

Review: Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory
Author(s): John C. Harsanyi
Review by: John C. Harsanyi
Source: *The American Political Science Review*, Vol. 69, No. 2 (Jun., 1975), pp. 594-606
Published by: American Political Science Association
Stable URL: http://www.jstor.org/stable/1959090
Accessed: 10-11-2015 17:28 UTC



# Can the Maximin Principle Serve as a Basis for Morality? A Critique of John Rawls's Theory

JOHN C. HARSANYI
*University of California, Berkeley**

## 1. Introduction

John Rawls's *A Theory of Justice*[1] is an important book. It is an attempt to develop a viable alternative to *utilitarianism*, which up to now in its various forms was virtually the only ethical theory proposing a reasonably clear, systematic, and purportedly rational concept of morality. I shall argue that Rawls's attempt to suggest a viable alternative to utilitarianism does not succeed. Nevertheless, beyond any doubt, his book is a significant contribution to the ongoing debate on the nature of rational morality.

Rawls distinguishes two major traditions of systematic theory in post medieval moral philosophy. One is the *utilitarian* tradition, represented by Hume, Adam Smith, Bentham, John Stuart Mill, Sidgwick, Edgeworth, and many others, including a number of contemporary philosophers and social scientists. The other is the *contractarian* (social-contract) tradition of Locke, Rousseau, and Kant. The latter has never been developed as systematically as the utilitarian tradition, and, clearly, one of Rawls's objectives is to remedy this situation. He regards his own theory as a generalization of the classical contractarian position, and as its restatement at a higher level of abstraction (p. 11).

Rawls argues that the "first virtue" of social institutions (i.e., the most fundamental moral requirement they ought to satisfy) is *justice* (or *fairness*). Suppose that all members of a society—or, more precisely, all "heads of families" (p. 128; *pace* Women's Lib!)—have to agree on the general principles that are to govern the institutions of their society. All of them are supposed to be rational individuals caring only about their own personal interests (and those of their own descendants). But, in order to ensure that they would reach a fair-minded agreement (p. 12), Rawls assumes that they would have to negotiate with each other under what he calls the *veil of ignorance*, i.e., without knowing their own social and economic positions, their own special interests in the society, or even their own personal talents and abilities (or their lack of them). This hypothetical situation in which all participants would have to agree on the most basic institutional arrangements of their society while under this veil of ignorance, is called by Rawls the *original position*. In his theory, this purely hypothetical—and rather abstractly defined—original position replaces the historical or semi-historical "social contract" of earlier contractarian philosophers. He considers the institutions of a given society to be *just* if they are organized according to the principles that presumably would have been agreed upon by rational individuals in the original position (p. 17).

What decision rule would rational individuals use in the original position in deciding whether a given set of institutions was or was not acceptable to them? In the terminology of modern decision theory, the initial position would be a situation of *uncertainty* because, by assumption, the participants would be uncertain about what their personal circumstances would be under any particular institutional framework to be agreed upon.

There are two schools of thought about the decision rule to be used by a rational person under uncertainty. One proposes the *maximin principle*, or some generalization or modification of this principle, as the appropriate decision rule.[2] From the mid-'forties (when the problem first attracted wider attention) to the mid-'fifties this was the prevailing opinion. But then came a growing realization that the maximin principle and all its relatives lead to serious paradoxes because they often suggest wholly unacceptable practical decisions.[3] The other—Bayesian—school of thought, which is now dominant, proposes *expected-utility maximization* as decision rule under uncertainty.[4]

In my opinion, the concept of the original posi-

* This paper has been supported by Grant GS-3222 of the National Science Foundation, through the Center for Research in Management Science, University of California, Berkeley.

[1] Cambridge, Mass.: Harvard University Press, 1971.

[2] See Abraham Wald, *Statistical Decision Functions* (New York: John Wiley & Sons, 1950); Leonid Hurwicz, "Optimality Criteria for Decision Making Under Ignorance," *Cowles Commission Discussion Paper,* Statistics #370 (1951, mimeographed); and Leonard J. Savage, "The Theory of Statistical Decision," *Journal of the American Statistical Association,* 46 (March, 1951), 55–67.

[3] See Roy Radner and Jacob Marschak, "Note on Some Proposed Decision Criteria," in R. M. Thrall, C. H. Coombs, and R. L. Davis, eds., *Decision Processes* (New York: John Wiley & Sons, 1954), pp. 61–68.

[4] See, e.g., Leonard J. Savage, *The Foundations of Statistics* (New York: John Wiley & Sons, 1954).

tion is a potentially very powerful analytical tool for clarifying the concept of justice and other aspects of morality. In actual fact, this concept played an essential role in my own analysis of moral value judgements,[5] prior to its first use by Rawls in 1957[6] (though I did not use the term "original position"). But the usefulness of this concept crucially depends on its being combined with a satisfactory decision rule. Unfortunately, Rawls chooses the maximin principle as decision rule for the participants in the original position. By the very nature of the maximin principle, this choice cannot fail to have highly paradoxical implications.

## 2. The Maximin Principle and its Paradoxes

Suppose you live in New York City and are offered two jobs at the same time. One is a tedious and badly paid job in New York City itself, while the other is a very interesting and well paid job in Chicago. But the catch is that, if you wanted the Chicago job, you would have to take a plane from New York to Chicago (e.g., because this job would have to be taken up the very next day). Therefore there would be a very small but positive probability that you might be killed in a plane accident. Thus, the situation can be represented by the following double-entry table:

| | If the N.Y.–Chicago plane has an accident | If the N.Y.–Chicago plane has no accident |
|---|---|---|
| If you choose the N.Y. job | You will have a poor job, but will stay alive | You will have a poor job, but will stay alive |
| If you choose the Chicago job | You will die | You will have an excellent job and will stay alive |

The maximin principle says that you must evaluate every policy available to you in terms of the *worst possibility* that can occur to you if you follow that particular policy. Therefore, you have to analyze the situation as follows. If you choose the New York job then the worst (and, indeed, the only) possible outcome will be that you will have a poor job but you will stay alive. (I am assuming that your chances of dying in the near future for reasons other than a plane accident can be taken to be zero.) In contrast, if you choose the Chicago job then the worst possible outcome will be that you may die in a plane accident. Thus, the worst possible outcome in the first case would be much better than the worst possible outcome in the second case. Consequently, if you want to follow the maximin principle then you must choose the New York job. Indeed, you must not choose the Chicago job *under any condition*—however unlikely you might think a plane accident would be, and however strong your preference might be for the excellent Chicago job.

