Bentham or Bergson? Finite Sensibility, Utility Functions and Social Welfare Functions
Author(s): Yew-Kwang Ng
Published by: Oxford University Press
Stable URL: http://www.jstor.org/stable/2296793
Accessed: 02/02/2015 01:02

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of
content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms
of scholarship. For more information about JSTOR, please contact support@jstor.org.

Oxford University Press and The Review of Economic Studies, Ltd. are collaborating with JSTOR to digitize,
preserve and extend access to The Review of Economic Studies.
Bentham or Bergson? Finite Sensibility, Utility Functions and Social Welfare Functions 1,2

YEW-KWANG NG
Monash University, Melbourne

1. INTRODUCTION AND SUMMARY
Since Bergson's famous paper [6] in 1938, the concept of social welfare function (SWF) has played an increasingly important role in welfare economics. This role, however, has been basically very formalistic. Typically, anything involving interpersonal comparison of utility is regarded as something entailing ethical consideration3 and is relegated to an abstract SWF whose form is unspecified. On the other hand, the attempt to derive a reasonable rule of forming our social ordering from individual preferences has been unsuccessful; the paradox of social choice is as real as in 1951 when Arrow published his celebrated work [4]. Can we derive a particular fully defined SWF based on widely acceptable value judgments? Can the paradox of social choice be resolved, even in principle? This paper attempts to answer both questions in the affirmative.

The central concept used in this paper is that of finite sensibility, the recognition of the fact that human beings are not infinitely discriminative.4 For example, suppose an individual prefers two spoons of sugar (x) to one (y) in his coffee. If we increase the amount of sugar continuously from one spoon, we will reach a point y' (say 1.8 spoons) for which the individual cannot tell the difference between x and y'. There may exist another point y" (say 1.6 spoons) for which the individual is indifferent to y' but he prefers x to y''. Hence, with finite sensibility, a perfectly rational individual may have intransitive indifference.

This concept of finite sensibility is not my discovery. It was touched on as far back as 1781 by Borda [7] and in 1881 by Edgeworth [9]. Edgeworth called it "minimum sensible" and took it as axiomatic, or, in his words "a first principle incapable of proof", that the "minimum sensible" or the just-perceivable increment of pleasure, of all pleasures for all persons, are equitable [9, pp. 7ff., 60ff.]. In 1951, Armstrong [2] gave a more elaborate discussion of a similar concept which he called "marginal preference". In psychological literature, this concept is usually called "just noticeable difference".5

1 First version received April 1974; final version accepted February 1975 (Eds.).
2 I am grateful to Nuffield Foundation for a grant for my visit to Nuffield College for the academic year 1973/74 (on leave from the University of New England) during which this paper was drafted. I have also benefited from seminar discussions at Oxford, Columbia, Harvard and Stanford Universities and wish to thank specifically Kenneth Arrow, Avinash Dixit, Murray Kemp, Jim Mirrlees, Avner Shaked, Mendel Weisser and the anonymous referees for very helpful comments.
3 I have argued elsewhere [27] that judgments involving interpersonal comparison of utility are not value judgments but subjective judgments of fact, and that economists are more qualified in making those subjective judgments of fact that are closely related to their field of study.
4 Infinite sensibility contradicts both common sense and psychological studies.
5 It will be seen later (Section 3) that, with the usual continuity assumption, for any alternative x, there does not exist an alternative y which is just marginally preferred to x. However, though we do not have "minimum sensible", "marginal preference" or "just noticeable difference", we still have the sister concept of "maximal insensible", "marginal indifference", or "just un-noticeable difference".
The main purpose of the present paper is to show that, employing the concept of finite sensibility, and using a value premise which is weaker than the usual Pareto Criterion, together with some other conventional technical assumptions, our social welfare function is of the Benthamite form of the unweighted summation of individual utilities (Section 4). The ethical acceptability of this result and the value premise is discussed in Section 5. Before we can make use of individual utility functions, we have to prove the existence of these functions (Section 3). This problem arises because it is not obvious that, with intransitivity of indifference, the preferences of an individual can be represented by a utility function.\footnote{Readers not interested in the technical details of such problems and prepared to accept Convention 2 (on p. 552) may, after reading the set of assumptions, skip Section 3 as well as Section 6.} The utility functions used in the proof of the Summation Theorem are not any arbitrary functions but those which satisfy a certain convention (Convention 1 or 2). Section 6 shows however that, if a utility function satisfies certain versions of expected utility maximization, it must also satisfy Convention 1. Section 7 argues that fully cardinal utility and welfare functions with fixed origins can be, at least in principle, meaningfully constructed to serve useful purposes. The implications of the arguments contained in this paper are outlined in Section 8. Though our conclusion that our SWF is of the additive form is not dependent on the practical possibility of measuring the number of units of marginal indifference, the conclusion that it is also of the unweighted form is. This practical measurement is, however, faced with the difficulties of possible false revelation of preferences, lack of feasible alternatives and the problems associated with the explicit introduction of the time element. These difficulties and the methods to overcome them are discussed in the final section.

The additive form of SWF may appear ethically objectionable. People with egalitarian ethics tend to prefer a strictly quasi-concave SWF whose welfare contours are convex to the origin of the utility space. It is argued in the next section that such belief in non-linear SWFs is usually due to what I shall call "utility illusion" which is the tendency of double discounting the social significance attached to the incomes of the well off.

2. UTILITY ILLUSION

"Economists believe that laymen have money illusion; they themselves have utility illusion."

If we spend a few minutes checking the literature of welfare economics, we will easily find out that a typical if not universal shape of social welfare contours is convex to the origin of the utility space. If a utility function is taken just as an ordinal indicator of preferences, then, in a sense, the shape of the welfare contours does not have any implication. A convex contour can be turned into a concave one by appropriate but strictly increasing transformation of the utility functions. However, the fact that welfare contours are drawn, almost without exception, as convex must have some reason. This, I believe, is due to egalitarian ethics. If we do not regard utility as just an ordinal indicator but make it cardinal and interpersonally comparable in some sense, most people still believe that one util to the worse-off is preferable to one util to the well-off. This belief could either be based on some ultimate preference for equality of utility as such or due to "utility illusion".

If we have to draw our welfare contours not on a utility space but on an income space with axes representing the income levels of different individuals (with say, a fixed set of output and pricing policies), most people would agree that the contours should be convex to the origin. This is due to the belief that, for any given individual, a marginal dollar meets more important needs when his income is low than when it is high. Hence, the marginal income of an individual is given a diminishing weight as his income increases. We thus have convexity in the welfare contours (in income space) irrespective of whether we assume equal capacity for enjoyment. If we use utility not only as an ordinal indicator, then we have, for each individual, diminishing marginal utility of income. Given inter-
personal comparison of utility and equal capacity for enjoyment, unequal distribution of a
given amount of total income diminishes total utility by denying more urgent needs and
satisfying less urgent needs. This egalitarian ethic may be, however, carelessly carried
over to the distribution of utilities. Since unequal distribution of income usually implies
unequal distribution of utilities the two are sometimes regarded as equivalent. Thus unequal
distribution of utilities is condemned along with unequal distribution of income.

Consider a simple example. Given a fixed total income of $100, we may prefer
($50 : $50) (i.e. $50 to each of our two individuals) to ($70 : $30), assuming similar capacity
for enjoyment. Then, when asked to choose in terms of utils we may then say that we prefer
(50 utils : 50 utils) to (70 utils : 30 utils), believing that the former is just a more equal
distribution of the same total income as the latter. But, given diminishing marginal utility
of income and similar capacity for enjoyment, the former must involve smaller total income.
If our preference for equality in the distribution of a given total income is based on the
diminishing marginal utility of income, it does not follow that a more unequal distribution
of a larger total income is inferior.

If the objection to unequal distribution of a given total income is based only on the
diminishing marginal utility of income, then the preference for a more equally distributed
but smaller aggregate *utility* over a larger aggregate utility must involve double-counting
(or double-discounting). A larger but less equally distributed total income is already
discounted by reckoning in terms of utilities rather than incomes, with diminishing marginal
utility and interpersonal comparability. If unequal distribution of utilities is again to be
discounted, this second level of discounting cannot be based again on the diminishing
marginal utility of income.

One could, of course, insist on a preference for equality in the distribution not only of
income but also of utilities and regard this as a basic value judgment which does not need
further explanation. It is also true that we cannot prove that a value judgment is incorrect.
However, we can show that certain value judgments imply others or that certain value
judgments imply the rejection of other value judgments [35, p. 59ff.; 29, pp. 20-24]. In
the following (Section 4 in particular), I will show that the acceptance of a very mild value
premise together with some other reasonable assumptions implies that we must be indifferent
with respect to the distribution of utilities and will be interested only in the aggregate of
utilities.1

3. INTRANSITIVE INDIFFERENCE AND THE EXISTENCE OF
UTILITY FUNCTIONS

A problem which has to be solved before we can proceed to the proof of our central Sum-
mation Theorem is the question of representation. With finite sensibility and the associated
intransitive indifference, does there exist, for each individual, a real valued utility function
that represents his preferences? This and related problems have been tackled by a number
of writers. In particular, Luce [22] defines a semiorder (which, in effect, requires that
(A.3), (A.4) and (A.5) below are satisfied). Scott and Suppes [34] prove that a semiorder
over a *finite* set is closed-interval representable (see [13, p. 92]) with constant intervals.
Fishburn [13, Theorem 6] provides necessary and sufficient conditions for a semiorder to
be closed-interval representable. These conditions involve some density requirement as
well as countability whose meaning is not intuitively clear. The approach used below to

---

1 My argument against utility illusion and for an additive SWF does not imply that I am in favour of a
more unequal distribution of income. It is true that studies of optimal income taxation using the Benthamite
SWF as done by Mirrlees [25] and his followers indicate a much lower degree of progressiveness than, say,
the present British system. These results are, however, based on certain assumptions on the particular shape
of the utility functions with given degrees of diminishing marginal utility of income and work incentives.
More important, I suspect, is the disregard (for the sake of mathematical simplicity) of the dependence
of utility on incomes of other people. If this is taken into account, the Benthamite SWF may be consistent
with rather progressive income taxation.
prove representation is to adopt the standard assumption on the traditional concepts of explicit preference and indifference and to show that the “underlying preference” must also satisfy certain conditions which are well known to ensure the existence of a utility function.  

