AMARTYA SEN

WELFARE INEQUALITIES AND RAWLSIAN AXIOMATICS*

ABSTRACT. This paper is concerned with ordinal comparisons of welfare inequality and its use in social welfare judgements, especially in the context of Rawls' 'difference principle'. In Section 1 the concept of ordinal inequality comparisons is developed and a theorem on ordinal comparisons of welfare inequality for distributional problems is noted. Section 2 is devoted to Harsanyi's (1955) argument that a concern for reducing welfare inequalities among persons must not enter social welfare judgements. In Section 3 an axiomatic derivation of Rawls' lexicographic maximim rule is presented; this relates closely to results established by Hammond (1975), d'Aspremont and Gevers (1975) and Strasnick (1975). In the last section the axioms used are examined and some alternative axioms are analysed with the aim of a discriminating evaluation of the Rawlsian approach to judgements on social welfare.

1. ORDINAL EQUALITY PREFERENCE

Usual measures of economic inequality concentrate on income, but frequently one's interest may lie in the inequality of welfare rather than of income as such.[1] The correspondence between income inequality and welfare inequality is weakened by two distinct problems: (i) welfare – even in so far as it relates to economic matters – depends not merely on income but also on other variables, and (ii) even if welfare depends on income alone, since it is not likely to be a linear function of it, the usual measures of inequality of income will differ from that of welfare. The first problem incorporates not merely the basic difficulties of interpersonal comparison of welfare, but also those arising from differences in non-income circumstances, e.g., age, the state of one's health, the pattern of love, friendship, concern and hatred surrounding a person. The second reflects the fact that the usual measures, such as the coefficient of variation, or the standard deviation of logarithm, or the Gini coefficient, or the inter-decile ratio, will not be preserved under a strictly concave transformation of income as the welfare function is typically assumed to be, when it is taken to be cardinally measurable. And when welfare is measurable only ordinally, then the usual measures of inequality are not even defined.

There is an obvious need for investigating inequality contrasts when welfare comparisons are purely ordinal. There is likely to be much greater agreement on the *ordering* of welfare levels of different persons than on a particular

Theory and Decision 7 (1976) 243–262.

numerical interpersonal welfare function unique up to a positive linear transformation. However, the meaning of more or less inequality is not altogether clear when comparisons of welfare levels are purely ordinal.

There are, nevertheless, some unambiguous cases. Let (x, i) stand for the position of being person i in social state x. Taking two persons 1 and 2 and two states x and y, consider the following strict descending orders:

1	2	3	4
$(y, 2)$	$(y, 1)$	$(y, 2)$	$(y, 1)$
$(x, 2)$	$(x, 1)$	$(x, 1)$	$(x, 2)$
$(x, 1)$	$(x, 2)$	$(x, 2)$	$(x, 1)$
$(y, 1)$	$(y, 2)$	$(y, 1)$	$(y, 2)$

Note that irrespective of the relative values of the 'differences', in each case y displays more inequality than x in an obvious sense. This type of comparison will be referred to as 'ordinal inequality comparison'.[2]

To formalize this criterion, let $\tilde{R}$ stand for an agreed 'extended ordering'[3] over the Cartesian product of X (the set of social states) and H (the set of individuals), i.e., over pairs of the form (x, i). The meaning of $(x, i)\, \tilde{R}\, (y, j)$ is that i is at least as well off in x as is j in y. $\tilde{P}$ and $\tilde{I}$ stand for the corresponding concepts of 'strictly better' and 'indifference'. Let ρ stand for any one-to-one correspondence between the pair of persons to itself.

Two-person ordinal inequality criterion (TOIC): For any pair of social states x, y, for a two-person community, if there is a one-to-one correspondence ρ from the pair of persons (i, j) to the same pair such that:

$$(y, i)\,\tilde{P}\,(x, \rho(i)),\ (x, \rho(i))\,\tilde{R}\,(x, \rho(j)),\ (x, \rho(j))\,\tilde{P}\,(y, j),$$

then x has less ordinal inequality than y, denoted $x\ \theta\ y$.[4]

The criterion can be extended to n-person communities also, by requiring the additional antecedent that all persons other than these two are equally well off under x and y. This implies an assumption of 'separability', which is however more debatable, and will be debated (see Section 4).

Strengthened two-person ordinal inequality criterion (STOIC): If for any n-person community with $n \geqslant 2$, for any two persons i and j, and any two social states x and y for some ρ: $(y, i)\,\tilde{P}\,(x, \rho(i))$, $(x, \rho(i))\,\tilde{R}\,(x, \rho(j))$, $(x, \rho(j))\,\tilde{P}\,(y, j)$, and for all $k \neq i, j$: $(x, k)\,\tilde{I}\,(y, k)$, then x has less ordinal inequality than y, denoted $x\ \theta^*\ y$. The transitive closure of θ^* is θ^{**}.

Of course, for a two person community $\theta = \theta^*$, and STOIC implies TOIC in this sense.

Note that no condition of constancy of total welfare has been used in the definitions, and indeed no such concept is definable given utility comparisons that are purely ordinal. It is, however, possible to use these definitions in the particular context of ranking alternative distributions of a fixed total income. Indeed, for that particular problem of 'pure distribution', ordinal inequality comparisons can be linked with some well-known results in the normative approach to inequality measurement based on Lorenz curve comparisons (see Kolm, 1966; Atkinson, 1970; Dasgupta *et al.*, 1973; Rothschild and Stiglitz, 1973). Our motivation here differs, however, from those exercises in the sense that our current concern is to look at welfare inequalities as such without necessarily saying anything about social welfare by invoking some group welfare function, and this contrasts with comparing values of social welfare given by a group welfare function, or a class of such functions.

Let the ranking relation λ stand for strict 'Lorenz domination', i.e., $x \lambda y$ if and only if the Lorenz curve of x is nowhere below that of y and somewhere strictly above it.

(T.1) In the 'pure distribution' problem with welfare rankings preserving the order of income rankings, if $x \lambda y$, then $x \theta^{**} y$.

