are not moral conceptions at all, but simply means of suppression. They have no place on a reasonable list of traditional conceptions of justice.\textsuperscript{17} Of course, this contention is not at all a matter of definition. It is rather a consequence of the conditions characterizing the original position, especially the conditions of the rationality of the parties and the veil of ignorance. That conceptions of right have a certain content and exclude arbitrary and pointless principles is, therefore, an inference from the theory.

26. THE REASONING LEADING TO THE TWO PRINCIPLES OF JUSTICE

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

Now consider the point of view of anyone in the original position. There is no way for him to win special advantages for himself. Nor, on the other hand, are there grounds for his acquiescing in special disadvantages. Since it is not reasonable for him to expect more than an equal share in the division of social primary goods, and since it is not rational for him to agree to less, the sensible thing is to acknowledge as the first step a principle of justice requiring an equal distribution. Indeed, this principle is so obvious given the symmetry of the parties that it would occur to everyone immediately. Thus the parties start with a principle requiring equal basic liberties for all, as well as fair equality of opportunity and equal division of income and wealth.

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

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

in these matters. Hence, they not only have final ends that they are in principle free to pursue or to reject, but their original allegiance and continued devotion to these ends are to be formed and affirmed under conditions that are free. Since the two principles secure a social form that maintains these conditions, they would be agreed to rather than the principle of utility. Only by this agreement can the parties be sure that their highest-order interest as free persons is guaranteed.

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

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


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

<table>
<thead>
<tr>
<th>Decisions</th>
<th>c₁</th>
<th>c₂</th>
<th>c₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>d₁</td>
<td>-7</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>d₂</td>
<td>-8</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>d₃</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

The maximin rule requires that we make the third decision. For in this case the worst of the worst that can happen is that one gains five hundred dollars, which is better than the worst for the other actions. If we adopt one of these we may lose either eight or seven hundred dollars. Thus, the choice of d₃ maximizes f (d,c) for that value of c, which for a given d, minimizes f. The term "maximin" means the maximum minimorum; and the rule directs our attention to the worst that can happen under any proposed course of action, and to decide in the light of that.
Now there appear to be three chief features of situations that give plausibility to this unusual rule. First, since the rule takes no account of the likelihoods of the possible circumstances, there must be some reason for sharply discounting estimates of these probabilities. Offhand, the most natural rule of choice would seem to be to compute the expectation of monetary gain for each decision and then to adopt the course of action with the highest prospect. (This expectation is defined as follows: let us suppose that \( g_{ij} \) represent the numbers in the gain-and-loss table, where \( i \) is the row index and \( j \) is the column index; and let \( p_j, j = 1, 2, 3 \), be the likelihoods of the circumstances, with \( \sum p_j = 1 \). Then the expectation for the \( i \)th decision is equal to \( \sum p_j g_{ij} \).) Thus it must be, for example, that the situation is one in which a knowledge of likelihoods is impossible, or at best extremely insecure. In this case it is unreasonable not to be skeptical of probabilistic calculations unless there is no other way out, particularly if the decision is a fundamental one that needs to be justified to others.

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

Let us review briefly the nature of the original position with these three special features in mind. To begin with, the veil of ignorance excludes all knowledge of likelihoods. The parties have no basis for determining the probable nature of their society, or their place in it. Thus they have no basis for probability calculations. They must also take into account the fact that their choice of principles should seem reasonable to others, in particular their descendants, whose rights will be deeply affected by it. These considerations are strengthened by the fact that the parties know very little about the possible states of society. Not only are they unable to conjure the likelihoods of the various possible circumstances, they cannot say much about what the possible circumstances are, much less enumerate them and foresee the outcome of each alternative available. Those deciding are much more in the dark than illustrations by numerical tables suggest. It is for this reason that I have spoken only of a relation to the maximin rule.

Several kinds of arguments for the two principles of justice illustrate the second feature. Thus, if we can maintain that these principles provide a workable theory of social justice, and that they are compatible with reasonable demands of efficiency, then this conception guarantees a satisfactory minimum. There may be, on reflection, little reason for trying to do better. Thus much of the argument, especially in Part Two, is to show, by their application to some main questions of social justice, that the two principles are a satisfactory conception. These details have a philosophical purpose. Moreover, this line of thought is practically decisive if we can establish the priority of liberty. For this priority implies that the persons in the original position have no desire to try for greater gains at the expense of the basic equal liberties. The minimum assured by the two principles in lexical order is not one that the parties wish to jeopardize for the sake of greater economic and social advantages (§§33–35).

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

So much, then, for a brief sketch of the features of situations in which the maximin rule is a useful maxim and of the way in which the arguments for the two principles of justice can be subsumed under them. Thus if the list of traditional views (§21) represents the possible decisions, these principles would be selected by the rule. The original position exhibits these special features to a sufficiently high degree in view of the fundamental character of the choice of a conception of justice. These remarks about the maximin rule are intended only to clarify the structure of the choice problem in the original position. I shall conclude this section by taking up an objection which is likely to be made against the difference principle and which leads into an important question. The

