Two interpretations of the difference principle in Rawls's theory of justice

by

PRAKASH P. SHENOY AND REX MARTIN (University of Kansas)

1. Introduction

THE PRINCIPAL SUBJECT of justice is the basic structure of a society, that is, the arrangement of major social institutions with regard to the way these institutions determine the division of advantages from social cooperation and, in particular, with regard to the distribution of social primary goods.

For simplicity, [Rawls says, let us] assume that the chief primary goods at the disposition of society are ... liberties, powers and opportunities, income and wealth. [Later on he adds self-respect.] These [Rawls continues] are the social primary goods. Other primary goods such as health and vigor, intelligence and imagination, are natural goods; although their possession is influenced by the basic structure [of society] they are not so directly under its control.¹

The ideas contained in this paper germinated during a series of discussion sessions on Rawls's *Theory of justice* with Joseph Pichler during the Spring of 1979. We are also indebted, for helpful comments, to Norman Bowie, Richard De George, Elizabeth Miller, John Rawls, Amartya Sen, and Michael Young.

Some of the leading ideas in the present paper, in particular the formalism, have been set out in shorter compass in Shenoy (1980). The development of the difference principle in the original position and, especially, its application to a going economic and social system have been discussed in Martin (1981). The appendix is the work of Shenoy.

This investigation was supported in part by University of Kansas General Research allocations No. 3478-XO-0038 and No. 3058-20-0038 and by the School of Business Research Fund provided by the Fourth National Bank and Trust Company, Wichita.

¹ John Rawls, A theory of justice. Cambridge, MA: Harvard University Press, 1971, p. 62; see also p. 303. (Hereafter: Rawls 1971, p. 62; see also p. 303.) A citation note listing Rawls's relevant writings, including all things cited in this paper, and giving full bibliographical data will be found at the end of this paper.

The Rawlsian problem of justice is to determine a basic structure that leads to a proper or just distribution of *social* primary goods.

To do this we need a fair procedure (for there are no antecedent or independent standards for such a distribution). Accordingly, we create a hypothetical bargaining situation (called 'the original position' by Rawls) in which certain significant constraints operate: these constraints include those required to discount or bracket off all special, peculiarly personal, or circumstantial facts and biases; those constraints embedded in objective circumstances (such as relative scarcity) or in our psychological orientation (such as mutual disinterest in one another's life plans); and as well a number of other presumed constraints (such as that the principles agreed to are to be public, that they constitue the ultimate or foundational standard, that the principles are to be chosen once and for all, that they are to be chosen unanimously, and so on). Then, whatever principles emerge from the original position, they are the principles of justice for the organization of the basic structure of society.

Rawls contends, further, that under the constraints involved in the original position (constraints imposed in order to make for a fair and realistic bargaining situation) it would be reasonable to define two preferred principles for the distribution of social primary goods. The principles chosen would be the Rawlsian two principles, in the following order:

- 1. Liberties are to be arranged so as to achieve the most extensive justifiable set (or aggregate) of equal basic liberties. The liberties envisioned are liberties of the person (speech, conscience, association) and political liberties (right to vote, etc.).
- 2. Social and economic inequalities are to be arranged so that they (a) are structured by social roles and offices which are open to all under conditions of 'fair equality of opportunity' and such inequalities in wealth, income, and position (b) serve to improve—ideally, to maximize—the life situation of the least advantaged group.³

² Rawls's own summary is given in 1971, pp. 146-147; see also pp. 126-127, 137.

³ See Rawls 1971, pp. 60, 83, 250, and especially 302-303 for his various statements of the principles. Rawls normally calls Principle 1 the Principle of Equal Liberty,

In this paper, we shall focus on the difference principle (principle 2b). This principle is examined in detail and alternate formulations of it are provided. Although the discussion here is limited to the difference principle, all claims made in this paper about the difference principle are subject to the assumption that the first principle (equal liberty) and the principle of fair equality of opportunity are valid and hold lexically prior to the difference principle. In the next section, the difference principle is discussed as a 'maximin principle'—the usual characterization of it in Rawls's own discussions.4 This also is the way in which many scholars have viewed the principle and on the basis of which it has been criticized and even rejected.⁵ In section 3, we present another way of formulating the difference principle, using the concepts of pareto efficiency and egalitarianism. However, this formulation of the difference principle is shown to be equivalent in effect to Rawls's statement of the difference principle. More precisely, in the case where society can be classified into two economic classes, our suggested version of the difference principle is shown to be exactly equivalent to Rawls's standard version of the difference principle. In the general case

Principle 2a the Principle of Fair Equality of Opportunity, and Principle 2b the Difference Principle. The Rawlsian two principles, when emerging as a set from the original position, will have what Rawls calls a 'lexical' order (see 1971, pp. 42-44)—that is, a serial order of importance: 1 > 2a > 2b.

In case of nonuniqueness (i.e., the maximum expectations of the least advantaged are not affected one way or the other by some changes in the expectations of the best-off although these changes benefit others), Rawls expresses a more general principle that he calls the *lexical difference principle*, as follows: in a basic structure with n relevant representatives, first maximize the welfare of the worst-off representative man; second, for equal-that is, unaffected—welfare of the worst-off representative, maximize the welfare of the second worst-off representative man, and so on until the last case; here, for equal welfare of all the preceding n-1 representatives, maximize the welfare of the best-off representative man. (See Rawls 1971, pp. 81–83, especially p. 83.)

⁴ See Rawls 1971, pp. 76-80 and 152-161; for his explicit description of it as a maximin principle, see, e.g., 1974a, pp. 639ff, 1974b, especially, pp. 141-142, and 1978, p. 69 (n. 4).

⁵ Examples of such criticisms are Arrow (1973), Harsanyi (1975), Nozick (1974, pp. 189–197), Rae (1979), and Sen (1970, pp. 135–141). For general discussion and a survey of the literature see Kaye (1980).

where society can be classified into n economic classes (where n > 2), the equivalence of the two versions of the difference principle is shown assuming that a regularity condition called chain connection holds. (The formal algebraic proofs of all the assertions made in section 3 are relegated to the Appendix.) The arguments, in the original position, for the pareto efficient-egalitarian version of the difference principle will be taken up in section 4. Finally, the many advantages of our restatement of the difference principle are discussed in section 5.

