

ALLAN GIBBARD

DISPARATE GOODS AND RAWLS'
DIFFERENCE PRINCIPLE:
A SOCIAL CHOICE THEORETIC TREATMENT

ABSTRACT. Rawls' Difference Principle asserts that a basic economic structure is just if it makes the worst off people as well off as is feasible. How well off someone is to be measured by an 'index' of 'primary social goods'. It is this index that gives content to the principle, and Rawls gives no adequate directions for constructing it. In this essay a version of the difference principle is proposed that fits much of what Rawls says, but that makes use of no index. Instead of invoking an index of primary social goods, the principle formulated here invokes a partial ordering of prospects for opportunities.

1. PRIMARY SOCIAL GOODS AND THE INDEXING PROBLEM

In *A Theory of Justice*, Rawls claims as one of the virtues of his theory that it does not require interpersonal comparisons of utility. Instead, the interpersonal comparisons needed for the theory are based on an 'index of primary social goods'. Primary goods are goods useful toward widely disparate ends, "things which it is supposed a rational man wants whatever else he wants" (p. 92).¹ The primary social goods include rights and liberties, powers and opportunities, income and wealth (pp. 62, 92). In Rawls' theory, then, the basis of interpersonal comparisons is overt: the comparisons are to be made on the basis of who gets how much of what.

Why might this be an advantage? To anyone who denies that interpersonal comparisons of utility are meaningful, the advantage will seem obvious: by not invoking such comparisons, the theory avoids a pseudo-concept. Even if such comparisons are meaningful, though, they should perhaps still be avoided in the formulation of a public conception of justice – a conception which is to be used in resolving conflicts of interest over the basic structure of society. For even if such comparisons can be made in principle, they will often be delicate, and the relevant psychological evidence will probably not be compelling. When interests conflict, and delicately based judgments are to be used to adjudicate them, each person's judgments are

likely to reflect his own interests. When that happens, there will be no agreement on how the standards of adjudication apply to the conflict in question. A public conception of justice should set up standards that are easy to apply, and interpersonal comparisons of utility, even if meaningful, do not pass this test (cf. pp. 90–93).

How clear, then, is Rawls' own standard? He speaks of an 'index' of primary social goods, which is to provide a clear standard for interpersonal comparison; my question concerns how that index is to be specified.² For it is this index that gives content to Rawls' *difference principle*: that the index of primary goods for the worst-off representative man is to be as high as is feasible (pp. 83, 90–95).

It might be thought that the specific index used does not greatly matter for the content of the difference principle. The index is used to identify the worst-off representative man, and those who are worst off in one primary good are likely to be worst off in all. In that case, all indices will agree on who is worst off, however differently they weigh the various primary goods.³

The index is used, though, not only to identify the worst-off representative man, but to compare various alternative arrangements of society from his standpoint: that social arrangement is just which accords the highest feasible index to the worst-off representative man. Now alternative social arrangements may differ vastly in the kinds of rewards they offer: capitalism with an income floor, for instance, might offer the worst-off representative man considerable income with few powers, whereas some alternative might offer him a lower income with more powers. How the index weighs income against powers will determine which social arrangement accords the worst-off representative man the higher index of primary social goods.

Some of the things Rawls says suggest that the index is not to be part of the difference principle itself, which is that "social and economic inequalities are to be arranged so that they are . . . to the greatest benefit of the least advantaged" (p. 83). True, to explicate the phrase 'to the greatest benefit of the least advantaged', we need an index, since the phrase means 'such that the least expected index of primary social goods in the society is as great as is feasible'. Rawls suggests, though, that whereas the difference principle is to be adopted in the original position, which is the first stage of a four-stage sequence of deliberation, the index that explicates it is to be left to a third, 'legislative' stage (1974, p. 642). In the legislative stage, as in the original

position, no one knows his own identity, abilities, and plan of life; but in the legislative stage, unlike the original position, “the full range of general economic and social facts” about the particular society in question can be brought to bear (p 199). The difference principle, then, is adopted by parties ignorant of the particular circumstances of their own society, but the index that interprets it is to be constructed after the parties have learned what their particular society is like. At that point, the index is to be constructed “by taking up the standpoint of the representative individual” from the worst-off group,

and asking which combination of primary social goods it would be rational for him to prefer. In doing this, we admittedly rely on our intuitive capacities. This cannot be avoided entirely, however. The aim is to replace moral judgments by those of rational prudence and to make the appeal to intuition more limited in scope, more sharply focused (p. 94).

More, though, needs to be said. The difference principle is of indeterminate meaning until we know how, given the circumstances of any particular society, to construct the index through which the principle applies to that society. The construction cannot come directly from judgments of rational prudence, since what matters is the rational preferences of ‘the representative individual’ of the worst-off group, and ‘the representative individual’ is not a person. Rather, statements about ‘the representative individual’ abbreviate more complex statements about a class of individuals. What is rationally prudent for the representative worst-off individual must in some way be a matter of what is rationally prudent for genuine individuals, or of what would be, under certain circumstances.

Perhaps we are to construct the index for a society by asking what would be rationally prudent for a person who knew that he would start out in the worst-off group in that society, who knew what the society was like, but who did not know particular facts about himself. Before that says much about how to construct the index, though, it must be joined with an account of how it is rational to choose with limited information. For what is rational under limited information has notoriously been a matter of controversy in discussions of Rawls’ theory. The difference principle has no clear content until directions are given for constructing the index of primary social goods in terms of which the principle is stated.

