THE POSSIBILITY OF A UNIVERSAL
SOCIAL WELFARE FUNCTION

Kenneth J. Arrow

P-41

26 September 1948

THE POSSIBILITY OF A UNIVERSAL SOCIAL WELFARE FUNCTION

By

Kenneth J. Arrow

1. Introduction

Traditional social philosophy of the Platonic realist variety has assumed that there exists an objective social good defined independently of individual desires; this social good was to be apprehended by the methods of philosophic inquiry. Such a philosophy could be and was used to justify government by élite, secular or religious.

To the nominalist temperament of the modern period, the assumption of the existence of the social ideal in some Platonic realm of being was meaningless. The utilitarian philosophy of Jeremy Bentham and his followers sought instead to ground the social good on the good of individuals. The hedonist psychology associated with utilitarian philosophy was further used to imply that each individual's good was identical with his desires. Hence, the social good was in some sense to be a composite of the desires of individuals. Clearly, some viewpoint of this type is basic both to political democracy and to laissez-faire economics or indeed to any economic system in which consumers are given free choice of goods and workers are given free choice of occupations.

W. S. Jevons[1] introduced Benthamite psychology into the field of economics in the form of the marginal utility theory of choice. It was natural to follow through with the utilitarian viewpoint by then identifying the social welfare with the sum of the individual utilities. This is the viewpoint of F. Y. Edgeworth[2] and is basic to

---

[1] W. S. Jevons, Theory of Political Economy, (1871). The marginal utility theory was developed independently by H. H. Gossen, The Laws of Human Need, (1854); K. Menger, Grundsatze der Wolkswirtschaftslehre, (1871); and L. Walras, Elements d'economie politique pure, (1874).

[2] F. Y. Edgeworth, Mathematical Psychics, (1871), ''The Pure Theory of Taxation'', Economic Journal, VII (1897), pp. 550-571.

A. Marshall's use of consumers' surplus to derive recommendation for economic policy.[3]

The addition of individual utilities implies, however, that such utilities are measurable, and moreover, measurable in some sufficiently objective way to be able to compare utilities of different individuals. (Thus, measurability of individual utility, as in the system of J. von Neumann and O. Morgenstern[4] is not sufficient to settle the question of social utility).[5] The reaction against measurable utility and in favor of a pure indifference-curve approach to the study of individual behavior, associated with the names of I. Fisher and V. Pareto, led to attempt to reformulate welfare propositions in economics on the basis of conditions which would fit into any ethical scheme. Thus, Pareto, Barone, and Hotelling, among others studied optimal states of economic welfare, where an optimal state was defined as one in which not everybody could be made better off by any reapportionment of resources.[6] Of course, this approach does not uniquely define the optimal point.

A. Bergson[7] has reintroduced the social welfare function and has pointed out that it need only depend on the preference schedules of individuals and not on the measurability of individual utility. Also, of course, no assumption need be made as to the measurability of social utility; the social welfare function need be unique only up to a monotone transformation. Bergson's approach has been accepted by Samuelson and Lange[8].

---

[3] A. Marshall; Principles of Economics, (1890), Book V, Chapter XIII.

[4] J. von Neumann and O. Morgenstern, Theory of Games and Economic Behavior, 2nd Ed., Princeton, (1946), pp. 15-31, 617-632.

[5] J. von Neumann and O. Morgenstern, op. cit., p. 604, "We have not only assumed that its utility is numerical - for which a tolerably good case can be made...- but also that it is substitutable and unrestrictedly transferable between the various players."

[6] P. A. Samuelson, Foundations of Economic Analysis, Harvard, (1947), Chapter VIII.

[7] A. Bergson, "A Reformulation of Certain Aspects of Welfare Economics", Quarterly Journal of Economics, LII, (1938), pp. 310-334.

[8] P. A. Samuelson, op. cit., pp. 219-252. O. Lange, "The Foundations of Welfare Economics", Econometrica, X, (1942) pp. 215-228.

The only concrete form that has been proposed for Bergson's social welfare function is the compensation principle developed by Hotelling.[9] Suppose the current situation is to be compared with another possible situation. Each individual is asked how much he is willing to pay to change to the new situation; negative amounts mean that the individual demands compensation for the change. The possible situation is said to be better than the current one if the algebraic sum of all the amounts offered is positive. Unfortunately, as pointed out by T. de Scitovsky,[10] it may well happen that situation B may be preferred to situation A when A is the current situation, while A may be preferred to B when B is the current situation.

Thus, the compensation principle does not provide a true ordering of social decisions. It is the purpose of this note to show that this phenomenon is very general. Under certain very reasonable restrictions, there is no method of aggregating individual preferences which leads to a consistent social preferences scale, apart from certain trivial methods which violate democratic principles.