20. Here I borrow from William Fellner, Probability and Profit (Homewood, Ill., R. D. Irwin, Inc., 1965), pp. 140–142, where these features are noted.
objection is that since we are to maximize (subject to the usual constraints) the prospects of the least advantaged, it seems that the justice of large increases or decreases in the expectations of the more advantaged may depend upon small changes in the prospects of those worst off. To illustrate: the most extreme disparities in wealth and income are allowed provided that they are necessary to raise the expectations of the least fortunate in the slightest degree. But at the same time similar inequalities favoring the more advantaged are forbidden when those in the worst position lose by the least amount. Yet it seems extraordinary that the justice of increasing the expectations of the better placed by a billion dollars, say, should turn on whether the prospects of the least favored increase or decrease by a penny. This objection is analogous to the following familiar difficulty with the maximin rule. Consider the sequence of gain-and-loss tables:

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

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

Part of the answer is that the difference principle is not intended to apply to such abstract possibilities. As I have said, the problem of social justice is not that of allocating \textit{ad libitum} various amounts of something, whether it be money, or property, or whatever, among given individuals. Nor is there some substance of which expectations are made that can be shuffled from one representative man to another in all possible combinations. The possibilities which the objection envisages cannot arise in real cases; the feasible set is so restricted that they are excluded.\(^{21}\) The reason for this is that the two principles are tied together as one conception of justice which applies to the basic structure of society as a whole. The operation of the principles of equal liberty and fair equality of opportunity prevents these contingencies from occurring. For we raise the expectations of the more advantaged only in ways required to improve the situation of the worst off. For the greater expectations of the more favored presumably cover the costs of training or answer to organizational requirements, thereby contributing to the general advantage. While nothing guarantees that inequalities will not be significant, there is a persistent tendency for them to be leveled down by the increasing availability of educated talent and ever widening opportunities. The conditions established by the other principles insure that the disparities likely to result will be much less than the differences that men have often tolerated in the past.

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

Contract theory agrees, then, with utilitarianism in holding that the fundamental principles of justice quite properly depend upon the natural facts about men in society. This dependence is made explicit by the description of the original position: the decision of the parties is taken in the light of general knowledge. Moreover, the various elements of the original position presuppose many things about the circumstances of human life. Some philosophers have thought that ethical first principles should be independent of all contingent assumptions, that they should take for granted no truths except those of logic and others that follow from these by an analysis of concepts. Moral conceptions should hold for all possible worlds. Now this view makes moral philosophy the study of the ethics of creation: an examination of the reflections an omnipotent deity might entertain in determining which is the best of all possible worlds. Even the general facts of nature are to be chosen. Certainly we have a natural religious interest in the ethics of creation. But it would

\(^{21}\) I am indebted to S. A. Marglin for this point.
appear to outrun human comprehension. From the point of view of contract theory it amounts to supposing that the persons in the original position know nothing at all about themselves or their world. How, then, can they possibly make a decision? A problem of choice is well defined only if the alternatives are suitably restricted by natural laws and other constraints, and those deciding already have certain inclinations to choose among them. Without a definite structure of this kind the question posed is indeterminate. For this reason we need have no hesitation in making the choice of the principles of justice presuppose a certain theory of social institutions. Indeed, one cannot avoid assumptions about general facts any more than one can do without a conception of the good on the basis of which the parties rank alternatives. If these assumptions are true and suitably general, everything is in order, for without these elements the whole scheme would be pointless and empty.

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

There are two reasons that justify this embedding of ideals into first principles. First of all, and most obviously, the utilitarian's standard assumptions that lead to the wanted consequences may be only probably true, or even doubtfully so. Moreover, their full meaning and application may be highly conjectural. And the same may hold for all the requisite general suppositions that support the principle of utility. From the standpoint of the original position it may be unreasonable to rely upon these hypotheses and therefore far more sensible to embody the ideal more expressly in the principles chosen. Thus it seems that the parties would prefer to secure their liberties straightway rather than have them depend upon what may be uncertain and speculative actuarial calculations. These remarks are further confirmed by the desirability of avoiding complicated theoretical arguments in arriving at a public conception of justice (§24). In comparison with the reasoning for the two principles, the grounds for the utility criterion trespass upon this constraint. But secondly, there is a real advantage in persons' announcing to one another once and for all that even though theoretical computations of utility always happen to favor the equal liberties (assuming that this is indeed the case here), they do not wish that things had been different. Since in justice as fairness moral conceptions are public, the choice of the two principles is, in effect, such an announcement. And the benefits of this collective profession favor these principles even though the utilitarian's assumptions should be true. These matters I shall consider in more detail in connection with publicity and stability (§29). The relevant point here is that while, in general, an ethical theory can certainly invoke natural facts, there may nevertheless be good reasons for embedding convictions of justice more directly into first principles than a theoretically complete grasp of the contingencies of the world may actually require.

27. THE REASONING LEADING TO THE PRINCIPLE OF AVERAGE UTILITY

I now wish to examine the reasoning that favors the principle of average utility. The classical principle is discussed later (§30). One of the merits of contract theory is that it reveals these principles to be markedly distinct conceptions, however much their practical consequences may coincide. Their underlying analytic assumptions are far apart in the sense that they are associated with contrasting interpretations of the initial situation. But first a word about the meaning of utility. It is understood in the traditional sense as the satisfaction of desire; and it admits of interpersonal comparisons that can at least be summed at the margin. I assume also that utility is measured by some procedure that is independent of choices involving risk, say by postulating an ability to rank differences between levels of satisfaction. These are the traditional assumptions; and while they are very strong, they will not be criticized here. As far as possible, I want to examine the historical doctrine on its own terms.