2. THE INDEXING PROBLEM FOR INCOME

In all but the last section of this paper, I shall discuss the indexing problem for only one primary good, income. Income closely fits Rawls' description of a primary good as something a rational person wants whatever else he wants. For an income is not an allotment of particular commodities, but rather an opportunity to choose among the most diverse combinations of goods. An income, then, can be used in the pursuit of a wide range of alternative sets of goals.

Restricting the initial discussion to income has a number of advantages. If conceptual problems arise with income, they will presumably remain when other primary social goods are included in the problem. We can perhaps best identify those problems by making simplifying assumptions. On the other hand, solutions to conceptual problems that arise in the case of income may turn out to apply to the general case of disparate primary social goods.

We may, if we wish, think of the special case of income as follows: we are now restricting ourselves to cases in which all primary goods other than income are distributed equally in a fixed amount, and then asking what would constitute maximizing the prospective incomes of those in the worst starting positions. For the sake of even more simplicity, I shall consider at the outset only cases of certainty, and suppose that we want to maximize the income of the worst-off person.⁴

Why might there be an indexing problem for incomes? An income, as I have said, amounts to a choice among diverse combinations of goods. Now although such a wide range of choices might be offered in other ways, the term 'income' suggests a particular way in which such a choice can be offered: one is faced with a quantity of income and a system of prices, and one may choose any combination of consumption goods the total price of which does not exceed one's income. For the sake of simplicity, I shall consider only such pure income-price systems at this point.

The indexing problem for incomes is this. Alternative economic policies may produce different relative prices. It may be that given one economic policy, the worst-off person would face one income-price combination, and given another policy, the worst-off person would face another income-price combination. More than one person may be worst-off, and those people may not agree on which income-price combination is preferable. In that case, how is it to be settled which policy leaves the worst-off people best off?

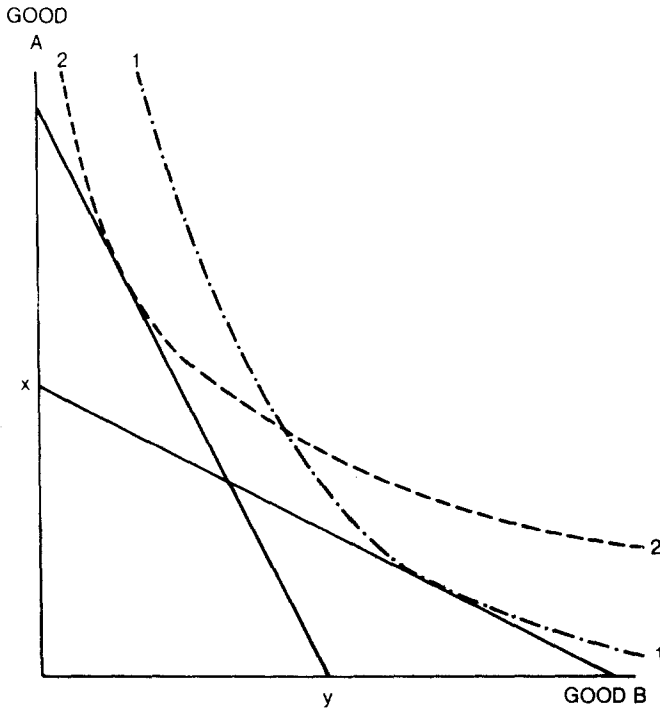


Fig. 1.

The situation can be represented graphically. Suppose there are only two commodities, *A* and *B*. Any point in a two-dimensional graph then represents some combination of goods *A* and *B*, or *commodity bundle*. The purchasing power of an income in the face of certain prices can be represented by a *budget constraint*, such as *x* in Figure 1: given that income and those prices, the person can purchase any commodity bundle on his budget constraint. He can also purchase less of any commodity. The set of commodity bundles he can purchase is his *budget set*.

I shall assume that each commodity is a good, in the sense that other things equal, each person prefers more of it to less. It follows from this assumption that a person's most preferred bundles lie on his budget

constraint. It also follows that given fixed prices, everyone agrees in preferring greater income to less.

Disagreement begins when alternative price ratios are considered. In Figure 1, the preferences of two people, 1 and 2, are represented by indifference curves in the space of commodity bundles. x is a budget constraint for someone who faces a high price for good A and a low price for good B ; y is a budget constraint for someone who faces a low price for A and a high price for B . Person 1 prefers x to y and person 2 prefers y to x . An example of the indexing problem for incomes, then, is this: if one economic policy would face the worst-off members of society with constraint x and another would face them with constraint y , which is the more just by the standards of the difference principle? Which policy, that is to ask, leaves the worst-off members of society better off?

Is this really a problem of justice? Questions of justice, on Rawls' view, concern the basic structure of society, and the details of relative prices do not seem part of the basic structure. Justice in prices will be procedural: once the basic structure is just, the prices which emerge from the procedures it sanctions are just simply because they emerge from those procedures (cf. pp. 86–9, 304–9).

With all this granted, however, the indexing problem remains one of justice. Whether the basic structure itself is just will depend in part on the incomes that could be expected to emerge from it, and different arrangements of the basic structure can be expected to produce different price ratios. Among the basic economic decisions to be made are the degree to which prices should be administered rather than set by markets, whether necessities such as food and shelter should be subsidized, whether recreational and cultural activities should be subsidized, and so forth. All these decisions involve making some commodities cheaper than would the market, and any taxes used to finance subsidies will make some commodities more expensive than would the untaxed market. Decisions about the basic economic structure of society affect price ratios as well as incomes, and we need some way of judging the alternatives by their expectable upshots.⁵

3. A MINIMAL DIFFERENCE PRINCIPLE

I want to propose not a way of constructing an index of budget sets, but a weak version of the difference principle that does without an index. The

proposal in this section is weak and preliminary, and later I shall consider how to strengthen it so as to rule out as unjust some economic states which this weak preliminary version admits as just. Even in this preliminary version, though, the difference principle has enough strength to be in conflict with the norms of efficiency: it may be that no Pareto efficient economic state is just by the difference principle, even in this weak formulation.