The proof follows immediately from a well-known result of Hardy, Littlewood and Polya (1934), which in this context implies that $x \lambda y$ holds if and only if x can be obtained from some inter-personal permutation y^0 of y through a finite sequence of transfer operations with income being transferred from a richer person to a poorer one without reversing their income ranking.[5] Since in each of these operations taking us from x^s to x^{s+1}, the incomes of others except the two involved (say, $1(s)$ and $2(s)$ respectively) in that operation remain the same, and since more income implies higher welfare, clearly:

$$(x^s, 1(s))\,\tilde{P}\,(x^{s+1}, 1(s)), (x^{s+1}, 1(s))\,\tilde{R}\,(x^{s+1}, 2(s)), (x^{s+1}, 2(s))\,\tilde{P}\,(x^s, 2(s)).$$

Thus $x^{s+1}\, \theta^*\, x^s$. Since x and y^0 are the two extreme members of this sequence, and by virtue of STOIC it makes no difference whether we start from y^0 or from y, clearly $x\, \theta^{**}\, y$.

Notice that (T.1) provides a welfare basis for comparisons of inequality which is not dependent on taking 'more social welfare' to be 'less unequal', and in this sense departs from the normative approach of Kolm (1966), Atkinson (1970) and others. Indeed, nothing is said about social welfare as such, and this welfare interpretation of inequality simply looks at the inequality of the welfare distribution (in ordinal terms).

Note also that no assumption of concavity (or quasi-concavity, or S-concavity) of welfare functions is required in establishing (T. 1) in contrast with the earlier results on Lorenz comparisons referred to above, e.g., Kolm (1966), Atkinson (1970), Dasgupta *et al.* (1973), Rothschild and Stiglitz (1973).

It is, however, possible to introduce the additional assumption that less welfare-inequality is socially preferred, or at least regarded to be as good. Let R stand for the weak relation of social preference, with P and I its asymmetric and symetric parts: 'strict preference' and 'indifference' respectively.

Two-person equality preference (TEP): In a 2-person community for any x, y, if $x \, \theta \, y$, then $x \, R \, y$.

Strengthened two-person equality preference (STEP): In any community, for any x, y, if $x \, \theta^* \, y$, then $x \, R \, y$.[6]

Again, STEP obviously implies TEP. Note, however, that STEP has been defined in terms of θ^* and not θ^{**}. Of course, if R is transitive, then the two are equivalent.

How appealing a condition is STEP? That would seem to depend on three types of considerations. The first applies to STEP only, while the last two apply to both STEP and TEP.

(1) STEP involves a 'separability' assumption being based on θ^* (rather than θ in a 2-person community). A reduction of ordinal inequality between two persons is rather more definitive for a community of those two persons than for a community where there are others also, even though they are equally well off under x and y.[7] In Section 4 we shall examine the far-reaching consequences of this extension from a 2-person to a n-person comparison in the presence of other conditions, e.g., 'independence of irrelevant alternatives'.

(2) TEP and STEP both give overriding importance to the reduction of welfare inequality without bringing in any consideration of relative gains and losses of different persons.[8] How disturbing this criticism is will depend partly on the 'informational basis' of welfare comparisons, i.e., on the mea-

surability and interpersonal comparability assumptions.[9] If individual welfare is not cardinal, or if interpersonal comparisons must be ordinal only (whether or not individual welfares are cardinal), then clearly the concept of 'gains' and 'losses' in welfare lose meaning. If, however, cardinal interpersonal comparisons can be made, then one can consider a choice in which x involves less ordinal inequality than y, but the loss of person 2 is so much and the gain of person 1 is so little, that a reasonable case can be made for the choice of y. With 'full comparability' this conflict of ethics (e.g., *vis-à-vis* utilitarianism) must be faced, but with 'level comparability' only, this objection to STEP or TEP cannot be sustained.[10]

(3) The approach of STEP or TEP shares with utilitarianism and other needs-based ethics, a disregard of the concept of desert (see Sen, 1973, Chapter 4 on the contrast between need-based and desert-based approaches). Arguments such as 'person i is better off than j both in x and in y and gains less than j loses, but the additional gain is his just desert', are not entertainable in this approach.[11]

2. HARSANYI'S CRITICISM OF CONCERN FOR WELFARE INEQUALITY

In the context of social evaluation taking note of welfare inequalities, we should consider an objection of John Harsanyi (1975) to attempts at using social welfare functions that are non-linear on individual welfares. We know, of course, that with some assumptions of interpersonal comparability (e.g., 'unit comparability'), individual welfare levels cannot be interpersonally compared even though gains and losses can be compared (see Sen, 1970, Chapter 7). Harsanyi's attack is, however, not based on any subtlety of the comparability assumption.[12] It takes mainly the form of quoting his justly celebrated result that if individual preferences and social preference can both be given von Neumann-Morgenstern cardinal representation, and if Pareto indifference must imply social indifference, then social welfare must be a *linear combination* of individual welfares (Harsanyi, 1955), and then of defending the acceptability of the von Neumann-Morgenstern axioms. Harsanyi thus sees social welfare simply as an average welfare (unweighted if a further assumption of symmetry is made), and there is no question of reflecting a concern for welfare equality in the value of social welfare by choosing a non-linear form.

The first question to ask is whether the von Neumann-Morgenstern axioms are acceptable, especially for social choice. Diamond (1967) raised this question effectively, especially questioning the use of the strong independence axiom. Harsanyi (1975) has analysed the issue (pp. 315–8), but seems to me to take little account of Diamond's main concern, viz, that our assessment of alternative policies from an *ethical* point of view, which is what Harsanyi means by 'social preference', may not depend only on the *outcomes* but also on the fairness in the *process* of interpersonal allocation. Harsanyi may well be right in claiming that "when we act on behalf of other people, let alone when we act on behalf of society as a whole, we are under an obligation to follow, if anything *higher* standards of rationality than when we are dealing with our own private affairs" (Harsanyi, 1975, p. 316), but the bone of contention surely is *whether* the strong independence axiom represents a 'higher' standard of rationality in the social context than a rule that takes note of the allocational *process.*

The strong independence axiom is, of course, not the only axiom of the von Neumann-Morgenstern system that has been questioned. The continuity postulate raises difficulties that are well-known, and even the assumption of there being a complete social ordering over all lotteries is a fairly demanding requirement.