2. The difference principle

The difference principle as formulated by Rawls states that social and economic inequalities are to be arranged so that they are to the greatest benefit to the least advantaged. To illustrate this principle let us consider a society that can be classified into two groups of individuals, a more favored group and the less advantaged group. Consider two individuals: the first, a representative of the more favored group, and the second, a representative of the less advantaged group. We can represent their possible expectations under various basic structure alternatives by means of an attainable set A in a two-dimensional space. That is, the attainable set A consists of all possible (practicable, feasible) pairs of expectations $x = (x_1, x_2)$ where x_1 is a real-valued aggregated index of social primary goods of the more favored representative and x_2 is the corresponding index of the less advantaged representative. Such an index, then, indicates the distribution of income, wealth, and position that the representative person would receive, or could reasonably be expected to receive, over his lifetime under some feasible basic structure arrangement. But to avoid clumsy language we will normally refer to (x_1, x_2) simply as a distribution of income, wealth, etc. Since we are concerned with the Rawlsian difference principle we, of course, exclude from the distribution space all pairs of expectations, arising from basic structure alternatives, that fail to satisfy the prior demands in justice for equal liberty and for fair equality of opportunity. In this two dimensional space, we can draw indifference curves, each one of which is a locus of distributions of income, wealth, and

position (x_1, x_2) that satisfy the relevant criterion of justice (the difference principle in the case at hand) to the same extent; the distributions on any one indifference curve satisfy the governing principle of justice to a lesser extent than the distributions on all the *higher* indifference curves. We realize, of course, that different criteria of justice would give different indifference curves. Under the difference principle, the indifference curves are as shown in figure 1 corresponding to the criterion of justice $\max\{\min\{x_1, x_2\}\}$ with higher indifference curves in the direction of the arrow. Note that two distributions would satisfy the above criterion to the same extent if the minimum expectation in each distribution is the same.

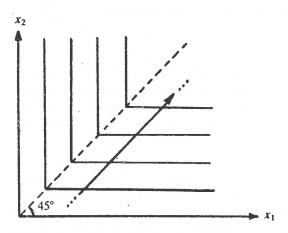


Figure 1. Indifference curves under the difference principle using $\max\{\min\{x_1, x_2\}\}$ as a criterion of justice.

Now, given the attainable set A and confining it to the distribution space marked out by the satisfaction of prior demands of justice, we can pick the distribution favored by this particular criterion of justice by selecting a distribution in the attainable set that lies on the highest indifference curve. For example, if the attainable set is as shown in figure 2 (the region on or below the curve oc), the difference principle would favor the distribution a.

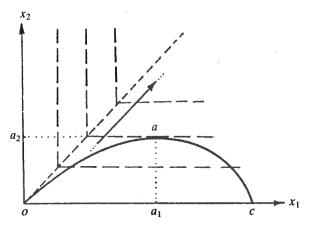


Figure 2. The attainable set and the difference principle.

The point o (the origin) represents the hypothetical state in which all primary goods are distributed equally and the point a is the distribution where the expectation of the least advantaged representative (min $\{x_1, x_2\} = x_2$) is maximized. Hence the difference principle is often referred to as a maximin principle.

One final point. Justice is, or should be, a virtue of society, specifically of its 'basic structure.' The problem for Rawls, then, is to choose a basic structure and not simply a favored distribution to individuals. Each basic structure alternative (among the set of all feasible basic structures satisfying principles 1 and 2a) gives rise to a particular distribution of income, wealth, and social position. Any one such arrangement of the basic structure can thus be mapped onto a single point in the distribution space; this point represents the distribution of economic and social primary goods to representative individuals under a particular basic structure. Point a, then, represents a distribution of income, wealth, etc., to such represent-

⁶ Both figures 1 and 2 appear in Rawls 1971, p. 76. Indeed, much of the material in this section is taken directly from Rawls 1971, pp. 76–80. It is presented here in detail, especially the discussion of the attainable set and of indifference curves, partly for completeness and partly to facilitate the exposition of pareto efficiency and egalitarianism which follows in section 3.

ative individuals under a basic structure arrangement that is just by Rawlsian principles (in particular, principle 2b).⁷

3. Pareto efficiency and egalitarianism

We will now attempt to provide an alternate characterization of the difference principle in terms of pareto efficiency (or optimality) and egalitarianism. Let us start by defining the concept of a pareto efficient arrangement of the basic structure. We say that a distribution of social primary goods (specifically, income, wealth, and social position) is pareto efficient relative to a set of distributions if and only if there does not exist another distribution in this set that makes some (at least one) representative person better off without at the same time making some (at least one) other representative person worse off.8 Also we shall say that a basic structure arrangement is pareto efficient if it gives rise to a pareto efficient distribution. (See Rawls 1971, p. 70; also 1968, pp. 56-57.) For a two-group society, if the attainable set is the region on or below the curve oc (as indicated in figure 2 and also in figure 3), a pareto efficient arrangement of the basic structure would lead to a distribution which lies on the north-east boundary of the attainable set, indicated by the curve ac in figure 3.

Thus there are several arrangements of the basic structure that are pareto efficient. Pareto efficiency by itself is insufficient to single out one of these distributions as *the* just distribution.

⁷ Included in the basic structure would be a society's political system and its economic system. Each of these in turn would be made up of a set of structural elements or institutions. We would require in the political sphere, for instance, a constitution, a form of government, modes of election. And in the economic, some form(s) of ownership of the means of production, specific sectors for determining capital investments, devices to provide 'public goods' and to deal with externalities, and so on. (See Rawls 1977, 1978, for elaboration of his theory of the basic structure. See Martin 1981, sects. 2 and 4, for the economic institutions, in particular, in a Rawlsian basic structure.)

⁸ This definition of efficiency derives from Vilfredo Pareto (see Pareto 1971, chap. 6, sects. 32–64, pp. 261–269; note especially sect. 33, p. 261, and Appendix, sect. 89, pp. 451–452). Since we are using *pareto* as a technical term, rather than as a proper name, we will employ the lower-case p throughout.

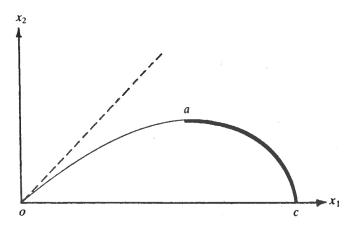


Figure 3. The pareto efficient distributions indicated by the curve ac.

In defining justice as fairness, it seems natural to appeal to egalitarianism to supplement the principle of efficiency. The principle of egalitarianism would pick a distribution that is the 'more nearly equal' or the 'least unequal.' For example, in our two-group society, the principle of egalitarianism would favor the distribution (x_1, x_2) that minimizes the difference between the expectation of the most favored representative and the expectation of the least advantaged representative, i.e., minimize $\{x_1-x_2\}$. The indifference curves under this criterion are shown in figure 4 where the arrow indicates the direction of the higher indifference curves. The indifference curve—for it is the line that marks out a strictly equal distribution between the two representative persons.