I begin with some definitions and notation. Since the range of choice an income represents is determined by prices, we must represent an individual's situation by giving not simply his nominal income, but the prices he faces. An *individual state*, then, will consist of an *income* and a *price vector*. The *income* is a non-negative number, and the *price vector* is an assignment of a non-negative number to each *commodity* in a finite non-empty set \mathcal{C} . A *social state* consists of a distribution and a price vector, where a *distribution* is an assignment of an income to each *person*, or member of a finite non-empty set I . Feasible social states⁶ will be represented by italic, bold-face letters w, x, y, z . Individual states will be represented by italic letters w, x, y, z , with or without subscripts. People will be represented by i, j , and k , and where w is a social state, w_i will be the individual state of person i in w : the individual state, that is, consisting of the income of i in w and the price vector of w .

Since where prices differ, incomes may not be comparable in any obvious way, we might do well to start with comparisons only of individual states with the same price vectors. Such individual states will be called *directly comparable*. Let $x > y$ iff x and y have the same price vector and the income of x is greater than that of y ; in that case, we shall call x *directly better* than y . The relation \succ , then, will be a *strict partial ordering* of individual states: transitive and irreflexive.

Although \succ gives few comparisons, we can use it to construct a very weak comparison of social states by a version of the difference principle. According to the difference principle, a social state x is more just than social state y iff the worst-off person in x is better off than the worst-off person in y , and a social state x is just iff there is no feasible social state that is more just than x . To say that x is more just than y , in other words, is to say that there is someone such that everyone is better off in x than he is in y . Say, then, that social state x is *directly more just* than social state y , or $x \succ^* y$, iff x and y have the same price vector and x has a higher minimum income than y ; in other words,

$$x \mathcal{J}^* y \quad \text{iff} \quad (\exists j)(\forall i)x_i > y_j.$$

A permissive version of the difference principle is that a social state x is just iff there is no feasible social state which is directly more just than x ; we say, then, that x is *minimally just* iff $\sim (\exists z)z \mathcal{J}^* x$. A social state is minimally just, in other words, iff there is no feasible social state with the same prices and a higher minimum income.

Now even the very restricted standard of comparison expressed by \mathcal{J}^* is incompatible with the weak Pareto principle: that if everyone prefers social state x to social state y , then x is better than y . The standard is incompatible with the Pareto principle not only in the sense that we may have x unanimously preferred to y without having x directly more just than y . It is incompatible in that if we try to combine the comparisons made by \mathcal{J}^* with comparisons made by the Pareto principle, we may have cycles.

Where each person i has an ordering P_i of individual states, we might set the following two conditions on a relation \mathcal{J} , to be read 'is better than' or 'is more just than'.

Unanimity: For any x and y , if $(\forall i)x_i P_i y_i$, then $x \mathcal{J} y$.

Minimal Difference Principle: For any x and y , if $x \mathcal{J}^* y$, then $x \mathcal{J} y$.

We can find patterns of preference such that Unanimity and the Minimal Difference Principle cause \mathcal{J} to cycle, and we can do so in the case of two goods and two people. A case is given in Figure 2; the idea behind it is this. Let person 1 prefer good A and person 2 prefer good B . Start 1 and 2 out in state z with equal incomes. Produce a Pareto improvement as follows. Raise the price of B and lower that of A in such a way as to please 1 and displease 2. Drop 1's income, but only slightly, so that 1 is still better off in the new state than he was in z . Raise 2's income enough to overcompensate him for the price change. Call the new state y ; then y is unanimously preferred to z . Now, leaving prices unchanged, form state x by returning 1 and 2 to their original incomes. That raises 1's income and lower 2's, producing, by the Minimal Difference Principle, a more just state. Thus x is more just than y by the Minimal Difference Principle, and y is unanimously preferred to z . Yet x differs from z only in that prices are changed to 1's advantage and 2's disadvantage. If a combination of Unanimity and the Minimal Difference Principle shows x to be better than z , then it is clear from the symmetry of the case that an analogous reverse argument will show that z is better than x .

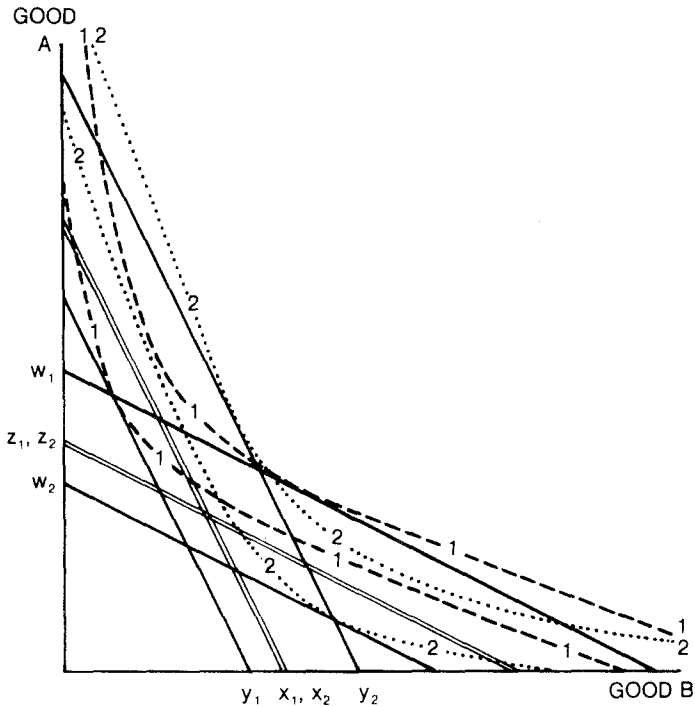


Fig. 2.