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

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


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

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

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

So far we have assumed that all individuals have similar preferences whether or not they belong to the same society. Their conceptions of the good are roughly the same. Once this highly restrictive assumption is dropped, we take the final step and arrive at a variation of the initial situation. Nothing is known, let us say, about the particular preferences of the members of these societies or of the person deciding. These facts as well as a knowledge of the structure of these societies are ruled out. The veil of ignorance is now complete. But one can still imagine that the hypothetical newcomer reasons much as before. He assumes that there is an equal likelihood of his turning out to be anyone, fully endowed with that person's preferences, abilities, and social position. Once again his prospect is highest for that society with the greatest average utility. We can see this in the following way. Let $n$ be the number of persons in a society. Let their levels of well-being be $u_1, u_2, \ldots, u_n$. Then the total utility is $\sum u_i$ and the average is $\frac{\sum u_i}{n}$. Assuming that one has an equal chance of being any person, one's prospect is: $\frac{1}{n} u_1 + \frac{1}{n} u_2 + \cdots + \frac{1}{n} u_n$ or $\frac{\sum u_i}{n}$. The value of the prospect is identical with the average utility.

Thus if we waive the problem of interpersonal comparisons of utility, and if the parties are viewed as rational individuals who have no aversion to risk and who follow the principle of insufficient reason in computing likelihoods (the principle that underlies the preceding probabilistic calculations), then the idea of the initial situation leads naturally to the average principle. By choosing it the parties maximize their expected well-being as seen from this point of view. Some form of contract theory provides, then, a way of arguing for the average over the classical view. In fact, how else is the average principle to be accounted for? After all, it is not a teleological doctrine, strictly speaking, as the classical view is, and therefore it lacks some of the intuitive appeal of the idea of maximizing the good. Presumably one who held the average principle would want to invoke the contract theory at least to this extent.

In the preceding discussion I have assumed that utility is understood in the traditional sense as the satisfaction of desire and cardinal interpersonal comparisons are regarded as possible. But this notion of utility has been largely abandoned by economic theory in recent decades; it is thought to be too vague and to play no essential role in explaining economic behavior. Utility is now understood as a way of representing the choices of economic agents and not as a measure of satisfaction. The main kind of cardinal utility presently recognized derives from the Neumann-Morgenstern construction, which is based on choices between prospects involving risks ($\S 49$). Unlike the traditional notion, this measure takes attitudes to uncertainty into account and it does not seek to provide a basis for interpersonal comparisons. Nevertheless, it is still possible to formulate the principle of average utility using this kind of measure: one supposes the parties in the original position, or some variant thereof, to
have Neuman-Morgenstern utility function and to assess their prospects accordingly. Of course, certain precautions must be taken; for example, these utility functions cannot take into account all kinds of considerations but must reflect the parties' estimate of what furthers their good. If they were influenced by other reasons, we would not have a teleological theory.

When these restrictions are observed, however, an average utilitarian view can be stated that takes into account the high level of risk aversion that it seems any normal person would have in the original position; and the greater this risk aversion the more this form of utility principle would resemble the difference principle, at least when the evaluation of economic benefits is in question. Of course, these two principles are not the same, since there are many important differences between them. But there is this similarity: risk and uncertainty from a suitably general perspective leads both views to weight more heavily the advantages of those whose situation is less fortunate. In fact, reasonable risk aversion may be so great, once the enormous hazards of the decision in the original position are fully appreciated, that the utilitarian weighting may be, for practical purposes, so close to the difference principle as to make the simplicity of the latter ($§49$) decisive in its favor.

28. SOME DIFFICULTIES WITH THE AVERAGE PRINCIPLE

Before taking up the arguments for the two principles of justice I wish to mention several difficulties with the average principle of utility. First, though, we should note an objection which turns out to be only apparent. As we have seen, this principle may be viewed as the ethics of a single rational individual prepared to take whatever chances necessary to maximize his prospects from the standpoint of the initial situation. (If there is no objective basis for probabilities, they are computed by the principle of insufficient reason.) Now it is tempting to argue against this principle that it presupposes a real and equal acceptance of risk by all members of society. At some time, one wants to say, everyone must actually have agreed to take the same chances. Since clearly there was no such occasion, the principle is unsound. Consider an extreme case: a slaveholder when confronted by his slaves attempts to justify his position to them by claiming that, first of all, given the circumstances of their society, the institution of slavery is in fact necessary to produce the greatest average happiness; and secondly, that in the initial contractual situation he would choose the average principle even at the risk of its subsequently happening that he is justifiably held a slave. Now offhand we are inclined to reject the slaveholder’s argument as beside the point, if not outrageous. One may think that it makes no difference what he would choose. Unless individuals have actually agreed to a conception of justice subject to real risks, no one is bound by its requirements.

On the contract view, however, the general form of the slaveholder’s argument is correct. It would be a mistake for the slaves to retort that his contentions are irrelevant since there has been no actual occasion of choice, no equal sharing of risk as to how things would turn out. The contract doctrine is purely hypothetical: if a conception of justice would be agreed to in the original position, its principles are the right ones to apply. It is no objection that such an understanding has never been nor ever will be entered into. We cannot have it both ways: we cannot interpret the theory of justice hypothetically when the appropriate occasions of consent cannot be found to explain individuals’ duties and obligations, and then insist upon real situations of risk-bearing to throw out principles of justice that we do not want. Thus in justice as fairness the way the theory of justice is to be interpreted (as discussed in the next section) is to establish that the balance of reasons favors the two principles of justice. In the next section I shall start on this task.