Clearly, this is a highly irrational conclusion. Surely, if you assign a low enough probability to a plane accident, and if you have a strong enough preference for the Chicago job, then by all means you should take your chances and choose the Chicago job. This is exactly what Bayesian theory would suggest you should do.

If you took the maximin principle seriously then you could not ever cross a street (after all, you might be hit by a car); you could never drive over a bridge (after all, it might collapse); you could never get married (after all, it might end in a disaster), etc. If anybody really acted this way he would soon end up in a mental institution.

Conceptually, the basic trouble with the maximin principle is that it violates an important continuity requirement: It is extremely irrational to make your behavior wholly dependent on some highly unlikely unfavorable contingencies *regardless of how little probability you are willing to assign to them.*

Of course, Rawls is right when he argues that in *some* situations the maximin principle will lead to reasonable decisions (pp. 154–156). But closer inspection will show that this will happen only in those situations where the maximin principle is essentially *equivalent* to the expected-utility maximization principle (in the sense that the policies suggested by the former will yield expected-utility levels as high, or almost as high, as the policies suggested by the latter would yield). Yet, the point is that in cases where the two principles suggest policies very dissimilar in their consequences so that they are far from being equivalent, it is

[5] See John C. Harsanyi, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," *Journal of Political Economy*, 61 (October, 1953), 434–435; and "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63 (August, 1955), 309–321.

[6] John Rawls, "Justice as Fairness," *Journal of Philosophy*, 54 (October, 1957), 653–662; and "Justice as Fairness," *Philosophical Review*, 67 (April, 1958), 164–194. The 1957 paper is a shorter version of the 1958 paper with the same title.

always the expected-utility maximization principle that is found on closer inspection to suggest reasonable policies, and it as always the maximin principle that is found to suggest unreasonable ones.

### 3. The Maximin Principle in the Original Position

In the last section I have argued that the maximin principle would often lead to highly irrational decisions in everyday life. This is already a sufficient reason for rejecting it as a decision rule appropriate for the original position. This is so because the whole point about the concept of the original position is to imagine a number of individuals ignorant of their personal circumstances and then to assume that under these conditions of ignorance they would act in a *rational manner*, i.e., in accordance with some decision rule which consistently leads to reasonable decisions under ignorance and uncertainty. But, as we have seen, the maximin principle is most definitely *not* a decision rule of this kind.

Yet, after considering the performance of the maximin principle in everyday life, I now propose to consider explicitly the more specific question of how well this principle would perform in the original position itself. In particular, do we obtain a satisfactory concept of justice if we imagine that the criteria of justice are chosen by people in the original position in accordance with the maximin principle?

As Rawls points out, use of the maximin principle in the original position would lead to a concept of justice based on what he calls the *difference principle*, which evaluates every possible institutional arrangement in terms of the interests of the *least advantaged* (i.e., the poorest, or otherwise worst-off) individual (pp. 75–78). This is so because in the original position nobody is assumed to know what his own personal situation would be under any specific institutional arrangement. Therefore, he must consider the possibility that he might end up as the worst-off individual in the society. Indeed, according to the maximin principle, he has to evaluate any particular institutional framework *as if* he were *sure* that this was exactly what would happen to him. Thus, he must evaluate any possible institutional framework by identifying with the interests of the worst-off individual in the society.[7]

Now, I propose to show that the difference principle often has wholly unacceptable moral implications. As a first example, consider a society consisting of one doctor and two patients, both of them critically ill with pneumonia. Their only chance to recover is to be treated by an antibiotic, but the amount available suffices only to treat one of the two patients. Of these two patients, individual A is a basically healthy person, apart from his present attack of pneumonia. On the other hand, individual B is a terminal cancer victim but, even so, the antibiotic could prolong his life by several months. Which patient should be given the antibiotic? According to the difference principle, it should be given to the cancer victim, who is obviously the less fortunate of the two patients.

In contrast, utilitarian ethics—as well as ordinary common sense—would make the opposite suggestion. The antibiotic should be given to A because it would do "much more good" by bringing him back to normal health than it would do by slightly prolonging the life of a hopelessly sick individual.

As a second example, consider a society consisting of two individuals. Both of them have their material needs properly taken care of, but society still has a surplus of resources left over. This surplus can be used either to provide education in higher mathematics for individual A, who has a truly exceptional mathematical ability, and has an all-consuming interest in receiving instruction in higher mathematics. *Or*, it could be used to provide remedial training for individual B, who is a severely retarded person. Such training could achieve only trivial improvements in B's condition (e.g., he could perhaps learn how to tie his shoelaces); but presumably it would give him some minor satisfaction. Finally, suppose it is not possible to divide up the surplus resources between the two individuals.

Again, the difference principle would require that these resources should be spent on B's remedial training, since he is the less fortunate of the two individuals. In contrast, both utilitarian theory and common sense would suggest that they should be spent on A's education, where they would accomplish "much more good," and would create a much deeper and much more intensive human satisfaction.[8]

Even more disturbing is the fact that the difference principle would require us to give *absolute* priority to the interests of the worst-off individual, *no matter what*, even under the most extreme conditions. Even if his interest were affected only in a very minor way, and all other individuals in society had opposite interests of the greatest impor-

[7] In cases where a more specific principle is necessary, Rawls favors the *lexicographical* difference principle: In comparing two possible societies, first compare them from the point of view of the *worst-off* individual. If they turn out to be equally good from his point of view, then compare them from the point of view of the *second-worst-off* individual. If this still does not break the tie, then compare them from the point of view of the *third-worst-off* individual, etc.

[8] This argument of course presupposes the possibility of interpersonal utility comparisons, at least in a rough and ready sense. I shall discuss the possibility of such comparisons in Section 8 on page 600.

tance, his interests would always override anybody else's. For example, let us assume that society would consist of a large number of individuals, of whom one would be seriously retarded. Suppose that some extremely expensive treatment were to become available, which could very slightly improve the retarded individual's condition, but at such high costs that this treatment could be financed only if some of the most brilliant individuals were deprived of all higher education. The difference principle would require that the retarded individual should all the same receive this very expensive treatment at any event—*no matter how many* people would have to be denied a higher education, and *no matter how strongly* they would desire to obtain one (and no matter how great the satisfaction they would derive from it).