In Figure 2, y is unanimously preferred to z , x is more just than y by the Minimal Difference Principle, w is unanimously preferred to x , and z is more just than w by the Minimal Difference Principle. The difference principle, even in its most weak and unproblematic form, cannot be reconciled with the weak Pareto principle.

The point might be put another way. In the case we have been considering, if w , x , y , and z are the only feasible social states, then only w and y are Pareto optimal, but according to the Minimal Difference Principle, at most x and z are just. Now any principle that might reasonably be called a version of the difference principle will at least say what the Minimal Difference Principle says: that prices equal, the economic system with the higher minimal income is more just. For the very idea of the difference principle is to use overt criteria such as income for making interpersonal comparisons, and to judge economic systems by how well off, by those overt criteria, they make the

worst-off members of society. What we have shown, then, is that any faithful explication of the difference principle will yield a criterion of economic justice which is incompatible with the Pareto principle.

4. STRENGTHENING THE PRINCIPLE: SOME PITFALLS

I turn now to strengthening the difference principle so as to rule out more economic systems as unjust. At this point, I shall make no attempt to reconcile the difference principle with the Pareto principle, for we have seen the two to be irreconcilable. Later, I shall discuss whether the conflict between the two principles is a bad thing, and how the difference principle could be modified to avoid the conflict if we wanted to do so.

Note first a pitfall. Since people differ in their preferences among individual states, we might want to compare not simply individual states, but people in individual states. Let a pair $\langle i, x \rangle$ consisting of a person i and an individual state x be called a *personal state*. Perhaps instead of ranking individual states, which in effect are simply budget sets, we should rank personal states. That will allow us to take into account the preferences of the people involved.

Now it is central to the difference principle that within a single social state, interpersonal comparisons are to be by income. For any individual state x , let $I(x)$ be the income in that state. Let $\langle i, x \rangle B \langle j, y \rangle$ mean 'person i is better off in individual state x than is person j in individual state y '. Then we must require the following.

Interpersonal Comparison by Income (CI): For any social state x and people i and j , if $I(x_i) > I(x_j)$, then $\langle i, x_i \rangle B \langle j, x_j \rangle$.

An apparent advantage of considering personal states rather than mere budget constraints is that we can now make intrapersonal comparisons by consulting the preferences of the person involved.

Intrapersonal Comparisons by Preference (CP): For any person i and individual states x and y , if $x P_i y$, then $\langle i, x \rangle B \langle i, y \rangle$.

Conditions CI and CP, though, will sometimes force the relation B to be cyclic. In Figure 3, we have

$$(1) \quad \langle 1, x_1 \rangle B \langle 2, x_2 \rangle \text{ by CI;} \\ \langle 1, y_1 \rangle B \langle 1, x_1 \rangle \text{ by CP;}$$

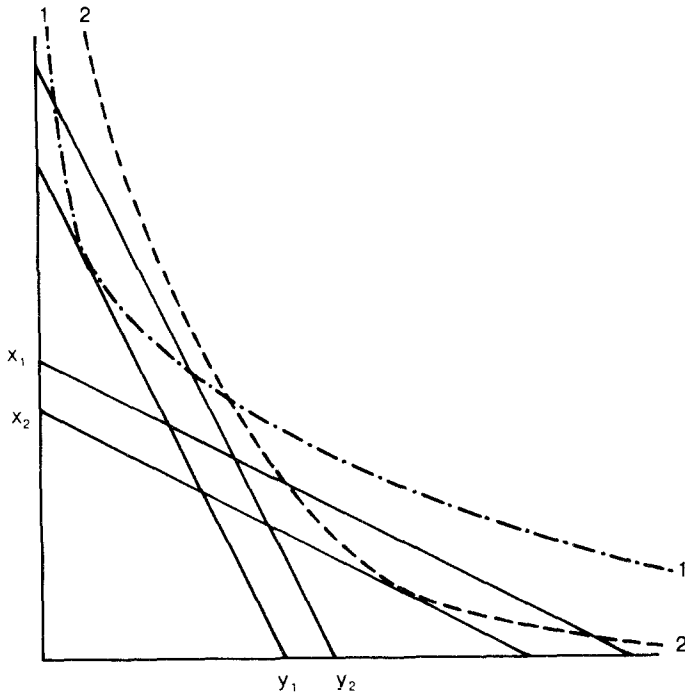


Fig. 3.

- (2) $\langle 2, y_2 \rangle B \langle 1, y_1 \rangle$ by CI;
- $\langle 2, x_2 \rangle B \langle 2, y_2 \rangle$ by CP.

Thus B is cyclic.