The set of social alternatives is denoted \( X \) and the set of individuals, \( T \). An alternative in \( X \) will be denoted by a small letter such as \( x, y, z, r \), etc. When individual \( i \) prefers \( x \) to \( y \) or is indifferent between them, we write \( xR_iy \). Moreover,

\[
xP_1y \iff xR_1y \land \sim yR_1x; \quad xI_1y \iff xR_1y \land yR_1x.
\]

Obviously, \( I = \text{“indifferent to”} \) and \( P = \text{“preferred to”} \). As usual, “\( \forall \)” is used as “for any”, “\( \exists \)” as “there exists”, “\( \lor \)” as the inclusive “or”, “:\( \vdash \)” as “such that”, “\( \varepsilon \)” as “belongs to”, and “\( \sim \)” as the Euclidean distance in the \( n \)-dimensional Euclidean space \( \mathbb{R}^n \). We are now ready to list the assumptions before commenting on them and proceeding to the proofs.

**Assumptions on Individual Preferences**

(A.1) \( X \) is taken to be \( \mathbb{R}^n \) or a connected subset of \( \mathbb{R}^n \).

(A.2) \( T \) is countable with number of members, \( s \). (The following is taken to apply over the set \( X \) and for all \( i \in T \). Hence the subscript \( i \) and the notations \( e \in X \), and \( \forall i \in T \) will be dropped except where emphasis is required.)

(A.3) **Connectivity:** \( (\forall x, y: x \neq y)(xRy \lor yRx) \).

(A.4) **Reflexivity of Indifference:** \( (\forall x)(xIx) \).

(A.5) **Weak Transitivity:** (i) \( (\forall x, y, z)(xPyPz \Rightarrow xPz) \); (ii) \( (\forall r, x, y, z)(rPxIyPz \lor rPxPyIz \Rightarrow rPz) \).

(A.6) **Continuity:** \( (\forall x) \), the sets \( \{y: yRx\} \) and \( \{y: xRy\} \) are closed.

(A.7) **Finite Sensibility:** \( (\forall x) (\exists \) some positive \( \varepsilon \) such that \( \forall x' \): \( \text{dis} (x', x) < \varepsilon, x'Ix) \).

(A.8) See Appendix II.

The last assumption (A.8) is required to prove the continuity of the underlying preferences. This involves only technical details and is discussed in an appendix. (A.1) is a standard assumption. It does subsume some form of divisibility. But for the general choice problem, we are dealing with all logically possible hypothetical situations. Hence, divisibility is not a very strong assumption to make.\(^2\) We shall, however, discuss a difficulty (practical, rather than technical) caused by indivisibility in a later section. It may also be noted that \( n \), the dimensionality of choice, can be any positive integer. It can comprise any relevant dimension of choice including (but not confined to) the \( ms \) allocations of the \( m \) commodities or services to the \( s \) individuals as well as any “non-economic” variable.

Any given alternative (or social state) \( x \), being a point in the set of alternatives, could mean a complete specification of the \( ms \) allocations plus the amounts of other economic (e.g. public goods) and non-economic variables (e.g. a set of laws and regulations, etc.).

(A.2) is hardly an assumption; by their very nature, individuals are countable (infinity is not ruled out). (A.3) is again standard; it expresses the requirement that, for any two alternatives, an individual has a definite preference (or indifference) between them. Being a condition of sanity, (A.4) needs no comment and (A.5) assumes something less than the traditional full transitivity. It may be noted that (i) of (A.5) is not an independent assumption as (A.4) and (A.5) (ii) \( \Rightarrow \) (i) of (A.5). However, it is retained to show that some of the propositions below (e.g. Proposition 3) depend only on the weaker (i) of (A.5). The standard

---

1 It must be admitted that, when writing the first draft of this paper, I was ignorant of the works of Luce, Fishburn and others which are published mainly in psychology journals. (See [10] for a survey.) Some of the propositions below have been anticipated by these writers. I have not deleted them partly to provide continuity for readers and partly because of the different approach and method of proofs used.

2 If we have indivisibility and \( X \) is a finite set, representability can be proved more readily (see Scott and Suppes [34]). In general, divisibility need not be assumed to prove representation if we make use of Cantor’s Theorem and assume that certain suitable subsets of \( X \) are countable (see [12], pp. 59-61).
assumption (A.6) requires that preference orderings are continuous. This is equivalent to the requirement that a (strict) preference between any two alternatives is not altered if either is altered by sufficiently small amounts \([5, \text{p.}78]\). This assumption is not needed here if one is prepared to assume the continuity of the underlying preferences instead. This is not really unreasonable since the traditional preference ordering involves perfect sensibility and the “underlying” preferences in this model precisely reflect this perfect ordering.

The only non-conventional assumption is (A.7) which is the basic feature of the model. It expresses the requirement that, for any given alternative, if it is altered in any direction, provided the amount of alteration is small enough, the individual will stay indifferent. Due to (A.7), we cannot have full transitivity such that indifference is also transitive. Otherwise all the points in \(X\) will be ranked indifferent to each other.

**Definitions of the Underlying Preferences**

(D.1) \(xBy\iff \exists r : xPrIy\) \(\text{ (B = "Better than".)}\)

(D.2) \(xNy\iff \sim xBy \iff \exists r: xPrIy\) \(\text{ (N = "No better than".)}\)

(D.3) \(xAy\iff xNy \& yNy\) \(\text{ (A = "As good as" or "Strictly indifferent to".)}\)

If \(x\) is preferred to an alternative \(r\) which is in turn indifferent to \(y\), then even if \(x\) itself is not preferred to \(y\) it is in a certain sense better than \(y\). Put it differently, \(xIyIr\), but \(xPr\). The reason why the individual is indifferent between \(x\) and \(y\) is that the difference is so small. If he had infinite sensibility, he would rank \(xPy\) which then leads us to full transitivity. Our notation \(B\) then denotes the “underlying” preference of the individual. \(B\) would be equivalent to \(P\) if the individual had infinite sensibility. Similarly, \(A\) is the counterpart of \(I\). But \(N\) is *not* the counterpart of \(R\) but rather the counterpart of its opposite. One could easily define a counterpart of \(R\) as \(B \lor A\), but this model goes more naturally with \(N\). Nothing of any consequence is involved in using \(N\) instead of \(B \lor A\) as the basic preference relation.

I wish to show that, with the preceding assumptions, there exists a real valued and continuous utility function representing the preference relation \(N\) such that

\[xNy\iff U(x) \leq U(y).\]

To establish this, I first prove some propositions relating to the completeness, transitivity, etc., of \(N\).

**Proposition 1.** \((\forall x, y), \text{one and only one of the following is true: } xPy, yPx, xIy \iff yIx) \text{ (from Connexity).}\)

**Proof.** This proposition follows trivially from (A.3), recognizing the fact that one cannot logically have \(xRy\) and \(\sim xRy\) at the same time.

**Proposition 2.** \((\forall x, y)(xNy\iff yRx) \text{ (from Connexity and Reflexivity).}\)

**Proof.** Suppose the reverse is true, then \(xPy\) (from (P.1), i.e. Proposition 1). Then there exist \(r(= y)\) such that \(xPrIy\) (Reflexivity). This violates the given fact that \(xNy\). So the proposition is true.

**Proposition 3.** The relation \(N\) is transitive, i.e. \((\forall x, y, z)(xNyNz\iff xNz) \text{ (from Connexity, Reflexivity, and Transitivity (i)).}\)

**Proof.** From the definition of \(N\),

\[xNy\iff \exists r : xPrIy\] \(\text{ ... (a)}\)

\[yNz\iff \exists v : yPvIz\] \(\text{ ... (b)}\)

If (P.3) is not true, then

\[\exists u : xPuIz.\] \(\text{ ... (c)}\)
From Connexity, \( yPu \lor uPy \lor yIu \). If \( yPu \), putting \( v = u \) violates (b), since \( uIz \) from (c). If \( uPy \), then \( xPy \) (A.5i), since \( xPu \) from (c). But this (i.e. \( xPy \)) violates the given fact that \( xNy \) (P.2). If \( uIy \), putting \( r = u \) violates (a), since \( xPu \) from (c). We have thus shown that (c) is inconsistent with the given (a) and (b). So the proposition is true.

**Proposition 4.** The relation \( N \) is reflexive, i.e. \( (\forall x)(xNx) \), (from Connexity).

**Proof.** If the proposition is not true, then \( \exists r : xPrIx \), violating (P.1) (derived from Connexity).

**Proposition 5.** The relation \( N \) is complete or connected, i.e. \( (\forall x, y)(xNy \lor yNx) \) (from Connexity and Transitivity (ii)).

**Proof.** If the proposition is not true, then

\[ \exists r : xPrIy, \]
and
\[ \exists v : yPvIx. \]

Then \( xPv \) from (A.5ii). This, together with \( xIv \), violates (P.1).

**Proposition 6.** \( (\forall x) \), the sets \( \{ y : yNx \} \) and \( \{ y : xNy \} \) are closed (from (A.6), (A.8)).

**Proof.** See Appendix.

**Proposition 7.** The underlying preference relation \( N \) of every preference ordering satisfying (A.3)-(A.6), and (A.8), and defined over a connected set can be represented by a continuous utility function with the property \( xNy \iff U(x) \leq U(y) \).

**Proof.** Such a representation is possible if \( N \) is transitive, complete, continuous, (P.3), (P.5), (P.6), and defined over a connected set (A.1). This has been proved by Debreu [8, p. 56] and by Arrow and Hahn [5, p. 87]. (A.3)-(A.6), (A.8) are the only assumptions used to prove (P.3), (P.5) and (P.6). So the proposition is true.

From the property \( xNy \iff U(x) \leq U(y) \), we can also easily derive the following:

\[ xAy \iff U(x) = U(y), \ xBy \iff U(x) > U(y), \ xPy \iff U(x) > U(y). \]

The last statement follows since one can easily show that \( xPy \implies xBy \) (by putting \( r = y \)). On the other hand, \( xIy \) is consistent with any of \( U(x) \equiv U(y) \). However, we can "standardize" this difference by adopting a convention discussed below.

First we have to prove some more propositions where the following notation will be used. \( P^x \) denotes the non-empty set of all points preferred to \( x \). \( B^x \) is the boundary of \( P^x \) with respect to \( X \). \( A^z \) is the set of all points as good as \( z \).