Rawls is fully aware that the difference principle has implications of this type. But he feels these are morally desirable implications because in his view they follow from Kant's principle that people should "treat one another not as means only but as ends in themselves" (p. 179). If society were to give priority to A's interests over B's on the utilitarian grounds that by satisfying A's interests "more good" or "more utility" or "more human satisfaction" would be produced (e.g., because A could derive a greater benefit from medical treatment, or from education, or from whatever else), this would amount to "treating B as means only, and not as end in himself."

To my own mind, this is a very artificial and very forced interpretation of the Kantian principle under discussion. The natural meaning of the phrase "treating B as a means only, and not as end in himself" is that it refers to using B's *person*, i.e., his mental or physical faculties or his body itself, as *means* in the service of other individuals' interests, without proper concern for B's own interests. One would have to stretch the meaning of this phrase quite a bit even in order to include an unauthorized use of B's material *property* (as distinguished from his person) in the service of other individuals.

This, however, is still not the case we are talking about. We are talking about B's merely being *denied* the use of certain resources over which he has no prior property rights, and this is done on the ground that other individuals have "greater need" for these resources, i.e., can derive greater utility from them (and let us assume, as may very well be the case, that almost all impartial observers would agree that this was so). But there is no question at all of using B's person or property for the benefit of other individuals. Therefore, it is very hard to understand how the situation could be described as "treating B as a means only, and not as end in himself."

In any case, even if we did accept such an unduly broad interpretation of the Kantian principle, the argument would certainly cut both ways—and indeed, it would go much more against the difference principle than in favor of it. For suppose we accept the argument that it would be a violation of the Kantian principle if we gave priority to a very important need of A over a relatively unimportant need of B, because it would amount to treating B as a mere means. Then, surely, the opposite policy of giving absolute priority to B's *unimportant* need will be an even stronger violation of the Kantian principle and will amount *a fortiori* to treating A now as a mere means rather than as an end.

### 4. Do Counterexamples Matter?

Most of my criticism of Rawls's theory up to now has been based on counterexamples. How much weight do arguments based on counterexamples have? Rawls himself seems to have considerable reservations about such arguments. He writes (p. 52): "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 be sure, counterexamples to some minor details of an ethical theory may not prove very much. They may prove no more than that the theory needs correction in some minor points, and this fact may have no important implications for the basic principles of the theory. But it is a very different matter when the counterexamples are directed precisely against the most fundamental principles of the theory, as are the maximin principle and the difference principle for Rawls's theory. In this case, if the counterexamples are valid, it can only mean that the theory is *fundamentally* wrong.

Admittedly, all my counterexamples refer to rather special situations. It is quite possible that, in *most* everyday situations posing no special problems, Rawls's theory would yield quite reasonable practical conclusions. Indeed, it is my impression that in most situations the practical implications of Rawls's theory would not be very different from those of utilitarian theories. But of course, if we want to *compare* Rawls's theory with utilitarian theories in order to see which of the two yields more reasonable practical conclusions, we have to concentrate on those cases where they yield significantly different conclusions.

Clearly, as far as Rawls's theory often has implications similar to those of utilitarian theories,

I must agree with his point that counterexamples do not prove that his theory does not have at least *approximate* validity in most cases. But my understanding is that Rawls claims more than approximate validity *in this sense* for his theory. Though he does not claim that his theory is absolutely correct in every detail, he does explicitly claim that at the very least the basic principles of his theory yield more satisfactory results than the basic principles of utilitarian theories do. Yet, in my opinion, my counterexamples rather conclusively show that the very opposite is the case.

### 5. An Alternative Model of Moral Value Judgments

All difficulties outlined in Section 3 can be avoided if we assume that the decision rule used in the original position would not be the maximin principle but would rather be the expected-utility maximization principle of Bayesian theory.

In the two papers already quoted,[9] I have proposed the following model. If an individual expresses his preference between two alternative institutional arrangements, he will often base his preference largely or wholly on his personal interests (and perhaps on the interests of his family, his friends, his occupational group, his social class, etc.). For instance, he may say: "I know that under capitalism I am a wealthy capitalist, whereas under socialism I would be at best a minor government official. Therefore, I prefer capitalism." This no doubt would be a very natural judgment of personal preference from his own point of view. But it certainly would not be what we would call a *moral* value judgment by him about the relative merits of capitalism and socialism.

In contrast, most of us will admit that he would be making a moral value judgment if he chose between the two social systems *without knowing* what his personal position would be under either system. More specifically, let us assume that society consists of $n$ individuals, and that the individual under consideration would choose between the two alternative social systems on the assumption that under either system he would have the same probability, $1/n$, of taking the place of the best-off individual, or the second-best-off individual, or the third-best-off individual, etc., up to the worst-off individual. This I shall call the *equiprobability assumption.* Moreover, let us assume that in choosing between the two social systems he would use the principle of expected-utility maximization as his decision rule. (This is my own version of the concept of the "original position.")

It is easy to verify that under these assumptions our individual would always choose that social system which, in his opinion, would yield the higher *average utility level* to the individual members of the society. More generally, he would evaluate every possible social arrangement (every possible social system, institutional framework, social practice, etc.) in terms of the average utility level likely to result from it. This criterion of evaluation will be called the *principle of average utility.*

Of course, in real life, when people express a preference for one social arrangement over another, they will often have a fairly clear idea of what their own personal position would be under both. Nevertheless, we can say that they are expressing a *moral value judgment*, or that they are expressing a *moral preference* for one of these social arrangements, if they make a serious effort to *disregard* this piece of information, and make their choice *as if* they thought they would have the same probability of taking the place of any particular individual in the society.

Thus, under this model, each individual will have two different sets of preferences: he will have a set of *personal preferences*, which may give a particularly high weight to his personal interests (and to those of his close associates); and he will have a set of *moral preferences*, based on a serious attempt to give the same weight to the interests of every member of the society, in accordance with the principle of average utility.

While Rawls's approach yields a moral theory in the contractarian tradition, my own model yields a moral theory based on the principle of average utility and, therefore, clearly belonging to the utilitarian tradition.

