A theory of justice

Richmond Campbell

Recommended Citation

Reviews

A THEORY OF JUSTICE. By John Rawls. Cambridge, Mass.: Belknap Press, 1971. Pp. xv, 607.

In *A Theory of Justice* John Rawls constructs a comprehensive social contract theory of justice to stand as a substantive alternative to utilitarianism. This work combines and develops the ideas of earlier essays, such as "Justice as Fairness" (1958), "The Sense of Justice" (1963), "Constitutional Liberty" (1963) and "Civil Disobedience" (1966), into a systematic moral and political philosophy of astonishing power and subtlety. I shall sketch its main principles, their derivation and justification, and then raise some questions about the supposed opposition between the standards of justice and utility.*

I

The guiding idea is that the principles of justice governing the institutions and social arrangements of society should be "principles that free and rational persons concerned to further their own interests would accept in an initial position of equality as defining the fundamental terms of their association" (11). Rawls calls this *justice as fairness.* The original position of equality is a hypothetical situation, not the state of nature in traditional theory of social contract. To guarantee equality in this situation, the principles are chosen behind a veil of ignorance: no one knows his place in society or his natural assets, such as intelligence or health; nor does anyone know his own conception of the good or his special psychological propensities. The parties to the agreement understand, however, that whatever their various conceptions of the good, each will be better able to realise his ends and ideals the greater his share of the social primary goods, such as income and wealth, power and authority. They understand also that they will live together in circumstances that will make human cooperation both

---

*Parenthetical page references are to the text.

†This a conceptual point which I argue in "The Pursuit of Happiness" (forthcoming in *The Personalist*).

possible and necessary. Rawls calls the latter the circumstances of justice. Finally, they accept certain formal constraints: the principles chosen must be formulated without reference to particular individuals, designed to hold for everyone (not merely restricted classes of individuals), adopted on the assumption that everyone knows that everyone accepts the principles, recognized as capable of ordering (with transitivity) all conflicting claims likely to arise in practice, and understood to take precedence over all other principles in practical reasoning. Being rational (in a morally neutral sense familiar in social theory) and aware of the circumstances of justice, the parties in the original position can agree only to principles that define a cooperative scheme that best serves to maximize their individual shares of the primary goods. Yet subject to the veil of ignorance and the formal constraints, they can agree only to principles that respect the interests of everyone and give no person an unfair advantage in pursuing his own ends.

It is assumed that the original parties know the general facts about human society. For example, they understand (within limits) the principles of economic theory, the laws of human psychology, and the basis of social organization. Even so, not knowing the likelihoods of the sets of circumstances in which they might find themselves, they cannot base their choice of principles on the mathematical expectation of gain in primary goods. Furthermore, in the original position it is not rational for anyone to gamble on a principle for the sake of a special advantage, since the potential losses are great. For these reasons Rawls argues that a conservative strategy of decision would be adopted in the choice of principles to minimize potential losses (on analogy with the "maximin" rule of game theory). The two chief principles of justice thus selected are: First Principle (of Equal Liberty) — Each person is to have an equal right to the most extensive basic liberty compatible with a like liberty for others. Second Principle — Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged and (b) attached to offices and positions open to all under conditions of fair equality of opportunity. Part (a) is called the Difference Principle and (b) the Principle of Fair Equality of Opportunity.

Much is packed into these principles. Among the basic liberties are the right to vote and to be eligible for public office,

freedom of speech and assembly, liberty of conscience and freedom of thought, freedom of the person, and the right to hold property. The Equal Liberty Principle allows these liberties to be restricted in certain cases provided that a less extensive liberty strengthens the total system of shared liberty or a less than equal liberty is acceptable to those with lesser liberty. This principle is "lexically" prior to the second in the sense that it must be satisfied before the other is applied. Hence, lesser liberties cannot be traded for greater social and economic benefits. Once the Equal Liberty Principle is satisfied, the Second Principle allows inequalities, measured by the distribution of primary goods other than liberties, provided that (a) a representative person from the least advantaged sector of society is better off than he would be if these inequalities did not exist and provided that (b) offices and positions are open to all in more than a merely formal sense. Within this principle part (b) is lexically prior to (a). Fair equality of opportunity requires that everyone should have a fair chance to attain offices and positions irrespective of the income class into which they are born. This requires, in turn, equal opportunities of education and other favourable social conditions on which depend the development of natural capacities and even the willingness to make an effort (74). Inequality of opportunity is permissible only when it increases the opportunities of those who otherwise, due to natural circumstances, would have lesser opportunities.