We have not shown that where B is cyclic, a \mathcal{L} defined from B must be cyclic. Nevertheless, the ease with which the relation of being better off can be made to cycle is grounds for caution. How should we proceed? One approach to making limited judgments of equity without psychologically based interpersonal comparisons is through a concept of 'envy'. The concept as originally formulated applied to bundles of goods: person i envies j 's bundle of goods if he would rather have it than his own. The same considerations could be applied to personal states. We could say that person i in state x_i envies person j in state y_j if he prefers y_j to x_i - if, in other words, $y_j P_i x_i$. Now there can easily be cases where $y_2 P_1 x_1$ and $x_1 P_2 y_2$, as in

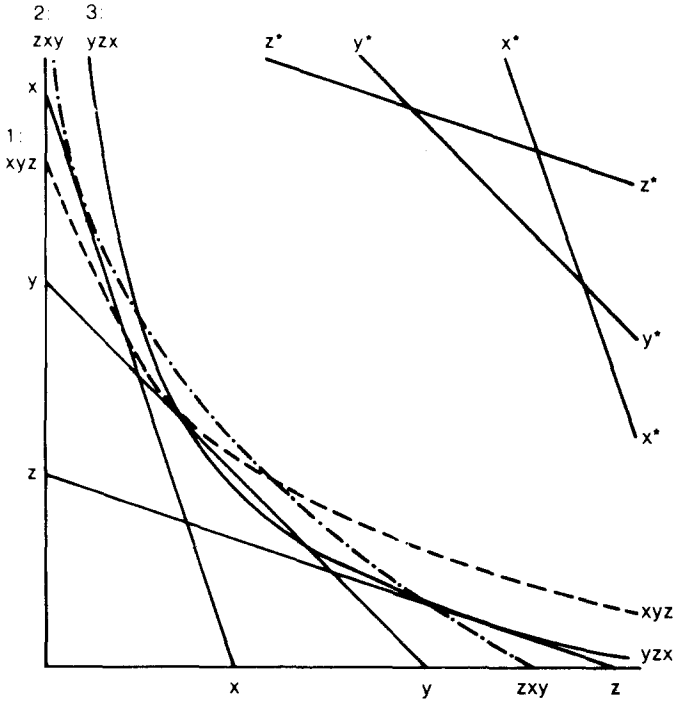


Fig. 4.

Figure 1 with $x_1 = x$ and $y_2 = y$. Perhaps the relation we should use is this: We say that i is *definitely better off* in x_i than is j in y_j if both i and j prefer x_i to y_j . I shall write this $x_i P_{ij} y_j$.

That suggests the following version of the difference principle. We do not hope for a complete weak ordering of personal states, but only for a partial ordering. B will be the relation *is definitely better off than*, and it holds between personal states. We have $\langle i, x \rangle B \langle j, y \rangle$ iff $x P_i y$ and $x P_j y$. We define the relation x is *more just than* y , or $x \mathcal{J} y$, as follows:

$$(4) \quad x \mathcal{J} y \quad \text{iff} \quad (\exists j)(\forall i)\langle i, x_i \rangle B \langle j, y_j \rangle.$$

Social state x is *just* iff no feasible social state is more just than x .

Now this new \mathcal{J} too is cyclic. To see this, note first that even when preferences are well-behaved in the usual ways, there will be triples of

income-price situations such that any preference ordering of them is possible. Moreover, where x and y are individual states, it will be possible to find an individual state x^* that is directly comparable to x such that x^*P_1y . That is, it is always possible to raise someone's income enough to compensate him for an unfavorable shift in prices. Hence we may suppose we have the following orderings of individual states: 1: xyz ; 2: zxy ; 3: yzx , where everyone prefers each of x^* , y^* , z^* to each of x, y, z , and moreover, x^* is directly comparable to x, y^* to y , and z^* to z . The situation is as shown in Figure 4. Let the social states be

$$\begin{aligned} x &= \langle x, x, x^* \rangle, \\ y &= \langle y, y^*, y \rangle, \\ z &= \langle z^*, z, z \rangle. \end{aligned}$$

Then $x \mathcal{S} y$, because xP_1y , $xP_{12}y$, and $x^*P_{13}y$. Like arguments show that $y \mathcal{S} z$ and $z \mathcal{S} x$, yielding a cycle.

5. A WAY OUT

The idea of this proof was to exploit an appropriate cyclic majority among three individual states. 1 prefers x to y , and so he prefers everyone's state in x to his in y . 2 also prefers x to y , and so prefers his state in x to 1's in y . Even though 3 prefers y to x , he is made rich enough in x to prefer his state in x to 1's in y . In that sense, x makes everyone better off than someone is made in y .

Since this relation is cyclic, we need to strengthen the difference principle as applied to \mathcal{S} so as to make fewer comparisons. Note that in the above example, although 1 and 2 agree that 2 is better off in x than 1 is in y , 3 does not: 3 prefers y to x . Perhaps, then, we should consider i better off in x than j is in y only if everyone prefers individual state x_i to individual state y_j . Our standard now makes no mention of the individuals involved, and so we can consider it a method for comparing individual states rather than personal states. We can now say xBy iff $(\forall i)xP_iy$; in that case, we say that *one is unequivocally better off* in individual state x than in individual state y . x is *more just* than y , or $x \mathcal{S} y$, iff there is some individual j such that one is unequivocally better off in anyone's state in x than one is in j 's state in y .

Formally put, we have $x \mathcal{L} y$ iff $(\exists j)(\forall i)x_i B y_j$. The *Difference Principle* will say that x is just iff there is no feasible state z such that $z \mathcal{L} x$.

\mathcal{L} is a strict partial ordering; that is, transitive and irreflexive. The simple lemma behind this claim will be useful in later discussion, and bears explicit statement. Note at the outset that B as defined here is a strict partial ordering, since it is simply a relation of unanimous preference among individual states.

LEMMA: Let B be any strict partial ordering of individual states, and define $x \mathcal{L} y$ as $(\exists j)(\forall i)x_i B y_j$. Then \mathcal{L} is a strict partial ordering of social states.