**Proposition 8.** \( (\forall x, z : z \in B^x)(B^x \subseteq A^z) \).

**Proof.** (Fig. 1 assists the presentation but is not necessary for the proof). If \( B^x \) is not a subset of \( A^z \), then there exists a point \( q \in B^x \) but \( q \notin A^z \). Then either \( U(q) > U(z) \) or \( U(z) > U(q) \). Take a point \( z' \) such that \( \text{dis}(z', z) < \text{any positive} e \) and \( z'Px \). This is possible as \( z \) is a boundary point of \( P^z \). Since \( q \in B^x \), \( q \) is not owned by \( P^z \) (continuity). So, either \( xIq \) or \( xPq \). Since \( z'PxPq \lor z'PxPq \) and \( z'Bq \) and \( U(z') > U(q) \). From the continuity of \( U \), \( U(z') \) is arbitrarily close to \( U(z) \). Hence we cannot have \( U(q) > U(z) \). Similarly, by taking a point \( q' \) arbitrarily close to \( q \) and \( q'Px \), we can rule out \( U(q) < U(z) \). So \( U(q) = U(z) \). Hence the proposition is true.

**Proposition 9.** \( (\forall x, y)(yPx \iff U(y) > U(x) > a \land xRy \iff U(y) > U(x) \leq a) \) where \( a \) is some positive number which may be a function of \( x \), and \( U \) is a utility function satisfying (P.7).

**Proof.** Take any point \( x \) (see Fig. 1). If \( P^x \) does not exist, then \( x \) belongs to the bliss set within which no point is contra-preferred to any other point. Take a point \( b \) in this bliss set with the lowest utility indicator \( U(b) \). We can then assign all points in the bliss set...
utility indices between $U(b)$ and $U(b) + a$ such that $xBy \Leftrightarrow U(x) > U(y)$. It is easily seen that (P.9) is satisfied. Now if $P^x$ is not empty, then from the connectedness of $X$ and the continuity of preferences, $B^x$ is not empty. From (A.7) (Finite Sensibility), $x \notin B^x$. Let $z$ be a point in $B^x$ which has a utility indicator no smaller than any other point in $B^x$. (Actually, any point will do, as it is shown in (P.8) that any point in $B^x$ is as good as any other.) Define $U(z) - U(x) = a$. (In general, $a$ may be a function of $x$.) It is easily shown that $a$ is positive. Now take any point $y \in P^x$. From (A.6) (Continuity), any point in $B^x$ is not owned by $P^x$. So, either $xIz$ or $xPz$. Hence, $\exists x : yPxIz$ or $yPz$. Therefore $yBz$. This leads to $U(y) - U(z) > 0$, and therefore $U(y) - U(x) > a$. This shows that $yPz \Rightarrow U(y) - U(x) > a$.

Now, to show the reverse direction. Let $y$ be any point for which $U(y) - U(x) > a$. This gives $U(y) - U(z) > 0$, and hence $yBz$. From the continuity of $N$ (P.6), $yBz'$ where $\text{dis} (z', z) < \varepsilon$ (where $\varepsilon$ is an arbitrarily small positive number), and $z'Px$. The last relation is possible as $z$ is a boundary point of $P^x$. So, $\exists r : yPrIz'$. Since $z'Px$, we have from (A.5i), $yPx$. This shows that $U(y) - U(x) > a \Rightarrow yPx$ and completes the proof of the first part of the proposition. The second part, i.e. $xRy \Leftrightarrow U(y) - U(x) \leq a$, follows immediately from the first part and Connexity (P.1).

**Proposition 10.** $(\forall x, u, z : uBxIz)$ (for any $v$, $vPu \Rightarrow vBz$).

*Proof.* (Figure 1 can still be used as a reference). Since $vPu$ and $\exists r : uPrIz, vPx$ from Transitivity (ii). Hence $vBz$, as $vPxIz$.

**Proposition 11.** $(\forall x, y : xAy)(xPz \Leftrightarrow zPy ; zIx \Leftrightarrow zIy)$.

*Proof.* If the proposition is not true, then $\exists z : xIzPy$ (or the other way round which can be taken care of by changing the notations of $x$ and $y$). Then, from (A.8), $\exists r : xPrIy$. So $xBy$, violating the given fact that $xAy$.

Proposition 11 may also be proved without using (A.8) but by assuming the continuity of the boundary $B^x$ as $x$ moves continuously. In effect, (P.11) amounts to the transitivity of strict indifference.

After proving (P.8)-(P.11), we are now in a position to adopt the following convention. For any utility function $U$ adopt the transformation: $V = f(U)$ where $f$ is continuous and $f' > 0$. Moreover, the transformation is such as to yield,

**Convention 1.** $(\forall x, y)(yPx \Leftrightarrow V(y) - V(x) > a ; xRy \Leftrightarrow V(y) - V(x) \leq a)$ where $a$ is a positive constant (i.e. constant for each individual but may vary over individuals).

The only difference between Convention 1 and (P.9) is that the positive number $a$ is
made independent of \( x \). That this can be done is ensured by (P.10) and (P.11). For any two points \( x, y \), either \( U(x) = U(y) \) or one is larger than the other. If \( U(x) = U(y) \), then \( a^* \) (the \( a \) associated with \( x \)) must equal \( a^* \). Otherwise \( \exists z : xIzP_y \), violating (P.11). If \( U(x) \neq U(y) \) and \( a^* \neq a^* \), we can transform the \( U \) to make \( a^* = a^* \) without affecting the representation. To see more precisely that this is possible, take any initial point (e.g. the worst point) \( x_1 \) and allot to it an arbitrary utility indicator, e.g. \( V(x_1) = 0 \). Take the boundary \( B^{x_1} \). From (A.8), \( B^{x_1} \subset A^{x_2} \) where \( x_2 \in B^{x_1} \). Hence all points on \( B^{x_1} \) may be allotted the same utility index \( V(x_2) = V(x_1) + a = 0 + 1 = 1 \). From (A.7), \( x_1 \notin B^{x_1} \). Hence we can cover the whole of \( X \) by a (possibly infinite) process indicated in Figure 2. For points better than \( x_1 \) but worse than \( x_2 \), we allot them utility indices between 0 and 1. For points better than \( x_2 \) but worse than \( x_3 \), we allot them utility indices between 1 and 2 and such that, if \( z \in B' \), \( U(z) = U(y) + 1 \), and so on. Due to (P.10), it is easy to see that this can be done. It is also easy to verify that for any two points

\[
x, y, \{ yP \times (y)^- (y) - V(x) > 1; xR \times (y) \times V(y) - V(x) \leq 1 \},
\]

which is Convention 1 with \( a = 1 \).

![Figure 2](image-url)

In a later section it will be shown that, the utility function \( V \) satisfying Convention 1 is not only a function representing the underlying preferences, any utility function representing the preferences will satisfy Convention 1, provided that it also satisfies certain assumptions similar to the postulates of the Expected Utility Hypothesis.

Convention 1 does not, however, rule out the possibility that the constant \( a \) may vary over individuals. Nevertheless, once we have Convention 1, it is only a matter of scaling the various individual utility functions to get

**Convention 2.** (\( \forall i, \forall x, y \)) \(( yP \times (y)^- V(i)(y) - V(i)(x) > a; xR \times (y) \times V(i)(y) - V(i)(x) \leq a \), where \( a \) is the same positive constant for all individuals.

**4. FROM UTILITY FUNCTIONS TO SOCIAL WELFARE FUNCTIONS**

After deriving a utility function \( V^i \) for each individual, we now proceed to derive the social welfare function. To do so, we need some more assumptions. If we define social welfare as something that ought to be pursued or maximized, then these assumptions are value premises, albeit very weak ones. Alternatively, we may take these assumptions merely
as the rules for defining a new concept "social welfare". In this case, no value judgment is involved, but the resulting social welfare function has no prescriptive content without the introduction of some value judgment. The former interpretation is a more interesting one and will be adopted in this paper.

Assumptions on Social Welfare

(A.9) General Utilitarianism. Social welfare $W$ is a function of individual utilities alone, i.e. $W = W'(U^1, \ldots, U^n) = W(V^1, \ldots, V^n)$.

(A.10) $W$ is continuous in the $V$'s and also differentiable.

(A.11) $W$ is either quasi-concave or quasi-convex. (Necessary only to rule out non-linearity of $W$ "in the small"; explained below.)

(A.12) Weak Majority Preference Criterion (WMP). For any two alternatives $x$ and $y$, if no individual prefers $y$ to $x$, and (i) if $x$, the number of individuals, is even, at least $s/2$ individuals prefer $x$ to $y$; (ii) if $s$ is odd, at least $(s-1)/2$ individuals prefer $x$ to $y$ and at least another individual's utility level has not decreased, then social welfare is higher in $x$ than in $y$.

SWFs satisfying (A.9) are sometimes called individualistic SWFs. The term "individualistic" is however misleading here since I do not assume that individual utility is a function of his own consumption vector alone. In other words, no form of externality including ethical consideration over other individuals is ruled out as a factor affecting individual preferences. In fact, (A.9) can be derived as an implication of (A.12) and other assumptions. But since (A.9) is very weak, little objection will be raised against its inclusion. (A.10) is also a reasonable assumption; one would not expect a big jump in social welfare with an infinitesimal change in any individual utility level. On the other hand, (A.11) is a bit stronger and also assumes more than is really needed. It requires either a non-increasing or non-decreasing marginal rate of substitution between utility levels for different individuals. In terms of Figure 3, social welfare contours of the form $W_2$, $W_3$ or $W_1$, $W_2$ are acceptable. A contour of the form $W_4$ is ruled out by (A.11). However, it will be seen later that $W_4$ is ruled out without (A.11), and the role of (A.11) is only to rule out such contours as $W_0$ which is linear "in the large" but not "in the small". Hence, what is required out by (A.11) is really very weak.

The Weak Majority Preference Criterion (A.12) requires that preference by a majority (including a bare majority) should prevail over indifference by the minority. This is the basic value assumption of this paper and is defended in the next section. Meanwhile, we proceed to state and prove our theorem.

1 However, with individual preferences (as distinct from satisfaction or happiness) being affected by such ethical consideration, the maximization of social welfare as a function of individual utilities (representing preferences) may not lead to the maximization of social welfare as a function of individual happiness; see [36, pp. 253-4]. In this case, one may wish, in utility calculus, to ask the individual to disclose his preferences based on his own happiness only.