The Difference Principle may appear somewhat paradoxical. How can an inequality in the distribution of primary goods benefit the least advantaged? The assumption is that some inequality is ineluctable. Men are born with different natural capacities into different social positions and consequently some will stand a better chance than others to realise their conceptions of a good life. This is a natural fact which is neither just nor unjust. But the distribution of social primary goods, such as income, wealth, power, and authority, can be controlled to some extent, and according to Rawls an unequal distribution of these goods may be more just than an equal distribution. For an unequal distribution, say, where a doctor is given special training and privileges, may improve the lives of those who have fared worst in the natural and social "lottery". Such inequalities are justified by the Difference Principle not

because some men "merit" more of the social primary goods than others or because the total (or average share) of such goods would in the end be greater; but only because the least advantaged would benefit.

These complex principles, subject to further qualifications, are used in the derivation of other principles, notably principles of natural duty and a principle of fairness. Their implications for institutions in a democratic society and for the issue of civil disobedience and conscientious refusal are elaborated at length. The principles are justified in part by these implications.

II

What arguments can be advanced to support this conception of justice and on what grounds can its principles be justified? It is no surprise that the justifying grounds are at least as complex as the derivation of the principles. For Rawls takes the derivation of the principles to be part of their justification. Derived as they are from a hypothetical agreement among *equals* concerned to further their own ends, *the principles should be to the advantage of everyone but partial to none.* This ideal appears to capture the essence of our intuitive understanding of justice and seems bound to be reflected in the two main principles just in virtue of their method of derivation. The derivation might be regarded, then, as an extended elaboration and analysis of our most basic intuitive understanding of justice.

It would be a serious mistake, however, to take justice as fairness to be merely an explication of a concept of justice devoid of any substantive moral judgments. Much less is it an account of how terms like "just" and "fair" are used in ordinary language. The theory of justice as fairness, in the first place, derives the main principles of justice not from definitions of concepts or appeal to usage, but rather from reflection on the kind of social contract that would be accepted in the original position. Secondly, the theory appeals throughout to empirical principles in moral psychology and political economy. In contrast to contemporary "meta-ethical" theories, justice as fairness is, in this respect, in the tradition of Bentham, John Stuart Mill, and Hume. The aim is not to formulate a priori principles that hold in all logically possible words — as in the ethics of creation (159) — for such principles, even if they could

be comprehended, would be too weak to provide an adequate account of justice for this world.

The principles of justice cannot, of course, be deduced from empirical facts. Appeal to such facts is made instead to draw out the implications of the principles of justice as they apply to political and social institutions generally and to special circumstances. In this way the content of the principles is clarified and their implications tested against the considered moral judgments we are inclined to make at various levels of generality. What is desired, in Rawls' view, is a fit between these reflective moral judgments and the first principles derived from a conception of justice. When the fit is not perfect, we may modify our considered judgments to bring them in line with the first principles — or we may qualify these principles to make them accord with our present convictions. If these principles are modified, there obviously will be further implications which may or may not match our considered judgments. This process of self-examination is said to reach a state of *reflective equilibrium* when, after considering every alternative conception of justice and all relevant philosophical arguments, we arrive at a conception whose implications are in harmony with our stable moral convictions. It is with this ideal state in mind that Rawls undertakes in the middle chapters to detail the implications of justice as fairness.

A relevant argument against a conception of justice would be that humans cannot be expected to act on its principles even if they endorse them and intend to act on them. Such a conception would not be feasible, however admirable from an ideal standpoint. To deal with this question, Rawls introduces the notion of a *well-ordered society* in which "everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions satisfy and are known to satisfy these principles" (454). He argues by appeal to moral psychology that a well-ordered society in which the principles of justice as fairness are accepted would tend to generate a sense of justice that would normally win out against propensities toward injustice. He argues further that in such a society a person's conception of his own good will support and affirm his sense of justice. In this way justice is said to be *congruent* with goodness. The idea here is clearly not that justice and egoism go hand and hand, as if justice placed no limitation on the pursuit

of one's own ends. It is rather that justice as fairness is compatible with, and in part makes possible, the realization of one's own good.

The most important alternative conception of justice is utilitarianism. Deliberately ignoring recent refinements in its formulation, Rawls takes the classical doctrine, as found in Adam Smith, Bentham, and Sidgewick, to imply that "society is rightly ordered, and therefore just, when its major institutions are arranged so as to achieve the greatest net balance of satisfaction summed over all the individuals belonging to it" (22). The refinements are thought to obscure its fundamental divergence from justice as fairness. The latter is more likely to generate a strong sense of justice, partly because it is more congruent with our good and partly because of the greater clarity of its principles. Clarity suffers even in refined utilitarianism because the complexity of the calculations involved in applying the doctrine make it difficult to know its real implications. Unless an adequate understanding of inter-personal utility comparisons is made out (so far none is at hand), even the theoretical possibility of accurate calculations is in doubt. For these reasons too it will be difficult to have a well-ordered society where basic institutions are known to satisfy the utilitarian principle and persons can be counted on to fulfill their duties.