Proof: \mathcal{L} is irreflexive, because $x \mathcal{L} x$ means $(\exists j)(\forall i)x_i B x_j$, which entails that for some j , $x_j B x_j$, contradicting the irreflexivity of B . To see that \mathcal{L} is transitive, suppose $x \mathcal{L} y$ and $y \mathcal{L} z$. Then there are a j such that $(\forall i)x_i B y_j$ and a k such that $(\forall i)y_i B z_k$. Therefore, $y_j B z_k$, and since $(\forall i)x_i B y_j$, by transitivity of B , we have $(\forall i)x_i B z_k$. Therefore $x \mathcal{L} z$.

Since \mathcal{L} is a strict partial ordering, there will always be at least one feasible social state which is just according to the Difference Principle. We have seen that in some cases, no feasible social state which is just according to the Difference Principle will be Pareto optimal, for the Minimal Difference Principle allowed all states as just that the Difference Principle does, as a check of the criteria will show, but it was Pareto-incompatible.

What the Difference Principle in its present form says is this. A social state x is just unless for some feasible social state z , everyone agrees that he would rather face the prices in z with the income of the poorest person in z than face the prices of x with the income of the poorest person in x . We have, then, a consistent version of the Difference Principle as restricted to income. That version is, I suspect, as strong as it can be made without introducing either an element of arbitrariness, or interpersonal comparisons of happiness, strength of preference, or the like.

6. PARETO COMPATIBILITY

How should we regard the Pareto-incompatibility of the Difference Principle? Does the joint inconsistency of the Difference Principle and the Pareto Principle disqualify the Difference Principle as a principle of justice? I think not. Remember that in Rawls' theory, a principle of justice is to serve as part

of a public conception of justice, to which people appeal in adjudicating conflicts over the basic structure of society. A principle of justice is not designed for an impartial, well-informed god with the power to institute whatever economic system he decides is most just. Rather, the application of a principle of justice should be understood by the people affected: whatever the principle in fact endorses as just ought to be seen as what the principle endorses by the people whose interests are involved. It is partly for that reason that the Difference Principle is put in terms of primary social goods rather than utilities.

Now conflicts between the Difference Principle and the Pareto Principle arise in cases like that of Figure 2. Suppose that in Figure 2, only social states w , x , y , and z are feasible. Then only w and y are Pareto efficient. Consider y : y contains unequal incomes, and is unjust according to the Difference Principle because a feasible alternative, x , has the same prices and equal incomes at an intermediate level. x , by a criterion that involves only incomes, is more just than y . Neither x nor y is a Pareto improvement over the other, but judged by income, x is the more egalitarian. The same could be said of w and z .

Should social state z , in which all incomes are equal, be rejected as unjust because it is Pareto inefficient? Questions of justice aside, everyone prefers y , with unequal incomes, to z with equal incomes: one group because y shifts price in a way favorable to them, and the other because they are over-compensated for an unfavorable price shift. All realize, though, that once they have shifted to y , there will be a conflict of interest over whether to shift further to x . They have agreed to resolve such conflicts by looking at income, and maximizing the minimal income level. By that principle, they will be committed to the further shift to x . Once people have agreed to make interpersonal comparisons by an overt standard of income, they should realize that to accept unequal incomes for the sake of a state everyone prefers to egalitarian state z may be to raise new questions of justice which, by agreement, will be resolved by moving to a state that differs from z only by a price shift – a price shift favorable to some and unfavorable to others. Realizing that, they may find it reasonable to forego a Pareto improvement for the sake of retaining an overt standard for the resolution of conflicts over the basic structure of society.

If all this leaves the reader unconvinced, he may wish to adopt a Paretian

version of the Difference Principle: that a social state is just iff either it is just according to the old, non-Paretian difference principle, or it is a weak Pareto improvement over some state which is just according to the non-Paretian difference principle. Liberalized slightly more, the principle might read as follows.

Paretian Difference Principle: A social state x is just iff for some feasible social state y ,

$$\begin{aligned} &\sim (\exists i)y_i P_i x_i, \\ &\sim (\exists z, j)(\forall i, k)z_i P_k y_j. \end{aligned}$$

This says that a social state is just iff there is a state which is just by the non-Paretian difference principle which no one likes any better.

7. ECONOMICALLY INFLUENCED PREFERENCES

So far, in formulating the difference principle, I have taken preferences as fixed independently of the choice of economic systems. Given the preferences P_1, \dots, P_n of all n members of society, I have proposed that xBy , individual state x is definitely better than individual state y , iff $(\forall i)x_i P_i y_i$. For two social states x and y , I have proposed that $x \mathcal{J} y$ iff for some person, everyone is definitely better off in x than he is in y . What the difference principle now says of basic economic structures is this: in conditions of certainty, one basic economic structure is definitely more just than another iff, where the first would lead to social state x and the second would lead to social state y , we have $x \mathcal{J} y$. A basic economic structure is just iff no alternative basic economic structure is definitely more just than it.

This principle needs to be modified. For preferences among bundles of commodities are clearly not fixed independently of the economic system: alternative basic economic structures will lead to different preferences.⁷ Perhaps, then, we should do the following. Let \mathcal{P} be the set of all preference orderings P over bundles of commodities such that for some basic economic structure, if that structure were instituted someone might have preference ordering P . Then for individual states x and y , say xBy iff $(\forall P \in \mathcal{P})xPy$. Alternatively, if we want a more discriminating criterion, let us say that xBy iff $(\exists P \in \mathcal{P})xPy$ and $\sim (\exists P \in \mathcal{P})yPx$. The new B is a strict partial

ordering, and so by the Lemma of Section 6, where $x \mathcal{S} y$ means $(\exists j)(\forall i)x_i B y_j$, \mathcal{S} is a partial ordering.