But suppose the von Neumann-Morgenstern axiom system is obeyed in social choice as well as in individual choices. In what sense does this rule out non-linear social welfare functions? Obviously, the von Neumann-Morgenstern values – let us call them the V-values – of social welfare will be a linear combination of the V-values of individual welfares. But when someone talks about social welfare being a non-linear function of individual welfares, the reference need not necessarily be to the V-values at all. The V-values are of obvious importance for predicting individual or social choice under uncertainty, but there is no obligation to talk about V-values only whenever one is talking about individual or social welfare.

What gives Harsanyi's (1955) concern with V-values a central role in his own model of social choice, is his concept of 'ethical' preference (the 'social preference of a person') being derived from the *as if* exercise (done by that person) of placing oneself in the position of everyone in the society with equal probability. (Note that Rawls' (1958) concern with 'ignorance' as opposed to 'equi-probability' is different and makes it impossible to define the von Neumann-Morgenstern 'lotteries' for social choice except with some additional axiom, e.g., 'insufficient reason'.) These are lotteries that apply to a person's 'social' (or 'ethical') preference only, and need not figure in his actual preferences – what Harsanyi calls their 'subjective' preferences. It will, of course, still remain true that if the social preference follows the von Neumann-Morgenstern axioms, then welfare numbers W_i will be attributed to individuals in the V-value system for social preference such that social choice will be representable in terms of maximizing $W = \frac{1}{n} \Sigma W_i$. But this linear form asserts very little, since $W(x)$ is simply the value of the lottery of being each person i with $1/n$ probability in state x, and the set of W_i need not necessarily have any other significance.

Consider the following conversation:

1: "Let (x, i) be the position of being person i in state x. Tell me how you would rank $(x, 1)$, $(x, 2)$, $(y, 1)$, $(y, 2)$, please."

2: "The best is $(y, 2)$. Then $(x, 2)$. Then $(x, 1)$. Worst $(y, 1)$."

1: "And the welfare gaps between each pair of adjacent positions? Scale them with $(y, 2)$ being 10 and $(x, 1)$ marked zero."

2: "I can't on weekdays, when I feel ordinal."

1: "So on weekdays you are lost and don't know whether to recommend x or y as your ethical judgement for society?"

2: "No, I would recommend x. I even accept TEP on weekdays."

1: "On weekends you are not so ordinal?"

2: "On weekends, on your normalization, I would put 10 for $(y, 2)$, 5 for $(x, 2)$, 2 for $(x, 1)$ and 0 for $(y, 1)$, though I don't like making the 'origin' quite so arbitrary."

1: "Never mind the origin! Since the welfare sum is 10 with y and 7 with x, you clearly will recommend y on weekends?"

2: "No, no, I would recommend x."

1: "So you don't follow von Neumann-Morgenstern axioms in these choices?"

2: "On Saturdays not. But on Sundays yes."

1: "But on Saturdays what do these cardinal welfare numbers stand for? What meaning can we attach to them since they are not von Neumann-Morgenstern numbers?"

2: "They reflect my views of the welfare levels and gaps. I can axiomatize them in many different ways.[13] The welfare numbers have quite nice properties."

1: "But I can't relate them to your observed behaviour."

2: "I should think not. Nor can I relate your von Neumann-Morgenstern numbers over *interpersonal* choices to your observed behaviour; there is not much to go by. No, these numbers reflect my introspection on the subject as do yours, I presume."

1: "Okay, forget the Saturdays. But on Sundays you say your hypothetical interpersonal choices satisfy the von Neumann-Morgenstern axioms. Then you must choose y since social welfare must be the *sum* of these individual welfare numbers."

2: "No, not of these; social welfare is non-linear over these values. It is linear over the V-values, of course. The V-values, which take my distributional attitude into account (to the extent it is possible to do this within the von Neumann-Morgenstern system), are, with the normalization suggested by you: 10 for $(y, 2)$, 7 for $(x, 2)$, 4 for $(x, 1)$ and 0 for $(y, 1)$."

1: "I am relieved. I thought you were about to take social welfare to be a non-linear function of the V-values in von Neumann-Morgenstern representation."

2: "You must be joking."

1: "Anyway, I am so glad that on Sundays you are a utilitarian as far as V-values are concerned."

2: "I am also glad that your pleasures are inexpensive."

I end this section with two final comments. First, 'Sen's utility-dispersion argument', to which Harsanyi (1975) makes extensive references (pp. 318–324), and which according to him "shows a close formal similarity ... to the view that the utility of a lottery ticket should depend, not only on its *expected (mean) utility*, but also on some measure of *risk*" (p. 320), and which "is an illegitimate transfer of a mathematical relationship for money amount, for which it does hold, to utility levels, for which it does not hold" (p. 321), is – in that form – a figment of Harsanyi's imagination. There is, alas, no 2-parameter "Sen's theory which would make social welfare depend, not only on the mean, but also on some measure of *inequality*, i.e., of *dispersion*" (Harsanyi, 1975, pp. 319–320). More importantly, there is no proposal, which would have been grotesque, to define a non-linear social welfare function on *von Neumann-Morgenstern utilities.*[14] Even the axioms for additive separability of the social welfare function over individuals was explicitly criticised (Sen, 1973, pp. 39–41).

Second, whether we use utilitarianism or not is an important moral issue,[15] and is not disposable by carefully defining individual utilities in such a way that the only operation they are good for is addition. An axiomatic justification of utilitarianism would have more content to it if it started off at a place somewhat more distant from the ultimate destination.[16]

3. AXIOMATIZATION OF THE LEXICOGRAPHIC MAXIMIN RULE

The Rawlsian (1958, 1971) 'maximin' rule ranks social states in terms of the welfare of the worst-off individual in that state. This rule can violate even the Pareto principle. The lexicographic version of the maximin rule (Rawls, 1971; Sen, 1970) does not. This rule, which for brevity, and not out of disrespect, I shall call 'leximin', can be formalized in the following way. Let the worst-off person in state x be called $1(x)$, the second worst-off $2(x)$, and in general the jth worst-off $j(x)$. When there are ties, rank the tied persons in *any* strict order. For an n-person community, for any x, y in X:

(i) $x P y$ if and only if there exists some r: $1 \leqslant r \leqslant n$ such that $(x, i(x)) \tilde{I} (y, i(y))$ for all i: $1 \leqslant i < r$, and $(x, r(x)) \tilde{P} (y, r(y))$;

(ii) $x I y$ if and only if $(x, i(x)) \tilde{I} (y, i(y))$ for all i: $1 \leqslant i \leqslant n$.