In the general case where society is classified into n groups according to economic well-being, there are correspondingly many ways one can define egalitarianism. We will define egalitarianism here as minimizing the difference between the expectations of the most favored representative and the least advantaged representative, i.e., minimizing the difference x_1-x_n where x_1 is an expectation of the most favored representative, x_2 is the corresponding expectation of the next most favored representative, and so on until

 x_n is the corresponding expectation of the least advantaged representative.

If the attainable set is the region on or below the curve oc (as indicated in figure 2 and again in figure 4), then the principle of strict egalitarianism would pick the distribution o. However, if we make the principle of pareto efficiency lexically prior to the principle of egalitarianism, this combination of the principles would result in the distribution a (from among the points on the pareto curve ac) because picking a distribution less unequal than a would violate the priority of the principle of pareto efficiency (see figure 4).

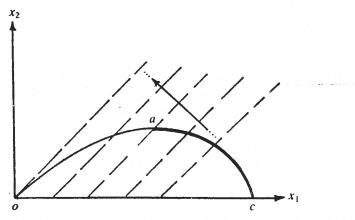


Figure 4. The most egalitarian pareto efficient ditribution in the attainable set.

⁹ Our suggested definition would, we think, be acceptable to Rawls. For he characteristically takes the well-being of the best-off class as the standard against which the well-being of the other classes, including the least well-off is measured (see the graphs in Ráwls 1971, p. 81), while at the same time stressing the importance of minimizing the (necessary) inequality in a society (see Rawls 1974a, p. 647, and 1974b, p. 145). Interestingly, Nozick too would measure the relevant inequality as that which holds between the best-off and the least well-off classes (see Nozick 1974, pp. 210–212). So, our definition is well motivated in light of the existing literature and is not at all eccentric.

For purposes of completeness, we add the following to our definition. In case of nonuniqueness, i.e., where there are two or more distributions with the same minimum inequality $x_1 - x_n$, we will state a more general egalitarian principle that we call the *lexical egalitarian principle*, as follows: in a basic structure with n relevant

Thus, at least in the two-group society, pareto efficiency and egalitarianism (with the former lexically prior to the latter) would result in the same distribution as the one that would result by maximizing the expectation of the least advantaged. This is not a coincidence and can be proven always to be true, as we show in the Appendix.

Rawls himself has noted that, where there are only two relevant classes, the difference principle (or 'maximin,' as he calls it there) 'selects the (Pareto) efficient point closest to equality.' He continues, 'Thus, in this instance at least, [the difference principle] has another interpretation . . .' (1974a, p. 648). But he adds, 'I do not know, however, whether the focal point can be defined sufficiently clearly to sustain the second interpretation . . . when there are three or more relevant classes' (1974a, p. 648, n. 7). It is to this particular problem that we now turn.

In the general case where society is classified into n groups, n > 2, the situation naturally becomes a bit more complicated. We can simplify the matter by saying that if the distributions in the attainable set exhibit a condition that Rawls calls 'chain connection' then, again, pareto efficiency and egalitarianism will lead to the same result as that obtained under the maximin version. (See the Appendix for a mathematical proof of this assertion.)

According to Rawls, chain connection means that whenever the expectation of the least advantaged group is increasing (as a result of increasing the expectation of the most favored group), the expectations of all the other intermediate groups are also increasing.¹¹

representatives, first minimize the inequality $x_1 - x_n$; second, for equal value of the difference $x_1 - x_n$, minimize the inequality $x_1 - x_{n-1}$, and so on, until the last case which is for equal value of all the preceding n-2 differences x_1-x_n , x_1-x_{n-1} , ... x_1-x_n , minimize the inequality x_1-x_2 .

¹⁰ Rawls 1974a, p. 648. Here (p. 648n) he cites Phelps (1973), pp. 334–337, and Atkinson (1973), pp. 105–108; but neither one has, in our judgment, stated the point unequivocally.

¹¹ See Rawls 1971, pp. 81-83, for a discussion of chain connection. Arrow (1973, p. 252) severely criticizes this condition claiming that '.. on the face of it, it seems clearly false.' However, Arrow's criticism is based on a misunderstanding of the chain connection condition (he also confuses chain connection with close-knittedness, an assumption tht Rawls makes to avoid nonuniqueness under the maximin criterion). Chain connection does not imply that as we raise the expectations of the

Interestingly, the coincidence of results which we have established as holding between the maximin criterion and that of pareto efficiency and egalitarianism, under the condition of chain connection, is not duplicated for the difference principle and utilitarian principles under that same condition. In figure 5, although chain connection holds, the difference principle, in its maximin version, selects distribution a whereas the principle of average utility and that of the sum of utilities would both select distribution b.

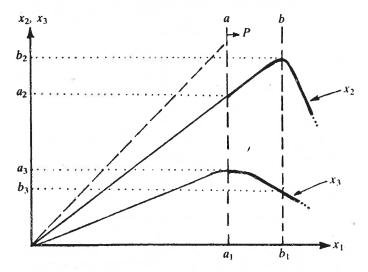


Figure 5. An attainable set that exhibits chain connection but where the maximin principle differs from utilitarianism. 12

more advantaged the situation of the worst off is continuously improved all the time. In fact, Rawls (1971, p. 158) is careful to qualify this statement with the statement that 'each such increase is in the latter's interest, up to a certain point anyway' (emphasis added). Hence, Arrow's example of reduction of the income tax for high brackets and, simultaneously, reduction of welfare payments is not an example against the chain condition.

¹² Rawls's comment that '... given these special assumptions, the difference principle has the same practical consequences as the [principle] of average utility ...' (1971, p. 82) and Arrow's concurrence 'if it [chain connection] holds, there is no difference in policy implication between the maximin principle and the sum of utilities ...' (1973, p. 252) are both clearly wrong, as is indicated in figure 5. Perhaps both Rawls and Arrow intended to say only that within the region of positive

It should be noted that Rawls claims (1971, p. 82) that the difference principle is not contingent on the chain connection condition being valid. So where chain connection does not hold, it is then possible that pareto efficiency and egalitarianism could lead to an outcome that is different from the outcome preferred by the maximin version of the difference principle. For example, in figure 6 where chain connection does not hold, pareto efficiency and egalitarianism would select distribution b whereas the maximin version of the difference principle would select distribution a.