The Difference Principle as now formulated says the following. A basic economic structure is just iff no alternative basic structure is definitely more just. In order to see whether a basic economic structure X is definitely more just than another, Y , do the following: For each basic structure, consider the incomes and prices to which it would lead; let these be given by social states x and y respectively. Consider the situation x_{\min} of a worst-off person in x ; x_{\min} consists of facing the prices of x with the least income of anyone in x . Compare that with y_{\min} , the situation of the worst-off person in y . x_{\min} is *definitely superior* to y_{\min} iff anyone, no matter what basic economic structure had influenced his preferences, would prefer facing situation x_{\min} to facing y_{\min} . x is *definitely more just* than y iff x_{\min} is definitely superior to y_{\min} .

This criterion is permissive, in that it allows policies that shift prices so long as anyone prefers the consequent situation of the worst-off – where ‘anyone’ here includes any sort of person who might emerge from a basic economic structure open for choice at the legislative stage. Suppose, for instance, in an impoverished society, we consider whether to tax everyone to heap lavish subsidies on grand opera. Let X be a basic economic structure, and Y that structure modified by the opera subsidy scheme. Under Y , the poorest people have a slightly lower income than the poorest people have under X , but under Y they have a chance to purchase tickets to grand spectacles that would consume more than their entire income under X . The tickets are expensive, let us suppose, but the poorest person could purchase one if he sacrificed enough in the way of other commodities. In that case, it may be that none of the poorest people do purchase opera tickets, and so by the test of their own preferences, they are worse off with the subsidy scheme than without. It may also be that none of the richer people who do buy tickets would do so if they were as poor as the worst off. None of these facts will settle the issue of whether X is definitely more just than Y , although they establish that Y is not definitely more just than X . The test is rather this: Would any basic economic structure open for choice at the legislative stage produce anyone so devoted to opera that, faced with a choice of being poorest under X and being poorest under Y , he would choose Y , preferring the combination of seeing opera at the cost of an expensive ticket and

additional taxes to the alternative of being unable to see it at all. If so, then X is not definitely more just than Y . If moreover, no other basic economic structure open for choice at the legislative stage is definitely more just than either X or Y , then according to the Difference Principle, both X and Y are just. It would then be both just to have the opera subsidy and just not to have it.

Is this criterion overly permissive in the subsidies it allows for the pursuits of the rich – or, for that matter, in the subsidies it allows for the objects of unusual tastes? Perhaps so, but the limits of what it allows should be noted. Call a basic economic structure *eligible* iff it is open for choice at the legislative stage. Which structures are eligible is constrained by feasibility, the priority of liberty, and the priority of having all positions of privilege open to all. Now the set \mathcal{P} consists of those preference orderings that would be formed under any basic economic structure that is eligible. The restriction to eligible structures is crucial. Eligible structures satisfy a constraint of liberty, and so cannot include such things as a compulsory course of brainwashing to determine preferences. The preferences in \mathcal{P} , then, do not consist of all preferences that might be produced by a suitable course of conditioning, but simply those preference orderings that would arise under circumstances of liberty given various basic economic structures compatible with liberty. Thus a scheme of opera subsidies may be unjust, since it may be that no one who forms his tastes freely will be so exclusively devoted to opera that even were he among the poorest of the poor, he would willingly sacrifice the price of a ticket and his share of the subsidy in order to see a lavish opera.

8. OTHER PRIMARY SOCIAL GOODS AND UNCERTAIN PROSPECTS

Return now to the problem of primary social goods in general. Just as an income constitutes a choice among alternative bundles of commodities, so does a combination of primary social goods – income and wealth, powers and opportunities – constitute a choice among alternative bundles; now, though, a bundle consists both of commodities and of exercises of powers and opportunities. Call a set of such bundles an *opportunity set*. A rational person certain of his preferences will prefer one opportunity set x to another y iff he prefers the most preferred bundle in x to all bundles in y . We may now

treat opportunity sets just as we have been treating individual situations, and formulate a version of the difference principle that applies to primary economic goods in general.

First, though, consider another problem. So far, I have talked as if the choice of a basic economic structure determines, for each person, precisely how he will fare. As Rawls emphasizes, though, such matters are not certain, and the difference principle as Rawls states it evaluates an economic system by the prospects it offers those in the worst starting position (see pp. 78, 96). What is really needed, then, is not an index of incomes or opportunity sets, but an index of prospects over opportunity sets – or, as I have been suggesting, a substitute for an index in the form of a partial ordering.

Even for prospects over income with fixed prices, a ranking by expected money value will not do. For suppose we judge prospects by their expected money payoff. Consider two economic systems: under system X , everyone in a worst-off starting position will get \$9000 per year, whereas under system Y , half will get \$20,000 and half will starve with no income at all. For those in the worst-off starting positions, then, the expected money value of system Y is higher than that of X . Surely, though, anyone would prefer starting out worst off in system X to starting out worst off in system Y . It would be preposterous to prefer system Y to system X out of a regard for the plight of those in the worst starting positions.

Now the formal method we have been using will apply to a choice among prospects over opportunity sets, and do so without measuring the prospects of those in the worst starting positions by anything like their expected money value. Let \mathcal{S} be the set of all eligible basic economic structures – all basic economic structures open for selection at the legislative stage. Let \mathcal{P} be the set of all prospects over opportunity sets that would be offered anyone under any structure in \mathcal{S} , and let variables \hat{x} , \hat{y} , \hat{z} take prospects in \mathcal{P} as values. Let \mathcal{P} be the set of all those preference orderings of members of \mathcal{P} that any one would have some likelihood of developing under any structure in \mathcal{S} . Then we can define what it is for one prospect to be definitely better than another:

$$\hat{x}B\hat{y} \quad \text{iff} \quad (\exists P \in \mathcal{P})\hat{x}P\hat{y} \ \& \ \sim (\exists P \in \mathcal{P})\hat{y}P\hat{x}.$$

Let the *social prospect* offered by a structure $S \in \mathcal{S}$ be the n -tuple $\hat{x} = \langle \hat{x}_1, \dots, \hat{x}_n \rangle$ of prospects offered by S to the members of the society.