Leximin has been recently illuminatingly analysed in axiomatic terms by Hammond (1975), d'Aspremont and Gevers (1975) and Strasnick (1975). The axiomatization presented here is on similar lines but it differs in some important respects. In particular, the strategy adopted here is first to propose axioms such that the lexicographic maximin rule emerges for 2-person communities, and then to ensure by additional axioms that the lexicographic maximin rule holding for 2-person communities should guarantee the same for n-person communities.

There are, it seems to me, two advantages in this procedure. First, in the 2-person case the axioms are easier to assess and the proof of the theorem is extremely brief. It is my belief that the rationale of the leximin comes out best in this case, and it is worth noting that. Second, this procedure permits the isolation of what appears to me to be the least acceptable feature of leximin, which emerges in the move from 2-person leximin to n-person leximin. The issues raised are discussed in Section 4.

Consider first a 2-person community with persons 1 and 2. The following axioms are defined for a GSWF (generalized social welfare function): $R = f(\tilde{R})$, where R is the social ordering over X and $\widetilde{R}$ the extended ordering over the product of X and H.

U(Unrestricted domain): Any logically possible $\widetilde{R}$ is in the domain of f.

I(Independence of irrelevant alternatives): If the restrictions of $\widetilde{R}$ and $\widetilde{R}'$ on any pair in X are the same, then the restrictions of $f(\widetilde{R})$ and $f(\widetilde{R}')$ on that pair are also the same.

J(Grading principle of justice): For any x, y in X, if for some one-to-one correspondence μ from (1,2) to (1,2): $(x, 1)\,\widetilde{R}\,(y, \mu(1))$ and $(x, 2)\,\widetilde{R}\,(y, \mu(2))$, then $x\,R\,y$. If, furthermore, one of the two $\widetilde{R}$'s is a $\widetilde{P}$, then $x\,P\,y$.

T(Two-person ordinal equity): For any x, y in X, if one person, say 1, prefers x to y, and the other prefers y to x, and if person 1 is worse off than 2 both in x and in y, then $x\,R\,y$.
U and I are standard parts of the Arrow framework applied to extended orderings for a 2-person community. J is proposed by Suppes (1966). T corresponds to Hammond's Equity Axiom E in the two-person case, without the separability requirement built into it in the n-person case (for $n > 2$). It corresponds to E in the same way as TEP corresponds to STEP.

(T.2) For a 2-person community, given at least three social states in X, leximin is the only generalized social welfare function satisfying U, I, J and T.

Proof. Since it is easily checked that leximin satisfies U, I, J and T, we need concentrate only on the converse. Suppose U, I, J and T are satisfied, but not leximin. Leximin can be violated in one of three alternative ways. For some x, y in X:

(I) $(x, 1(x))\,\widetilde{P}\,(y, 1(y))$, but not $x\,P\,y$.
(II) $(x, 1(x))\,\widetilde{I}\,(y, 1(y))$ and $(x, 2(x))\,\widetilde{P}\,(y, 2(y))$, but not $x\,P\,y$.
(III) $(x, i(x))\,\widetilde{I}\,(y, i(y))$ for $i = 1, 2$, but not $x\,I\,y$.

Since (II) and (III) contradict J directly, we need be concerned only with (I). Suppose (I) holds.

If $(x, 2(x))\,\widetilde{R}\,(y, 2(y))$, then $x\,P\,y$ by J. Hence it must be the case that $(y, 2(y))\,\widetilde{P}\,(x, 2(x))$. *A fortiori*, $(y, 2(y))\,P(x, 1(x))$. Consider now $\widetilde{R}'$ reflecting the following strict descending order involving x, y and a third social state z:

$$(y, 2(y)), (x, 2(x)), (x, 1(x)), (z, 2(z)), (z, 1(z)), (y, 1(y)).$$

By T and J, $z R' y$, where $R' = f(\widetilde{R}')$, and by $J, x P' z$. Hence $x P' y$. But then by $I, x P y$. So (I) is impossible. This establishes (T.2).

Consider now a family of GSWFs, one for each subset of the community H. In what follows only those for pairs and for H will be used.

In addition to the axioms for 2-person communities, two axioms with a wider scope are now introduced. Let the social states x and y be called 'rank-equivalent' for $\widetilde{R}$ if everyone's relative welfare rank is the same in x as in y, i.e., $i(x) = i(y)$ for all i.

B(Binary build-up): For any $\widetilde{R}$, for any two rank-equivalent social states, for a set π of pairs of individuals in the community H such that $\cup \pi = H$, if $x R y$ (respectively, $x P y$) for each pair in π, then $x R y$ (respectively, $x P y$) for H.

J^* (Extended grading principle): If $\widetilde{R}'$ is obtained from $\widetilde{R}$ by replacing i by $\mu(i)$ in all positions (x, i) for some x and all i, where $\mu(\cdot)$ is a one-to-one correspondence from H to H, then $f(\widetilde{R}) = f(\widetilde{R}')$ for H.

J^* is an extension of J and is in the same spirit. Notice that it is not satisfied by many conditions, e.g., the method of majority decision; the majority method does not satisfy J either. J^* stipulates essential use of interpersonal comparison information in an anonymous way, e.g., taking note of $(x, i) \widetilde{R} (y, i)$ but in the same way as $(x, i) \widetilde{R} (y, k)$.

(T.3) Given at least three social states, if for each pair of persons in H, there is a 2-person GSWF satisfying U, I, J and T, then the only GSWF for H satisfying U, J^* and B is leximin.

Proof. It is clear from (T.2) that the GSWF for each pair of individuals is leximin. If the community H has only two members, then (T.3) is trivial. In general for any community H, leximin clearly satisfies U and J^*. It remains to be established that it must satisfy B also, and then to establish the converse proposition.