2 Suppose we have some other factor (or some vector of factors) $E$ entering our welfare function, $W = W(V^1, \ldots, V^n, E)$. Using WMP and (A.10), we can show that $E$ does not really enter $W$ by proving that $W_k = \partial W/\partial E = 0$ for all values of $E$ and the $V$'s. Take two alternatives $x$ and $y$ such that $V(y) = V(x) = 0$ for half of the population and $V(x) - V(y) = g + h$ for the other half, where $g$ is a small positive number (if the number of individuals is odd, hold $V$ constant for the last individual.) According to WMP, $W(x) > W(y)$ irrespective of what happens to $E$; whether it increases one thousand times or vanishes altogether, we still have $W(x) > W(y)$. From (A.10), we can make $W(x) - W(y)$ as small (absolutely) as we like if $E$ does not enter $W$. If $E$ does enter $W$ and $W_k$ is not zero throughout, some sufficiently large change in $E$ should reverse the effect of $h$ on $W$ to give $W(x) = W(y)$. But this contradicts WMP. This argument is true at whatever combination of $V$'s we start with. Hence $E$ does not enter $W$.

3 It may be noted that the summation social choice function discussed by Fishburn [14] is a much broader concept than the Benthamite SWF in our theorem below.
Summation Theorem. Any social welfare function satisfying (A.9)-(A.12) must possess the form \( \sum_{i=1}^{s} V^i \) or its positive monotonic transformation.

Proof. Take the case where \( s \) is even. (The case with \( s \) odd can be taken care of by holding one \( V \) constant.) Consider some pair of alternative \( x \) and \( y \) for which \( V^j(y) = V^j(x) + b, b > a \) for all \( j = 1, \ldots, s/2 \); and \( V^k(y) = V^k(x) - a \) for all \( k = s/2 + 1, \ldots, s \). From Convention 2, \( yP_jx \) and \( yP_kx \) for all \( j = 1, \ldots, s/2 \) and \( k = s/2 + 1, \ldots, s \). According to WMP, \( W(y) > W(x) \). More precisely,

\[
W\{V^1(x) + b, \ldots, V^{s/2}(x) + b, V^{s/2+1}(x) - a, \ldots, V^s(x) - a\} > W\{V^1(x), \ldots, V^s(x)\}. \quad \text{(d)}
\]

Similarly we have

\[
W\{V^1(x) + a, \ldots, V^{s/2}(x) + a, V^{s/2+1}(x) - b, \ldots, V^s(x) - b\} < W\{V^1(x), \ldots, V^s(x)\} \quad \text{(e)}
\]

By making \( b \) approach \( a \), the LHS of (d) approaches that of (e) (from (A.10)). But as long as \( b > a \), both the above inequalities hold. Hence, taking the limit, we have

\[
W\{V^1(x) + a, \ldots, V^{s/2}(x) + a, V^{s/2+1}(x) - a, \ldots, V^s(x) - a\} = W\{V^1(x), \ldots, V^s(x)\}. \quad \text{(f)}
\]

By a similar process (subtracting \( a \) from the utility index of each of the second to the \((s/2)\)th as well as the last individual and adding \( a \) to each of the rest), we can derive,

\[
W\{V^1(x) + a, \ldots, V^{s/2}(x) + a, V^{s/2+1}(x) - a, \ldots, V^s(x) - a\} = W\{V^1(x) + 2a, V^{s/2}(x), \ldots, V^{s-1}(x), V^s(x) - 2a\}. \quad \text{(g)}
\]

From (f) and (g), the RHS of (g) must equal that of (f). This equality holds irrespective of the initial point \( x \) and for any two individuals we take to vary the \( V \)'s by plus \( 2a \) and minus \( 2a \). Hence, we have

\[
\int_{\lambda}^{\lambda + 2a} W_{\lambda} \delta V^j = \int_{\mu}^{\mu + 2a} W_{\mu} \delta V^k \quad \text{...(h)}
\]

for all \( j, k \) and all \( \lambda, \mu \), where the two integrals are taken consecutively. This means that, if we increase any \( V \) by \( 2a \) and reduce any other by \( 2a \) while holding all other \( V \)'s constant, we will remain on the same iso-welfare contour. Consider the welfare contour map of Figure 4 over the utility space of \( V^j \) and \( V^k \) with all other \( V \)'s held constant. If \( V^j \) and \( V^k \) are labelled on the same scale with \( a \) represented by the same distance, any welfare contour
passing through any given point (e.g. A) must also pass through all points (i.e. B, C, etc.) on the negatively sloped 45° line with distance \( \sqrt{4a} \) or its multiple from itself (A). Since a utility difference of \( a \) does not give rise to a preference we may regard \( 2a \) as "small". Hence equation (h) requires that the welfare function must be linear "in the large". But equation (h) does not rule out all social welfare functions non-linear "in the small". For example, a function with periodic welfare contours\(^1\) such as those depicted in Figure 4 is consistent with (h). It is at this point that (A.11) is required. Since \( W \) has to be either quasi-concave or quasi-convex, \(- W_j/W_k = \partial V^j/\partial V^k \) with \( W \) and all other \( V \)'s held constant\(^\ast\) either stays constant and/or changes in one direction only as we travel along a given contour. However, equation (h) holds for all \( j, k \) and all values of \( \lambda, \mu \). If \(- W_j/W_k \) changes in any one direction (i.e. either decreases or increases), we cannot have such an equality throughout. So \( W_j/W_k \) must stay constant throughout. From (h) we have, therefore, \( W_i = G \) (a positive constant) for all \( i \).

\[ V^j \]
\[ 2a \]
\[ 45° \]
\[ V^k \]

\[ A \]
\[ B \]
\[ C \]
\[ D \]

\[ \text{FIGURE 4} \]

Now totally differentiate \( W \) as stated in (A.9),

\[ dW = \sum_{i=1}^{s} W_i dV^i. \]  

... (i)

Substitute \( W_i = G \) into (i) and integrate both sides, yielding

\[ W = G \sum_{i=1}^{s} V^i + H, \]  

... (j)

where \( H \) is the sum of the constants of integration. Up to this point, we have not placed any restriction on \( W \) making it a cardinal function. Hence any positive monotonic transformation of \( W \) is also an acceptable substitute, since the ordering is preserved. Thus, we lose no generality by putting \( G = 1 \) and \( H = 0 \), giving \( W = \sum_{i=1}^{s} V^i \). QED

\(^1\) At one stage, I mistakenly believed that equation (h) was sufficient to ensure linearity. The periodic contours were advanced by Mirrlees as a counter-example during a discussion in his workshop on economic theory from which I have benefited greatly.
5. ETHICAL CONSIDERATIONS

The Summation Theorem is no doubt a strong result and many readers will probably try to dismiss some of our assumptions as unacceptable in order to reject our theorem. It seems that all the assumptions except (A.12) are very weak and/or involve purely technical details. The Summation Theorem, therefore, hinges on the acceptability of the Weak Majority Preference Criterion.

The Weak Majority Preference Criterion (WMP) seems to me extremely reasonable. It is a combination of Majority Rule and the Pareto Criterion. Roughly it says that, if at least half of the people say "yes" and no one says "no", then a change must be recommended. As a sufficiency criterion, it is much more acceptable than Majority Rule and the Pareto Criterion. The latter says that, if some individual prefers $x$ to $y$ and no individual prefers $y$ to $x$, then social welfare is higher in $x$ than in $y$. In addition to this, WMP requires that at least half of the population prefers $x$ to $y$. Anyone who accepts either the Pareto Criterion or Majority Rule must logically accept WMP.

It may be objected that, for models with finite sensibility, the Pareto Criterion should really refer to the underlying preferences and not the explicit preferences. The Weak Majority Preference Criterion is not necessarily weaker than this interpretation of the Pareto Criterion. However, I wish to argue that WMP is still a very acceptable criterion on its own.

First, we have to clarify our meaning of "preference". Ideally, as a variable in the SWF, the preference of an individual should refer to his actual feeling of well-being not as an ex-ante revelation of preference or actual choice. The revealed preference may differ from actual well-being due to imperfect foresight, irrational choice, or choice influenced by a regard to other people's welfare. If the people of Australia reveal their preference for an easy money policy ($x$) rather than a cautious policy ($y$), believing that there is a permanent trade-off between unemployment and inflation, and if the trade-off is really only a temporary one, one may not be prepared to say that social welfare is higher in $x$ than in $y$. However, if we are concerned with actual well-being, our concept of preference may be largely non-operational. Hence, we may have to be content mainly with revealed preference and to make adjustments only where divergences are clear and significant. However, the problem of imperfect foresight, etc., is a separate issue which may be abstracted away in this paper. In other words, we may either take "preference" to refer to the actual feeling of well-being and ignore the question of operationality or we may assume that imperfect foresight, etc., does not exist so that revealed preference coincides with actual well-being.

From the above, it is clear that our concept of finite sensibility is different from the threshold in choice as discussed by Georgescu-Roegen [16, 17]. That threshold is influenced by imperfect foresight and is a function of the amount of time allowed for the choice. "The greater this interval of time, the smaller will be the psychological threshold. At the limit when the time of experimenting is infinite, the threshold is zero." [16, p. 152]. In our model, the problems of imperfect foresight, time, etc., have been abstracted away and the threshold exists due to the fundamental limitation in human capability of feeling.

Coming back to consider the acceptability of WMP. In comparing two social states $x$ and $y$, if no individual feels (using the broad sense of the word "feeling") worse off at $x$ and a majority of people feel better off at $x$, there seem to be no grounds for denying that social welfare is higher at $x$. If a majority prefers $x$ to $y$, we may yet choose $y$ provided the minority more strongly prefers $y$ to $x$. But for the case satisfying WMP, there is no individual in the minority who feels any worse off, not to mention "more strongly worse off".

1 I have argued elsewhere [28] that the usual objection to the Pareto Criterion is based on confusing necessity and sufficiency.
If WMP cannot be rejected as it stands, one may like to reject it indirectly by condemning its implication, the Summation Theorem. This theorem implies that a marginal indifference or a discrimination level of each individual has the same effect on social welfare irrespective of his status. This has been objected to by a number of writers [4, pp. 116-118; 35, p. 94; 29, p. 150]. First, consider the following.