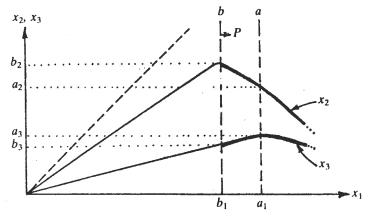


Figure 6. An attainable set that does not exhibit chain connection and where the maximin principle differs from pareto efficiency and egalitarianism.

But, clearly, Rawls also believes that the normal situation in which the difference principle is to be applied is one in which chain connection does hold. In particular, chain connection can be expected when the other principles of justice—equal basic liberty and fair equality of opportunity—are fulfilled (see Rawls 1971, p. 82). We can, accordingly, regard cases where chain connection does not hold as unusual ones and the application of the difference principle in such cases as, in a sense, problematic. And we believe that Rawls

contributions (i.e., the region to the left of a in figure 5) there is a coincidence between Rawls's maximin criterion and utilitarianism.

would be reluctant to see the difference principle applied in precisely those situations.

Indeed, he stresses that a well-ordered society should 'operate only on the upward rising part of the contribution curve (including of course the maximum)' (Rawls 1971, p. 104). Or, as he puts the same point later, it is 'desirable when the justice of the basic structure is involved' that 'we should stay in the region of positive contributions' (Rawls 1971, p. 105; see also pp. 79, 81–82, 585). This particular emphasis is wholly in line with Rawls's representation of the difference principle as one of 'reciprocal advantage' (see 1971, pp. 177–179; also p. 510).

In short, chain connection does hold in a just basic structure. Thus, Rawls is committed to chain connection and constrains his maximin version of the difference principle by it in a well-ordered society. Hence, our use of chain connection in a reformulation of that principle is wholly legitimate insofar as we are attempting to explicate the Rawlsian difference principle.¹³

We now propose (and shall argue shortly) that the combination of pareto efficiency and egalitarianism is more plausible than the maximin version of the difference principle and that the second principle of justice should be reformulated as follows.

Second Principle: Social and economic institutions in the basic structure are to be arranged so that (a) the offices and positions are open to all under conditions of fair equality of opportunity, (b) the resulting distribution of economic primary goods is pareto efficient, and (c) the inequality between the most favored and the least advantaged is minimized.

(Regarding priorities, fair equality of opportunity is lexically prior to pareto efficiency which in turn is lexically prior to egalitarianism.)¹⁴

¹³ See the Appendix for a rigorous definition of chain connection. We shall not, however, attempt a proof of chain connection since that would be unnecessary for our purposes in the present paper.

¹⁴ This statement is to be taken as an initial formulation. At the end of section 4 we introduce the notion of relative pareto efficiency. The statement here, then, would have to be revised to include that notion. See n. 18 (below) for our suggested reformulation.

4. Pareto efficiency, egalitarianism and justice

In this analysis we have assumed that there will be several pareto efficient (or optimal) points on the boundary of the attainable set A. But only one of these points can qualify as 'more nearly equal' (that is, as minimizing the difference x_1-x_n). This unique pareto efficient-egalitarian point marks the criterion for distribution under the difference principle in the version we have been developing. In the present section we will attempt to show the peculiar relevance of the pareto efficient-egalitarian criterion to Rawls's theory of justice. We will begin by recounting the main considerations developed in the original position that tend to favor the difference principle. Then we will show how these considerations lead to the pareto efficient-egalitarian criterion we have just sketched.

We start with the fact that people have different natural endowments and are born into and grow up in different social circumstances. No one can be said to deserve or be responsible for these factors in his own case. Nonetheless, such factors affect a person's life prospects, advantageously for some and disadvantageously for others. And these particular end results cannot be said to be a deserved or fair return either.

Now, though we can reduce the differences of people in respect of natural endowment and social circumstance, we can never really nullify or remove them. Nor should we attempt to do so. Rather, some further step—on down the line-will be in order

Since these initial differences are both arbitrary and undeserved on the part of the individual, each person can agree that the natural endowment of each, that is, the use of the assets each person has in the circumstances described, can and should benefit everyone. This might not be a view we could expect individuals to take in the everyday world; it is a view that it would be reasonable to take in what Rawls calls the 'original position,' a conjectured arena for discussion and decision about principles of justice in which all individuals, by hypothesis, are ignorant of a vast number of particular facts pertaining to themselves and others, in particular, such things as their natural endowments and actual social circumstances.

In short, society can be arranged (and should be, from the

perspective of the original position) so that no one is hurt and no one unduly helped by his own 'luck' in the 'natural lottery' (as measured by his initial 'draw' of natural endowments and social circumstances). Rather, all are to use their natural assets and to exploit their social circumstances not merely for their own advantage but also for the common benefit.¹⁵

The idea of collective asset, then, is the Rawlsian solution to the problem posed by differential starting points and by the fact that these initial differences can never be reduced to allow for more than a passable measure of equal opportunity. For if we can assure that everyone will be made better off, we have not abandoned anyone to luck and we have taken a significant step toward reducing the initial inequality of people to a tolerable and manageable level. And it is this intuition that the difference principle responds to.

What we want to see next is how these considerations, in particular the collective asset idea, would actually work as an argument for the difference principle—that is, as an argument for the version we have identified, where the criterion for the distribution of income, wealth, and social position to representative individuals from various groups is marked out by (1) that pareto optimal point which is (2) the most egalitarian one. We shall take up these two points in turn.

We will assume from the beginning that the persons in the deliberative group have already accepted the principle of equal liberty and the principle of fair equality of opportunity (with the lexical priority of the former over the latter). In short, we have an attainable set A in which distributions are constrained by the acceptance of these prior demands of justice.

Now consider a simple problem where an individual in the original position is asked to choose between exactly two basic structure alternatives labelled c and d such that d is not pareto efficient relative to c; i.e., alternative c results in a distribution that is at least as good as the distribution resulting from alternative d for each and every economic group, and c results in a strictly better outcome for

¹⁵ The idea of regarding 'the distribution of natural abilities as a collective asset' is from Rawls 1971, p. 179. See also 1967, p. 68; 1971, pp. 101–102, 523; 1980b, p. 551.

at least one economic group as compared to d. Assume that both c and d satisfy the principle of equal liberty and the principle of fair equality of opportunity. We shall assert that regardless of the other attributes of alternatives c and d, a person in the original position will prefer c to d. The reasoning is quite obvious. Alternative c results in an outcome that is at least as good as the outcome resulting from alternative d for each and every segment of society. Furthermore, c is actually better than d for at least one economic group. Then, clearly, one should not choose alternative d. ¹⁶

This argument for not preferring d to c is quite strong and holds even if more alternatives are available. Thus, we can assert that an individual in the original position will never choose a basic structure alternative that results in a distribution of economic primary goods that is not pareto efficient relative to other feasible basic structure alternatives. The reason for this would appear to be given in the set of considerations which led up to the adoption of the collective asset idea in the original position. The idea states that inequalities are justified only if certain conditions are met, among them the condition of mutual benefit. This condition can be met in the case of choosing between a pareto efficient outcome and one that is not (but cannot be met in choosing between two different pareto efficient points—as, for example, in moving from one pareto efficient point to another). It follows, then, that inequality—the perpetuation or increase of inequality—can be justified only in the former sort of case. Thus, we restrict ourselves to such cases.