The first difficulty with the average principle I have already mentioned in discussing the maximin rule as a heuristic device for arranging the arguments favoring the two principles. It concerns the way that a rational individual is to estimate probabilities. This question arises because there seem to be no objective grounds in the initial situation for assuming that one has an equal chance of turning out to be anybody. That is, this


assumption is not founded upon known features of one's society. In the early stages of the argument leading to the average principle, the hypothetical newcomer does have some knowledge of his abilities and of the design of the societies among which he is choosing. The estimates of his chances are based upon this information. But at the last stage there is complete ignorance of particular facts (with the exception of those implied by the circumstances of justice). The construction of the individual's prospect depends at this stage solely upon the principle of insufficient reason. This principle is used to assign probabilities to outcomes in the absence of any information. When we have no evidence at all, the possible cases are taken to be equally probable. Thus Laplace reasoned that when we are drawing from two urns each containing a different ratio of black to red balls, but we have no information as to which urn we are faced with, then we should assume initially that the chance of drawing from each of these urns is the same. The idea is that the state of ignorance on the basis of which these prior probabilities are assigned presents the same sort of problem as the situation where one has a lot of evidence showing that a particular coin is unbiased. What is distinctive about the use of the principle is that it enables one to incorporate different kinds of information within one strictly probabilistic framework and to draw inferences about probabilities even in the absence of knowledge. Prior probabilities however arrived at are part of one theory along with estimates of chances based on random sampling. The limiting case of no information does not pose a theoretical problem. As evidence accumulates the prior probabilities are revised anyway and the principle of insufficient reason at least insures that no possibilities are excluded from the outset.

Now I shall assume that the parties discount likelihoods arrived at solely on the basis of this principle. This supposition is plausible in view of the fundamental importance of the original agreement and the desire to have one's decision appear responsible to one's descendants who will be affected by it. We are more reluctant to take great risks for them than for ourselves; and we are willing to do so only when there is no way to avoid these uncertainties, or when the probable gains, as estimated by objective information, are so large that it would appear to them irresponsible to have refused the chance offered even though accepting it should actually turn out badly. Since the parties have the alternative of the two principles of justice, they can in large part sidestep the uncertainties of the original position. They can guarantee the protection of their liberties and a reasonably satisfactory standard of life as the conditions of their society permit. In fact, as I argue in the next section, it is questionable whether the choice of the average principle really offers a better prospect anyway, waiving the fact that it is based on the principle of insufficient reason. It seems, then, that the effect of the veil of ignorance is to favor the two principles. This conception of justice is better suited to the situation of complete ignorance.

There are, to be sure, assumptions about society that, if they were sound, would allow the parties to arrive at objective estimates of equal probability. To see this one can convert an argument of Edgeworth for the classical principle into one for average utility. In fact, his reasoning can be adjusted to support nearly any general standard of policy. Edgeworth's idea is to formulate certain reasonable assumptions under which it would be rational for self-interested parties to agree to the standard of utility as a political principle to assess social policies. The necessity for such a principle arises because the political process is not a competitive one and these decisions cannot be left to the market. Some other method must be found to reconcile divergent interests. Edgeworth believes that the principle of utility would be agreed to by self-interested parties as the desired criterion. His thought seems to be that over the long run of many occasions, the policy of maximizing utility on each occasion is most likely to give the greatest utility for any person individually. Consistent application of this standard to taxation and property legislation, and so on, is calculated to give the best results from any one man's point of view. Therefore by adopting this principle self-interested parties have reasonable assurance that they will not lose out in the end and, in fact, will best improve their prospects.

The flaw in Edgeworth's idea is that the necessary assumptions are extremely unrealistic, especially in the case of the basic structure. To

---

26. See William Fellner, Probability and Profit, pp. 27 f. The principle of insufficient reason in its classical form is known to lead to difficulties. See J. M. Keynes, A Treatise on Probability (London, Macmillan, 1921), ch. IV. Part of Rudolf Carnap's aim in his Logical Foundations of Probability, 2nd ed. (Chicago: University of Chicago Press, 1962), is to construct a system of inductive logic by finding other theoretical means to do what the classical principle was intended to do. See pp. 344 f.


The pervasive and continuing influence of our initial place in society and of our native endowments, and of the fact that the social order is one system, is what characterizes the problem of justice in the first place. We must not be enticed by mathematically attractive assumptions into pretending that the contingencies of men's social positions and the asymmetries of their situations somehow even out in the end. Rather, we must choose our conception of justice fully recognizing that this is not and cannot be the case.

It seems, then, that if the principle of average utility is to be accepted, the parties must reason from the principle of insufficient reason. They must follow what some have called the Laplacian rule for choice under uncertainty. The possibilities are identified in some natural way and each assigned the same likelihood. No general facts about society are offered to support these assignments; the parties carry on with probabilistic calculations as if information had not run out. Now I cannot discuss here the concept of probability, but a few points should be noted. First of all, it may be surprising that the meaning of probability should arise as a problem in moral philosophy, especially in the theory of justice. It is, however, the inevitable consequence of the contract doctrine which conceives of moral philosophy as part of the theory of rational choice. Considerations of probability are bound to enter in given the way in which the initial situation is defined. The veil of ignorance leads directly to the problem of choice under complete uncertainty. Of course, it is possible to regard the parties as perfect altruists and to assume that they reason as if they are certain to be in the position of each person. This interpretation of the initial situation removes the element of risk and uncertainty.

