Review
Reviewed Work(s): Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice by John Rawls
Review by: Kenneth J. Arrow
Source: *The Journal of Philosophy*, Vol. 70, No. 9 (May 10, 1973), pp. 245-263
Published by: Journal of Philosophy, Inc.
Stable URL: http://www.jstor.org/stable/2025006
Accessed: 07-03-2017 03:00 UTC

THE JOURNAL OF PHILOSOPHY

VOLUME LXX, NO. 9, MAY 10, 1973

A THEORY OF JUSTICE BY JOHN RAWLS: TWO REVIEW ARTICLES *

SOME ORDINALIST-UTILITARIAN NOTES ON RAWLS'S THEORY OF JUSTICE †

RAWLS'S major work has been widely and correctly acclaimed as the most searching investigation of the notion of justice in modern times. It combines a genuine and fruitful originality of viewpoint with an extraordinary systematic evaluation of foundations, implications for action, and connections with other aspects of moral choice. The specific postulates for justice that Rawls enunciates are quite novel, and yet, once stated, they clearly have a strong claim on our attention as at least plausible candidates for the foundations of a theory of justice. The arguments for accepting these postulates are part of the contractarian tradition, but have been developed in many new and interesting ways. The implications of these postulates for specific aspects of the institutions of liberty, particularly civil liberty, and for the operations of the economic order are spelled out in considerable and thoughtful detail (as an economist accustomed to much elementary misunderstanding of the nature of an economy on the part of philosophers and social scientists, I must express my gratitude for the sophistication and knowledge which Rawls displays here). Finally, the relations between justice of social institutions and the notion of morally right behavior on the part of individuals is analyzed at considerable and intelligent length.

It will become clear in the sequel that I have a number of questions and objections to Rawls's theory. Indeed, it is not surprising that no theory of justice can be so compelling as to forestall some

* Cambridge, Mass.: The Belknap Press of the Harvard University Press, 1971. xv, 607 p. Cloth $15.00, paper $3.95.

† This note was prepared with assistance from Grant N.S.F. No. GS28626X.

objections; indeed, that very fact is disturbing to the quest for the concept of justice, as I shall briefly note in the last section of this paper. These questions are a tribute to the breadth and fruitfulness of Rawls's work.

My critical stance is derived from a particular tradition of thought: that of welfare economics. In the prescription of economic policy, questions of distributive justice inevitably arise (not *all* such questions arise, only some; in particular, justice in the allocation of freedoms rather than goods is not part of the formal analysis of welfare economics, though some economists have made strong informal and unanalyzed commitments to some aspects of freedom). The implicit ethical basis of economic policy judgment is some version of utilitarianism. At the same time, descriptive economics has relied heavily on a utilitarian psychology in explaining the choices made by consumers and other economic agents. The basic theorem of welfare economics: that, under certain conditions, the competitive economic system yields an outcome that is optimal or efficient (in a sense which requires careful definition), depends on the identification of the utility structures that motivate the choices made by economic agents with the utility structures used in judging the optimality of the outcome of the competitive system. As a result, the utility concepts which, in one form or another, underlie welfare judgments in economics as well as elsewhere (according to Rawls's and many other theories of justice) have been subjected to an intensive scrutiny by economists. There has been more emphasis on their operational meaning, but perhaps less on their specific content; philosophers have been more prone to analyze what individuals should want, where economists have been content to identify "should" with "is" for the individual (not for society).

I do not mean that all economists or even those who have concerned themselves with welfare judgments will agree with the following remarks, but I do want to suggest the background out of which these concerns originated.

In section I, I will highlight the basic assumptions of Rawls's theory and stress those aspects which especially intersect my interests. I will be brief, since by now the theory is doubtless reasonably familiar to the reader. In section II, I raise some specific questions about different aspects of the theory, in particular, the logic by which Rawls proceeds from the general point of view of the theory (the "original position," the "difference principle" in its general form) to more specific implications, such as the priority of liberty and the maximin principle for distribution of goods. Section III is the central section of this paper; in it I raise a number of the epistemological issues that seem to me to be crucial in the development of most kinds of ethical theory and in particular Rawls's: How do we know other peoples' welfare enough to apply a principle of justice? What knowledge is assumed to be possessed by those in Rawls's original position when they agree to a set of principles? In section IV, I state more explicitly what may be termed an *ordinalist* (i.e., epistemologically modest) version of utilitarianism and argue that, in these terms, Rawls's position does not differ sharply. A brief section V discusses the role of majority and other kinds of voting in a theory of justice, especially in light of the discussion in section IV. Section VI turns to a different line, an examination of the implication of Rawls's theory for economic policy. Finally, in section VII, some of the preceding discussions are applied and extended to raise some questions about the possibility of any theory of justice; the criterion of universalizability may be impossible to achieve when people are really different, particularly when different life experiences mean that they can never have the same information.