Nonetheless, there may well be several arrangements of the basic structure that are pareto efficient. But, clearly, it would not be reasonable, from the perspective of the original position where all participants have an equal status, to prefer a pareto efficient out-

¹⁶ See Rawls 1968, p. 57. The principle relied on here is a special case of the one implicated in the idea of collective asset—the principle of mutual benefit. To bridge from mutual benefit to the present principle one would have to resort to a scheme of lexical ordering (analogous to the one employed in note 3) in which (i) the situation in which each group is benefited is preferred to (ii) the case where one group has its situation unchanged and all the other groups improve their situation, etc., on down to (n) the situation in which only one group improves its situation and all the others have theirs unchanged. Our point, then, is that (n) would be preferred to a situation of no improvement at all.

come that allowed for greater inequality than did another eligible pareto efficient outcome. Therefore, that pareto optimal outcome which minimizes inequality (that is, minimizes x_1-x_n) is selected. To do otherwise would be to opt for a surplus of inequality, that is, an inequality without compensating benefit(s). Such a choice would perpetuate or increase inequality unnecessarily. Any such choice would be unreasonable and unjustifiable. Thus, individuals in the original position will settle on the most nearly egalitarian pareto efficient basic structure alternative.

One problem remains however. For it is not clear how the principle we have identified would govern the choice between non-pareto outcomes—the choice of options that stayed strictly within the nonpareto zone. It seems to us, though, that the argument already developed would cover these cases. (See Martin 1985 chap. 5, for an elaboration of the argument.)

The choice problem we faced earlier can be generalized. Initially it was the problem of moving from some nonpareto-efficient point to any of several pareto efficient points; the choice problem in that case was solved by selecting that one pareto efficient point which is most egalitarian. By the same token, when one moves from some nonpareto-efficient point, the preferred choice between *relative* pareto efficient points (where yet further improvement, beyond them, is still possible) is that one which is more egalitarian. The number of relative pareto efficient points could be increased, but so long as the choice in the nonpareto zone assumes—with respect to some fixed point—a finite set of discrete options, all of them relatively pareto efficient, then one and only one of these points can be selected as the most egalitarian one. Thus, the principle we have formulated can be extended to choices within the nonpareto zone, with only a slight change in wording. ¹⁸

¹⁷ Rawls's clearest statement of this line of argument, stressing the equal status of participants in the original position, is found in 1980a, pp. 205–206 in particular. See also Rawls 1971, p. 547; 1974a, p. 637; 1974b, pp. 142–143; and 1980b, pp. 529, 532, 546–550.

¹⁸ Thus, the (b) part of our principle as stated at the end of section 3 would now read: 'the resulting distribution of economic primary goods is pareto efficient (either absolutely or relatively).'

^{9 -} Theoria 3:1983

The solution we have hit upon, while verbally different from his maximin criterion, is one that Rawls could accept. In the next section we will examine the advantages of our formulation over the maximin version of the difference principle that Rawls has provided.

5. Pareto efficiency and egalitarianism versus the maximin version of the difference principle

One way of viewing our reformulation is to regard it as another characterization of the difference principle. Our reformulation (in terms of pareto efficiency and egalitarianism) can hardly be considered a radical departure from Rawls's usual version. For the operational equivalence of these two formulations is demonstrated in the Appendix (subject to the chain connection condition in the general case of n groups in society). The main difference between them, then, must come at another point.

To bring this out, let us distinguish a criterion for distribution (e.g., Rawls's maximin version of the difference principle) from an argument for that criterion. (Such a distinction is carefully drawn by Rawls; see 1974b, p. 141.) Many people have objected to Rawls's characterization of the difference principle because it appeared to rest on a particular argument, specifically, that a rational individual in the original position would use the maximin rule for decision making under uncertainty. 'The maximin rule tells us to rank alternatives by their worst possible outcomes: we are to adopt the alternative, the worst outcome of which is superior to the worst outcomes of the others' (Rawls 1971, pp. 152-153). In the course of conducting his maximin argument Rawls rules out the use of objective probability values in the original position, decrees that it would be unreasonable to arbitrarily assign equal probability values to each possible life situation under the indeterminate conditions prevailing behind the veil of ignorance, assumes (implicitly) that individuals in the original position are extremely risk-averse, and so on. 19

¹⁹ See Rawls 1971, pp. 150–161, especially pp. 153–154. The main arguments against computing risks and, as well, against using a principle of insufficient reason (to assign equal probability values to possible cases) is found in Rawls 1971, pp. 165, 167–173,

This argument is perhaps the weakest link in his chain of reasoning that utilitarians such as Arrow (1973) and Harsanyi (1975) have attacked in defending their own—utilitarian—doctrine.²⁰

In sum, these critics have reasoned that, since the maximin argument was inherently suspect or since it did not apply under the conditions envisioned, or was not uniquely applicable, the criterion supported by that argument was itself defective or precarious. Thus, doubts about the maximin *argument* tended to translate into doubts about the maximin *criterion*, that is, into doubts about Rawls's favored way of characterizing the difference principle.

We will not say that Rawls's preferred criterion for distribution—his favored version of the difference principle—requires the support of the maximin argument. We will say only that it is difficult to see how this particular formulation of the criterion of distribution would have been chosen except in the context of the maximin argument.²¹ The formulation we have hit upon sets a different criterion for distribution (that of pareto efficiency constrained by egalitarianism). More important, a different argument—that of mutual benefit as implicated in the idea of collective asset—is used to

^{175;} see also 1974a, p. 649. Rawls also argues that individuals, behind the veil of ignorance, are unaware of their *personal* propensities for risk-taking or -aversion. (See 1971, p. 172; also pp. 152–157, 168, 530; and also Rawls 1974a, pp. 649–650, 653; 1974b, p. 143.) What he contends instead is that *all* individuals, as a matter of rationality, would be extremely risk-averse under these conditions of extreme uncertainty (see especially Rawls 1971, p. 172).