### 6. Rawls's Objection to Using Probabilities in the "Original Position"

Rawls discusses my model primarily in Chapters 27 and 28 of his book. One of his critical comments is directed against my use of probabilities in the original position, in the form of the equiprobability assumption. He does not object to the equiprobability assumption as such *if* probabilities are to be used at all. He accepts Laplace's principle of indifference in the limited sense that in a situation of complete ignorance, *if* we want to use probabilities at all, *then* it is reasonable to assign equal probabilities to all possibilities (p. 169).[10] What he objects to is the very use of probabilities in the original position, and

[9] Harsanyi, "Cardinal Utility . . . ," and Harsanyi, "Cardinal Welfare. . . ."

[10] My equiprobability assumption obviously can be regarded as an application of the principle of indifference. But it also has another possible interpretation. It may be regarded as an expression of the purely *moral* principle that, in making basic moral value judgments, we must give the same *a priori* weight to the interests of all members of the society.

in all those cases where these probabilities are not based on empirical evidence. That is, he objects to using *subjective* probabilities or even *logical* probabilities,[11] in the absence of *empirical* probabilities estimated on the basis of empirical facts. (He does not insist, however, that these empirical probabilities should be estimated on the basis of observed statistical frequencies. He is willing to accept more indirect empirical evidence.)

The need and justification for using subjective probabilities have been extensively discussed by Bayesian decision theorists.[12] But Rawls makes no attempt to refute their arguments. Here I shall make only two points.

(a) The only alternative to using subjective probabilities, as required by Bayesian theory, would be to use a decision rule chosen from the maximin-principle family; and, as I have argued (in Section 2), all these decision rules are known to lead to highly irrational decisions in important cases.

(b) Bayesian decision theory shows by rigorous mathematical arguments that any decision maker whose behavior is consistent with a few—very compelling—rationality postulates simply *cannot help* acting *as if* he used subjective probabilities. (More precisely, he cannot help acting *as if* he tried to maximize his expected utility, computed on the basis of some set of subjective probabilities.) I shall quote only two of these rationality postulates: (1) "If you prefer $A$ to $B$, and prefer $B$ to $C$, then consistency requires that you should also prefer $A$ to $C$"; (2) "You are better off if you are offered a *more valuable* prize with a given probability, than if you are offered a *less valuable* prize with the same probability." The other rationality postulates of Bayesian theory are somewhat more technical, but are equally compelling.

To illustrate that a rational decision maker simply *cannot help* using subjective probabilities, at least implicitly, suppose I offered you a choice between two alternative bets and said: "*Either*, I shall pay you \$100 if candidate *X wins* the next election, and shall pay you nothing if he does not. *Or* I shall pay you \$100 if he does *not* win, and pay you nothing if he does. Which of the two bets do you choose?"

First of all, it would be clearly irrational for you to refuse both bets, because *some* chance of obtaining \$100 is surely better than no chance at all—since you can get this chance for free. So, if you are rational, you will choose one of the two bets. Now, if you choose the first bet then I can infer that (at least implicitly) you are assigning a subjective probability of 1/2 or *higher* to Mr. X's winning the next election. On the other hand, if you choose the second bet then I can infer that (at least implicitly) you are assigning a subjective probability of 1/2 or *lower* to Mr. X's winning the election. Thus, whichever way your choice goes, it will amount to choosing a subjective probability for Mr. X's winning the election—either a probability in the range [1/2, 1], or one in the range [0, 1/2].

By the same token, if a decision maker follows the maximin principle, he is not really avoiding a choice of subjective probabilities, at least implicitly. Of course, he may not think explicitly in terms of probabilities at all. But, whether he likes it or not, his behavior will really amount to assigning probability one (or nearly one) to the worst possibility in any given case. He may very well regard the task of choosing subjective probabilities as a rather burdensome responsibility: but he has no way of escaping this responsibility. For instance, if his reliance on the maximin principle results in a foolish decision because it amounts to grossly overestimating the probability of the worst possibility, then he cannot escape the consequences of this foolish decision. (He certainly cannot escape the consequences by saying that he has never explicitly assigned any numerical probability to the worst possibility at all; and that in actual fact he acted in this foolish way only because he wanted to avoid any explicit choice of numerical probabilities.)

Rawls also argues that a given individual's actions in the original position will be easier to justify to other people, including his own descendants, if these actions are based on the maximin principle, than if they are based on the equiprobability assumption (p. 169). But it seems to me that the exact opposite is the case.

As we have seen (cf. Footnote 10), the equiprobability assumption can be justified by the principle of indifference, and also by the moral principle of assigning the same *a priori* weight to every individual's interests. On the other hand, using the maximin principle in the original position is equivalent to assigning unity or near-unity probability to the possibility that one may end up as the worst-off individual in society; and, as far as I can see, there cannot be any rational justification whatever for assigning such an extremely high probability to this possibility.

Rawls's argument becomes much more convincing if it is turned around. If the original position were an historical fact, then any person, other than the worst-off individual in society, would have a legitimate complaint against his ancestor if the latter in the original position voted for an institutional arrangement giving undue priority to

[11] Following Carnap, by logical probabilities I mean subjective probabilities completely determined by symmetry considerations (if appropriate symmetry postulates are added to the standard rationality postulates of Bayesian theory).

[12] See footnote 4.

the interests of the worst-off individual. (For instance, to take the examples discussed in Section 3, he would have a legitimate complaint if his ancestor's vote in the original position now had the effect of depriving him of some life-saving drug, or of a much-desired higher education, etc.)

### 7. Do von Neumann-Morgenstern Utility Functions have any Place in Ethics?

In my model, every person making a moral value judgment will evaluate any institutional arrangement in terms of the average utility level it yields for the individual members of the society, i.e., in terms of the arithmetic mean of these individuals' von Neumann-Morgenstern (=vNM) utility functions.[13] This means that, under my theory, people's vNM utility functions enter into the very definition of justice and other moral values. Rawls objects to this aspect of my theory on the ground that vNM utility functions basically express people's attitudes toward risk-taking, i.e., towards gambling—and these attitudes have no moral significance. Therefore, Rawls argues, vNM utility functions should not enter into our definitions of moral values (pp. 172 and 323).

This objection is based on a misinterpretation of vNM utility functions, which is unfortunately fairly widespread in the literature. To be sure, the vNM utility function of any given individual is estimated from his choice behavior under risk and uncertainty. But this does not mean that his vNM utility function is *merely* an indication of his attitudes toward risk taking. Rather, as its name shows, it is a utility function, and more specifically, it is what economists call a cardinal utility function. This means that the primary task of a vNM utility function is *not* to express a given individual's attitudes toward risk taking; rather, it is to indicate how much utility, i.e., how much subjective *importance*, he assigns to various goals.