I. SOME BASIC ASPECTS OF RAWLS'S THEORY

The central part of Rawls's theory is a statement of fundamental propositions about the nature of a just society, what may be thought of as a system of axioms. On the one side, it is sought to justify these axioms as deriving from a contract made among rational potential members of society; on the other side, the implications of these axioms for the determination of social institutions are drawn.

The axioms themselves can be thought of as divided into two parts: one is a general statement of the notion of justice, the second a more detailed elaboration of more specific forms.

The general point of view is a strongly affirmed egalitarianism, to be departed from only when it is in the interest of all to do so. "All social values—liberty and opportunity, income and wealth, and the bases of self-respect—are to be distributed equally unless unequal distribution of any, or all, of these values is to everyone's disadvantage" (p. 62; parenthetical page references are to Rawls's book). This *generalized difference principle*, as Rawls terms it, is no tautology. In particular, it implies that even natural advantages, superiorities of intelligence or strength, do not in themselves create any claims to greater rewards. The principles of justice are "an agreement to regard the distribution of natural talents as a common asset and to share in the benefits of this distribution" (101).

Personally, I share fully this value judgment; and, indeed, it is implied by almost all attempts at full formalization of welfare economics.[1] But a contradictory proposition: that an individual is entitled to what he creates, is widely and unreflectively held; when teaching elementary economics, I have had considerable difficulty in persuading the students that this *productivity principle* was not completely self-evident.

It may be worth stressing that the assumption of what may be termed *asset egalitarianism*: that all the assets of society, including personal skills, are available as a common pool for whatever distribution justice calls for, is so much taken for granted that it is hardly argued for. All the alternatives to his principles of justice that Rawls considers imply asset egalitarianism (though some of them are very inegalitarian in result, since more goods are to be assigned to those most capable of using them). The productivity principle is not even considered. It must be said, on the other hand, that asset egalitarianism is certainly an implication of the "original position" contract. (The practical implications of asset egalitarianism are, however, severely modified in the direction of the productivity principle by incentive considerations; see section VI below).

But Rawls's theory is a much more specific statement of the concept of justice. This consists of two parts. First, among the goods distributed by the social order, liberty has a priority over others; no amount of material goods is considered to compensate for a loss of liberty. Second, among goods of a given priority class, inequalities should be permitted only if they increase the lot of the least well off. The first principle will be referred to as the *priority of liberty*, the second as the *maximin* principle (*max*imizing the welfare at its *min*imum level; Rawls himself refers to this as the *difference principle*).

Rawls argues for these two principles as being those which would be agreed to by rational individuals in a hypothetical *original position*, where they have full general knowledge of the world, but do not know which individual they will be. The idea of this "veil of ignorance" is that principles of justice must be universalizable; they must be such as to command assent by anyone who does not take account of his individual circumstances. If it is assumed that rational individuals under these circumstances have some degree of aversion to uncertainty, then they will find it desirable to enter into an insurance agreement: that the more successful will share

[1] See A. Bergson, *Essays in Normative Economics* (Cambridge, Mass.: Harvard, 1966), ch. I; P. A. Samuelson, *The Foundations of Economic Analysis* (Cambridge, Mass.: Harvard, 1947), pp. 230–248; or F. Y. Edgeworth (London: Kegan Paul, 1881), pp. 56–82.

with the less, though not so much as to make them both worse off. Thus, the original-position argument does lead to a generalized view of justice. Rawls then further argues that his more specific principles (priority of liberty and the maximin principle) also follow from the original-position argument, at least in the sense of being preferable to other principles advanced in the philosophical literature, such as classical utilitarianism.

Two final remarks on the general nature of Rawls's system: (1) The principles of justice are intended to apply to the choice of social institutions, not to the actual allocative decisions of society separately. (2) The principles are supposed to characterize an ideal state of justice. If the ideal state is not achieved, they do not in themselves supply any basis for deciding that one non-ideal state is more or less just than another. "Questions of strategy are not to be confused with those of justice. . . . The force of opposing attitudes has no bearing on the question of right but only on the feasibility of arrangements of liberty" (231). It is intended of course that a characterization of ideal or optimal states of justice is a first step in a complete ordering of alternative institutional arrangements as more or less just.

II. THE DERIVATION OF RAWLS'S SPECIFIC RULES

From the viewpoint of the logical structure of the theory, a central question is the extent to which the assumption of the original position really implies the highly specific forms of Rawls's two rules. Let me take the priority of liberty first. This is given a central place in presentation, and at a number of points the fact that the theory puts such emphasis on liberty is used to distinguish it favorably from utilitarianism; the latter, it is argued, might easily lead to sacrificing the liberty of a few for the benefit of many. "Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override. For this reason justice denies that the loss of freedom for some is made right by a greater good shared by others" (3/4).