But Rawls' main objection to utilitarianism is that it allows the interests of some to be sacrificed to maximize the net balance of satisfactions. This injustice is compatible with Bentham's formula "everybody to count for one, nobody for more than one", since the formula implies only that no person's satisfactions should be given a special weight. The result is that some persons are treated as means only, and not as ends in themselves in the Kantian sense. Rawls considers the example (450) where the larger part of society has an abhorrence for certain religious or sexual practices. Suppose that the thought that these things are going on is enough to arouse the majority to extreme anger and hatred, even though no social injury has been done. In this situation the principle of utility can justify repressive measures which Rawls believes would not accord with our considered moral judgments. Similar examples are easy to produce; they appear to show that utilitarianism is not a

conception of justice in reflective equilibrium with our most stable moral convictions.

I shall now try to place this attack in perspective by considering the relation between the theories of right and good in justice as fairness.

III

Rawls is fully aware that to account for the primary goods without circularity he must develop a "thin" theory of goodness in which principles of right are not presupposed. For ethical principles about how institutions ought to be structured and how individuals ought to act — what Rawls calls principles of right — are to be established by assuming that each party in the original position is motivated to maximize his share of the primary goods. If the primary goods themselves are determined at least in part by an appeal to principles of right, Rawls would be moving in a vicious circle. A thin theory of goodness, which does not presuppose such principles, is needed therefore to account for the primary goods. Once accounted for, however, they may be used to arrive at principles of right, and these in turn may be employed, as in the last chapters, to construct a "full" theory of goodness to explain the moral virtues, the moral worth of persons, and related concepts. A thin theory is similarly required to argue for the "congruence" of justice and goodness. For if this argument is to be significant, it must do more than show that a sense of justice, defined by principles of right, is desirable from the standpoint of human values in which a sense of justice is already imbedded.

The price of avoiding circularity by the use of two theories of goodness is that the deontological character of justice as fairness is seriously weakened. Rawls defines a deontological (non-teleological) theory as one that "either does not specify the good independently from the right, or does not interpret the right as maximizing the good" (30). Justice as fairness, in contrast with classical utilitarianism, may appear to be deontological in both ways. As regards the first, Rawls maintains that "in justice as fairness the concept of right is prior to that of the good" (31, 396). This means:

> The principles of right, and so of justice, put limits on
> which satisfactions have value; they impose restric-

tions on what are reasonable conceptions of one's good. . .the interests requiring the violation of justice have no value. Having no merit in the first place, they cannot override its claims. (31)

Thus an "individual who finds that he enjoys seeing others in positions of lesser liberty understands he has no claim whatever to this enjoyment. The pleasure he takes in other's [sic] deprivations is wrong in itself" (31). By contrast, in "utilitarianism the satisfaction of any desire has some value which must be taken into account in deciding what is right. . .We are to arrange institutions so as to obtain the greatest sum of satisfactions; we ask no questions about their source or quality. . ." (30).

Clearly there is a problem here. It is only with respect to concepts and implications peculiar to the full theory of goodness that the concept of right is prior to that of the good. In the thin theory, which is to account for the primary goods, the concept of right cannot be prior on pain of circularity. Hence, when Rawls says that "the interests requiring the violation of justice have no value" we must interpret this to mean that such interests have no value in the full theory of goodness, although they *may* have value in the thin theory. But then justice as fairness is no more deontological than utilitarianism with respect to the priority of right. In both, a theory of goodness is developed *independently* of the concept of right, and principles of right are formulated in terms of this prior theory of goodness. Moreover, once these principles are at hand either theory may be used to argue that certain desires are wrong and that others ought to be developed instead. The utilitarian could argue, for example, that the desire to discriminate against another is wrong because it will tend to diminish the total net balance of satisfactions that is otherwise possible. Apparently, the only significant difference is that in justice as fairness a desire can be wrong *in itself* if its object is wrong according to a principle of right, whereas in utilitarianism a desire can be wrong only in view of the consequences of having it in a given situation. But this difference reflects the difference in the principles of right; it does not show that the principles of right somehow have a greater priority in justice as fairness.