For example, suppose we find that a given individual is willing to gamble at very unfavorable odds—say, he is willing to pay $5 for a lottery ticket giving him a 1/1000 chance of winning $1000. This allows us the inference that his vNM utility function assigns (at least) 1000 times as much utility to $1000 as it assigns to $5. Thus, the theory of vNM utility functions suggests the following explanation for this individual's willingness to gamble at unfavorable odds: he is acting this way because he is attaching unusually *high* importance to getting $1000, and is attaching unusually *low* importance to losing $5. More generally, people are willing to gamble at unfavorable odds, if they feel they would need a large sum of money *very badly* (but do not care too much about losing a small sum of money).

Consequently, vNM utility functions have a completely legitimate place in ethics because they express the subjective importance people attach to their various needs and interests. For example, I cannot see anything wrong with a concept of justice which assigns high priority to providing university education for a given individual partly on the ground that he attaches very high *utility* to receiving such an education (i.e., wants to receive one very badly)—as shown by the fact that he would be prepared to face very considerable personal and financial *risks*, if he had to, in order to obtain a university education.

### 8. Do Interpersonal Utility Comparisons Make Sense?

Rawls objects to the use of interpersonal utility comparisons in defining justice (p. 173). In contrast, my own model makes essential use of such comparisons in the sense that it requires any person making a basic moral value judgment to try to visualize what it would be like to be in the shoes of any other member of the society. That is, he must try to estimate what utility level he would enjoy if he himself were placed in the *objective* physical, economic, and social conditions of any other individual—and if at the same time he also suddenly acquired this individual's *subjective* attitudes, taste, and preferences, i.e., suddenly acquired his utility function.

Admittedly, the idea of evaluating another individual's personal circumstances in terms of *his* utility function, and not in terms of our own, is a difficult concept. But it is a concept we cannot avoid in any reasonable theory of morality. Clearly, if I want to judge the fairness of a social policy providing a diet very rich in fish for a given group of individuals (e.g., for students living in a certain dormitory), I obviously must make my judgment in terms of these individuals' liking or disliking for fish, and not in terms of my own.

As I tried to show in my 1955 paper,[14] the ultimate logical basis for interpersonal utility comparisons, interpreted in this way, lies in the postulate that the *preferences and utility functions of all human individuals* are *governed by the same basic psychological laws*. My utility function may be very different from yours. But, since both of our utility functions are governed by the very same basic psychological laws, if I had your personal characteristics—and, in particular, if I had your biological inheritance and had your life history behind me—then presumably I would now have a utility function exactly like yours.[15] This means

[13] As defined by John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, 2nd ed. (Princeton, N.J.: Princeton University Press, 1947), pp. 15–31.

[14] Harsanyi, "Cardinal Welfare. . . ."

[15] This statement would admittedly require appropriate qualifications if the psychological laws governing

that any *inter*personal comparison I may try to make between your present utility level and my own, reduces to an *intra*-personal utility comparison between the utility level *I* myself *do* now enjoy, and the utility level *I* myself *would* enjoy under certain hypothetical conditions, namely if I were placed in your physical, economic, and social position, and also had my own biological and biographical background replaced by yours.

This means that interpersonal utility comparisons have a completely specific theoretical meaning, in the sense that, "under ideal conditions," i.e., if we had full knowledge of the psychological laws governing people's preferences and their utility functions, and also had sufficient information about other people's personal characteristics, then we could make perfectly error-free interpersonal utility comparisons. Of course, in actual fact, our knowledge of psychological laws and of other people's personal characteristics is very limited, and, therefore, interpersonal utility comparisons are often subject to considerable error—but, of course, so are many other judgments we have to make before we can reach practical decisions, whether these are moral decisions or purely pragmatic ones. Nevertheless, in many *specific* cases, we may have enough background information to be quite confident in our judgments of interpersonal utility comparison—and this confidence is often justified by the fact that in many of these cases there is a reasonable agreement between the conclusions reached by different competent observers when they try to make such comparisons.

In any case, we all make, and cannot help making, interpersonal utility comparisons all the time. We have to decide again and again which particular member of our family, or which particular friend of ours, etc., has a more urgent need for our time or our money, or could derive greater satisfaction from a present, and so on. Likewise, as voters or public officials, we have to decide again and again which particular social group would derive the greatest benefit from government help, etc. To my mind, it makes no sense to deny the legitimacy of a mental operation we all perform every day, and for which a completely satisfactory logical analysis can be provided.[16]

Rawls expresses considerable doubts about the validity of interpersonal utility comparisons (pp. 90 and 321–324). But he makes no attempt to refute my theory of such comparisons, stated in my 1955 article (and briefly summarized above). Instead, he concentrates his criticism on two highly artificial procedures suggested in the literature for making interpersonal utility comparisons (pp. 321–323). One is based on equating the smallest noticeable utility differences of different people. The other is based on equating all individuals' highest possible utility levels, and then again equating their lowest possible utility levels. Of course, he has no trouble showing that, in order to use either procedure in moral philosophy, we would have to introduce some highly arbitrary and implausible moral postulates. But none of these criticisms applies to my own theory of interpersonal utility comparisons.

This completes my discussion of Rawls's objections to my own version of utilitarian theory. I shall now discuss some objections of his to utilitarian theories in general.

### 9. Utilitarianism and Supererogatory Actions

Commonsense morality distinguishes between morally good actions we have a duty to perform, and morally good actions which go beyond the call of duty (supererogatory actions). But, as Rawls points out (p. 117), classical utilitarianism cannot accommodate this distinction because it claims that our duty is always to perform the actions likely to produce the *greatest* good for society. This would mean that, even if we were constantly engaged in the most heroic acts of altruistic self-sacrifice, we would merely do our duty, and no human action could ever be correctly described as supererogatory. I agree with Rawls that it is a serious shortcoming of classical utilitarianism that it cannot admit the existence of supererogatory actions, and draws the line between morally permissible and impermissible conduct at an absurdly high level of moral perfection.