²⁰ See also Sen (1970, pp. 139–141) on this point. He writes, the link between the concept of 'fairness' and the two principles of 'justice' that identify the maximin rule lies in the belief that in a 'fair' agreement these two principles will be chosen. Is this argument acceptable? ... Rawls's maximin solution is a very special one and the assertion that it must be chosen in the original position is not altogether convincing. ²¹ It has been argued, for example, by Kaye (1980, pp. 34, 41) and Ihara (1981), that other rational decision strategies—such as minimax regret (see Baumol 1977, pp. 460–466, especially 463–466)—would support Rawls's difference principle and that those other strategies would have as well the effect of ruling out utilitarian criteria as viable alternatives to that principle. Nonetheless, our point is that if the maximin strategy were thought to be unsound or inapplicable then both Rawls (at least in his book) and his critics would regard the difference principle in its maximin version as unsupported.

support this criterion; and this argument, while itself Rawlsian, does not draw, directly or indirectly, on the maximin argument.

We would stress, parenthetically, that the main use of the maximin argument in Rawls's hands is to rule out or to constrain utilitarian alternatives to his preferred two principles of justice. It is important in this regard to note that Rawls intends the maximin strategy to support the *two* principles *per se* against utilitarianism; the strategy is not specifically intended to support the difference principle (see Rawls 1971, pp. 152–157).

Interestingly, in some of his later writings, Rawls suggests a two-stage argument for the two principles (see 1974a, sections 3 and 5 especially and also p. 652; 1974b, pp. 142–143). In the first stage, the two principles are contrasted, as a set, to a utilitarian alternative (say, that of maximizing average utility). Here Rawls alleges that the two principles will be preferred because they protect the equal basic liberties and fair equality of opportunity from utilitarian social calculations in which the well-being of some persons, in these respects, might be sacrificed in the interest of maximizing average utility. It is principally this notion, that citizens want to protect liberty and opportunity against risk, that Rawls tries to capture with his maximin strategy. Thus, maximin rules out utilitarian alternatives because they threaten, or fail to guarantee, these particular social primary goods.

Now we move to the second stage. Here principles 1 (equal basic liberties) and 2a (fair equality of opportunity) are taken as accepted across the board. So the competition between the two principles and utilitarianism is reduced to the consideration whether the difference principle (2b) is to be accepted for the distribution of the other social primary goods—specifically income, wealth, social and economic position—or, alternatively, a principle of maximizing utility (again let us say average utility) is to be chosen. 2b or not 2b, that is the question.

At this point Rawls shifts from the maximin strategy (for it has already successfully done its work in seeing to the protection of the equal basic liberties and fair equality of opportunity) to introduce another argument specifically in support of the difference principle—the argument from collective asset (see Rawls 1974a, pp. 647–648;

also 1968, pp. 59, 67, 71). It is this argument that he calls the 'compelling' one for the difference principle (see Rawls 1974b, pp. 144–145), and it is this argument that we have been emphasizing throughout our paper.

Rawls's use of a two-stage argument makes it clear, then, that the maximin strategy does not support the difference principle in particular and that the argument specifically and peculiarly designed for that purpose is the one based on the idea of collective asset, namely, the argument from mutual benefit as it would be developed in the original position.

Once the two-stage argument is in place, the next step—it seems clear to us—should be to replace Rawls's preferred version of the difference principle (with its strong verbal affinities and resonance to the maximin strategy) with a version more attuned to the supporting two-stage argument, in particular its second stage. The version we have been developing in this paper has precisely this character.

In any event, one signal advantage of our reformulation (in terms of pareto efficiency under the constraint provided by egalitarianism) is that it does away with any need to rely on the maximin principle as the main support for the difference principle. For the criterion of distribution we have provided is not dependent on the maximin argument and, hence, would not be brought down by its failure. Accordingly, insofar as there are doubts about the efficacy of the maximin argument in the conditions imposed by the original position, our formulation of the difference principle criterion would escape that difficulty.

Another point. The maximin principle is understood, specifically, as a principle for decision making under conditions of uncertainty. Hence its formulation and choice, as a decision procedure, presuppose a thick veil of ignorance; extreme uncertainty is its home ground. But the real social world, in which such a principle would have effect, probably does not exhibit a correspondingly high level of uncertainty. It seems anomalous, then, actually to employ the maximin *criterion* in conditions where the degree and range of knowledge is sufficiently great to make the maximin *argument* unnecessary. Under real world conditions the maximin criterion

seems not so much artificial as simply inappropriate.

Our formulation avoids these difficulties. Its principal conditions of support, in the original position, would not require the radical and stringent restrictions on knowledge that the maximin principle does. Accordingly, the conditions for the original formulation and selection of our criterion are much more like the conditions, in the real social world, of its actual use. This is a virtue in itself, and the anomaly we have cited would not arise.

There are other signal advantages as well.

The formulation we have provided and the argument that lies behind it—the principle of mutual benefit and the idea of collective asset—are more clearly linked to other important features of Rawls's system than is the maximin criterion. Thus, an additional advantage of the criterion of pareto efficiency-egalitarianism is that it serves to integrate the Rawlsian account of justice. Here we suggest consideration, in particular, of two problematic points: (1) the connection of Rawls's two principles with what he calls the general conception of justice and (2) the relationship of the difference principle to the principle of equal basic liberty and to fair equality of opportunity. We shall take these up in turn.

One of the more compelling reasons for the adoption of our reformulation of the second principle of justice is that it is more amenable for derivation from the original position construct that Rawls uses to define justice as fairness. (See Rawls 1971, pp. 17-22). Among the important features of this construct is Rawls's general conception of justice, which he expresses as follows: 'All social values—liberty and opportunity, income and wealth and the bases of self-respect—are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage' (1971, p. 62; see also p. 303). Thus, Rawls's general point of view on justice is a strongly affirmed egalitarianism to be departed from only when it is not pareto efficient. Our reformulation of the second principle of justice expresses precisely this idea in specific terms. And, if one starts from Rawls's general conception of justice, then the derivation of the second principle in terms of pareto efficiency and egalitarianism is immediate.²²

²² In Rawls's earliest formulation of the two principles, the second asserted that