In justice as fairness, however, there is no way to avoid this question entirely. The essential thing is not to allow the principles chosen to depend on special attitudes toward risk. For this reason the veil of ignorance also rules out the knowledge of these inclinations: the parties do not know whether or not they have an unusual aversion to taking chances. As far as possible the choice of a conception of justice should depend on a rational assessment of accepting risks unaffected by peculiar individual preferences for taking chances one way or the other. Of course, a social system may take advantage of these varying propensities by having institutions that permit them full play for common ends. But ideally anyway, the basic design of the system should not depend on one of these attitudes.

Secondly, I have simply assumed that judgments of probability, if they are to be grounds of rational decision, must have an objective basis, that is, a basis in knowledge of particular facts (or in reasonable beliefs). This evidence need not take the form of reports of relative frequencies but it should provide grounds for estimating the relative strength of the various tendencies that affect the outcome. The necessity for objective reasons is supported these assignments; the parties carry on with probabilistic calculations as if information had not run out. Now I cannot discuss here the concept of probability, but a few points should be noted. First of all, it may be surprising that the meaning of probability should arise as a problem in moral philosophy, especially in the theory of justice. It is, however, the inevitable consequence of the contract doctrine which conceives of moral philosophy as part of the theory of rational choice. Considerations of probability are bound to enter in given the way in which the initial situation is defined. The veil of ignorance leads directly to the problem of choice under complete uncertainty. Of course, it is possible to regard the parties as perfect altruists and to assume that they reason as if they are certain to be in the position of each person. This interpretation of the initial situation removes the element of risk and uncertainty.

In justice as fairness, however, there is no way to avoid this question entirely. The essential thing is not to allow the principles chosen to depend on special attitudes toward risk. For this reason the veil of ignorance also rules out the knowledge of these inclinations: the parties do not know whether or not they have an unusual aversion to taking chances. As far as possible the choice of a conception of justice should depend on a rational assessment of accepting risks unaffected by peculiar individual preferences for taking chances one way or the other. Of course, a social system may take advantage of these varying propensities by having institutions that permit them full play for common ends. But ideally anyway, the basic design of the system should not depend on one of these attitudes.

Secondly, I have simply assumed that judgments of probability, if they are to be grounds of rational decision, must have an objective basis, that is, a basis in knowledge of particular facts (or in reasonable beliefs). This evidence need not take the form of reports of relative frequencies but it should provide grounds for estimating the relative strength of the various tendencies that affect the outcome. The necessity for objective reasons is

---

all the more urgent in view of the fundamental significance of the choice in the original position and the fact that the parties want their decision to appear well founded to others. I shall assume, therefore, to fill out the description of the original position, that the parties discount estimates of likelihoods not supported by a knowledge of particular facts and that derive largely if not solely from the principle of insufficient reason. The requirement of objective grounds does not seem to be in dispute between neo-Bayesian theorists and those adhering to more classical ideas. The controversy in this case is how far intuitive and imprecise estimates of likelihoods based on common sense and the like should be incorporated into the formal apparatus of the theory of probability rather than used in an ad hoc way to adjust the conclusions reached by methods that leave this information out of account. Here neo-Bayesians have a strong case. Surely it is better when possible to use our intuitive knowledge and common sense hunches in a systematic and not in an irregular and unexplained manner. But none of this affects the contention that judgments of probability must have some objective basis in the known facts about society if they are to be rational grounds of decision in the special situation of the original position.

The last difficulty I shall mention here raises a deep problem. Although I cannot deal with it properly, it should not be passed over. The trouble arises from the peculiarity of the expectation in the final step of the reasoning for the average principle. When expectations are computed in the normal case, the utilities of the alternatives (the \( u \) in the expression \( \sum p u \)) are derived from a single system of preferences, namely those of the individual making the choice. The utilities represent the worth of the alternatives for this person as estimated by his scheme of values. In the present case, however, each utility is based on the preferences of a different person. There are as many distinct persons as there are utilities. Of course, it is clear that this reasoning presupposes interpersonal comparisons. But leaving aside for the moment the problem of defining these, the point to notice here is that the individual is thought to choose as if he has no aims at all which he counts as his own. He takes a chance on being any one of a number of persons complete with each individual’s system of ends, abilities, and social position. We may doubt whether this expectation is a meaningful one. Since there is no one scheme of aims by which its estimates have been arrived at, it lacks the necessary unity.

satisfaction; and so granted the acceptance of these comparison-rules, average satisfaction can be defined and the parties are assumed to maximize their expected satisfaction so understood. Thus everyone thinks of themselves as having the same deep utility function, so to speak, and regards the satisfaction that others achieve as legitimate entries into their own expectations as seen from the perspective of the original position. The same unified expectation holds for all and (using the Laplacean rule) the agreement on the principle of average utility follows.

It is crucial to note that this reasoning presupposes a particular conception of the person. The parties are conceived as having no definite highest-order interests or fundamental ends by reference to which they decide what sorts of persons they care to be. They have, as it were, no determinate character of will. They are, so we might say, bare-persons: as settled by certain comparison rules, they are equally prepared to accept as defining their good whatever evaluations these rules assign to the realization of their, or anyone else’s, final ends, even if these evaluations conflict with those required by their existing fundamental interests. But we have assumed that the parties do have a determinate character and will, even though the specific nature of their system of ends is unknown to them. They are, so to speak, determinate-persons: they have certain highest-order interests and fundamental ends by reference to which they would decide the kind of life and subordinate aims that are acceptable to them. It is these interests and ends, whatever they are, which they must try to protect. Since they know that the basic liberties covered by the first principle will secure these interests, they must acknowledge the two principles of justice rather than the principle of utility.