"Assume that there are two persons with equal capacity for feeling in the sense that the range between the two extreme levels of well-being—for the sake of convenience we may call these bliss and abject misery—is the same for both persons. But assume that one of the persons has a fine sense of discrimination so that he has a large number of discretion levels between bliss and abject misery whereas the other person has only a few discretion levels between these two extremes. In this case to declare that the social significance of a movement from one discretion level to another is the same for both individuals will be regarded by many as being unfair to the person with fewer discretion levels" [29, p. 150].

If it is true that the second person has much fewer levels of discretion between bliss and misery than the first person, it seems to me contradictory to assume that the range between the two extremes is the same for both persons.

Another objection is valid only against the scheme of interpersonal comparison proposed by Goodman and Markowitz [18]. They define a discretion level in the sense of a marginal preference: "A change from one level to the next represents the minimum difference which is discernible to an individual" [18, p. 259]. However, recognizing the practical difficulty of finding out the exact number of discretion levels, they propose to use the practically available levels of preference to approximate the levels of discretion. For example, if there are only five candidates for the US presidency, the largest discernible number of approximate discretion levels is four. The number of discretion levels thus measured is therefore a very crude measure and is not independent of the availability of other alternatives [4, p. 116; 35, p. 93]. This defect of Goodman-Markowitz's proposal does not apply to the concept of discretion level as such. It is true that the practical difficulty due to the lack of feasible intermediate alternatives is quite real. But I shall argue in a later section that this difficulty can be overcome by using indirect measurement.

Arrow [4, pp. 116-118] also argues against the principle of maximizing the number of discrimination levels by showing that, a small difference in sensibility between individuals will lead to complete inequality in income distribution. This demonstration is based on the assumption that the amount of income differential giving rise to a discrimination level is independent of the level of income. One would expect that, as income is reduced, an individual will be more sensitive to a given change in his income level. (Diminishing marginal utility of income.) Arrow acknowledges that "it would not be difficult to construct examples . . . for which . . . increasing ability to discriminate at lower levels of income is sufficient to prevent complete inequality" [4, p. 118n.] However, he sustains his objection by saying that a moderate difference in sensibility may still lead to a very great inequality of income, even taking account of the diminishing marginal utility of income.

Apart from incentive effects, I doubt that the differences in sensibility will lead to much inequality of income if we apply the Bentham SWF in practice. Psychological studies of pain sensation show that pain thresholds are very close for different individuals (e.g. averaging 230 ± 10 standard variation), as are the number of just noticeable differences. (See Hardy [39], pp. 88, 157.) If there are more differences in the capacity to enjoy income, these are probably due to "learning by doing", and a long-run SWF will take account of that. If some inequality still persists after taking all effects (including externality) into account, I cannot see why this is not an optimal distribution if it maximizes aggregate utility. Many people (with utility illusion, I believe) wish to give more weights to the preferences of the poor than those of the rich. If we change the word "preferences" into "incomes", I could not agree more. But giving different weights to preferences means that we are prepared to have a majority of strict preferences overruled by a minority of indifference. I do not think that this is ethically acceptable.
If we ask ourselves why we wish to give greater weight to the poor, we may have a number of answers. The most obvious one is that incomes of the poor meet more urgent needs. But this is taken care of by reckoning in terms of utilities instead of incomes. Secondly, it may be said that the consumption of the rich is self-defeating due, e.g., to the snob effects, the desire to keep up with the Joneses, etc., while the consumption of the poor (if spent on education and health, e.g.) may have very beneficial long-run effects. This again is taken care of in our model where all forms of externality can be allowed.

The only logical (but not necessarily acceptable) argument I can think of for not accepting the additive SWF of the Summation Theorem is the following. According to such a SWF, the less sensitive person will end up with lower total utility. If we place some ultimate value on equality of utility as such (not due to the undesirable effect of inequality on social harmony, etc.), we may be prepared to accept a larger decrement of utility by the well-off for the sake of a smaller increment in utility to the not-so-well-off. This is apparently a very persuasive argument. What is not generally recognized is that it implies the rejection of the Weak Majority Preference Criterion. To show that the persuasiveness of the above-mentioned argument is more apparent than real, let us look at the problem from a different angle. Abstracting from the issues of incentives, externality, etc., let us start from a position of complete equality in income distribution. If I am less sensitive than Mr A, transferring $x from me to him will not make me any worse off in actual experience of feeling but will make him better off. Why shouldn’t I agree that it be so transferred? I think we tend to emphasize (perhaps correctly) the need for the well-off to have more regard for the worse-off to the (incorrect) neglect of the need for the worse-off to have some regard for the well-off. This is partly why I find the Rawlsian Maximin Criterion of social justice ethically unacceptable. It has some persuasiveness looking from the viewpoint of the well-off since it appeals to the altruistic motive (perhaps partly explained by guilt feelings of the well-off). This might partly explain the overwhelming attention to Rawls in recent years. If we view it from the point of the worse-off and especially the worst-off, the Maximin Criterion becomes very objectionable. First, why should society pay sole attention to the worst-off to the exclusion of the rest of the badly-off? Secondly, if one puts oneself in the position of the worst-off, one will be, I believe, compelled to say that he would not like the society to sacrifice enormous amounts of other people’s welfare for a small improvement in his own well-being. This seems to put advocates of the Rawlsian Criterion into a dilemma. If people are required to sacrifice enormously for the small benefit of the worst-off, similar (not exactly the same) ethics seem to require that the worst-off should not accept such a sacrifice.1

Consider the much-cherished principle, “From each according to his ability; to each according to his needs” (which I personally accept, assuming no disincentive effect). Why doesn’t it read, “An equal amount of work from each; an equal amount of income to each”? If a weak man is tired by four hours of work, it is better for a stronger man to work longer to relieve him. Similarly, if a less sensitive man does not enjoy much the extra purchasing power, it is better that the more sensitive man receives more of it. It is our utility illusion plus perhaps a guilt conscience that prevents us from seeing such a simple analogy.

6. EXPECTED UTILITY AND UTILITY FUNCTIONS

In this section I wish to show that, by adopting a set of assumptions similar to the postulates for the Expected Utility Hypothesis [26, pp. 26-29; 24], we can show that any preference-representing utility function satisfying our version of expected utility will necessarily satisfy

1 It is true that Rawls attempts to justify his maximin criterion by hypothesizing a voluntary contractual arrangement at a preconstitutional stage. At the “original position”, no one knows who will be the worst-off. At this pre-constitutional level, it seems to me irrational to accept the maximin criterion. Moreover, I should think that one will agree that, whoever is the worst-off at the post-constitutional stage, he should not tolerate enormous sacrifice for his small welfare.
Convention 1 (without having to transform $U$ into $V$). In fact, I will present two alternative sets of assumptions, which are both sufficient for the purpose, one defined on the traditional concepts preference and indifference, the other defined on our concepts “better than” and “as good as”.

In the following, a prospect or lottery $L = (x, y; \alpha, \beta)$ denotes the mutually exclusive alternatives $x$ and $y$ with $\alpha (0 \leq \alpha \leq 1)$ the probability of $x$ and $\beta (1-\alpha)$ the probability of $y$. Other lotteries are similarly denoted, e.g. $\beta' = 1-\alpha'$, $\beta^x = 1-\alpha^x$, etc.

Assumptions on Choices Involving Risk

(A.15) $(\forall x, y : xPy)(L^1PL^2 & x > \alpha^x) \Rightarrow \alpha^x \Rightarrow (A.18)$

(A.16) $(\forall x, y, z : xRyRz)(x, z; \alpha, \beta)Iy$ for some $\alpha : 0 \leq \alpha \leq 1$.

(A.17) $(\forall x, y : xIL^a & yIL^b)(x, y; \alpha, \beta)I(L^a, L^b; \alpha, \beta)$.

(A.18) For all $x, y, \{(x, y; \alpha, \beta)I(L^a, L^b; \alpha, \beta) \text{ where } L^a = (b, w; \alpha^x, \beta^x), L^b = (b, w; \alpha^x + \beta^x, \alpha^x + \beta^x)\}.$

(A.19) $(\forall u, x, y, z : uIx & zPy)(u, z; \frac{1}{2}, \frac{1}{2})p(x, y; \frac{1}{2}, \frac{1}{2}).$

Assumptions (A.15)-(A.18) are also used in the Expected Utility Hypothesis. Actually, the ones used here are weaker. For example, (A.15) only assumes one way relation. It says that if $xPy$, then for two lotteries with $x$ and $y$ as the two alternative payoffs, the one lottery which is preferred to the other must have a higher probability of winning $x$. (A.16) assumes some form of continuity. If $xPy$, then there exists some probability mix of $x$ and $x$ which is indifferent to $y$. The objection to this postulate by Alchian, [1] has been cleverly answered by Green [19, p. 218]. I cannot resist the temptation of quoting. If you prefer two candy bars $(2c)$ to one candy bar $(1c)$, and one candy bar to being shot in the head $(S)$, Alchian doubts that there is any positive probability $\alpha$ of being shot in the head such that $(S, 2c; \alpha, \beta)(1c)$. Green replies that “I should regard the second candy bar as compensation for the (positive) probability that someone in the middle of the Sahara desert firing a revolver in the direction of my head in Toronto could hit his target”. For an ordinary revolver (not a guided missile), I would take the extra candy bar in exchange for letting someone in North Oxford fire at me in the centre of Oxford (assuming no one else would be hurt) or any other probability of death no larger than $(0.1)^9$. It is not clear that Green's answer is superior.