Despite its importance, the definitive argument for the priority of liberty is postponed to very late in the book (541–548). The key argument is that the priority of liberty is desired by every individual. In technical terms, each individual has a *lexicographical* (or "lexical" in Rawls's simplification) ordering of goods of all kinds, with liberty coming first; of any two possible states, an individual will always prefer that with the most liberty, regardless of other goods (such as income), and will choose according to income only among states with equal liberty. "The supposition is that . . . the persons

. . . will not exchange a lesser liberty for an improvement in their economic well-being, at least not once a certain level of wealth has been attained. . . . As the conditions of civilization improve, the marginal significance for our good of further economic and social advantages diminishes relative to the interests of liberty" (542).

The argument is clearly an empirical judgment, and the reader can decide for himself how much weight it will bear. I want to bring out another aspect, the relation to utilitarianism. If in fact each individual assigns priority to liberty in the lexicographical sense, then the most classical sum-of-utilities criterion will do the same for social choice; the rule will be for society to maximize the sum of individuals' liberties and then, among those states which accomplish this, choose that which maximizes the sum of satisfactions from other goods.

Let me now turn to the maximin rule (this is to be applied separately to liberty and to the nonpriority goods). The justification appears most explicitly on pages 155–158; it is mainly an argument for maximin as against the sum-of-utilities criterion. It should first be noted that the original-position assumption had also been put forth by the economists W. S. Vickrey[2] and J. C. Harsanyi[3]; but they use it to supply a contractarian foundation to a form of utilitarianism (discussed at considerable length by Rawls, 161–175). They start from the position, due to F. P. Ramsey, and J. von Neumann and O. Morgenstern, that choice under risky conditions can be described as the maximization of expected utility. In the original position, each individual may with equal probability be any member of the society. If there are n members of the society and if the ith member will have utility u_i under some given allocation decision, then the value of that allocation to any individual is $\Sigma u_i(1/n)$, since $1/n$ is the probability of being individual i. Thus, in choosing among alternative allocations of goods, each individual in the original position will want to maximize this expectation, or, what is the same thing for a given population, maximize the sum of utilities.

[2] "Measuring Marginal Utility by Reactions to Risk," *Econometrica*, XIII (1945): 319–333, p. 329; "Utility, Strategy, and Social Decision Rules," *Quarterly Journal of Economics*, LXXIV (1960): 507–535, pp. 523f.

Vickrey's 1945 statement has been overlooked by all subsequent writers, not surprisingly, since it received relatively little emphasis in a paper overtly devoted to a seemingly different subject. I read the paper before I was concerned with the theory of social choice; the implications for that theory were so easy to overlook that they did not occur to me at all when they would have been relevant.

[3] "Cardinal Utility in Welfare Economics and the Theory of Risk-taking," *Journal of Political Economy*, LXI (1953): 434/5; "Cardinal Welfare, Individualistic Ethics, and Personal Comparisons of Utility," *ibid.*, LXIII (1955): 309–321.

Rawls, however, starting from the same premises, derives the statement that society should maximize min u_i. The argument seems to have two parts: first, that in an original position, where the quality of an entire life is at stake, it is reasonable to have a high degree of aversion to risk, and being concerned with the worst possible outcome is an extreme form of risk aversion; and, second, that the probabilities are in fact ill defined and should not be employed in such a calculation. The first point raises some questions about the meaning of the utilities and does not do justice to the fact that, at least in Vickrey and Harsanyi, the utilities are already so measured as to reflect risk aversion (see some further discussion in section IV). The second point is a version of a recurrent and unresolved controversy in the theory of behavior under uncertainty; are all uncertainties expressible by probabilities? The view that they are has a long history and has been given an axiomatic justification by Ramsey[4] and by L. J. Savage.[5] The contrary view has been upheld by F. H. Knight[6] and by many writers who have held to an objective view of probability; the maximin theory of rational decision-making under uncertainty was set forth by A. Wald[7] specifically in the latter context. Among economists, G. L. S. Shackle[8] has been a noted advocate of a more general theory which includes maximin as a special case. L. Hurwicz and I[9] have given a set of axioms which imply that choice will be based on some function of the maximum and the minimum utility.

It has, however, long been remarked that the maximin theory has some implications that seem hardly acceptable. It implies that any benefit, no matter how small, to the worst-off member of society, will outweigh any loss to a better-off individual, provided it does not reduce the second below the level of the first. Thus, there can easily exist medical procedures which serve to keep people barely alive but with little satisfaction and which are yet so expensive as to reduce the rest of the population to poverty. A maximin principle would apparently imply that such procedures be adopted.

[4] F. P. Ramsey, "Truth and Probability," in *The Foundations of Mathematics and Other Logical Essays* (London: K. Paul, Trench, Trubner, 1931), p. 156–198.

[5] *The Foundations of Statistics* (New York: Wiley, 1954).

[6] *Risk, Uncertainty, and Profit* (New York: Houghton Mifflin, 1921).