The second way in which justice as fairness appears to be deontological is that the right is not interpreted as maximizing the good in the way that utilitarianism interprets it. For in justice as fairness the main principles explicitly preclude sacrificing the interests of one party to maximize the net balance of satisfactions on the whole. Notice, however, that the parties in the original position are to agree upon principles that maximize (with minimum risk) their individual shares of the primary goods. So justice as fairness cannot be claimed to be deontological on the ground that it is not in *any* sense a maximizing theory. What can be said is that in justice as fairness the maximizing character of the principles of right is limited by qualifications that are built into the formulation of the principles and that there are no parallel qualifications in classical utilitarianism. The fact that these qualifications are explicit in its principles of right makes justice as fairness more deontological in this respect than utilitarianism.

But now it must be asked whether this is a *substantive* difference: Is there a difference in the practical implications of the two theories for the structure of institutions and the conduct of individuals? This question cannot be answered without assuming some theory of goodness. Since utilitarianism as such is not tied to any one theory of goodness, suppose that it is formulated in terms of Rawls' thin theory. If, given this common theory of goodness and true empirical generalizations, the substantive implications of utilitarianism are even close to those of justice as fairness, then this "deontological" theory does not provide the kind of substantive alternative to utilitarianism that Rawls wants.

I shall not argue that Rawls has failed to construct such an alternative. I am not sure he has. But I shall discuss some important ways in which his account of justice as fairness leaves the issue unsettled. (1) It is not at all clear that a definite set of primary goods, much less the ones Rawls lists, can be derived from the thin theory of goodness. (2) Once derived, they may be used in utilitarianism to overcome some of Rawls' objections to utilitarianism regarding the possibility of interpersonal utility comparisons. (3) Given a plausible extension of Rawls' set of primary goods, each party in the original position could rationally choose utilitarian principles without fear that his own interests would be sacrificed on some occasions to benefit the

majority. To defend these points it is necessary to consider how the primary goods are supposed to be established within the thin theory.

IV

The primary goods are "things that every rational man is presumed to want" and "normally have use whatever a person's rational plan of life" (62). "They are things which it is supposed a rational man wants whatever else he wants. Regardless of what an individual's rational plans are in detail, it is assumed that there are various things which he would prefer more of rather than less" (92). "[R]ational plans . . . all require for their execution certain primary goods. . .whatever one's system of ends, primary goods are the necessary means" (93). (See also 174, 410-11.) But what is a rational plan of life? A person's plan of life or system of ends describes in a general way that he intends to do in his life. Although it is not a detailed blueprint for action stretching over the whole course of life, it "establishes the basic point of view from which all judgments of value relating to a particular person are to be made" (409). A rational plan is defined by two conditions:

> (1) It is one of the plans that is consistent with the principles of rational choice when these are applied to all the relevant features of his situation, and (2) it is that plan among those meeting this condition which would be chosen by him with full deliberative rationality, that is, with full awareness of the relevant facts and after a careful consideration of the consequences. (408)

The three relatively uncontroversial principles of rational choice which Rawls defines turn out not to be always sufficient to rank all the plans open to a given person (415), but when deliberative rationality is applied exactly one plan is understood to be selected as "the objectively rational plan" for the individual and this plan determines the individual's "real good" (417). Rawls tells us that we can think of a person as being happy (with certain qualifications) "when he is in the way of a successful execution (more or less) of a rational plan of life drawn up under (more or less) favorable conditions, and he is

reasonably confident that his plan can be carried through"
(409).

The primary goods are the necessary means to the
execution of anyone's rational plan of life in which lies his good
— indeed, his happiness. But does anyone know what his
rational plan is? Rawls concedes: "Often we do not know what
is the rational plan for us; the most that we can have is a
reasonable belief as to where our good lies, and sometimes we
can only conjecture" (417). Rawls is too optimistic. To know
one's rational plan requires exercising one's "full deliberative
rationality" and this demands perfect information about *all* the
relevant consequences of *all* the various courses of conduct
open to one (417). Furthermore, one can be under no
misconceptions as to what one really wants or will want, so that
when one achieves one's aim, there will never be any regret
about the course taken (417, 421). It is plain, I think, that no
person has reasonable beliefs about the nature of his rational
plan of life as long as this is, by definition, the plan a person
would choose with full deliberative rationality. But if this is so,
it is hard to see how one could ascertain which goods are
necessary for the success of *any* rational plan.

Here it will be objected that some goods, such as food and
shelter, necessary to sustain life itself, can be known to be
necessary for the success of any plan of life, even when nothing
is known about the content of these plans. This is true. But an
adequate theory of how institutions ought to be structured and
how individuals ought to act must do more than guarantee a fair
share of such basic goods. Presumably the parties in the original
position are motivated to maximize their share of the primary
goods not merely because they desire to live, but because they
want to be in a position to realize their conceptions of a good
life. The primary goods should include, therefore, *all* the goods
which are necessary for executing any rational plan.

