are.

A useful comparison here is with the problem of describing the sense of grammaticalness that we have for the sentences of our native language.25 In this case the aim is to characterize the ability to recognize well-formed sentences by formulating clearly expressed principles which make the same discriminations as the native speaker. This undertaking is known to require theoretical constructions that far outrun the ad hoc precepts of our explicit grammatical knowledge. A similar situation presumably holds in moral theory. There is no reason to assume that our sense of justice can be adequately characterized by familiar common sense precepts, or derived from the more obvious learning principles. A

correct account of moral capacities will certainly involve principles and theoretical constructions which go much beyond the norms and standards cited in everyday life; it may eventually require fairly sophisticated mathematics as well. Thus the idea of the original position and of an agreement on principles there does not seem too complicated or unnecessary. Indeed, these notions are rather simple and can serve only as a beginning.

So far, though, I have not said anything about considered judgments. Now, as already suggested, they enter as those judgments in which our moral capacities are most likely to be displayed without distortion. Thus in deciding which of our judgments to take into account we may reasonably select some and exclude others. For example, we can discard those judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside. All these judgments are likely to be erroneous or to be influenced by an excessive attention to our own interests. Considered judgments are simply those rendered under conditions favorable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain. The person making the judgment is presumed, then, to have the ability, the opportunity, and the desire to reach a correct decision (or at least, not the desire not to). Moreover, the criteria that identify these judgments are not arbitrary. They are, in fact, similar to those that single out considered judgments of any kind. And once we regard the sense of justice as a mental capacity, as involving the exercise of thought, the relevant judgments are those given under conditions favorable for deliberation and judgment in general.

I now turn to the notion of reflective equilibrium. The need for this idea arises as follows. According to the provisional aim of moral philosophy, one might say that justice as fairness is the hypothesis that the principles which would be chosen in the original position are identical with those that match our considered judgments and so these principles describe our sense of justice. But this interpretation is clearly oversimplified. In describing our sense of justice an allowance must be made for the likelihood that considered judgments are no doubt subject to certain irregularities and distortions despite the fact that they are rendered under favorable circumstances. When a person is presented with an intuitively appealing account of his sense of justice (one, say, which embodies various reasonable and natural presumptions), he may well revise his judgments to conform to its principles even though the theory does not fit
his existing judgments exactly. He is especially likely to do this if he can find an explanation for the deviations which undermines his confidence in his original judgments and if the conception presented yields a judgment which he finds he can now accept. From the standpoint of moral theory, the best account of a person’s sense of justice is not the one which fits his judgments prior to his examining any conception of justice, but rather the one which matches his judgments in reflective equilibrium. As we have seen, this state is one reached after a person has weighed various proposed conceptions and he has either revised his judgments to accord with one of them or held fast to his initial convictions (and the corresponding conception).

There are, however, several interpretations of reflective equilibrium. For the notion varies depending upon whether one is to be presented with only those descriptions which more or less match one’s existing judgments except for minor discrepancies, or whether one is to be presented with all possible descriptions to which one might plausibly conform one’s judgments together with all relevant philosophical arguments for them. In the first case we would be describing a person’s sense of justice more or less as it is although allowing for the smoothing out of certain irregularities; in the second case a person’s sense of justice may or may not undergo a radical shift. Clearly it is the second kind of reflective equilibrium that one is concerned with in moral philosophy. To be sure, it is doubtful whether one can ever reach this state. For even if the idea of all possible descriptions and of all philosophically relevant arguments is well-defined (which is questionable), we cannot examine each of them. The most we can do is to study the conceptions of justice known to us through the tradition of moral philosophy and any further ones that occur to us, and then to consider these. This is pretty much what I shall do, since in presenting justice as fairness I shall compare its principles and arguments with a few other familiar views. In light of these remarks, justice as fairness can be understood as saying that the two principles previously mentioned would be chosen in the original position in preference to other traditional conceptions of justice, for example, those of utility and perfection; and that these principles give a better match with our considered judgments on reflection than these recognized alternatives. Thus justice as fairness moves us closer to the philosophical ideal; it does not, of course, achieve it.

This explanation of reflective equilibrium suggests straightway a number of further questions. For example, does a reflective equilibrium (in the
sense of the philosophical ideal) exist? If so, is it unique? Even if it is unique, can it be reached? Perhaps the judgments from which we begin, or the course of reflection itself (or both), affect the resting point, if any, that we eventually achieve. It would be useless, however, to speculate about these matters here. They are far beyond our reach. I shall not even ask whether the principles that characterize one person’s considered judgments are the same as those that characterize another’s. I shall take for granted that these principles are either approximately the same for persons whose judgments are in reflective equilibrium, or if not, that their judgments divide along a few main lines represented by the family of traditional doctrines that I shall discuss. (Indeed, one person may find himself torn between opposing conceptions at the same time.) If men’s conceptions of justice finally turn out to differ, the ways in which they do so is a matter of first importance. Of course we cannot know how these conceptions vary, or even whether they do, until we have a better account of their structure. And this we now lack, even in the case of one man, or homogeneous group of men. If we can characterize one (educated) person’s sense of justice, we might have a good beginning toward a theory of justice. We may suppose that everyone has in himself the whole form of a moral conception. So for the purposes of this book, the views of the reader and the author are the only ones that count. The opinions of others are used only to clear our own heads.