[7] "Contributions to the Theory of Statistical Estimation and Testing Hypotheses," *Annals of Mathematical Statistics*, x (1939): 299–326.

[8] *Expectations in Economics* (Cambridge: University Press, 1949) and subsequent works.

[9] "An Optimality Criterion for Decision-making under Ignorance," in C. F. Carter and J. L. Ford, eds., *Uncertainty and Expectation in Economics* (Oxford: Basil Blackwell, 1972), pp. 1–11.

Rawls considers this argument, but rejects it on the ground that it will not occur in practice. He fairly consistently assumes that the actual society has the property he calls *close-knittedness*: "As we raise the expectations of the more advantaged the situation of the worst off is continuously improved. . . . For the greater expections of the more favored presumably cover the costs of training and encourage better performance" (158). It is hard to analyze this argument fairly in short compass. On the face of it, it seems clearly false; there is nothing easier than to point out changes that benefit the well-off at the expense of the poor, including the least advantaged, e.g., simultaneous reduction of the income tax for high brackets and of welfare payments. Rawls holds that one must consider his principles in their totality, in particular, a strongly expressed demand for open access to all positions. But, even with perfect equality of opportunity, there will presumably remain inequalities due to biological and cultural inheritance (Rawls nowhere advocates abolition of the family) and chance events, and, once inequalities do exist, the harmony of interests seems to be less than all-pervasive. In any case, the assumption of close-knittedness undermines all the distinctions that Rawls is so careful to make. For, if it holds, there is no difference in policy implication between the maximin principle and the sum of utilities; if all satisfactions go up together, the conflict between the individual and the society disappears.

III. EPISTEMOLOGICAL ISSUES IN THE THEORY OF JUSTICE

Many theories of justice, including both Rawls's and utilitarianism, imply that the social institutions or their creators have access to some kinds of knowledge. This raises the question whether they can in fact or even in principle have such knowledge. In this section, two epistemological questions are raised, though there are others: (1) How can interpersonal comparisons of satisfaction be made? and (2) What knowledge is available in the original position?

1. The problem of interpersonal comparison of utilities seems to bother economists more than philosophers. As already indicated, utility or satisfaction or any other similar concept appears in economic theory as an explanation of individual behavior, for example, as a consumer. Specifically, it is hypothesized that the individual chooses his consumption so as to maximize his utility, subject to the constraints imposed by his budget. But, for this purpose, a quantitatively measurable utility is a superfluous concept. All that is needed is an ordering, that is, a statement for each pair of consumption patterns as to which is preferred. Any numerical function over

the possible consumption patterns having the property that it assigns larger numbers to preferred bundles could be thought of as a utility function. Clearly, then, any monotonic transformation of a utility function is also a utility function.

To turn the matter around, it might be asked, How can we have any evidence about the magnitude of utility? The only evidence on an individual's utility function is supplied by his observable behavior, specifically the choices he makes in the course of maximizing the function. But such choices are defined by the preference ordering and must therefore be the same for all utility functions compatible with that ordering. Hence, there is no quantitative meaning for utility for an individual. (This *ordinalist* position was introduced into economics by V. Pareto and I. Fisher and has become fairly orthodox in the last thirty years.)

If the utility of an individual is not measurable, then *a fortiori* the comparison of utilities of different individuals is not meaningful. In particular, the sum-of-utilities criterion becomes indefensible as it stands. Rawls's maximin criterion also implies interpersonal comparison, for we must pick out the least advantaged individual, and that requires statements of the form, "individual *A* is worse off than individual *B*." Unlike the sum-of-utilities approach, however, this does not require that the units in which different individuals' utilities are measured be comparable, only that we be able to rank different individuals according to some scale of satisfaction. But we do not have any underlying numerical magnitude to use for this purpose, and the question still remains, What is the operational meaning of the interpersonal comparison?

If one is to take the sum-of-utilities criterion seriously, then it would have to be considered possible for individuals to have different utility functions; in particular, they might derive different amounts of satisfaction from the same increments to their wealth. Then, the utilitarian would have to agree that the sum of utilities would be increased by shifting wealth to the more sensitive individuals. This does not occur in Rawls's theory, but something parallel to it does. Consider an individual who is incapable of deriving much pleasure from anything, whether because of psychological or physical limitations. He may well be the worst-off individual and, therefore, be the touchstone of distribution policy, even though he derives little satisfaction from the additional income.