The difficulty of deriving a definite set of primary goods is
complicated by two further considerations. First, happiness in
life is impossible unless one enjoys at least some of one's
activities.† But it is not clear that a rational plan of life, as
Rawls defines it, need include any enjoyment at all. It is
therefore unclear whether the primary goods necessary for
rational plans are those necessary for happy lives. Second,
which goods are necessary for either one will depend partly on

what institutions are assumed to exist already. Only very limited individual income and wealth, for example, may be necessary for happiness within some possible social and economic structures. But Rawls is in danger of arguing in a circle if he first assumes the existence of certain kinds of institutions to determine the primary goods and then uses the latter to determine what kinds of institutions ought to exist. If, on the other hand, no institutions are assumed to exist, reasonable beliefs about rational plans of life become even more difficult. For one must try to imagine what one's desires and needs would be apart from any social or economic context.

It is noteworthy that Rawls presents no argument to show that his list, or indeed any list, of primary goods can be derived within the thin theory of goodness. He apparently regards the derivation as evident to common sense (434). In any case, there is the following dilemma. Either some set of primary goods can be derived or not. If not, the two main principles, presupposing a definite set of primary goods, must be replaced, presumably by a more general principle by which all social goods are to be distributed equally unless an unequal distribution of any, or all, of these goods is to everyone's advantage (62). The social goods, as opposed to the natural ones, will be those interests and aims which are encouraged by each rational life plan and controlled by the basic structure of society. But it is entirely unclear, of course, what the practical implications of such a principle would be compared with those of utilitarianism formulated with respect to the same interests and aims. This is unclear not only because we do not know what these interests and aims are (apart from the bare necessities of life), but also because the determination of an equal distribution, not to say a comparison of the advantages of an equal with an unequal distribution, requires the sort of interpersonal utility comparisons that are so troublesome for the utilitarian.

If, to take the other horn, we assume that some set of primary goods can be derived, it is open to the utilitarian to interpret his first principle of right as maximizing these goods. He thus avoids many of the problems of interpersonal utility comparisons. Such comparisons would not be eliminated altogether in either theory. Utilitarianism, moreover, lacks the advantage of the Difference Principle. For once the representative least-advantaged man is located, application of this

principle requires only ordinal comparisons among combina-
tions of primary goods. Still, a substantial difficulty in the way
of an adequate understanding of the implications of utilitari-
anism would be reduced if a complete set of primary goods
could be determined.

It might be thought, however, that given almost any set of
primary goods and a method of comparing them, the
distributions dictated by utilitarianism are bound to be unfair in
many cases. For in maximizing the total (or average share) of
primary goods possessed by everyone, the interests of some
persons would be sacrificed unfairly to benefit the majority. Let
us suppose with Rawls that the chief social primary goods are
self-respect, rights and liberties, powers and opportunities,
income and wealth. Self-respect occupies a central place among
these goods (62, 67, 178ff, 440ff). Rawls says: "It is clearly
rational for men to secure their self-respect. A sense of their
own worth is necessary if they are to pursue their conception of
the good with zest and to delight in its fulfillment" (178).
Rawls believes that the public recognition of utilitarianism
entails some loss of self-esteem (181). But he also says that
"those who respect themselves are more likely to respect each
other and conversely. Self-contempt leads to contempt of
others and threatens their good as much as envy does" (179).
The latter is surely plausible at the level of common-sense
empiricism and suggests that respect for others, especially an
appreciation of their worth as equal to one's own, is also a
primary good. Indeed, respect for and appreciation of others
must be a primary good if we can assume that rational plans will
include some form of social cooperation among individuals who
recognize their needs as equally worthy of fulfillment. Now it
must be asked whether maximizing the total (or average) of
primary goods will lead to sacrifices by minorities. It is not
obvious that it would, especially if respect for others is given
priority in utility comparisons. For unfair distributions, at least
in a well-ordered society, will automatically prevent the
advantaged from sharing this primary good, and probably the
disadvantaged too.

V

Should there be little difference in the practical import of

utilitarianism and justice as fairness it would not follow from this alone that utilitarianism is an adequate theory or that justice as fairness is not. What would be true is that Rawls has definitely misconceived the nature of justice as fairness, but I have not argued for this conclusion either. Instead I have expressed reservations about the supposed opposition between the theories. This issue cannot be settled, it appears, until the underlying theories of goodness are clarified. Only then will the actual implications of utilitarianism be clear enough for us to make a choice, if there is one to be made.

Richmond Campbell,
Department of Philosophy,
Dalhousie University.