Suppose the GSWF is leximin, but B is violated. This is possible only if x and y are rank-equivalent, and

(I) $x\,R\,y$ for all pairs in π, but not $x\,R\,y$ for H.
(II) $x\,P\,y$ for all pairs in π, but not $x\,P\,y$ for H.

Consider (I) first. Since each R is an ordering, $y\,P\,x$ must hold for H. Given the leximin nature of the GSWF for H, this is possible only if there is some rank r such that: $(y, r(y))\,\tilde{P}\,(x, r(x))$, and $(y, i(y))\,\tilde{I}\,(x, i(x))$, for all $i < r$. Given rank-equivalence, $i(x) = i(y) = i$, say, for all i. Thus: $(y, r)\,\tilde{P}\,(x, r)$, and $(y, i)\,\tilde{I}\,(x, i)$, for all $i < r$. Since r must belong to at least one pair included in π, for that pair, by leximin, $y\,P\,x$. So the supposition (I) leads to contradiction.

Next consider (II). If not $x\,P\,y$ for H, then either $y\,P\,x$, which leads to the same problem as (I), or $x\,I\,y$, which is now considered. For leximin this implies, given rank equivalence, $(x, i)\,\tilde{I}\,(y, i)$ for all i. Clearly then $x\,P\,y$ is impossible for any pair contained in π, thus contradicting (II).

Now the converse. Let the stated axioms hold. To establish that the GSWF for the community H must be leximin, we have to show that:

(III) If $(x, i(x))\,\tilde{I}\,(y, i(y))$ for all i, then $x\,I\,y$ for H.
(IV) If there is some r such that $(x, r(x))\,\tilde{P}\,(y, r(y))$, and for all $i < r$: $(x, i(x))\,\tilde{I}\,(y, i(y))$, then $x\,P\,y$ for H.

Let the antecedent in (III) hold. Take the one-to-one correspondence μ such that $i(x) = \mu(i(y))$ for all i, and let this transformation applied to the y-invariant elements convert $\tilde{R}$ to $\tilde{R}'$. Note that x and y are rank-equivalent for $\tilde{R}'$. Note also that $f(\tilde{R}) = f(\tilde{R}')$ for all subset of H by J^*. Consider now any set π of pairs of persons in H such that $\cup\pi = H$. Leximin guarantees $x\,I'\,y$ for all such pairs with $R' = f(\tilde{R}')$. By Binary build-up B: $x\,I'\,y$ for H. By J^*: $x\,I\,y$.

Finally, let the antecedent of (IV) hold. Consider μ and $\tilde{R}'$ as defined in the last paragraph with x and y rank-equivalent for $\tilde{R}'$. Consider now the set π of pairs $(r(x), i)$ for all $i \neq r(x)$. Since the GSWF for each pair is leximin, clearly $x\,P'\,y$ for all pairs in π. Furthermore $\cup\pi = H$. Hence $x\,P'\,y$ for H by Binary build-up. By J^*: $x\,P\,y$, which completes the proof.

4. RAWLSIAN AXIOMS: A DISCRIMINATING ASSESSMENT

The axiom structure used in the last section to derive the Rawlsian leximin rule was not chosen to provide an axiomatic 'justification' of the rule. Rawls himself did not seek such a justification (see especially his 'Concluding Remarks on Justification', Rawls (1971, pp. 577–587)), and was much more concerned with being able 'to see more clearly the chief structural features' of the approach chosen by him (p. viii). In (T.2) and (T.3) axioms have been chosen with a view to distinguishing between different aspects of Rawlsian ethics, which would permit a discriminating evaluation.

Before examining the axioms one by one, it is important to clarify the type of aggregation that is involved in the exercise as a whole. Social choice problems can be broadly divided into the aggregation of personal 'interests' and that of 'judgements' as to what is good for society, and as I have tried to argue elsewhere (Sen, 1975), the 'theory of social choice' seems to have suffered persistently from a failure to make clear which particular problem is being tackled. It seems reasonable to take leximin as a proposed solution to the interest aggregation exercise. The contrast between giving priority to the welfare ranking of the worst off person as opposed to the welfare ranking of the person who 'gains more' (as under utilitarianism) is a contrast between two alternative approaches to dealing with interest conflicts. The problem of aggregating people's different judgements on what should be done (e.g., aggregating different 'views' on the 'right' public policy), while central to Arrow's (1951) analysis of social choice, is not a problem to which leximin can be sensibly addressed.

It is perhaps easiest to think of a generalized social welfare function GSWF as an exercise by a person of deriving ethical judgements from his assessment of everyone's interests implicit in the particular $\widetilde{R}$ in terms of which he does the exercise. (This is the sense in which Harsanyi (1955) also uses 'social preference': "When I speak of preferences 'from a social standpoint', often abbreviated to social preferences and the like, I always mean preferences based on a given individual's value judgements concerning social welfare" (p. 310).) The exercise can be institutional also, e.g., taking a person with a lower money income to be invariably worse off as a 'stylized' assumption in a poverty programme (see Atkinson, 1969). These exercises are done with one given $\widetilde{R}$ in each case. The problem of basing a 'social judgement' on the n-tuple of 'extended orderings' $\{\widetilde{R}_i\}$ – one for each person – is a different issue, raising problems of its own (see Sen, 1970, Theorems 9*2, 9*2.1, and 9*3, pp. 154–156, and Kelly, 1975, 1975a).

The fact that a GSWF is defined as a function of $\widetilde{R}$, an extended *ordering,* without any information on preference intensities, is of some importance, since this rules out the possibility of varying the ethical judgements with cardinalization. Formally the axiom in question is 'unrestricted domain' U, since if cardinalization made any difference in any particular case, R will not be a *function* of $\tilde{R}$ in that case, and such an $\tilde{R}$ will not be an element in the domain of $f(\cdot)$. However, even if R were not defined as a function of $\tilde{R}$, and the possibility of using intensities of welfare differences were kept open, no essential difference will be made in the axiom structure used in the derivation of leximin. In the 2-person case in (T.2), axioms J(Grading principle of justice) and T(Two-person ordinal equity) along with I (independence of irrelevant alternatives) do 'lock' the social preferences leaving no room for cardinal intensities to exert themselves. In (T.3) there is formally a bit of room which is, however, easily absorbed by a slight variation of the axioms. For the responsibility of elimination of intensity considerations we must critically examine axioms other than U.