In the usual applications of the sum-of-utilities approach, the problem of differing utilities is dodged by assuming it away; it is postulated that everyone has the same utility function. This avoids not only what may be thought of as the injustice of distributing income in favor of the more sensitive, but also the problem of ascertaining in detail what the utility functions are, a task which might be thought impossible, as argued above, or at least very difficult in practice, if the ordinalist position is not accepted. Rawls criticizes this utilitarian evasion, though cautiously; he does not wish to reject interpersonal comparisons (90/1). But in fact he winds up with a somewhat similar approach. He introduces the interesting concept of *primary goods,* those goods which are needed whatever an individual's preference relation ("rational plan of life," in Rawls's terms) is. These might be liberties, opportunities, and income and wealth. Then, even though individuals might have very different uses for these primary goods, we need consider only some simple index of them for purposes of interpersonal comparison. Thus, the fact that one individual was satisfied with water and soy flour, while another was desperate without pre-phylloxera clarets and plovers' eggs, would have no bearing on the interpersonal comparison; if they had the same income, they would be equally well off.

If this comparison appears facetious, consider the haemophiliac who needs about $4000 worth per annum of coagulant therapy to arrive at a state of security from bleeding at all comparable to that of the normal person. Does equal income mean equality? If not, then, to be consistent, Rawls would have to add health to the list of primary goods; but then there is a trade-off between health and wealth which involves all the conceptual problems of differing utility functions.

The restriction to some list of primary goods is probably essential. I have but two comments: (1) so long as there is more than one primary good, there is an index-number problem in commensurating the different goods, which is in principle as difficult as the problem of interpersonal comparability with which we started; (2) if we could resolve the problem of interpersonal comparability in Rawls's system by reducing everything in effect to a single primary good, we could do the same in the sum-of-utilities approach. To the last statement, however, there is a qualification: the maximin criterion requires only interpersonal ordinality, whereas the classical view requires interpersonally comparable units; to that extent, the Rawls system is epistemologically less demanding.

2. Let us turn from the epistemological problems of the current decision-maker for society to those in the original position. Individuals are supposed to know the laws of the physical and the

social worlds, but not to know who they are or will be. But empirical knowledge is after all uncertain, and even in the original position individuals may disagree about the facts and laws of the universe. For example, Rawls argues for religious toleration on the grounds that one doesn't know what religion one will have, and therefore one wants society to tolerate all religions. Operationally, a Catholic would have to recognize that in the original position he wouldn't know he would be a Catholic and would therefore have to tolerate Protestants or Jews or whatever, since he might well have been one. But suppose he replies that in fact Catholicism is the true religion, that it is part of the knowledge which all sensible people are supposed to have in the original position, and that he insists on it for the salvation of all mankind. How could this be refuted?

Indeed, just this sort of argument is raised by writers like Marcuse, not to mention any totalitarian state and, within wider limits, any state. Only those who correctly understand the laws of society should be allowed to express their political opinions. I feel I know that Marxism (or laissez-faire) is the truth; therefore, in the original position, I would have supported suppressing other positions. Even Rawls permits suppression of those who do not believe in freedom.

I hope it is needless to say that I am in favor of very wide toleration. But I am not convinced that the original position is a sufficient basis for this argument, for it transfers the problem to the area of factual disagreement.

There is another kind of knowledge problem in the original position: that about social preferences. Rawls assumes that individuals are egoistic, their social preferences being derived from the veil of ignorance. But why should there not be views of benevolence (or envy) even in the original position? All that is required is that they not refer to named individuals. But if these are admitted, then there can be disagreement over the degree of benevolence or malevolence, and the happy assumption, that there are no disagreements in the original position, disappears.

IV. SOME REMARKS ON UTILITARIANISM

It will already have been seen that my attitude toward utilitarianism is ambivalent. On the one hand, I find it difficult to ascribe operational meaning to the utilities to be added. On the other hand, I have suggested that the practical differences between the maximin and the sum-of-utilities criteria are not great, and indeed that the maximin principle would lead to unacceptable consequences if the world were such that they really differed.

In this section, I will take up several different points raised by Rawls, and try to defend utilitarianism against them.

First, let me extend a little the discussion of the Vickrey-Harsanyi position, which Rawls calls *average utilitarianism*. In part, this discussion continues the epistemological considerations of the last section. As Ramsey and von Newmann and Morgenstern have shown, if one considers choice among risky alternatives, there is a sense in which a quantitative utility can be given meaning. Specifically, if choice among probability distributions satisfies certain apparently natural rationality conditions, then it can be shown that there is a utility function (unique up to a positive linear transformation) on the outcomes such that probability distributions of outcomes are ordered in accordance with the mathematical expectation of the utility of the outcome.