Assumption (A.17) says that, in any lottery, any components alternative can be replaced by a lottery indifferent to it, and the resulting lottery is indifferent to the original one. The reasonableness of a stronger version of (A.17) (the Strong Independence Postulate) is explained in Samuelson [33, pp. 133-134]. (A.18) just means that we may apply the usual rules of combining probabilities. The lottery $(L^a, L^b, \alpha, \beta)$ involves only the alternatives $b$ and $w$ as possible outcomes. What are the probabilities of $b$ and $w$? These are given by the combination of probabilities as stated in (A.18). This assumption is sometimes queried on the ground that people's preferences may not only depend on the final outcomes and probabilities, but also on the way these probabilities are determined. Consider the case of the slot machine. "Why are there three wheels with many items on each wheel. Why not one big wheel, and why are the spinning wheels in sight? . . . Does seeing the wheels go round or seeing how close one comes to winning, affect the desirability? " [1, p. 39]. Green [19, p. 200] meets this objection by using the dichotomy of "pleasure-oriented" gambling and "wealth-oriented" gambling. (A.18) may be violated in the former but not likely to be violated in the latter. A more logical defence of (A.18) is to interpret cases like slot machines as not only involving monetary payoffs ($b$ and $w$ in (A.18)), but also the payoff of watching the working of the spinning wheels. Hence the combined lottery (three wheels) does not strictly involve the same alternatives as the simple lottery (one covered wheel). (A.19) is not used in the Expected Utility Hypothesis. However, it can be shown that

1 This assumption is not needed if the relevant set of alternatives is finite. See Fishburn [12, pp. 46-47]
(A.19) can be deduced as an implication of the set of postulates (which is somewhat stronger than (A.15)-(A.18) giving rise to the Hypothesis. Moreover, (A.19) is quite reasonable even in the context of finite sensibility. If someone prefers \( z \) to \( y \) and is indifferent between \( u \) and \( x \), he will obviously prefer \((u, z; 1, \frac{1}{2})\) to \((x, y; 1, \frac{1}{2})\), even though the difference in expected utility between these two lotteries is smaller than the difference between two alternatives between which he is indifferent. (This would be the case if \( z \) is slightly preferred to \( y \) and \( x \in B^* \)).

From the relevant set of alternatives, take a best (or rather a most preferred) alternative \( b \) (to which no alternative in the set is preferred) and a worst (least preferred) alternative \( w \) (which is not preferred to any other alternative in the set).\(^1\) If we arbitrarily allot the utility number "1" to \( b \), "0" to \( w \), and "\( x \)" to any \( x \bar{I}(b, w; x^* \bar{b}^*) \) and define the expected utility of a lottery \( L = (x, y; \alpha, \beta) \) as \( E(L) = \alpha U(x) + \beta U(y) \),\(^2\) it can be shown that, for any two lotteries \( L \) and \( L' \), \( E(L') \geq E(L) \). To show this, consider any \( L = (x, y; \alpha, \beta) \) and \( L' = (u, v; \alpha', \beta') \). From (A.16), \( xI(b, w; \alpha^* \bar{b}^*) \equiv L^*; yI(b, w; \alpha^* \bar{b}^*) \equiv L^b \) for some \( \alpha^* \) and \( \alpha^b \), \( 0 \leq \alpha^* \leq 1 \). From (A.17), \( L(L^*, L^b; \alpha, \beta) \). From (A.18), \( L(b, w; \alpha x^* + \beta b^* \alpha^* + \beta^* \beta^b) \). Similarly, we can derive \( L'(b, w; \alpha' x^* + \beta' b^* \alpha^b + \beta^b \beta^b) \). According to (A.15),

\[ LPL' \Rightarrow \alpha x^* + \beta b^* > \alpha' x^* + \beta' b^* \]

But, \( E(L) = \alpha x^* + \beta b^* \) and \( E(L') = \alpha' x^* + \beta' b^* \). Hence

\[ LPL' \Rightarrow E(L') \geq E(L). \]

This is our abridged version of Expected Utility Hypothesis. Together with (A.19), it ensures that Convention 1 will be satisfied, as shown below.

For any four alternatives \( u \bar{I}x, z \bar{I}y \), we have, from (k) and (A.19),

\[ \frac{1}{2}U(u) + \frac{1}{2}U(z) > \frac{1}{2}U(x) + \frac{1}{2}U(y). \]

Hence, \( U(z) - U(y) = U(x) - U(w) \). This is so as long as \( z \bar{I}y \) and \( x \bar{I}u \), for any \( u, x, y, z \). Hence, from the continuity of \( U \), we have

\[ (\forall x, y)[U(b') - U(x) = U(b') - U(y) = a], \]

where each \( a' \) is a positive constant.

From (I) and (P.7) we have Convention 1 (for \( V \) read \( U \)) without having to transform the utility functions. In other words, any utility function representing the underlying preferences will satisfy Convention 1 as long as it also satisfies our abridged version of the Expected Utility Hypothesis.

It must be noted that, with (A.19), the alternatives mentioned in Convention 1 refer only to certain alternatives and not to lotteries. With (A.19), \( LPL' \) only leads to \( E(L) \geq E(L') \)

\(^1\) The assumption that there exists a most and a least preferred alternative is not really needed. Whenever we wish to consider the range of preferences outside the limits set by \( b \) and \( w \), we can always choose a new pair \( b' \) and \( w' \). Alternatively, we may adopt the convention discussed by Samuelson [33, p. 135].

\(^2\) Green [19, pp. 223-226] has argued that, if expected utility is not defined in this additive form but in some other form, e.g. \( E^*(L) = U(x)U(y) \), the utility function derived need not be unique up to a linear transformation with diminishing marginal utility of income for a risk-averse individual. Though this is true, it is also true that, in the transformation of the utility function, we also only have two degrees of freedom (in the additive case). Once the utility levels for any two alternatives are fixed, the others are also fixed since, for the case \( E^*(L) = U(x)U(y) \), the ratio of utilities of any two alternatives is equal to a constant power of any other ratio, e.g. \( U(z)/U(x) = (U(y)/U(x))^a \). It is true that a utility index derived from the non-additive form of expected utility has the same predictive content even for choices involving risk, but it does not have any subjective sense (which it is not claimed to possess). One may however still accept the concept of diminishing marginal utility as meaningful for "utility" defined in some subjective sense. A sufficient condition (in conjunction with other usual postulates) for this is the assumption that the individual, when faced with choices involving risky prospects, does in fact maximize the sum of his subjective satisfaction weighted by probabilities. This assumption seems reasonable as the alternatives involved are mutually exclusive. Alternatively, one may define this as a requirement for rationality and hence conclude that a rational, risk-averse individual has diminishing marginal utility. The confusion as to whether utility is ordinal or cardinal is due to the use of the same term "utility" both as an objective indication of choice and as a measure of subjective satisfaction or happiness.
not necessary \(E(L)-E(L') > a\). As explained above, (A.19) is not unreasonable. But (A.19) is relaxed in the following new version of expected utility.

**Alternative Assumptions on Choices Involving Risk**

(A.15') \((\forall x, y : xBy)(L^1BL^2 \Rightarrow \alpha^1 > \alpha^2)\) where \(L^1 = (x, y; \alpha^1, \beta^1)\), \(L^2 = (x, y; \alpha^2, \beta^2)\).

(A.16') \((\forall x, y, z : zNyNx)((x, z; \alpha, \beta)Ay \text{ for some } \alpha: 0 \leq \alpha \leq 1}\).

(A.17') \((\forall x, y: xAL^a & z\alphaL^bL^c)((x, y; x, \alpha, \beta)A(L^a, L^b, \alpha, \beta)\)}.

(A.18') For all \(x, y\), \((x, y; \alpha, \beta)A(L^a, \alpha, \beta)\) where \(L^a = (b, w; \alpha^2, \beta^2)\), \(L^b = (b, w; \alpha^2, \beta^2)\Rightarrow(x, y; \alpha, \beta)A(b, w; \alpha^2 + \beta^2 + \beta'2)\).

(A.19') \((\forall u, x, y, z : uIx & zPy)((x, y; 1, 1)N(u, z; 1, 1))\).

Comparing the above set of assumptions with the previous set, it can be seen that (A.19') is much weaker than (A.19). (A.19') only says that, if \(uIx\) and \(zPy\), then \((x, y; 1, 1)\) cannot be better than \((u, z; 1, 1)\). This weakening of (A.19) is bought at the cost of strengthening (A.15') (in comparison to (A.15)). However, (A.15') is still very reasonable. Since the two alternatives are exactly the same in both lotteries \((L^1\) and \(L^2\)), and since \(x\) is better than \(y\), the only conceivable reason for \(L^1\) to be better than \(L^2\) is for \(\alpha^1\) to be larger than \(\alpha^2\). If \(\alpha^1 > \alpha^2\), the two lotteries are exactly the same. But if \(\alpha^1 < \alpha^2\), then there is no reason for \(L^1\) to be better than \(L^2\). Now compare (A.16')-(A.18') with (A.16')-(A.18). It seems that (A.16')-(A.18') are stronger than (A.16')-(A.18'). This, in a certain sense, is true. But it must be noted that (A.16')-(A.18') are no stronger than the corresponding postulates for the conventional model of infinite sensibility. This is so because, with infinite sensibility, \(xIy \Rightarrow xAy\). Moreover, even with finite sensibility, (A.16')-(A.18') are also very reasonable.

With (A.15')-(A.18'), it can be shown that, for all \(L\) and \(L'\), \(LBL' \Rightarrow E(L) > E(L')\). (The demonstration is similar to the previous model.) With (A.19'), it can then be shown that Convention I will necessarily be satisfied by any preference-representing utility function satisfying this second version of the expected utility hypothesis.

### 7. EXPECTED WELFARE

In Section 4 it was shown that an acceptable SWF must possess the form \(\Sigma V^i\) or its positive monotonic transformation. If each individual utility function satisfies any of our two models of expected utility, we need not transform \(U^i\) into \(V^i\). However, the constant \(a\) will then differ between individuals. So, we have \(W = \Sigma U^i/a\) or its positive monotonic transformation. If we re-scale each \(U^i\) to give a uniform \(a\), then we have \(W = \Sigma U^i\) or its positive monotonic transformation. However, if we adopt assumptions similar to the postulates for the Expected Utility Hypothesis, (A.15)-(A.18) or (A.15')-(A.18'), *for social choice*, we can restrict the admissible SWF into \(\Sigma U^i\) or its positive linear transformation, so that \(W = GΣU^i + H\) for some positive \(G\) and some \(H\).\(^1\) \(G\) is of course only a scale factor and the existence of \(H\) denotes the indeterminacy of the origin. The origin could, at least in principle, be fixed in the following meaningful way.