This shortcoming, however, can be easily remedied without going beyond the principles of utilitarianism. The mistake of the classical utilitarians was to overlook the fact that people attach considerable utility to *freedom* from unduly burdensome moral obligations. It may be true (though this is by no means a foregone conclusion) that society will reach a higher level of economic prosperity and cultural excellence if its moral code requires all people all the time to act in the most public-spirited manner, and to set themselves the highest possible standards in their economic and cultural activities. But most people will prefer a society with a more relaxed moral code, and will feel that such a society will achieve a higher level of average utility—even if adoption of such a moral code should lead to some losses in economic and cultural accomplishments (so long as these losses remain within tolerable limits). This means that utilitarianism, if correctly inter-

people's utility functions were found to be probabilistic, rather than deterministic. But this would not affect the basic validity of my analysis, though it would necessitate its restatement in a more complicated form.

[16] For a more detailed discussion of the epistemological problems connected with interpersonal utility comparisons, see my 1955 paper cited in footnote 5.

preted, will yield a moral code with a standard of acceptable conduct very much below the level of highest moral perfection, leaving plenty of scope for supererogatory actions exceeding this minimum standard.

### 10. Vagueness Versus Simplemindedness in Moral Philosophy

As Rawls correctly states (p. 320), the utilitarian concept of morality inevitably shows some degree of vagueness or indeterminacy because of its dependence on—more or less uncertain—interpersonal utility comparisons. Other authors have pointed out another source of indeterminacy, no less important, in the dependence of utilitarian morality on uncertain predictions about the short-run and long-run consequences of alternative social policies and institutional arrangements. As a result, two equally well-intentioned and well-informed, and equally intelligent utilitarians may very well disagree in many specific situations about what is socially useful or socially harmful and, therefore, also about what is right or wrong, and just or unjust, etc.

Rawls's own theory, of course, cannot completely escape such ambiguities either, but it is certainly much less affected by them than utilitarian theories are. First of all, Rawls's basic postulate, the difference principle, is much less dependent on interpersonal utility comparisons than the basic utilitarian principles (for example, the principle of average utility) are; therefore, it yields more specific practical conclusions than the latter do in many cases. In addition, Rawls supplements the difference principle by second-order rules, which are supposed to rank the major values of human life according to their relative moral importance. Thus, for example, according to Rawls, people's basic liberties should always be given absolute priority over their economic and social interests, etc. Clearly, if we are willing to accept such rigid second-order rules of priority, then they will often go a long way toward deciding our moral uncertainties in a fairly unambiguous manner.

Yet, I very much doubt that this is really an advantage. It seems to me that the uncertainties of utilitarian morality merely reflect the great complexity and the unavoidable dilemmas of real-life moral situations. Simple minded rigid mechanical rules cannot possibly do justice to the complexity of moral problems; and they cannot resolve our moral dilemmas satisfactorily, because they cannot help choosing the wrong horn of the dilemma in many important cases.

For example, there are good reasons to believe that in an underdeveloped country in many cases economic growth cannot be set in motion without concentrating a good deal of power in the hands of the government and perhaps even without *some* curtailment of civil liberties (though this does not mean that there is any need or justification for a complete suppression of civil liberties as practiced by the arbitrary dictatorial governments now existing in many of these countries).

Who is the moral philosopher to lay down the law for these countries and tell them that no amount of economic and social development, however large, can ever justify any curtailment of civil liberties, however small? Should we not rather say, with the utilitarian philosophers, that judgments about any particular policy must always depend on the balance of the advantages and disadvantages it is likely to yield, and that the main task of the moral philosopher is to ensure that people will not overlook any major advantage or disadvantage in reaching a decision?

### 11. Saving as a Moral Duty to Future Generations

What proportion of national income ought to be saved as a matter of moral duty (as a matter of justice) to future generations? As Rawls rightly argues (p. 286), utilitarianism (at least as it is usually interpreted) gives an unsatisfactory answer to this question, in that it seems to require unreasonably high savings. The mathematical problem of computing the morally optimal amount of savings under utilitarian criteria was solved by Keynes's friend, the brilliant economist-philosopher Frank P. Ramsey, in 1928.[17] Of course, the numerical answer depends on the utility functions used. But Ramsey showed that, if we use reasonable-looking utility functions, then the utilitarian model may easily yield optimal savings amounting to much more than one half of national income, which is clearly an unacceptable conclusion.

How well does Rawls's own theory deal with this problem? It is easy to verify that the difference principle would suggest *zero* net savings from one generation to another. This is so because, even without any net savings, as a result of mere technological progress, future generations will be much better off than the present generation is, anyhow (provided the population explosion can be brought under control). Therefore, any positive net saving would be inconsistent with the difference principle since it would amount to a transfer of economic resources from a much poorer generation to much richer generations. Thus, while utilitarian theory seems to require unduly high savings, Rawls's difference principle would certainly require unduly low (viz., zero) savings.

Rawls is aware that the difference principle would have this undesirable implication (p. 291).

[17] Frank P. Ramsey, "A Mathematical Theory of Saving," *Economic Journal*, 38 (December, 1928), 543–559.

Nevertheless, surprisingly enough, he seems to imply that his theory handles the saving problem much *better* than utilitarian theory does (pp. 297–298). The truth is that he can avoid the zero-savings conclusion only by giving up the difference principle altogether in dealing with the saving problem, and by replacing it with a completely *ad hoc* motivational assumption. (Whereas in all other respects he makes the participants of the original position complete egoists, in this one respect, viz., in relation to future generations, he endows them with considerable altruism.) Of course, by introducing *ad hoc* assumptions of its own, utilitarianism could just as easily avoid the unwelcome logical necessity of enjoining excessive savings.

In actual fact, in order to obtain a reasonable solution for the problem of optimal savings in terms of utilitarian principles, we have no need for *ad hoc* assumptions. All we have to do is to take a second look at Ramsey's utility functions. The utility functions he postulates seem to be reasonable enough if they are meant to measure the utility that a given individual in the present generation would derive from higher income levels (on the assumption that other people's incomes would remain more or less unchanged). But they greatly overstate the extra utility that future generations are likely to derive from higher incomes as a result of substantially increased saving and investment by the present generation. There are at least three reasons for this:

(1) The *risk effect:* there is always a considerable risk that future changes in technology and in social customs will drastically reduce the benefit that future generations would derive from investments undertaken by the present generations. (For instance, the United States and some European countries invested very large amounts of money in building canals just before the railway age. These huge investments almost completely lost their usefulness very soon thereafter as a result of railway construction.)

(2) The *relative-income effect:* a rise in a given person's income, when other people's incomes remain largely the same, will tend to increase his social status. But if his income rises as a result of a general increase in society's income then of course this effect will be lost. Therefore, in the former case the rise in his income will produce a much greater increase in his utility than it will do in the latter case.