By itself, this theorem does not establish any welfare implications for this utility function; after all, the choice among probability distributions of outcomes could equally well be described by any monotonic transformation of the expected utility. When I first wrote on this matter,[10] I therefore denied the welfare relevance of expected-utility theory. But the Vickrey-Harsanyi argument puts matters in a different perspective; if an individual assumes he may with equal probability be any member of society, then indeed he evaluates any policy by his expected utility, *where the utility function is specifically that defined by the von Neumann-Morgenstern theorem.* Rawls therefore errs when he argues that average utilitarianism assumes risk neutrality (165); on the contrary, the degree of risk aversion of the individuals is already incorporated in the utility function. This point may be given further strength by noting that the maximin criterion, far from being opposed to average utilitarianism, can be regarded as a limiting case of it. For let U be any utility function, in the sense of a function that represents preferences without uncertainty. Then, for any $a > 0$, $-U^{-a}$ is an increasing function of U and so also is a utility function. Any member of this family could be the von Neumann-Morgenstern utility function, i.e., that utility function for which it is true that the individual seeks to maximize expected utility. It is easy to see that, the larger the value of a, the higher the degree of risk aversion. Then, according to Vickrey, the value of a policy to an individual with a random stake in society would be

$$V = \Sigma(-U_i)^{-a} = -\Sigma U_i^{-a}$$

[10] *Social Choice and Individual Values* (New York: Wiley, 1951), first ed., pp. 9/10.

But a social-welfare function is only an index of choice and can itself be subject to monotonic transformation; hence, another criterion that would yield the same choice is

$$W = (-V)^{-1/a} = (\Sigma U_i^{-a})^{-1/a}$$

It can, however, easily be proved that, as a approaches infinity, representing increasing degrees of risk aversion, W approaches min U_i.

I do not wish to argue that average utilitarianism meets all the problems that can be raised. Rawls very properly points out that each individual may have a different utility function, so that, although each wishes to maximize a sum of utilities, each individual has a different utility function in his maximand (173); in addition, the use of equiprobability in this case is certainly not beyond cavil.

A second of Rawls's objections to utilitarianism is that it may require that some individuals sacrifice for the benefit of others, so that other men appear to be means, not only ends (181, 183). But I don't follow this argument at all. A maximin principle certainly seems to imply that the better off should sacrifice for the less well off, if that will in fact help. The talents of the more able are, in Rawls's system (and in my value judgments), to be used on behalf of the less able; is this not using some people as means?

A third criticism of classical utilitarianism is that it makes an illegitimate analogy between individuals and society. "The classical view results, then, in impersonality, in the conflation of all desires into one system of desire" (188). But it would appear to me a purely formal requirement of any theory of justice that it act as such a conflation. A theory of justice is presumably an ordering of alternative social states, and therefore is formally analogous to the individual's ordering of alternative social states. Further, Rawls and Bentham and I would certainly all agree that justice should reflect individual satisfactions; hence, the social choice made in accordance with any of these theories of justice is "a conflation of all desires." No doubt perfectionist theories or those based on religious considerations would not be so characterized; but Rawls is not defending *them.*

V. A REMARK ON VOTING

The expression and aggregation of individual preferences through voting does not have a high place in Rawls's system: "There is nothing to the view, then, that what the majority wills is right" (356). The legislators or voters are thought of as experts in justice and are not to vote in self-interest. The assumption seems to be simply that the workings of justice will not always be clear and that a pooling of opinions is worth while; a majority makes more sense from this point of view than a minority.

Clearly, there is something to Rawls's position, which indeed he shares with many political philosophers, as he notes. A political system in which there is no other-regardingness will not function at all. Further, Rawls is right in saying that the analogy with the market is imperfect. In the market, he agrees that selfish behavior is socially correct, but holds that the political process can never lead to perfect justice if based on self-seeking behavior. But I would argue that the analogy, though imperfect, is not completely wrong either. Political competition does serve some of the same functions in its sphere as economic competition. Further, the expression of one's own interests in voting seems to me an essential part of the information process needed for voting. Unless voters express their interests, how is anyone going to know if the ends of justice are in fact being carried out? "If I am not for myself, then who is for me?," said Hillel, though he continued in more Rawlsian terms, "and if I am not for others, then who am I?"

To put the matter more emotionally, I would hold that the notion of voting according to one's own beliefs and then submitting to the will of the majority represents a recognition of the essential autonomy and freedom of others. It recognizes that justice is a pooling of irreducibly different individuals, not the carrying out of policies already known in advance.

VI. ECONOMIC IMPLICATIONS OF RAWLS'S PRINCIPLES

Rawls's views have implications most directly for the redistribution of income, both among contemporaries and across generations. The maximin rule would seem on the face of it to lead to radical equalization of income. Indeed, so would the sum-of-utilities rule, if it is assumed that all individuals have the same utility function which displays decreasing marginal satisfactions from additional increments of income. Rawls, however, holds that the close-knittedness of members of the society means that perfect equality of income is not to the advantage of the least well-off, but that typically they will benefit by an increase in income to some higher up in the income scale. Rawls is rather brief on why one might expect this kind of relation, but economists have laid considerable stress on the *incentive* effects of taxation. Assume that each individual can produce a certain amount per hour worked, but that this productivity varies from individual to individual. In the absence of taxation, the least productive individual will be the worst off. Therefore, a Rawlsian (or even an old-fashioned utilitarian) may advocate a tax on the income of the more able to be paid out to the less able. This is, in fact, essentially the widespread proposal for a negative income tax. But since the effort to produce may in itself detract from satisfaction, an income tax will lead individuals to reduce the number of hours they work and therefore the amount they produce. If the tax rate on the more able is high enough, the amount of work will go down so much that the amount collected in taxes for redistribution to the worst off will actually decrease. It is at this stage that the economy becomes close knit.