Then where $S, T \in \mathcal{S}$, we say that S is definitely more just than T iff, where \hat{x} is the social prospect offered by S and \hat{y} is the social prospect offered by T ,

$$(\exists j)(\forall i)\hat{x}_i B_i \hat{y}_j.$$

Then basic economic structure S is just iff no alternative structure $T \in \mathcal{S}$ is definitely more just than S . What this says is that a structure is just so long as there is no eligible structure T which improves the prospect of the worst-off in the following sense: that some prospect \hat{x} offered by S is so unappealing compared to all the prospects offered by T that anyone, no matter how his preferences had developed in circumstances of liberty, would prefer each of the prospects offered by T to prospect \hat{x} .⁸

I do not know how Rawls would find this as an explication of the difference principle. Clearly it conflicts with the letter of what he says: it evades the construction of an index rather than constructing it, and yields a principle in conflict with the norms of efficiency. It differs from Rawls in its treatment of relevant social positions (pp. 95–100), and avoids aggregation of expectations where Rawls permits it. It may be more permissive than Rawls would like, in that a system is saved from being condemned as less just than an alternative so long as for each prospect it offers, there might be one person whose preferences developed under conditions of liberty and who prefers that prospect to some prospect offered by the alternative.

On the other hand, the principle I have proposed captures a number of aspects of Rawls' theory. It takes seriously the dictum that primary goods are things it is rational to want whatever else one wants, by explaining them as opportunities to choose among bundles of more specific goods. It avoids interpersonal comparisons of satisfaction, and instead looks at prospects for overtly observable income, powers, and opportunities. It avoids the supposition that preferences are fixed independently of the economic structure, and while it respects the constraints of human nature on preferences formed in conditions of liberty, it makes comparisons of prospects independently of the kinds of preferences that would be formed in any particular economic system. Finally, it captures the idea of making the worst off as well off as possible. In a just society, on the explication in this paper, even a person in the worst starting position should realize that, in a sense, it would be impossible to make everyone better off than he in fact is. "In every alternative basic economic structure", he may be told with truth, "there is a starting position

that is no better than yours, in the sense that someone with preferences formed under liberty might like that starting position no more than the starting position you actually have".⁹

University of Michigan

NOTES

¹ Page references, unless otherwise indicated, are to Rawls (1971).

² Plott (1978) discusses the problem of indexing primary social goods. His approach differs from mine in a number of respects that are discussed in Note 8.

³ Rawls speaks at one point (p. 97) as if when greater powers and income go together in a society, the indexing problem is avoided. Earlier, he argues only that in that case, the problem is simplified (p. 94).

⁴ Rawls speaks not of making the worst off person as well off as possible, but of making the worst-off 'representative man' as well off as possible. He ponders the question of what group's expectations should be aggregated in order to define the worst-off representative man. We are to consider the 'starting places' in society 'properly generalized and aggregated' (p. 96). In this paper I try to avoid aggregation, but in Section 8 I try to do at least part of what Rawls wants to accomplish by aggregation, by taking up his point that we are to apply the maximin criterion to starting positions and the expectations that attach to them, rather than to achieved income levels.

⁵ Rawls writes that the difference principle holds among other things "for income and property taxation, for fiscal and economic policy" (1975, p. 97. See also 1977, p. 164.).

⁶ Whether a social state is feasible will depend on individuals' preferences, since in a feasible social state, the demand for each commodity must equal its supply. (I owe this observation to Roy Gardner.) Here preferences are taken as fixed; the case of malleable preferences is taken up in Sections 7 and 8.

⁷ This problem was brought to my attention by John Bennett. The point is important in Rawls' thought; see 1974, p. 641 and 1975, p. 95.

⁸ The approach I have taken to the indexing problem differs from Plott's in a number of respects. Plott considers the problem of indexing bundles of disparate goods, whereas I consider the problem of indexing sets of such bundles, or prospects over sets of such bundles. Plott makes assumptions from which it follows that social states can be weakly ordered by the level of welfare of their worst-off people; I require only a partial ordering. Finally, Plott interprets the difference principle as a requirement that social institutions be designed so as to achieve a most just state whatever well-behaved preferences people may have. I am simply inquiring whether the difference principle can be intelligibly formulated in a way that guarantees that at least one feasible state will be just. Plott shows that the rest of the conditions he imposes are incompatible with Pareto principle, and so in that respect, his conclusion is similar to the conclusion that the weak difference principle is incompatible with the Pareto principle.

⁹ I am grateful to John Bennett for extensive and extremely helpful discussion of this paper.

BIBLIOGRAPHY

- Plott, Charles R.: 1978, 'Rawls' theory of justice: An impossibility result', in Hans W. Gottinger and Werner Leinfellner (eds.), *Decision Theory and Social Ethics: Issues in Social Choice* (D. Reidel, Dordrecht, Holland).
- Rawls, John: 1971, *A Theory of Justice* (Harvard University Press, Cambridge, Mass.).
- Rawls, John: 1974, 'Reply to Alexander and Musgrave', *Quarterly Journal of Economics* 88, pp. 633–55.
- Rawls, John: 1975, 'A Kantian conception of equality', *Cambridge Review*, pp. 94–99.
- Rawls, John: 1977, 'The basic structure as subject', *American Philosophical Quarterly* 14, pp. 159–65.