Rawls's first principle states that inequalities in some things, the basic liberties, will not be tolerated. It would be unreasonable and uniust for the basic liberties to be unequal. (See Rawls 1971, especially sects. 40, 77, 82; also pp. 211, 478, 496, 583.) In addition, he requires 'fair equality' of opportunities. (See Rawls 1971, sects. 12-14, 46.) Thus, two of the primary goods—liberty and opportunity-must be distributed equally to all if the distribution is to be a just one. But Rawls does not, similarly, require equality in the distribution of such things as social position and authority or income and wealth. Provided that unequal distribution allows everyone to be better off than he would have been if some original or primoridial distribution had been absolutely equal and provided that the constraints of the difference principle are met (as well as the lexically prior demands of justice), unequal distribution of these primary goods is just. Now the problem is that, if the difference principle is formulated, as Rawls usually puts it, in maximin version, the deviation from equality which is allowed by the general conception is not seen in any explicit way to serve an underlying equality or to be constrained by it. Thus, the two principles seem to incorporate radically different standards and the connection between them is obscure. In our formulation, the difference principle is interpreted in such a way that the deviations allowed in order to achieve pareto efficiency, either absolute or relative, are constrained by the minimizing of the necessary inequality. Thus equality continues to operate explicitly in our version of the second principle and the connection of the difference principle with the egalitarian first principle is thereby mediated.

In sum, our restatement of the second principle of justice has several virtues, as we have indicated. Chief among these is the degree of simplification that it affords Rawls's overall theory of

^{&#}x27;inequalities are arbitrary unless it is reasonable to expect that they will work out for everyone's advantage ...' (1958, p. 165). Thus, the connection of this earlier formulation with the so-called general conception is simply that of special case to general principle. One minor virtue of our analysis is that, by linking the difference principle (in the version we have provided) to the general conception, we have also made perspicuous the historical linkage of the difference principle (in 1971) to its earliest formulation by Rawls (in 1958).

justice. For it integrates main elements of the theory more fully by showing, in intuitive fashion, how the two principles follow from features and considerations developed in the original position; by exhibiting the connection of the two principles with the so-called 'general conception of justice'; and, finally, by making explicit the commitment to equality that operates throughout the formulation of the two principles. In contributing a greater coherence to the theory and in protecting the difference principle from some of the more effective criticisms directed against it, our reformulation has, we think, made the theory of justice as fairness a more compelling and persuasive one.

Appendix

Here, we will formally prove the equivalence of the two versions of the difference principle.

Consider a society that can be classified into n groups of individuals according to their economic well-being. The economic well-being of an individual will be represented by a real-valued index (for example, an index of primary goods) as a function of the individual's expected distribution of income, wealth, social position, etc. over the individual's lifetime. Each of these n groups will be represented by one hypothetical representative endowed with the average characteristics of the group he represents. The attainable set, A, will then be the set of all feasible (practicable) distributions of expectations of these n representatives under various basic structures consistent with the principles of equal liberty and fair equality of opportunity.

Without loss of generality, let group 1 consist of the most favored individuals, group 2 consist of the next most favored individuals, and so on until group n consists of the least advantaged individuals. With this labeling of the groups, the attainable set A will satisfy the condition

$$A \subset \{x = (x_1, \ldots, x_n) \colon x_1 \geq x_2 \geq \ldots \geq x_n\}$$

One should not interpret this condition to imply that, for example,

the least advantaged group is restricted to remain least advantaged under any feasible basic structure. Instead, since x_n represents the expectation of the hypothetical representative of the least advantaged group (regardless of the actual membership of this group), by definition x_n will be smaller than the expectations of the representatives of other groups, and so on. Hence, the above condition does not constitute a restriction on the attainable set.

Let the unique distribution favored by the maximin version of the difference principle (including the lexical version in case of non-uniqueness) be denoted by $a = (a_1, \ldots, a_n)$, i.e.,

(i)
$$\max \{\min \{x_i : i = 1, ..., n\} : x \in A\} = a_n;$$

(ii) max
$$\{\min \{x_i : i = 1, ..., n-1\}: x \in A, x_n = a_n\} = a_{n-1};$$

and so on until

(n) max {min
$$\{x_i : i = 1\}: x \in A, x_n = a_n, \ldots, x_2 = a_2\} = a_1$$
.

The set of all pareto efficient distributions in A will be denoted by P, i.e.,

 $P = \{x \in A : \text{there does not exist a distribution } (y_1, \ldots, y_n) \in A \text{ such that } y_i \ge x_i \text{ for each } i = 1, \ldots, n \text{ and } y_j > x_j \text{ for some } j = 1, \ldots, n\}.$

The combination of pareto efficiency and egalitarianism (with the lexical priority of the former over the latter) will lead to a unique distribution in A (see Lemma 2 below). Let us denote this unique distribution by $b = (b_1, \ldots, b_n)$, i.e.,

(i) min
$$\{x_1-x_n: x \in P\} = b_1-b_n;$$

(ii) min $\{x_1-x_{n-1}: x \in P, x_1-x_n = b_1-b_n\} = b_1-b_{n-1};$

and so on until

$$(n-1)$$
 min $\{x_1-x_2: x \in P, x_1-x_n=b_1-b_n, \ldots, x_1-x_3=b_1-b_3\}=b_1-b_2$.

Before we prove the equivalence of the two versions of the difference principle, we need to state two lemmas (the proofs of which are omitted).

Lemma 1: The unique distribution a favored by the maximin version of the difference principle is pareto efficient, i.e., $a \in P$.

LEMMA 2: The combination of pareto efficiency and egalitarianism (with the lexical priority of pareto efficiency over egalitarianism) will lead to a unique distribution in A.

THEOREM 1: If society can be classified into two groups, the more favored group and the less advantaged group (i.e., n = 2), then the distribution a favored by the maximin version of the difference principle will coincide with the distribution b favored by the combination of pareto efficiency and egalitarianism (with the lexical priority of the former over the latter), i.e., a = b.

Proof: By the definition of a, $a_2 \ge b_2$, and since $a \in P$, then by the definition of b, $a_1 - a_2 \ge b_1 - b_2$. Adding these two inequalities, we get $a_1 \ge b_1$. Since $b \in P$ (by the definition of b), $a_1 \ge b_1$ and $a_2 \ge b_2$ together imply that $a_1 = b_1$ and $a_2 = b_2$, i.e., a = b (because either $a_1 > b_1$ or $a_2 > b_2$ will imply that $b \notin P$). Q.E.D.

To show the equivalence of the maximin difference principle with the combination of pareto efficiency and egalitarianism in the general case of n groups n > 2, we need a regularity condition called 'chain connection.'