The conflict between incentive and equity occurs in a utilitarian framework and was already noted by Edgeworth (who was really very conservative and was glad to escape from the rigorous egalitarianism to which his utilitarianism led). The mathematical problem of choosing a tax schedule to maximize the sum of utilities, taking account of the adverse incentive effects, is a very difficult one; it was broached by Vickrey in his 1945 paper (*op. cit.*) and analyzed by Mirrlees,[11] Fair,[12] and Sheshinski,[13] among others. More recently, the tax implications of the Rawls criterion have been analyzed along similar lines in forthcoming papers by Atkinson, Phelps, and Sheshinski. The practical implications of this research are as yet dubious, primarily because too little is known about the magnitude of the incentive effects, particularly in the upper brackets.

As I have indicated, Rawls is inexplicit about the incentive effects and so does not give clear guidance to the determination of tax rates. On pages 277–279 he argues for progressive income and inheritance taxes to achieve justice, but there is no indication how the rates should be chosen. Clearly, the philosophy of justice is under no obligation to tell us what the rates should be in a numerical sense; but it is supposed to define the rule that translates any given set of facts into a tax schedule. The maximin rule would, on the face of it, lead to perfect equalization, i.e., 100 per cent taxation above a certain level, with corresponding subsidies below it. As far as I can see, it is only the incentive question that prevents us from carrying this policy out.

The incentive question raises another issue with regard to the obligation of an individual to perform justice (Rawls has much to

[11] J. A. Mirrlees, "An Exploration in the Theory of Optimal Income Taxation," *Review of Economic Studies*, XXXVIII (1971): 175–208.

[12] R. C. Fair, "The Optimal Distribution of Income," *Quarterly Journal of Economics*, LXXXV (1971): 551–579.

[13] E. Sheshinski, "The Optimum Linear Income Tax," *Review of Economic Studies*, XXXIX (1972): 297–302.

say on the notion of duties and obligations on individuals, though I have slighted this discussion in this review). If each individual revealed his productivity (the amount he could produce per unit of time), it would be possible to achieve a perfect reconciliation of justice and incentives; namely, tax each individual according to his ability, not according to his actual output. Then he could not escape taxes by working less, and so the tax system would have no adverse incentive effects. Practical economists would reject this solution, because it would be taken for granted that no individual would be truthful if the consequences of truth-telling were so painful. But Rawls, like most social philosophers, takes it for granted that individuals are supposed to act justly, at least in certain contexts. For example, as legislators or voters, it is an obligation or duty to judge according to the principles of justice, not according to self-interest. If, then, an individual is supposed to assess his own potential for earning income, is there an obligation to be truthful?

One of the most difficult questions in allocative justice is the distribution of wealth over generations. To what extent is one generation obligated to save, so as to increase the welfare of the next generation? The traditional economic problem has been the general act of investment in productive land, machines, and buildings which produce goods in the future; more recently, we have become especially concerned with preservation of undisturbed environments and natural resources. The most straightforward utilitarian answer is that the utilities of future generations enter equally with those of the present. But since the present generation is a very small part of the total number of individuals over a horizon easily measurable in thousands of years, the policy conclusion would be that virtually everything should be saved and very little consumed, a conclusion which seems offensive to common sense. The most usual formulation has been to assert a criterion of maximizing a sum of *discounted* utilities, in which the utilities of future generations are given successively smaller weights. The implications of such policies seem to be more in accordance with common sense and practice, but the foundations of such a criterion seem arbitrary.

Rawls argues that the maximin criterion, properly interpreted, can be applied to the determination of a just rate of savings (284–292). In the original position, individuals do not know which generation they belong to and should therefore judge of a just rate on that basis. That is, they agree to leave a fixed fraction of their income to the next generation in return for receiving an equal fraction of the previous generation's income. There are two difficulties with this argument: (1) Why should they agree on a *fraction* rather than some more complicated rule, for example, an increasing fraction as wealth increases? (2) More serious, it would appear that the maximin rule would most likely lead to zero as the agreed-on savings rate; for the first generation would lose under any positive savings rate, whereas the welfare of all future generations would increase. This point is reinforced strongly if one adds the empirical fact of technological progress, so that even in the absence of savings the successive generations are getting better off. Then a maximin policy would call for improving the lot of the earlier generations, which can only be done by negative saving (running down existing capital equipment) if at all possible. (To be precise, the above argument is valid only in the absence of population growth. If population is growing, then zero saving would mean less capital per person and therefore a falling income per capita. Hence, a maximin rule in the absence of technological progress would call for positive saving; it can easily be shown that the rule would be that the rate of savings equals the rate of population growth multiplied by the capital-output ratio.)