Axiom J (Grading principle of justice) is, however, quite harmless in this respect since it operates on utilitizing dominance. Indeed, the preference relation generated by J is not merely a subrelation of the Rawlsian leximin relation, it is a subrelation also of the utilitarian preference relation (see Sen, 1970, pp. 159–160).[17] J incorporates the Pareto relation but also all similar dominance relations obtained through interpersonal permutations.[18]

The eschewal of intensities of welfare differences as relevant considerations is, however, an important aspect of T(Two-person ordinal equity). There is no dominance here, and person i's preference for x over y is made to override j's for y over x if i is worse off in each of the two states without any reference to the relative magnitudes of i's gain and j's loss. This was one reason for our hesitation with TEP also (Section 1) and the same applies to T. Both give priority to reducing welfare inequality in the ordinal sense without any concern for 'totals' and for comparing welfare 'differences'. If cardinal welfare comparisons were ruled out *either* because of ordinality of individual welfare, *or* because of level comparability, T like TEP would be a lot more persuasive. In so far as T is a crucial aspect of the Rawlsian approach, this point about 'informational contrast' is of great importance.[19] The hazier our notion of welfare differences, the less the bite of this criticism of the Rawlsian rules.

The usual criticisms of Arrow's use of the independence of irrelevant alternatives (see Sen, 1970, pp. 39, 89–92, and Hansson, 1973), also apply to the use of I for a GSWF. It rules out postulating cardinal measures of intensity based on rank positions (as in 'positional rules' discussed by Gärdenfors (1973), Fine and Fine (1974) and others). This, as it were, puts the last nail in the coffin of using welfare difference intensities.[20]

Turning now to (T.3), axioms J^* and B have to be considered. J^* uses the interpersonal permutation approach pioneered by Suppes (1966). While J uses it for 'dominance' only, J^* uses this more generally, to the extent of not discriminating between two extended orderings where positions of individuals for some social state are switched around. The objections that apply to the usual 'anonymity' postulates apply here too, and it is particularly serious when considerations of personal liberty are involved (see Sen, 1970, Chapter 6, 1975b; Kelly, 1975).[21]

Binary build-up B is in some ways the least persuasive of the axioms used. It permits a lexicographic pattern of dictatorship of *positions* (being least well off) as opposed to persons (as in Arrow's (1951) theorem). The worst off person rules the roost not merely in a 2-person community, but in a community of any size no matter how many persons' interests go against his.[22]

Leximin can be derived without using B (see Hammond, 1975; d'Aspremont and Gevers, 1975 and Strasnick, 1975), and using instead conditions that look less narrow in their focus (e.g., d'Aspremont and and Gevers' 'Elimination of Indifferent Individuals'). However, leximin must satisfy B, as we establish in (T.3). And no matter how we 'derive' leximin, B is an integral part of the Rawlsian set-up. This seems to bring out a rather disagreeable feature of leximin. In a 2-person community in the absence of information on welfare difference intensities, it might seem reasonable to argue that the worse off person's preference should have priority over the other's, but does this really make sense in a billion-person community even if everyone else's interests go against that of this one person? The transition from 2-person leximin to n-person leximin (making use of Binary build-up) is a long one.

It may be interesting to observe how Binary build-up creeps into axioms that look rather mild. Consider Hammond's Equity axiom E. It differs from the 2-person equity axiom T used here in being extendable to n-member communities also if all others are *indifferent* between x and y. This may look innocuous enough, but in the presence of U and I, this 'separability' assumption is quite overpowering. The 'elimination of indifferent individuals', as d'Aspremont and Gevers (1975) call it, is not merely (as it happens) spine-chilling in the choice of words, but also quite disturbing in its real implications. The condition is defined below in a somewhat different form (to permit ready comparability with Hammond's E).

EL (Elimination of the influence of indifferent individuals): For any $\tilde{R}$, for any x, y, if $(x, i)\,\tilde{I}\,(y, i)$ for all i in some subset G of H, and if f^H and f^{H-G} are the GSWF's for the communities H and $(H-G)$ respectively, then $x\, f^H(\tilde{R})$ y if and only if $x\, f^{H-G}(\tilde{R})\, y$.

Notice that our T (Two-person ordinal equity) and EL together imply Hammond's Equity axiom E. What may, however, not be obvious is that for the class of GSWF satisfying unrestricted domain (U) and independence (I), T and EL together eliminate the influence not merely of indifferent individuals but also of non-indifferent ones, leading to a single-minded concern with one person. The same effect is achieved by Hammond's E itself in the presence of the other axioms.

SFE(n) *(Single-focus equity for n-member communities)*: If for an n-member community, for any x and y, $\tilde{R}$ involves the strict extended order: (y, n), $(x, n), (y, n-1), (x, n-1), \ldots, (y, 2), (x, 2), (x, 1), (y, 1)$, then $x\, R\, y$.

SFE(2) is equivalent to two-person equity T (and trivially to Hammond's E also), and may not be thought to be exceptionally objectionable (especially in the absence of preference intensity information). But SFE(n) for relatively large n is very extreme indeed, since everyone other than 1 is better off under y than under x, and still $x\, R\, y$.

(T.4) Given at least three social states, if U, I and EL hold for the GSWF for each subset of the community H, then T implies SFE(k) for GSWF for each subset of the community including the one for the entire community H (i.e., $k \leqslant n$).

Proof. Suppose SFE(m) holds for some $m < n$. We first show that SFE ($m+1$) holds. Consider the following extended order (with indifference written as =) for the triple x, y and some z: $(y, m+1), (x, m+1) = (z, m+1)$, $(y, m) = (z, m), (x, m), (y, m-1), (z, m-1), (x, m-1), \ldots, (y, 2), (z, 2), (x, 2)$, $(x, 1)$, $(z, 1)$, $(y, 1)$. By U, this is admissible. By SFE(m), for the m-member community $(1, \ldots, m)$, $x\, R\, z$. By EL, for the $(m+1)$-member community $(1, \ldots, m+1)$, $x\, R\, z$ also. Again, by SFE(m), for the m-member community

$(1, ..., m-1, m+1)$, $z R y$. By EL, for the $(m+1)$-member community $(1, ..., m+1)$, $z R y$ also. By the transitivity of R, $x R y$. By independence I, this must be due solely to the restriction of $\tilde{R}$ over the pair (x, y), and this establishes SFE$(m+1)$.