Start with individual utility functions. What could an origin (i.e. zero utility) reasonably mean in the preference scale of an individual? This is answered by the following quotation: "There can be little doubt that an individual, apart from his attitude of preference or indifference to a pair of alternatives, may also desire an alternative not in the sense of preferring it to some other alternative, or may have an aversion towards it not in the sense of contra-prefering it to some other alternative. There seem to be pleasant situations that are intrinsically desirable and painful situations that are intrinsically repugnant. It

\(^1\) Cf. [20]. Harsanyi uses the expected utility postulates (of both individual and social preferences) to derive an additive but *weighted* SWF; we use them to restrict the admissible transformation after we have derived an unweighted SWF. Fleming [15] also derived a SWF similar to Harsanyi's, based on somewhat different postulates.
does not seem unreasonable to postulate that welfare is +ve in the former case and −ve in the latter" [2, p. 269]. If we are thinking in terms of a utility function for the whole life-span of the individual, positive and negative utilities are still obvious in such comments as “If I had to lead such a miserable life, I'd wish not to have been born into this world at all!”

If we have fixed a definite origin for each utility function, the origin for $W$ can be fixed by adopting the convention $W = 0$ if $U_i = 0$ for all $i$. Then we have a fully determined cardinal welfare function restricted to a proportionate transformation.

We usually do not need the full cardinality in the welfare (or utility) function to represent, predict or prescribe choices. For choices concerning certain (non-risky) alternatives, any monotonic transformation of a function is as good as the original function. For choices involving risky prospects, then functions restricted to a linear transformation begin to make sense. For risky prospects involving the annihilation of the relevant community (mankind for the global welfare function and the individual for the individual utility function), then functions restricted to a proportionate transformation begin to be useful. For example, suppose that the sending of a spaceship to Mars and back is expected to have the following two possible consequences: (i) 99-99 per cent chance of a certain advancement in our scientific knowledge; (ii) 0-01 per cent chance of bringing back an unknown form of super-poisonous matter which will kill mankind within a second, before we have time to realize it or even to sense the suffering. Then a rational choice can be assisted by comparing the welfare significance we attach to that scientific advancement as a proportion of our total welfare without such advancement (the negative of which is the change in welfare associated with the annihilation of mankind). It is true that a choice can be made without using such a cardinal comparison by just asking whether we are prepared to accept the 0-01 per cent risk with the 99-99 per cent benefit. However, such a decision is very difficult to make. If we agree that a rational choice is to maximize the expected welfare, then a decision can be made more sensibly by reckoning in terms of cardinal welfare functions restricted to a proportionate transformation.

It may be noted that the validity of the previous sections in no way depends on the argument of this section. In fact, if one believes that it is (logically or practically) impossible to fix meaningful origins for utility functions, the arguments of the previous sections become even more compelling since, in this case, there is less ground for rejecting the Weak Majority Preference Criterion.

8. IMPLICATIONS

I shall discuss the implications of the results on two levels. First, we have implications that are not dependent on the practical possibility of measuring the number of marginal differences involved. If we can derive a utility function for each individual satisfying one of our models of expected utility, then we know that the SWF must possess the form $\Sigma U_i^j/a_i = \Sigma k_i^jU^j$ (or its positive monotonic transformation) where the $k_i$'s are constants. The values of these constants are not determined without comparing the marginal differences. But we know at least that the choice is limited to the weights $k_i^j$; a SWF which is not linear in the $U_i$'s is not an acceptable function. In particular, any strictly quasi-concave (or convex, for that matter) function with welfare contours convex to the origin of a utility space (which is so much in vogue in the literature of welfare economics) is not an acceptable function.

For choices involving risk, there is a problem as to whether we should maximize welfare as a function of expected utilities i.e.

$$WE = W(E^1, ..., E^n) = W(\Sigma_i \theta_i U_i^j, ..., \Sigma_i \theta_i U_j^i),$$

or expected welfare as a function of ex-post utilities, i.e. $EW = \Sigma \theta_i W(U_i^j, ..., U_j^i)$, where $\theta_i$ is the probability of state $j$. There seem to be good grounds for adopting either method. If welfare is a sum (unweighted or weighted with constant individual weights) of utilities,
then the two methods are equivalent. Thus, \( EW = \sum \theta_j \sum k^j U^j = \sum k^j \Sigma \theta_j U^j = WE \).
Hence, our summation theorem frees us from the agonizing choice between maximizing \( WE \) or \( EW \), and we may use either one as convenience dictates.

A by-product of our analysis is the derivation of utility functions for cases of intransitive indifference. Moreover, these utility functions can be meaningfully confined to be cardinal functions. If we can further measure the levels of marginal indifference, these utility functions also become interpersonally comparable on a non-ethical basis, in a sense. This is what Sen [35, p. 106] calls unit comparability. If we fix the origins as discussed in Section 7, full comparability can also be achieved. Since this cardinality and comparability are derived as implications of weak assumptions, our analysis casts serious doubts on the belief that interpersonal comparison of utility is scientifically impossible, that “every mind is inscrutable to every other mind and no common denominator of feeling is possible”.\(^1\)

It is true that the Weak Majority Preference Criterion, though very acceptable, is still a value assumption. But if we are interested in interpersonal comparison of utility as such and not in the derivation of a social welfare function, we can replace WMP with a more positive assumption or convention. For example, we may agree to assign the utility number “one” to each level of marginal indifference. The comparability thus achieved is then not based on any normative consideration. It is true that it does not have any moral force unless some value assumption (e.g. WMP) is introduced (Hume’s Law). But the interpersonally comparable unit is based on the same concept of preference (marginal indifference) for all individuals and hence is not scientifically meaningless. Whether the Brahmin is ten times as capable of happiness as the untouchable [31, p. 636] is refutable after all.\(^2\)

Readers will have noticed that we have already gone into the “second level” of implications for which levels of marginal indifference are assumed to be known. With this knowledge, we do not even have to select weights for individual utilities in our SWF which is simply the unweighted sum of individual utilities. Bentham’s ethics is then fully vindicated with very weak assumptions. Those who find our assumptions reasonable (if not compelling) cannot logically reject Bentham or believe in any SWF apart from the unweighted summation of utilities. I leave it to those who still wish to retain the “egalitarian” (strictly) quasi-concave SWF to find any possible escape.

There is evidence in psychological studies that, apart from extremal values, the just noticeable increment to any stimulus value is a constant proportion of that value. Written as the Weber-Fechner Law, Sensation = \( k \log \text{Stimulus} \).\(^3\) If we accept this general law as applicable to the utility of an individual as a function of his income, then \( U = k \log c \) where \( c \) is his consumption. This provides some justification for using this simple and manageable utility function. If we further assume equal capacity or regard the consideration of interpersonal differences as beyond our present capability, then \( k \) can be taken to be the same for all individuals. Then social welfare can be written as a simple function of \( c^j \), i.e. \( W = \Sigma \log c^j \). A major weakness of this is the disregard of interdependency and long-run effects.

Another implication of our analysis is the possibility of resolving Arrow’s paradox of social choice. If we adopt the rule of basing our social ordering on the summation of individual utilities, it can be seen that the rule will satisfy all Arrow’s Axioms and Conditions (except the ordering aspect\(^4\) of the Independence of Irrelevant Alternatives) as well as some other reasonable conditions such as anonymity, neutrality and path-independence. (On the last, see Plott [30].) Since the ordering aspect of the Independence Condition is not

---

\(^1\) Jevons quoted with approval by Robbins [31, p. 637].

\(^2\) It may be noted that the method of interpersonal comparison using such proxies as suicide rates proposed by Simon [37] is quite different from our method based on marginal indifference.

\(^3\) Luce and Edwards’ [23] criticism of Fechner does not apply to this particular function.

\(^4\) For the penetrating observation that Arrow’s Condition 3 involves two quite different aspects (the ordering and the irrelevance aspects), see [35, pp. 89-92].
satisfied, this rule does not constitute a counter-example to Arrow’s Theorem. However, having taken account of the intensities of preferences, this ordering aspect becomes itself irrelevant.

The above rule satisfies the condition of anonymity since every individual is treated similarly in the summation of utilities. With the fulfilment of anonymity, the condition of non-dictatorship is satisfied a fortiori. That Arrow’s condition $P$ (the weak Pareto Principle with $xP_iy$ for all $i$) is satisfied is obvious since $xP_iy \implies U'(x) > U'(y)$. It may also be mentioned that, with our rule, instead of the majority rule, the case where “unattractive social choices may result whenever lotteries are not allowed to compete” [38, p. 696] does not occur.1

Our analysis shows that the paradox of social choice can be, at least in principle, resolved on a rational basis satisfying reasonable conditions. However, the practical difficulties of doing so remain very real. In this sense, Arrow’s Impossibility Theorem is still very significant. However, these practical difficulties are not insurmountable, as the following section shows.

9. PRACTICAL MEASUREMENT

While the concept of finite sensibility and the associated concept of marginal indifference seems logically compelling, the practical measurement of the levels of marginal indifference is beset with difficulties. However, most measurements are some forms of approximation; we cannot have perfect accuracy even in the measurement of physical magnitudes like length. Moreover, I wish to argue that the practical difficulties, to a significant extent, can be overcome by using the appropriate methods of measurement.

An obvious difficulty in practical measurement is the problem of sincerity in revaluation of preferences. An individual may try to exaggerate his degree of sensibility for his own benefit. For example, if he knows that sugar is being continuously added to his coffee, he may say that he prefers the one to the other even if he actually cannot notice the difference. This suggests that a principle in actual measurement is not to let the individual know the direction of change. Sugar is sometimes added and sometimes deducted without his knowledge. If he still tries to cheat, he may contradict himself. This also suggests that any measure having an explicit direction of change is not suitable.

Secondly, there is the element of time. We have been concerned with a model in which the element of time has not been explicitly taken into account. But, as pointed out by Armstrong, “experience (and the satisfaction experienced) always involves duration and is given to us as made up of time parts” [3, p. 173]. With the introduction of time duration, the Weak Preference Criterion (and a fortiori the Pareto Criterion) has to be defined in reference to some time unit. However, a full consideration of the problems associated with the introduction of the time factor cannot be undertaken here.2 It may be noted, nevertheless, that the difficulties in measurement created by the time factor can, at least to a certain extent, be overcome by using indirect methods of measurement discussed below.

Thirdly, there is the problem of indivisibility or the lack of feasible alternatives to ascertain marginal indifference precisely. Whereas we can vary the degree of sweetness of a cup of coffee more or less continuously, it is much more difficult, if not impossible, to have a close, not to mention continuous, ranking of candidates for the British Premiership.