I wish to stress that in its initial stages at least a theory of justice is precisely that, namely, a theory. It is a theory of the moral sentiments (to recall an eighteenth century title) setting out the principles governing our moral powers, or, more specifically, our sense of justice. There is a definite if limited class of facts against which conjectured principles can be checked, namely, our considered judgments in reflective equilibrium. A theory of justice is subject to the same rules of method as other theories. Definitions and analyses of meaning do not have a special place: definition is but one device used in setting up the general structure of theory. Once the whole framework is worked out, definitions have no distinct status and stand or fall with the theory itself. In any case, it is obviously impossible to develop a substantive theory of justice founded solely on truths of logic and definition. The analysis of moral concepts and the a priori, however traditionally understood, is too slender a basis. Moral theory must be free to use contingent assumptions and general facts as it pleases. There is no other way to give an account of our considered judgments in reflective equilibrium. This is the conception of the subject
adopted by most classical British writers through Sidgwick. I see no reason to depart from it.26

Moreover, if we can find an accurate account of our moral conceptions, then questions of meaning and justification may prove much easier to answer. Indeed some of them may no longer be real questions at all. Note, for example, the extraordinary deepening of our understanding of the meaning and justification of statements in logic and mathematics made possible by developments since Frege and Cantor. A knowledge of the fundamental structures of logic and set theory and their relation to mathematics has transformed the philosophy of these subjects in a way that conceptual analysis and linguistic investigations never could. One has only to observe the effect of the division of theories into those which are decidable and complete, undecidable yet complete, and neither complete nor decidable. The problem of meaning and truth in logic and mathematics is profoundly altered by the discovery of logical systems illustrating these concepts. Once the substantive content of moral conceptions is better understood, a similar transformation may occur. It is possible that convincing answers to questions of the meaning and justification of moral judgments can be found in no other way.

I wish, then, to stress the central place of the study of our substantive moral conceptions. But the corollary to recognizing their complexity is accepting the fact that our present theories are primitive and have grave defects. We need to be tolerant of simplifications if they reveal and approximate the general outlines of our judgments. Objections by way of counterexamples are to be made with care, since these may tell us only what we know already, namely that our theory is wrong somewhere. The important thing is to find out how often and how far it is wrong. All theories are presumably mistaken in places. The real question at any given time is which of the views already proposed is the best approximation overall. To ascertain this some grasp of the structure of rival theories

26. I believe that this view goes back in its essentials to Aristotle’s procedure in the *Nicomachean Ethics*. See W. F. R. Hardie, *Aristotle’s Ethical Theory*, ch. III, esp. pp. 37–45. And Sidgwick thought of the history of moral philosophy as a series of attempts to state “in full breadth and clearness those primary intuitions of Reason, by the scientific application of which the common moral thought of mankind may be at once systematized and corrected.” *The Methods of Ethics*, pp. 373f. He takes for granted that philosophical reflection will lead to revisions in our considered judgments, and although there are elements of epistemological intuitionism in his doctrine, these are not given much weight when unsupported by systematic considerations. For an account of Sidgwick’s methodology, see J. B. Schneewind, “First Principles and Common Sense Morality in Sidgwick’s Ethics,” *Archiv für Geschichte der Philosophie*, Bd. 45 (1963).
is surely necessary. It is for this reason that I have tried to classify and to
discuss conceptions of justice by reference to their basic intuitive ideas,
since these disclose the main differences between them.

In presenting justice as fairness I shall contrast it with utilitarianism. I
do this for various reasons, partly as an expository device, partly because
the several variants of the utilitarian view have long dominated our philo-
sophical tradition and continue to do so. And this dominance has been
maintained despite the persistent misgivings that utilitarianism so easily
arouses. The explanation for this peculiar state of affairs lies, I believe, in
the fact that no constructive alternative theory has been advanced which
has the comparable virtues of clarity and system and which at the same
time allays these doubts. Intuitionism is not constructive, perfectionism is
unacceptable. My conjecture is that the contract doctrine properly worked
out can fill this gap. I think justice as fairness an endeavor in this direc-
tion.

Of course the contract theory as I shall present it is subject to the
strictures that we have just noted. It is no exception to the primitiveness
that marks existing moral theories. It is disheartening, for example, how
little can now be said about priority rules; and while a lexical ordering
may serve fairly well for some important cases, I assume that it will not
be completely satisfactory. Nevertheless, we are free to use simplifying
devices, and this I have often done. We should view a theory of justice as
a guiding framework designed to focus our moral sensibilities and to put
before our intuitive capacities more limited and manageable questions for
judgment. The principles of justice identify certain considerations as
morally relevant and the priority rules indicate the appropriate prece-
dence when these conflict, while the conception of the original posi-
tion defines the underlying idea which is to inform our deliberations. If
the scheme as a whole seems on reflection to clarify and to order our
thoughts, and if it tends to reduce disagreements and to bring divergent
convictions more in line, then it has done all that one may reasonably ask.
Understood as parts of a framework that does indeed seem to help, the
numerous simplifications may be regarded as provisionally justified.