The proof is completed by noting that SFE(2) holds since it is equivalent to Two-person equity T, and then obtaining SFE(k) for all $k \leqslant n$ by induction.

While the Rawlsian leximin can be established from axioms that look more appealing, it must end up having the extreme narrowness of focus that is represented by SFE(n) for large n. This in itself is obvious enough, since leximin clearly does satisfy SFE(n). What (T.4) does is to show precisely how it comes about that such apparently broad-focussed conditions together produce such a narrow-focussed property.

It should be observed that the force of the Rawlsian approach as a critique of utilitarian ethics stands despite the limitation of SFE(n). SFE(2) is equivalent to T – an appealing requirement – and while Rawlsian leximin satisfies it, utilitarianism may not. As large values of n are considered, SFE(n) becomes less appealing and so does – naturally – Rawlsian leximin, but the criticism of utilitarianism is not thereby wiped out. In this paper a 'warts-and-all' view of Rawlsian leximin has been taken, choosing a set of axioms with the focus on transparency rather than on immediate appeal. This 'warts and all' axiomatization does not, however, give any reason for disagreeing with Rawls' (1971) own conclusions about his theory: (i) "it is not a fully satisfactory theory", and (ii) "it offers ... an alternative to the utilitarian view which has for so long held the pre-eminent place in our moral philosophy" (p. 586). Rawls was, of course, referring to his theory in its broad form including his contractual notion of fairness and justice, but the observations seem to apply specifically to leximin as well.

London School of Economics

NOTES

* Based on the text of a lecture delivered at the International Congress of Logic, Methodology and Philosophy of Science in London, Ontario, Canada, on August 29th, 1975. Thanks are due to Peter Hammond and Kevin Roberts for helpful comments and criticisms.

[1] See Hansson (1975).

[2] Cf. the concept of 'ordinal intensity' in comparisons of preference intensity used by Blau (1975) and Sen (1975b).
[3] See Sen (1970), Chapters 9 and 9*, for a discussion of the concept of extended ordering.
[4] Note that this criterion permits $(x, \rho(i))\ \tilde{I}(x, \rho(j))$, in contrast with the examples noted above. But obviously if $\tilde{I}$ holds, then there is no inequality in x at all, and this must be less than whatever inequality there is in y.
[5] See Dasgupta *et al.* (1973), or Sen (1973), pp. 53–56.
[6] Note that STEP subsumes Hammond's (1975) axiom of 'Equity' (E), which extends and generalizes Sen's (1973) 'weak equity axiom' (WEA). Preferring inequality reduction irrespective of any consideration of the 'total' (as in STEP) has the effect of giving priority to the person who is going to be worse off anyway (as in these equity axioms).
[7] Note, however, that the characteristic of 'separability' is shared by STEP with many other criteria, e.g., utilitarianism, or the lexicographic maximin. Cf. the condition of 'elimination of indifferent individuals' (El) of d'Aspremont and Gevers (1975).
[8] In the special case of 'pure distribution' problem however, there will be no conflict between the welfare sum and the equality of the welfare distribution if everyone shares the same concave welfare function on individual income. This is, however, a very special case, and the condition can be somewhat relaxed without introducing a conflict; on this see Hammond (1975a).
[9] See Sen (1970, 1973), d'Aspremont and Gevers (1975), and Hammond (1975).
[10] However, even with *partial* unit comparability there will be a quasi-ordering of the total (see Sen, 1970; Fine, 1975; Blackorby, 1975). With level comparability only, this quasi-ordering will shrink only to the weak n-person version of Suppes' 'grading principle of justice', which will never contradict θ or θ^*.
[11] Contrast Nozick (1973). See also Williams (1973, pp. 77–93).
[12] Indeed the interpretation of Harsanyi's (1955) own theorems on social choice is seriously hampered by his silence on the precise comparability assumption (on which and related issues, see Pattanaik (1968)).
[13] See, for example, Krantz *et al.* (1971).
[14] The only page reference Harsanyi gives on this point, which he discusses so extensively, is to p. 18 of Sen (1973). I see nothing there that justifies Harsanyi's presumption that I had non-linear designs on utilities in the *von Neumann-Morgenstern representation*, let alone a 2-parameter non-linear design on them.
[15] For an illuminating debate on this see Smart and Williams (1973).
[16] For some extremely interesting recent contributions in this direction, see d'Aspremont and Gevers (1975), Hammond (1975a), and Maskin (1975), even though more work may still need to be done in terms of starting off from individual welfare functions which are not necessarily confined precisely to the class of positive affine transformations.
[17] Blackorby and Donaldson (1975) demonstrate that with cardinal interpersonal comparability the convex hull of the 'at least as good as' set according to the grading principle is a subset of the intersection of the utilitarian and leximin 'at least as good as' sets, and in the 2-person case exactly equals the intersection.
[18] Hammond's axioms for leximin include the symmetric part of the grading relation (his S) as well as the asymmetric part in the case coinciding with the Pareto strict preference (his P^*), but the remainder of J follows from his remaining axioms U, I, P^* and S (as he notes in Theorem 5.1).
[19] See Sen (1973, 1974), and d'Aspremont and Gevers (1975).
[20] Furthermore, combined with the 'separability' assumption implicit in Binary build-up B (or in Hammond's E, or in d'Aspremont and Gevers' EL, to be defined below), independence of irrelevant alternatives can be very demanding from the point of view of inter-pair consistency (see (T.4) below). In this its role here is not dissimilar to that in

the class of possibility theorems on social welfare functions without interpersonal comparisons.

[21] In this sense it seems a bit misleading to call Rawls' theory a 'liberal theory of justice' (see, for example, Barry (1973), which is very helpful contribution otherwise).