The second and the third difficulties can be largely evaded by the use of indirect

---

1 For social choices with lotteries, see also [11]. It may also be mentioned that though the admission of lotteries increases the “ex-ante Pareto efficiency” of majority rule, it does not ensure social welfare maximization. Intriligator’s probabilistic model [21] is different from the admission of lotteries as alternatives but defines a rule of determining the social probabilities of choosing among the certain alternatives. While it has some interesting features, it does not lead to welfare maximization. In particular, an alternative which is mildly preferred by one individual but strongly disliked by all others has a positive probability of being actually chosen by the society.

2 Time permitting, I hope to pursue these problems and the related comments by Rothenberg [32, Ch. 7] in another paper.
measurement. We first choose some aspect of preference which is most easily measured and find out the number \((D)\) of units of marginal indifference between some two alternatives \(x\) and \(y\) involving a difference in this particular aspect only. We now wish to measure the number \((D')\) of units of marginal difference between \(x\) and \(z\) involving differences in other aspects of preference where direct measurement is difficult. Several methods of indirect measurement are possible.

First, we may use the amount of money necessary for full compensation as a means of comparison. Suppose our individual is at \(x\) which he prefers to both \(y\) and \(z\). We may then find out the maximum amount of money \((\varepsilon g)\) he is willing to pay\(^1\) in order not to move to \(y\), and the corresponding amount of money \((\varepsilon h)\) he is willing to pay in order not to move to \(z\) (i.e. from the position \(x\) before paying \(\varepsilon g\)). If we ignore differences in preference of no more than a marginal indifference, we have the following equations.

\[
U(x-\varepsilon g) = U(x) + \int_{\varepsilon g}^{M} MUMdM = U(y)
\]

\[
U(x-\varepsilon h) = U(x) + \int_{\varepsilon h}^{M} MUMdM = U(z),
\]

where \(MUM\) is the marginal utility of money. Assuming \(MUM\) to be fairly constant, we can then infer that \(\{U(x) - U(y)\}/\{U(x) - U(z)\} = g/h\), or \(D/D' = g/h\). \(D'\) can then be calculated from the values of the other three variables.

In the above method of estimation, we do not really need the assumption that the individual cares only about his own consumption. In the above, \((x - \varepsilon g)\) need not be taken as his original consumption minus \(\varepsilon g\), but may be taken as the social state arrived at by the transferring of \(\varepsilon g\) from the individual. In this general form, the use to be made of the money transferred has to be specified since the individual may have preference over this matter. The \(MUM\) is then the marginal utility he attaches to this transfer (plus the use made of the money transferred). The shortcoming of the method is that it involves a margin of error due to the possible changing marginal utility of money if \(g\) and \(h\) differ significantly. However, we can make this margin as small as we like, subject to practical difficulty, by selecting the alternatives such that the value of \(g\) is close to \(h\).

Secondly, we may utilize the principle of expected utility maximization (see Section 6 above). Suppose the individual prefers \(z\) to \(y\) and \(y\) to \(x\). We let him choose between the certainty of \(y\) and the prospect \((x, z; \alpha, \beta)\) where \(\beta = 1 - \alpha\). Adjust the value of \(\alpha\) until he is indifferent between the two (or, more strictly, until he finds one as good as the other). We have,

\[
U(y) = \alpha U(x) + \beta U(z).
\]

From which,

\[
U(y) - U(x) = \beta\{U(z) - U(x)\}.
\]

Hence we also have \(D = D'\beta\).\(^2\)

Our comparison need not be confined to three alternatives with a linking alternative \(x\). For example, suppose we know the number \((D)\) of units of marginal indifference between \(x\) and \(y\) and wish to estimate that \((D'\alpha)\) between \(z\) and \(r\). Using the above method, we can establish, in turn, the following equations (supposing \(rPzPyPx\)).

\[
U(y) = \alpha U(x) + \beta U(r)
\]

\[
U(z) = \alpha' U(x) + \beta' U(r).
\]

\(^1\) If \(y\) and \(z\) are preferred to \(x\), we have to find out the amount of money he has to be paid.

\(^2\) The possibility of such indirect measurement does not depend on the definition of expected utility in the additive form. If some other form is used, indirect estimation is still possible although the calculation is made more complex.
From which,

\[ U(y) - U(x) = \beta [U(r) - U(x)] \]
\[ U(r) - U(z) = \alpha' [U(r) - U(x)]. \]

Hence,

\[ \{U(y) - U(x)\}/\beta = \{U(r) - U(z)\}/\alpha' \]

or,

\[ D/\beta = D'/\alpha'. \]

From the above it can be concluded that, even for those aspects of preference which are not easily susceptible to direct measurement, the number of units of marginal indifference involved may still be estimated by using indirect measurement. Hence, in practical measurement of marginal indifference, the best method may be to select a few aspects of preferences and measure the relevant numbers to a fair degree of precision by intensive and repetitive measurement. Other aspects of preferences can then be estimated by using indirect measurement.

I do not attempt to deny that, despite the possibility of indirect measurement, the practical measurement and comparison of utility differences are still beset with many difficulties. But three hundred years ago, it was also difficult to measure the temperature of the atmosphere and to compare it with that of another area. Our proposed scheme seems logically valid and operationally meaningful. But I cannot tell whether we will live long enough to see its actual application for policy formation, though I hope to see research experiments on its practicability. In closing, may I say presumptuously (if not provocatively), while apologizing to Friedrich Engels, “Only when mankind is using the principle of maximizing the summation of units of marginal indifference in the pursuit of their interests and in the resolution of their conflicts, can mankind proudly declare that it has passed through the era of irrationality into the era of rationality”.

**APPENDIX**

*On the Continuity of Preferences*

This appendix is provided because I need some assumption, (A.8), to show the continuity of the underlying preferences, (P.6), from the continuity of the explicit preferences, (A.6). Without using (A.8), I can show that, for any \( x \), the set \( \{ y : yNx \} \) is closed, (P.6a), but I cannot show that the set \( \{ y : xNy \} \) is also closed, (P.6b).

**Proposition 6a.** (\( \forall x \)) the set \( \{ y : yNx \} \) is closed.

**Proof.** Suppose the proposition is not true. Then there exists a \( z \) not belonging to the set \( \{ y : yNx \} \) but in any neighbourhood of \( z \), there exists a \( z' \) belonging to the set. Since \( z \) is not in the set, we have,

\[ \exists u : zPuIx. \]

From (A.6), \( zPu \) implies that there exist a neighbourhood of \( z \) in which all points are preferred to \( u \). Hence,

\[ \exists u : z'PuIx. \]

for all \( z' \) in that neighbourhood. This means that there does not exist a \( z' \) in that neighbourhood belonging to the set \( \{ y : yNx \} \), contrary to the supposition at the beginning of this proof. So the proposition must be true.

To establish (P.6b), I need some additional assumption. If we assume that there exists neither a bliss set for which no point is preferred to any point in that set nor a hell set in which no point is preferred to any other point, or if we are only concerned with the intermediate zone (for which choice is usually concerned), the following seems to be the weakest assumption sufficient for our purpose.

(A.8i) **Symmetry:** (\( \forall x, y \))((\( \exists r : xPrIy \))\( \iff \) (\( \exists t : xItPy \))).
Whereas (A.8i) seems perfectly reasonable for the intermediate zone (it is implied by the transitivity requirement of the infinite sensitive model), it obviously cannot hold in the bliss set and the hell set. In fact, if there exists a hell set, the definitions of "B" and "N" in the text, (D.1) and (D.2), is not strictly appropriate, as they render all points in the hell set as "as good as" the other, even if for some two points x, y in the hell set, \( \exists t : x \nleq_P y \). Hence, we may like to define \( xPy \Leftrightarrow \exists r : xPrly \vee \exists t : x\nleq_P y \), and define \( xNy \) correspondingly. With this, the following assumption is still sufficient to establish (P.6a) and (P.6b).

(A.8ii) \( \forall (x, y) (\exists r : xPrly) \Rightarrow \exists \) some positive \( \varepsilon \) such that

\[
(\forall y' \colon \text{dis } (y', y) < \varepsilon) (\exists t : xPtly');
\]

\[
(\exists u : yIuPx) \Rightarrow \exists \) some positive \( \varepsilon \) such that
\]

\[
(\forall y' \colon \text{dis } (y', y) < \varepsilon) (\exists v : y'IoPx).\]

One may also adopt the compromise of assuming (A.8i) for the intermediate zone and (A.8ii) for the bliss and hell zones. The proofs of all the propositions in the text need only to be slightly revised. We now proceed to prove (P.6b) using (A.8i).

**Proposition 6b.** (\( \forall x \), the set \( \{ y : xNy \} \) is closed.

**Proof.** Suppose the proposition is not true. Then there exists a \( z \) not belonging to the set \( \{ y : xNy \} \) but in any neighbourhood of \( z \), there exists a \( z' \) belonging to the set. Since \( z \) is not in the set, we have

\[
\exists u : xPuIz.
\]

Using (A.8i), we have

\[
\exists v : xIoPz.
\]

From (A.6), \( vPz \) implies the existence of a neighbourhood of \( z \) in which all points are contra-preferred to \( V \). Or for all \( z' \) in that neighbourhood,

\[
\exists v : xIoPz'.
\]

Using (A.8i) again, \( \exists v : xIoPz' \Rightarrow \exists u : xPuIz' \). This means that there does not exist a \( z' \) in that neighbourhood belonging to the set \( \{ y : yNx \} \), contrary to the supposition at the beginning of this proof. So the proposition must be true.

The proof of (P.6a) and (P.6b) using (A.8ii) can easily be shown. In fact, (A.8ii) amounts to assuming half of the continuity requirement. However, it can be seen to be very reasonable. For one thing, it is implied by the transitivity and continuity requirements of the infinite sensitivity model where \( xPrly \Rightarrow xPy \). For those who do not wish to assume (A.8ii), they may assume the following for points in and around the bliss and hell sets, while retaining (A.8i) for the intermediate zone.

\[
\{ xNy & \text{dis } (y', y) < \text{any positive } \varepsilon \} \Rightarrow \exists x' \colon \text{dis } (x', x) < \text{any positive } \varepsilon \& x'Iy'.
\]

I hesitate to assume this because it rules out thick "as good as" curves.

**REFERENCES**


Yeew-Kwang Ng

Social Welfare Functions

569