Thus to sum up: I have argued that the expectation on which the reasoning for the average principle relies is faulty on two counts. First, since there are no objective grounds in the original position for accepting equal likelihoods, or indeed, any other probability distribution, these likelihoods are merely as-if probabilities. They depend solely on the principle of insufficient reason and provide no independent reason for accepting the utility principle. Instead, the appeal to these likelihoods is, in effect, an indirect way of stipulating this principle. Second, the utilitarian argument assumes that the parties have no definite character or will, that they are not persons with determinate final interests, or a particular conception of their good, that they are concerned to protect. Thus taking both counts together, the utilitarian reasoning arrives at a purely formal expression for an expectation, but one that lacks an appropriate meaning. It is as if one continued to use probabilistic arguments and ways of making interper-

sonal comparisons long after the conditions for their legitimate use had been ruled out by the circumstances of the original position.

29. SOME MAIN GROUNDS FOR THE TWO PRINCIPLES OF JUSTICE

In this section I use the conditions of publicity and finality to give some of the main arguments for the two principles of justice. I shall rely upon the fact that for an agreement to be valid, the parties must be able to honor it under all relevant and foreseeable circumstances. There must be a rational assurance that one can carry through. The arguments I shall adduce fit under the heuristic schema suggested by the reasons for following the maximin rule. That is, they help to show that the two principles are an adequate minimum conception of justice in a situation of great uncertainty. Any further advantages that might be won by the principle of utility are highly problematical, whereas the hardship if things turn out badly are intolerable. It is at this point that the concept of a contract has a definite role: it suggests the condition of publicity and sets limits upon what can be agreed to.

The first confirming ground for the two principles can be explained in terms of what I earlier referred to as the strains of commitment. I said (§25) that the parties have a capacity for justice in the sense that they can be assured that their undertaking is not in vain. Assuming that they have taken everything into account, including the general facts of moral psychology, they can rely on one another to adhere to the principles adopted. Thus they consider the strains of commitment. They cannot enter into agreements that may have consequences they cannot accept. They will avoid those that they can adhere to only with great difficulty. Since the original agreement is final and made in perpetuity, there is no second chance. In view of the serious nature of the possible consequences, the question of the burden of commitment is especially acute. A person is choosing once and for all the standards which are to govern his life prospects. Moreover, when we enter an agreement we must be able to honor it even should the worst possibilities prove to be the case. Otherwise we have not acted in good faith. Thus the parties must weigh with care whether they will be able to stick by their commitment in all circumstances. Of course, in answering this question they have only a general knowledge of human psychology to go on. But this information is enough to tell which conception of justice involves the greater stress.
In this respect the two principles of justice have a definite advantage. Not only do the parties protect their basic rights but they insure themselves against the worst eventualities. They run no chance of having to acquiesce in a loss of freedom over the course of their life for the sake of a greater good enjoyed by others, an undertaking that in actual circumstances they might not be able to keep. Indeed, we might wonder whether such an agreement can be made in good faith at all. Compacts of this sort exceed the capacity of human nature. How can the parties possibly know, or be sufficiently sure, that they can keep such an agreement? Certainly they cannot base their confidence on a general knowledge of moral psychology. To be sure, any principle chosen in the original position may require a large sacrifice for some. The beneficiaries of clearly unjust institutions (those founded on principles which have no claim to acceptance) may find it hard to reconcile themselves to the changes that will have to be made. But in this case they will know that they could not have maintained their position anyway. In any case, the two principles of justice provide an alternative. If the only possible candidates all involved similar risks, the problem of the strains of commitment would have to be waived. This is not the case, and judged in this light the two principles seem distinctly superior.

A second consideration invokes the condition of publicity as well as that of the constraints on agreements. I shall present the argument in terms of the question of psychological stability. Earlier I stated that a strong point in favor of a conception of justice is that it generates its own support. When the basic structure of society is publicly known to satisfy its principles for an extended period of time, those subject to these arrangements tend to develop a desire to act in accordance with these principles and to do their part in institutions which exemplify them. A conception of justice is stable when the public recognition of its realization by the social system tends to bring about the corresponding sense of justice. Now whether this happens depends, of course, on the laws of moral psychology and the availability of human motives. I shall discuss these matters later on (§§75–76). At the moment we may observe that the principle of utility seems to require a greater identification with the interests of others than the two principles of justice. Thus the latter will be a more stable conception to the extent that this identification is difficult to achieve. When the two principles are satisfied, each person's basic liberties are secured and there is a sense defined by the difference principle in which everyone is benefited by social cooperation. Therefore we can explain the acceptance of the social system and the principles it satisfies by the psychological law that persons tend to love, cherish, and support whatever affirms their own good. Since everyone's good is affirmed, all acquire inclinations to uphold the scheme.