(3) The *inherited-wealth effect:* inherited wealth often has a very powerful influence on human motivation. Some of this influence may be beneficial. (People born into very rich families often develop highly idealistic and altruistic attitudes, and may take a strong interest in social causes or in political, philanthropic, and cultural activities.) But some of this influence may be highly detrimental for a person's chances of leading a happy and socially useful life. (People born into very rich families often lack all interest in serious work, including altruistic or intellectual work; and they may be offered so many opportunities to amuse themselves that they may lose all ability to enjoy the normal pleasures of human life.) It is not unreasonable to assume that if society as a whole inherits very high levels of material abundance, so that there is very little pressure on the average man to earn a living by serious work, then the negative effects are likely to predominate. (We can already see some indications of this in our own society.) Therefore, the net benefit that future generations are likely to derive from increased saving and investment by the present generation may be much smaller than at first one might think.

Thus, if the likely utility of much higher incomes to future generations is reassessed in a more realistic manner then utilitarian theory will yield much lower levels of optimal savings, and in fact will furnish a completely satisfactory solution for this problem, without any need for *ad hoc* assumptions.

### 12. The Stability of a Just Society

Rawls raises a very interesting problem, so far largely neglected by moral and political philosophers. Suppose there is a society with a strong sense of justice among its citizens, and with completely (or almost completely) just institutions. Would such a society be stable? He strongly argues that the answer is in the affirmative (pp. 490–504). He also suggests that a society based on his own conception of justice would be more stable than one based on a utilitarian conception (p. 498).

The just society he describes in this connection, however, is not merely an improved version of the best societies now existing; rather, it is unlike any society known to political scientists or historians or other competent observers. It is a society where citizens and legislators are never motivated by their own selfish interests or (in the case of the legislators) by the selfish interests of their constituents, but rather are always motivated by their strong sense of justice. As such, this society is almost the opposite of the society pictured in Anthony Downs's *An Economic Theory of Democracy*.[18]

Of course, Rawls is quite right in rejecting Downs's motivational assumptions as a fully realistic picture of human motivation. It is certainly not true that ordinary citizens never care about

[18] Anthony Downs, *An Economic Theory of Democracy* (New York: Harper and Brothers, 1957).

anything but their narrow economic (and perhaps other) self-interest, or that politicians never care about anything but their chances for election or reelection. Indeed, it is quite clear that under *some* conditions many rich people will strongly support legislation benefiting the poor but greatly increasing their own taxes (though it is much less clear under what conditions this will or will not happen). Again, we have all seen elected officials follow their own moral and political convictions and make highly unpopular decisions, greatly endangering their prospects for reelection (and sometimes we wished they did not, while at other times we were glad they did). Indeed, it is quite obvious that Downs does not claim that his oversimplified motivational assumptions are literally true; all he claims is that our political system operates most of the time *as if* these motivational assumptions *were* correct.

Nevertheless, the fact that Downs's motivational assumptions come so close to being true, should make us stop to think before accepting Rawls's theory of stability. Should we not take this fact as an indication that the very high levels of public-spirited motivation that Rawls assumes for his just society, would be intrinsically *unstable?* Indeed, our historical experience seems to show that whole societies can achieve such motivational states only for rather short periods (e.g., during revolutions or some very popular wars). The same experience also shows that these highly idealistic—and often highly fanatical and intolerant—motivational states of a society are far from being an unmixed blessing.

It seems to me that any healthy society needs a proper balance between egoistic and altruistic motivation. Without political leaders fighting for altruistic objectives, or without private citizens giving them political support, present-day democratic societies would not have achieved even that, no doubt imperfect, level of social justice and of good government they currently enjoy.

On the other hand, political movements based largely or wholly on well-understood self-interest are an equally essential component of any political system. Citizens pressing for their sectional economic interests may be very biased judges of the public interest, but at least they are well-informed judges in most cases. In contrast, citizens pursuing highly altruistic objectives might often fight for causes about which they know very little, or about which they have strikingly one-sided information. Steelworkers pressing for their own economic interests will at least know what they are talking about. But faraway benevolent millionnaires fighting for the steelworkers' interests might have very mistaken ideas about what these steelworkers really want or need. A society where everybody neglects his own interests, and is busily looking after everybody else's interests, probably would not be a very stable society—and certainly would not be a very happy one.

Accordingly, it seems to me that a just society with a reasonable prospect for social stability would *not* be a society where ordinary citizens and legislators would be primarily motivated by their sense of justice. Rather, it would be a society where most people would be motivated by the normal mixture of egoistic and altruistic interests. Of course, it would have to be a society where people have a strong sense of justice—but this does not mean that a pursuit of justice would have to be their main and continual preoccupation. It only means that they would have to show enough respect for justice so as to stop pressing their own egoistic—and altruistic—objectives *beyond* the point where they would violate the just legal and moral rights of other people; and so as to fight for restoring these rights if they have been violated by injustices of the past.

### 13. Conclusion

To conclude, in spite of my numerous disagreements with Rawls's theory, I strongly recommend his book to all readers interested in moral and political philosophy. He raises many interesting and, to my mind, highly important problems, even though some of us may question the solutions he proposes. The author's serious concern for truth and justice is evident on every page of the book. He makes a real effort to look at both (or all) sides of every difficult or controversial problem, and to reach a fair and balanced conclusion. Where he touches on problems of topical interest, he does not hesitate for a moment to express unpopular views, for example, by pointing out the possible destabilizing effects that very widespread civil disobedience might have on democratic institutions (p. 374). In the political climate of Harvard in the late 'sixties or early 'seventies it must have required no little moral courage to express such an opinion.

We live in an age where our moral attitudes are rapidly changing, and so are many of our social institutions, with end results very hard to predict; where traditional world views are more and more replaced by a world view based on science and depriving man of his privileged position in nature; where the fast progress of technology poses very difficult moral dilemmas and is likely to pose incomparably more difficult ones in the not-too-distant future (e.g., when it may become feasible to double the present human life span, opening up new dimensions for the problem of overpopulation; or when it may become possible to undertake large-scale genetic and reproductional engineering; or when robots and computers truly competitive with humans may become available,

and so on). In an age like this, any investigation into the criteria of rational choice between alternative moral codes is of much more than merely theoretical significance.