According to Rawls (1971, pp. 81-83), chain connection means that whenever the expectation of the least advantaged group is increasing as a result of increasing the expectation of the most favored group, the expectations of all the other intermediate groups are also increasing. Formally we will say that the attainable set A exhibits chain connection if, and only if, for any two distributions x, $y \in A$, if $x_n \ge y_n$ and $x_1 \ge y_1$, then $x_i \ge y_i$ for each i = 2, ..., n-1. Although Rawls never really defines chain connection very precisely, the above definition captures the essence of Rawls's discussion of chain connection.

In the general case of n>2 groups, the equivalence of the maximin difference principle and the combination of pareto efficiency and egalitarianism can be shown if we assume chain connection. This is done in the following theorem.

THEOREM 2: If the attainable set exhibits chain connection, then the maximin difference principle favors the same distribution as that favored by the combination of pareto efficiency and egalitarianism (with the lexical priority of the former over the latter), i.e., a = b.

Proof: Assume that chain connection holds. By the definition of a, $a_n \ge b_n$, and since $a \in P$, then by the definition of b, $a_1 - a_n \ge b_1 - b_n$. Adding these two inequalities, we get $a_1 \ge b_1$. Since chain connection holds, we can then conclude that $a_i \ge b_i$ for $i = 2, \ldots, n-1$. Since $b \in P$ (by the definition of b), we must have $a_i = b_i$ for $i = 1, \ldots, n$, i.e., a = b (because $a_i > b_i$ for some i would imply that $b \notin P$).

Q.E.D.

Bibliography

Arrow, K. J. (1973). 'Some ordinalist-utilitarian notes on Rawls' theory of justice,' *The journal of philosophy* vol. 70, pp. 245-263.

ATKINSON, A. B. (1973). 'How progressive should income-tax be?' pp. 90-109 in M. Parkin, editor. *Essays in modern economics*. London: Longmans.

BAUMOL. WILLIAM J. (1977). Economic theory and operations analysis. 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

HARSANYI, J. C. (1975). 'Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory,' *American political science review*, vol. 69, pp. 594-606.

IHARA, CRAIG. (1981) 'Maximin and other decision principles,' Philosophical topics, vol. 12, pp. 59-72.

KAYE, DAVID H. (1980). 'Playing games with justice: Rawls and the maximin rule,' Social theory and practice, vol. 6, pp. 33-51.

MARTIN, REX. (1981). 'Rawlsian economic justice and the proper bounds of government regulation,' pp. 114-132 in Norman Bowie, editor. Ethical issues in government. Philadelphia: Temple University Press.

MARTIN, REX. (1985). Rawls and rights. Lawrence: University Press of Kansas, forthcoming.

NOZICK, ROBERT. (1974). Anarchy, state, and utopia. New York: Basic Books.

- PARETO, VILEREDO. (1971). Manual of political economy. Translated by Ann S. Schwier; edited by Ann S. Schwier and Alfred N. Page. New York: Augustus M. Kelley. (Pareto 1971, is a translation of the French edition; Geneva, Switzerland: Librairie Droz, 1927. Pareto's book originally appeared as Manuel d'economie politique. Paris: Giard and Brière, 1909. This book is described on its title page as 'translated from the Italian edition by Alfred Bonnet and reviewed by the author.')
- PHELPS, E. S. (1973). 'Wage taxation for economic justice,' Quarterly journal of economics, vol. 87, pp. 331–354.
- RAE, DOUGLAS W. (1979). 'A principle of simple justice,' pp. 134-154 in Peter Laslett and James Fishkin, editors. *Philosophy, politics and society*. 5th series. New Haven: Yale University Press.
- RAWLS, JOHN. (1958). 'Justice as Fairness,' Philosophical review, vol. 67, pp. 164–194.
- RAWLS, JOHN. (1967). 'Distributive justice,' pp. 58-82 in P. Laslett and W. G. Runciman, editors. *Philosophy, politics, and society*. 3rd series. Oxford: B. Blackwell.
- Rawls, John. (1968). 'Distributive justice,' Some addenda,' *Natural law forum*, vol. 13, pp. 51–71.
- RAWLS, JOHN. (1971). A theory of justice. Cambridge, MA: The Belknap Press of Harvard University Press.
- RAWLS, JOHN. (1973). 'Distributive justice,' pp. 319-362 in E. S. Phelps, editor. *Economic justice*. Baltimore, MD: Penguin. (Rawls 1973, includes all of Rawls 1967, as sections 1-4 and 11-14 and most of Rawls 1968, as sections 5-10.)
- RAWLS, JOHN. (1974a). 'Reply to Alexander and Musgrave,' Quarterly journal of economics, vol. 88, pp. 633–655.
- RAWLS, JOHN. (1974b). 'Some reasons for the maximin criterion,' American economic review, vol. 64, pp. 141–146.
- RAWLS, JOHN. (1975). 'Fairness to goodness,' Philosophical review , vol. 84, pp. 536–554.
- RAWLS, JOHN. (1977). 'The basic structure as subject,' *American philosophical quarterly*, vol. 14, pp. 159–165.
- RAWLS, JOHN. (1978). 'The basic structure as subject,' pp. 47-71 in Alvin I. Goldman and J. Kim, editors. *Values and morals*. Dordrecht, Holland: Reidel. (This essay is a considerable revision of Rawls 1977; sections II and III are new.)
- RAWLS, JOHN. (1980a). 'A Kantian conception of equality,' pp. 198–208 in Virginia Held, editor. *Property, profits, and economic justice*. Belmont, CA: Wadsworth. (Rawls 1980a, draws upon Rawls 1974a, in sections 1, 3, and 4. Rawls 1980a, is reprinted, slightly revised, from an essay of the same name in Cambridge Review [February 1975], pp. 94–99; it has also been reprinted under the title 'A well-ordered society,' pp. 6–20 in Peter Laslett and James Fishkin, editors. [1979]. *Philosophy, politics and society*. 5th series. New Haven: Yale University Press.
- RAWLS, JOHN. (1980b). 'Kantian constructivism in moral theory,' *Journal of philoso-phy*, vol. 77, pp. 515–572. (Published version of three lectures given as the John Dewey Lectures at Columbia University, 14–16 April 1980).

SEN, A. K. (1970). Collective choice and social welfare. San Francisco: Holden-Day. SHENOY, PRAKASH P. (1980). 'Efficiency, egalitarianism and the difference principle in Rawls's theory of justice,' presented at the Second seminar on game theory and mathematical economics (Held in Hagen and Bonn, Federal Republic of Germany, 7-10 October 1980), mimeographed.

Received on April 15, 1983