When the principle of utility is satisfied, however, there is no such assurance that everyone benefits. Allegiance to the social system may demand that some, particularly the less favored, should forgo advantages for the sake of the greater good of the whole. Thus the scheme will not be stable unless those who must make sacrifices strongly identify with interests broader than their own. But this is not easy to bring about. The sacrifices in question are not those asked in times of social emergency when all or some must pitch in for the common good. The principles of justice apply to the basic structure of the social system and to the determination of life prospects. What the principle of utility asks is precisely a sacrifice of these prospects. Even when we are less fortunate, we are to accept the greater advantages of others as a sufficient reason for lower expectations over the whole course of our life. This is surely an extreme demand. In fact, when society is conceived as a system of cooperation designed to advance the good of its members, it seems quite incredible that some citizens should be expected, on the basis of political principles, to accept still lower prospects of life for the sake of others. It is evident then why utilitarians should stress the role of sympathy in moral learning and the central place of benevolence among the moral virtues. Their conception of justice is threatened with instability unless sympathy and benevolence can be widely and intensely cultivated. Looking at the question from the standpoint of the original position, the parties would reject the principle of utility and adopt the more realistic idea of designing the social order on a principle of reciprocal advantage. We need not suppose, of course, that in everyday life persons never make substantial sacrifices for one another, since moved by affection and ties of sentiment they often do. But such actions are not demanded as a matter of justice by the basic structure of society.

Furthermore, the public recognition of the two principles gives greater support to men's self-respect and this in turn increases the effectiveness of social cooperation. Both effects are reasons for agreeing to these principles. It is clearly rational for men to secure their self-respect. A sense of their own worth is necessary if they are to pursue their conception of the good with satisfaction and to take pleasure in its fulfillment. Self-respect is not so much a part of any rational plan of life as the sense that one's plan is worth carrying out. Now our self-respect normally depends upon the respect of others. Unless we feel that our endeavors are respected by
them, it is difficult if not impossible for us to maintain the conviction that our ends are worth advancing (§67). Hence for this reason the parties would accept the natural duty of mutual respect which asks them to treat one another civilly and to be willing to explain the grounds of their actions, especially when the claims of others are overruled (§51). Moreover, one may assume that those who respect themselves are more likely to respect each other and conversely. Self-contempt leads to contempt of others and threatens their good as much as envy does. Self-respect is reciprocally self-supporting.

Thus a desirable feature of a conception of justice is that it should publicly express men’s respect for one another. In this way they insure a sense of their own value. Now the two principles achieve this end. For when society follows these principles, everyone’s good is included in a scheme of mutual benefit and this public affirmation in institutions of each man’s endeavors supports men’s self-esteem. The establishment of equal liberty and the operation of the difference principle are bound to have this effect. The two principles are equivalent, as I have remarked, to an undertaking to regard the distribution of natural abilities in some respects as a collective asset so that the more fortunate are to benefit only in ways that help those who have lost out (§17). I do not say that the parties are moved by the ethical propriety of this idea. But there are reasons for them to accept this principle. For by arranging inequalities for reciprocal advantage and by abstaining from the exploitation of the contingencies of nature and social circumstance within a framework of equal liberties, persons express their respect for one another in the very constitution of their society. In this way they insure their self-respect as it is rational for them to do.

Another way of putting this is to say that the principles of justice manifest in the basic structure of society men’s desire to treat one another not as means only but as ends in themselves. I cannot examine Kant’s view here. Instead I shall freely interpret it in the light of the contract doctrine. The notion of treating men as ends in themselves and never as only a means obviously needs an explanation. How can we always treat everyone as an end and never as a means only? Certainly we cannot say that it comes to treating everyone by the same general principles, since this interpretation makes the concept equivalent to formal justice. On the contract interpretation treating men as ends in themselves implies at the very least treating them in accordance with the principles to which they would consent in an original position of equality. For in this situation men have equal representation as moral persons who regard themselves as ends and the principles they accept will be rationally designed to protect the claims of their person. The contract view as such defines a sense in which men are to be treated as ends and not as means only.

But the question arises whether there are substantive principles which convey this idea. If the parties wish to express this notion visibly in the basic structure of their society in order to secure each man’s rational interest in his self-respect, which principles should they choose? Now it seems that the two principles of justice achieve this aim: for all have equal basic liberties and the difference principle interprets the distinction between treating men as a means only and treating them also as ends in themselves. To regard persons as ends in themselves in the basic design of society is to agree to forgo those gains which do not contribute to everyone’s expectations. By contrast, to regard persons as means is to be prepared to impose on those already less favored still lower prospects of life for the sake of the higher expectations of others. Thus we see that the difference principle, which at first appears rather extreme, has a reasonable interpretation. If we further suppose that social cooperation among those who respect each other and themselves as manifest in their institutions is likely to be more effective and harmonious, the general level of expectations, assuming we could estimate it, may be higher when the two principles of justice are satisfied than one might otherwise have thought. The advantage of the principle of utility in this respect is no longer so clear.

The principle of utility presumably requires some who are less fortunate to accept even lower life prospects for the sake of others. To be sure, it is not necessary that those having to make such sacrifices rationalize this demand by having a lesser appreciation of their own worth. It does not follow from the utilitarian doctrine that it is because their aims are trivial or unimportant that some individuals’ expectations are less. But the parties must consider the general facts of moral psychology. Surely it is natural to experience a loss of self-respect, a weakening of our sense of the value of accomplishing our aims, when we are already less favored. This is particularly likely to be so when social cooperation is arranged for the good of individuals. That is, those with greater advantages do not claim that they are necessary to preserve certain religious or cultural values which everyone has a duty to maintain. We are not here considering a doctrine of traditional order nor the principle of perfectionism, but
rather the principle of utility. In this instance, then, men's self-respect hinges on how they regard one another. If the parties accept the utility criterion, they will lack the support to their self-respect provided by the public commitment of others to arrange inequalities to everyone's advantage and to guarantee the basic liberties for all. In a public utilitarian society men, particularly the least advantaged, will find it more difficult to be confident of their own worth.