Therefore, there is no question whatever in my mind that Rawls poses problems of the greatest importance. But this is precisely the reason why I feel it is important to *resist* the solutions he proposes for these problems. We should resist any moral code which would force us to discriminate against the legitimate needs and interests of many individuals merely because they happen to be rich, or at least not to be desperately poor; or because they are exceptionally gifted, or at least are not mentally retarded; or because they are healthy, or at least are not incurably sick, etc. We should resist such a moral code, because an alternative moral code, the utilitarian one, is readily available to us; and the latter permits us to give equal *a priori* weight to every person's legitimate interests, and to judge the relative importance of any given need of a particular person in each case by its merits, as assessed by commonsense criteria —rather than forcing us to judge them according to rigid, artificial, and often highly discriminatory rules of priority.

**Postscript**

This paper was written in May 1973. In the meantime, John Rawls has tried to answer some of my criticisms in a paper entitled "Some Reasons for the Maximin Criterion."[19] His defense to the counterexamples I have put forward against using the maximin principle as a moral principle (in Section 3 of the preceding paper) is that "the maximin criterion is not meant to apply to small-scale situations, say, to how a doctor should treat his patients or a university its students. . . . Maximin is a macro not a micro principle" (p. 142). Regretfully, I must say that this is a singularly inept defense.

First of all, though my counterexamples do refer to small-scale situations, it is very easy to adapt them to large-scale situations since they have intrinsically nothing to do with scale, whether small or large. For example, instead of asking whether a doctor should use a life-saving drug in short supply for treating patient A or patient B, we can ask whether, in allocating scarce medical manpower and other resources, society should give priority to those patients who could best benefit from medical treatment, or should rather give priority to the most hopelessly sick patients—a policy problem surely affecting several hundred thousand individuals in any major country at any given time. Or, again, instead of asking whether scarce educational resources should be used for the benefit of individual A or individual B, we can ask whether, in allocating educational expenditures, society should give priority in certain cases to several hundred thousand highly gifted students, who could presumably benefit most, or to several hundred thousand seriously retarded individuals, who could derive only minor benefits from additional education, etc. I am really astonished that a distinguished philosopher like Rawls should have overlooked the simple fact that the counterexamples I have adduced (and the many more counterexamples one could easily adduce) have nothing whatever to do with scale at all.

In fact, it would be *a priori* rather surprising if, at the most fundamental level, the basic principles of morality should take different forms for large-scale and for small-scale situations. Does Rawls seriously think that there is a certain number $x$, such that a situation involving *more* than $x$ people will come under moral principles basically different from a situation involving *fewer* than $x$ people?

In any case, what moral considerations will determine this curious boundary number $x$ itself? More fundamentally, what are the basic logical reasons that should make large-scale and small-scale situations essentially different from a moral point of view? I cannot see how anybody can propose the strange doctrine that scale is a fundamental variable in moral philosophy, without giving credible answers to these questions at the same time.

I have argued that in *most* situations Rawls's theory will have much the same policy implications as utilitarian theory does, but that there are *some* important situations where this is not the case. Moreover, I have tried to show that, in those situations where the two theories do have quite dissimilar policy implications, Rawls's theory consistently yields morally highly *unacceptable* policy conclusions whereas utilitarian theory consistently yields morally fully *acceptable* ones (Sections 3 and 4 of the preceding paper).

Arrow has expressed a similar view.[20] After saying that in the real world the maximin principle and the utilitarian principle would have very similar practical consequences, he adds: " . . . the maximin principle would lead to unacceptable consequences if the world were such that they [these consequences] really differed." My only disagreement with Arrow is that I think the world is in fact so constituted that these two principles *do* have very different practical consequences in some important cases. (In effect, in some parts of

[19] John Rawls, "Some Reasons for the Maximin Criterion," *American Economic Review*, 64, Papers & Proc. (May, 1974), 141–146.

[20] Kenneth J. Arrow, "Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice," *The Journal of Philosophy*, 70 (May 10, 1973), 255.

his paper, Arrow himself seems to admit that much—pp. 251–252.) But we do agree on the main point, viz., on the conditional statement that, *if* such differences exist, they all speak very strongly against the maximin principle.

In my opinion, if this criticism is valid, then it completely disqualifies Rawls's theory as a serious competitor to utilitarian theory. (Why should anybody choose a theory that often does much worse, and never does any better, than utilitarian theory does?) For this reason, I find it rather unfortunate that Rawls's paper does not even try to answer this criticism at all.

To be sure, the maximin principle does have its valuable uses, and we must be grateful to Rawls for calling our attention to it. Even if it cannot serve as a *basic* principle of moral *theory*, it can be used as a principle of approximate validity in practical *applications*, such as the theory of optimal income distribution or of optimal taxation. In such applications, its relative independence of detailed interpersonal utility comparisons, and of the actual mathematical form of people's von Neumann-Morgenstern utility functions for money, is an important advantage, and can be fruitfully exploited in economic studies.[21]

Of course, from the point of view of a utilitarian observer, the results of a study of, e.g., optimal income tax rates, based on the maximin principle, will have only approximate validity. For example, if the study finds that, owing to the disincentive effect of very high marginal tax rates, the marginal income tax for the highest income group should be (say) 50 per cent, then a utilitarian observer can infer that this tax rate should certainly be *no more* than 50 per cent. Indeed, he can infer that, if the study had been based on the average utility principle instead of the maximin principle, then the marginal tax rate at the top would have come out presumably *a little lower* than 50 per cent, though perhaps not very much lower. (Sensitivity analysis may even enable us to estimate the actual percentage points by which studies based on the maximin principle are likely to overestimate the optimal tax rates for various income groups.)

It is regrettable that Rawls has ever made the untenable claim that he is proposing a moral *theory* superior to utilitarian theory. This claim can only obscure the practical merits of the maximin principle as an easily applicable postulate of approximate validity. These practical merits of course do not in any way provide a reason for abandoning utilitarian moral philosophy. (Basic philosophical principles must be exactly right, and not merely approximately right.) But they do provide a reason, even for a utilitarian moral philosopher, to use the maximin principle as an admissible approximation in many cases. Had Rawls only made this more modest, but much more realistic, claim for the maximin principle, few people would have contradicted him.

One thing that all of us must have learned in the last fifty years is that we must never commit ourselves seriously to moral principles or political ideologies that are bound to lead to morally utterly *wrong* policies from time to time—however great the advantages of these principles or ideologies may be in terms of administrative convenience, ease of application, and readier understandability.

[21] Arrow, p. 259.