[22] It should be remarked that utilitarianism does not satisfy Binary build-up B. For example it is possible that in a 3-person community with 1 preferring x to y and the others y to x that $W_1(x) - W_1(y) > W_i(y) - W_i(x)$, for $i = 2, 3$, but $W_1(x) - W_1(y) < \sum_{i=2,3} [W_i(y) - W_i(x)]$. So $x P y$ for the 2-person communities (1, 2) and (1, 3), but not for their union (1, 2, 3). However, utilitarianism satisfies another – and in some ways weaker – binary build-up condition B^*, viz, if $x R y$ for a set of pairs which *partition* the community (with no person belonging to more than one pair), then $x R y$ for the community (and similarly with P). Strasnick's (1975) condition of 'unanimity' incorporates B^* for any partition of the community (not necessarily into pairs), and is more reasonable and much less demanding than B, in this sense.

BIBLIOGRAPHY

Arrow, K. J., 1951, *Social Choice and Individual Values*, Wiley, New York, 2nd edition, 1963.

Arrow, K. J., (1973, 'Some Ordinalist-Utilitarian Notes on Rawls' Theory of Justice', *Journal of Philosophy* **70**.

Atkinson, A. B., 1969, *Poverty in Britain and the Reform of Social Security*, Cambridge University Press.

Atkinson, A. B., 'On the Measurement of Inequality', *Journal of Economic Theory* **2**.

Barry, B., 1973, *The Liberal Theory of Justice*, Clarendon Press, Oxford.

Blackorby, C., 1975, 'Degrees of Cardinality and Aggregate Partial Orderings', *Econometrica* **43**.

Blackorby, C. and Donaldson, D., 1975, 'Utility vs. Equity: Some Plausible Quasi-orderings', mimeographed.

Blau, J. H., (1975), 'Liberal Values and Independence', *Review of Economic Studies* **42**.

Daniels, N. (ed.), 1975, *Reading Rawls*, Basil Blackwell, Oxford.

Dasgupta, P., Sen, A., and Starret, D., 1973, 'Notes on the Measurement of Inequality of Incomes', *Journal of Economic Theory* **6**.

d'Aspremont, C. and Gevers, L., 1975, 'Equity and the Informational Basis of Collective Choice', presented at the Third World Econometric Congress, mimeographed.

Diamond, P., 1967, 'Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility: A Comment', *Journal of Political Economy* **61**.

Fine, B., 1975, 'A Note on "Interpersonal Comparison and Partial Comparability" ', *Econometrica* **43**.

Fine, B. and Fine, K., 1974, 'Social Choice and Individual Ranking', *Review of Economic Studies* **42**.

Gärdenfors, P., 1973, 'Positionalist Voting Functions', *Theory and Decision* **4**.

Hammond, P. J., 1975, 'Equity, Arrow's Conditions and Rawl's Difference Principle', mimeographed; forthcoming *Econometrica*.

Hammond, P. J., 1975a, 'Dual Interpersonal Comparisons of Utility and the Welfare Economics of Income Distribution', mimeographed, Essex University.

Hansson, B., 1973, 'The Independence Condition in the Theory of Social Choice', *Theory and Decision* **4**, 25.

Hansson, B., 1975, 'The Measurement of Social Inequality', text of lecture at the Congress of Logic, Philosophy and the Methodology of Science, London, Ontario.

Hardy, G., Littlewood, J. and Polya, G., 1934, *Inequalities*, Cambridge University Press.
Harsanyi, J.C., 1955, 'Cardinal Welfare, Individualistic Ethics and Interpersonal Comparisons of Utility', *Journal of Political Economy* 63.
Harsanyi, J.C., 1976, 'Nonlinear Social Welfare Functions', *Theory and Decision* 6.
Kelly, J.S., 1975, 'The Impossibility of a Just Liberal', mimeographed; forthcoming *Economica*.
Kelly, J.S., 1975a, 'Arrow's Theorem, Grading Principle of Justice and the Axiom of Identity', mimeographed, University of Minnesota.
Kolm, S.Ch., 1966, 'The Optimum Production of Social Justice', paper presented in the Biarritz Conference on Public Economics; published in J. Margolis (ed.), *Public Economics*, Macmillan, London.
Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A., 1971, *Foundations of Measurement*, Academic Press, New York.
Maskin, E., 1975, 'A Theorem on Utilitarianism', mimeographed, Cambridge University.
Nozick, R., 1973, *Anarchy, State and Utopia*, Basil Blackwell, Oxford.
Pattanaik, P.K., 1968, 'Risk, Impersonality and the Social Welfare Function', *Journal of Political Economy* 76.
Pattanaik, P.K., 1971, *Voting and Collective Choice*, Cambridge University Press.
Phelps, E.S. (ed.), 1973, *Economic Justice*, Penguin, Harmondsworth.
Rawls, J., 1958, 'Justice as Fairness', *Philosophical Review* 67.
Rawls, J., 1971, *A Theory of Justice*, Harvard University Press, Cambridge, Mass., and Clarendon Press, Oxford.
Rothschild, M. and Stiglitz, J.E., 'Some Further Results on the Measurement of Inequality', *Journal of Economic Theory* 6.
Sen, A.K., 1970, *Collective Choice and Social Welfare*, Holden-Day, San Francisco, and Oliver & Boyd, Edinburgh.
Sen, A.K., 1973, *On Economic Inequality*, Clarendon Press, Oxford, and Norton, New York.
Sen, A.K., 1975, 'Social Choice Theory: A Re-examination', text of lecture at the Third World Econometric Congress; forthcoming *Econometrica*.
Sen, A.K., 1975a, 'Interpersonal Comparisons of Welfare', mimeographed; forthcoming in a festschrift for Tibor Scitovsky.
Sen, A.K., 1975b, 'Liberty, Unanimity and Rights', mimeographed, London School of Economics; forthcoming *Econometrica*.
Smart, J.J.C. and Williams, B., 1973, *Utilitarianism: For and Against*, Cambridge University Press.
Strasnick, S., 1975, 'Arrow's Paradox and Beyond', mimeographed, Harvard University.
Suppes, P., 1966, 'Some Formal Models of Grading Principles', *Synthese* 6; reprinted in P. Suppes, *Studies in the Methodology and Foundations of Science*, Dordrecht, 1969.
Williams, B., 1973, 'A Critique of Utilitarianism', in Smart and Williams (1973).