The utilitarian may answer that in maximizing the average utility these matters are already taken into account. If, for example, the equal liberties are necessary for men's self-respect and the average utility is higher when they are affirmed, then of course they should be established. So far so good. But the point is that we must not lose sight of the publicity condition. This requires that in maximizing the average utility we do so subject to the constraint that the utilitarian principle is publicly accepted and followed as the fundamental charter of society. What we cannot do is to raise the average utility by encouraging men to adopt and apply non-utilitarian principles of justice. If, for whatever reasons, the public recognition of utilitarianism entails some loss of self-esteem, there is no way around this drawback. It is an unavoidable cost of the utilitarian scheme given our stipulations. Thus suppose that the average utility is actually greater should the two principles of justice be publicly affirmed and realized in the basic structure. For the reasons mentioned, this may conceivably be the case. These principles would then represent the most attractive prospect, and on both lines of reasoning just examined, the two principles would be accepted. The utilitarian cannot reply that one is now really maximizing the average utility. In fact, the parties would have chosen the two principles of justice.

We should note, then, that utilitarianism, as I have defined it, is the view that the principle of utility is the correct principle for society's public conception of justice. And to show this one must argue that this criterion would be chosen in the original position. If we like, we can define a different variation of the initial situation in which the motivation assumption is that the parties want to adopt those principles that maximize average utility. The preceding remarks indicate that the two principles of justice may still be chosen. But if so, it is a mistake to call these principles—and the theory in which they appear—utilitarian. The motivation assumption by itself does not determine the character of the whole theory. In fact, the case for the principles of justice is strengthened if they would be chosen under different motivation assumptions. This indicates that the theory of justice is firmly grounded and not sensitive to slight changes in this condition. What we want to know is which conception of justice characterizes our considered judgments in reflective equilibrium and best serves as the public moral basis of society. Unless one maintains that this conception is given by the principle of utility, one is not a utilitarian.

The strains of commitment and the publicity condition, both of which we have discussed in this section, are also important. The first arises from the fact that, in general, the class of things that can be agreed to is included within, but smaller than, the class of things that can be rationally chosen. We can decide to take a chance and the same time fully intend that, should things turn out badly, we shall do what we can to retrieve our situation. But if we make an agreement, we have to accept the result; and so to give an undertaking in good faith, we must not only intend to honor it but with reason believe that we can do so. Thus the contract condition excludes a certain kind of randomizing. One cannot agree to a principle if there is a real possibility that it has any outcome that one will not be able to accept. I shall not comment further on the publicity condition except to note that it ties in with the desirability of embedding ideals in first principles (end of §26), with simplicity (§49), and with stability. The latter is examined further in what I have called the second part of the argument (§§79–82).

The form of the argument for the two principles is that the balance of reasons favors them over the principle of average utility, and assuming transitivity, over the classical doctrine as well. Thus the agreement of the parties depends on weighing various considerations. The reasoning is informal and not a proof, and there is an appeal to intuition as the basis of the theory of justice. Yet, as I have remarked (§21), when everything is tallied up, it may be clear where the balance of reasons lies. If so, then to the extent that the original position embodies reasonable conditions used in the justification of principles in everyday life, the claim that one would agree to the principles of justice is perfectly credible. Thus they can serve as a conception of justice in the public acceptance of which persons can recognize one another's good faith.

It may be helpful at this point to list some of the main grounds in favor of the two principles of justice over the principle of average utility. That
the conditions of generality of principle, universality of application, and limited information are not sufficient by themselves to require these principles is clear from the reasoning for the utility principle (§27). Further assumptions must, therefore, be incorporated into the original position. Thus, I have assumed that the parties regard themselves as having certain fundamental interests that they must protect if they can; and that, as free persons, they have a highest-order interest in maintaining their liberty to revise and alter these ends (§26). The parties are, so to speak, persons with determinate interests rather than bare potentialities for all possible interests, even though the specific character of these interests is unknown to them. They must try to secure favorable conditions for advancing these finite ends, whatever they are (§28). The hierarchy of interests and its relation to the priority of liberty is taken up later (§§39, 82), but the general nature of the argument for the basic liberties is illustrated by the case of liberty of conscience and freedom of thought (§§33–35).

In addition, the veil of ignorance (§24) is interpreted to mean not only that the parties have no knowledge of their particular aims and ends (except what is contained in the thin theory of the good), but also that the historical record is closed to them. They do not know, and cannot enumerate, the social circumstances in which they may find themselves, or the array of techniques their society may have at its disposal. They have, therefore, no objective grounds for relying on one probability distribution rather than another, and the principle of insufficient reason cannot be invoked as a way around this limitation. These considerations, together with those derived from regarding the parties as having determinate fundamental interests, imply that the expectation constructed by the argument for the utility principle is unsound and lacks the necessary unity (§28).

30. CLASSICAL UTILITARIANISM, IMPARTIALITY, AND BENEVOLENCE

I now want to compare classical utilitarianism with the two principles of justice. As we have seen, the parties in the original position would reject the classical principle in favor of that of maximizing average utility. Since they are concerned to advance their own interests, they have no desire to maximize the total (or the net balance) of satisfactions. For similar reasons they would prefer the two principles of justice. From a contractarian point of view, then, the classical principle ranks below both of these alter-