Rawls, however, modifies the motivations in the original position at this point in the argument. "The parties are regarded as representing family lines, say, with ties of sentiment between successive generations" (292). This is a major departure from the egoistic assumptions held up to this point about behavior and choice in the original position. It should be noted that so long as fathers think more highly of themselves than of their sons or even more highly of their sons than of subsequent generations, the effect of this modification is very much the same as that of discounting future utilities. Although my guess is that any justification for provision for the future will run somewhat along these lines, it cannot be said that the solution fully escapes all difficulties. (1) It introduces an element of altruism into the original position; if we introduce family sentiments, why not others (nation, tribal)? And why not elements of envy? (2) One might like a theory of justice in which the role of the family was derived rather than primitive. In a reexamination of social institutions, why should the family remain above scrutiny, its role being locked into the original assumptions? (3) Anyway, the family argument for saving has an implication that should be displayed and might be questioned. Presumably the burden of saving should fall only on those with children and perhaps in proportion to the number of children. Since education and public construction are essentially forms of saving, taxes to support them should fall only on those with children. In the original position, this is just the sort of contract that would be arrived at if the concern for the future were based solely on family ties.

VII. A CRITICAL NOTE ON THE POSSIBILITY OF JUSTICE

Rawls's work is based on the hypothesis that there is a meaningful universal concept of justice. If there is, it surely must, as he says, be universalizable in some sense, that is, based on principles that are symmetric among the particular accidents that distinguish one individual's position from another. But as I look around at the many conflicts that plague our humanity, I find many for which I can imagine no argument of a symmetric nature which would convince both sides.

One problem is that any actual individual must necessarily have limited information about the world, and different individuals have different information. Hence, they cannot possibly argue themselves back into an original position with common information, even if they succeed in "forgetting" who they are. Indeed, one of the most brilliant passages in Rawls's book is that on what he calls "social union" (520–530). He argues that no human life is enough to encounter more than a small fraction of the experiences needed for completeness, so that individuals have a natural complementarity with each other (a more mundane version of this idea is Adam Smith's stress on the importance of the division of labor). The social nature of man springs from this variegation of experience. But precisely the same differentiations imply differing and incompletely communicable life experiences and therewith the possible impossibility of agreeing on the just action in any concrete situation.

Indeed, the thrust of Rawls's work, particularly in its latter passages, is highly harmonistic; the principles of justice are stable, according to Rawls, because the moral education they induce reinforces them. But if the specific application of the principles is judged to be different according to different life experiences (and of course different genetic experiences), even as between parent and child, then the needed concordance of views may evaporate.

To put the matter somewhat differently, many sociologists would hold that, in a world of limited information, conflict unresolved by appeal to commonly accepted principles may have a positive value; it is the means by which information about others is conveyed. In its own sphere, this is the role assigned to competition by economists; if everyone attempted to act justly at every moment in his economic life, it might be difficult ever to find out what the true interests of anyone were.

To the extent that individuals are really individual, each an autonomous end in himself, to that extent they must be somewhat mysterious and inaccessible to each other. There cannot be any rule that is completely acceptable to all. There must, or so it now seems to me, be the possibility of unadjudicable conflict, which may show itself logically as paradoxes in the process of social decision-making.

KENNETH J. ARROW

Economics, Harvard University

DUTY AND OBLIGATION IN THE NON-IDEAL WORLD

THERE is no need to summarize the argument of this philosophical epic. In its basic outline it is sufficiently well known to the readers of this journal from Rawls's articles over the last twenty years. In this book Rawls has filled in gaps in the argument, answered numerous critical objections, applied his theory to problems of justice in politics, economics, education, and other important areas, and buttressed it with a theory of moral psychology and other argumentative reinforcement. The result is a remarkably thorough treatise which well deserves to be called a philosophical classic.

Rawls's primary aim, he tells us, is to provide a "workable and systematic moral conception" (viii) to oppose utilitarianism. Until now, the opponents of utilitarianism have been unable to provide an equally systematic alternative of their own, and have contented themselves with a series of *ad hoc* amendments and restrictions to utilitarianism designed to bring it into closer harmony with our spontaneous moral sentiments, at whatever cost in theoretical tidiness. They are likely to concede that *one* of the prime duties of social policy makers is to promote social utility, but then insist that one may not properly pursue that commendable goal by grinding the faces of the poor, framing and punishing the innocent, falsifying history, and so on. On the level of personal ethics, such moralists as W. D. Ross admit utilitarian duties of beneficence and nonmaleficence, but supplement them with quite nonutilitarian duties of veracity, fidelity, and the like, and there is no way of telling in advance which duty must trump the others when circumstances