2. The Nature of Preference Relations

This section is a brief discussion of the language which will be used to describe preference relations. It is assumed that the behavior of each individual can be expressed by saying that, given any set of alternative actions, he chooses the one or ones which he prefers to all others in that set. In the present essay, the alternatives in question are taken to be social decisions. The end product of the process of aggregating individual preferences is to be a social preference scale, such that the decision to be made among any given set of alternatives is to be one which is preferred to any other alternative in the set according to that scale.

---

[9] H. Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates", Econometrica, VI (1938), pp. 242-269. The idea is implicit in the earlier work of Pareto and Barone; see Samuelson, op. cit., pp.213-217.

[10] T. de Scitovsky, "A Note on Welfare Propositions in Economics", Review of Economic Studies, IX (1941), pp. 77-88.

It is customary in economic literature to work with the two relations of preference and indifference. It is slightly more convenient to discuss instead the relation, "preferred or indifferent". Let x,y,z,... be various possible social decisions, referred to as <u>alternatives</u>. The relation, x R y, read, "x is preferred or indifferent to y", is assumed to obey the following axioms:

I.   For all x and y, either x R y or y R x. (Connexity)

II. For all x,y and z, x R y and y R z implies x R z. (Transitivity)

These two axioms are precisely those for a weak ordering relation. Preference and indifference can be defined in terms of the relation R.

<u>Definition 1.</u>   x P y means x R y but not y R x.

<u>Definition 2.</u>   x I y means x R y and y R x.

The following lemma brings together a number of obvious consequences of the above axioms and definitions which will be used subsequently.

<u>Lemma 1.</u>   (1)   x R x for all x.

(2)   If x P y, then x R y.

(3)   If x P y and y P z, then x P z.

(4)   For all x and y, either x R y or y P x.

(5)   If x P y and y R z, then x P z.

(6)   If x R y and y P z, then x P z.

Alternatively, we may describe the preference pattern in terms of the behavior of an individual or society when confronted with a set of alternatives. Let C(S) be what is chosen from the set of alternatives S.

I'.   C(S) is a subset of S.

We do not wish to prescribe that C(S) contains only a single element; for example, S may contain two elements between which the chooser is indifferent. Nor do we wish to prescribe that C(S) is always defined, and non-null; for example, S may be the sequence of alternatives of which the nth is that the chooser gets 1 - (1/n) dollars.

As a convention, we will say that C(S) is the null set under such circumstances.

II'. If S is finite, C(S) is non-null.

III'. If S is infinite, x belongs to S, and x belongs to C(S') for every finite S' containing x and included in S, then x belongs to C(S).

Finally, we wish to prescribe a certain degree of rationality or consistency in the choice.

IV'. If S' is a subset of S, and the intersection $S' \wedge C(S)$ is non-null, then $C(S') = S' \wedge C(S)$.

A set-function C(S) satisfying I'-IV' may be termed a rational choice function. It is not hard to see that a rational choice function and a weak ordering relation are simply alternative descriptions of the same phenomena. Each can be defined in terms of the other. Let [x,y] mean the set consisting of the alternatives x and y.

Definition 3. x R y means that x belongs to C([x,y]).

Definition 4. C(S) is the set of all x's such that x belongs to S and x R y for all y in S.

For subsequent use, some consequences of axioms I'-IV' will be set forth.

Lemma 2. (1) If $S' \subset S$, and x belongs to $S' \wedge C(S)$, then x belongs to C(S').

(2) If $S' \subset S$, S' has at least one point in common with C(S), and x belongs to C(S'), then x belongs to C(S).

Lemma 2 is an obvious consequence of IV'.

Lemma 3. A necessary and sufficient condition that x belong to C(S) is that x belong to S and x belong to C(S') for every finite S' containing x and included in S.

Proof of Sufficiency: If S is infinite, Lemma 3 coincides with III'. If S is finite, Lemma 3 is a tautology, since S is a finite subset of itself containing x.

Proof of Necessity: Let S' be any finite subset of S containing x. Then x belongs to $S' \wedge C(S)$, and therefore to C(S') by Lemma 2.

Lemma 4. If S is finite, x belongs to $C(S)$ if and only if x belongs to $C(S')$ for every two-element subset S' of S containing x.

Proof of Necessity: Lemma 2.

Proof of Sufficiency: We proceed by induction on n, the number of elements in S. The lemma is clearly true for n=1, if one-element subsets are included with two-element subsets. Suppose the lemma is true for n. Let S have n+1 elements, and let S_1 be a subset of S containing x and having n elements. Then x belongs to $C(S_1)$ by the induction hypothesis. Let y be the one element of S which is not in S_1. If S_1 did not intersect $C(S)$, $C(S)$ would contain the single element y by I' and II'. Then $[x,y]$ intersects $C(S)$; as x belongs to $C([x,y])$, x belongs to $C(S)$ by Lemma 2, which is a contradiction. Therefore, S_1 intersects $C(S)$, so that x belongs to $C(S)$ by Lemma 2.

Lemma 5. A necessary and sufficient condition that x belong to $C(S)$ is that x belong to S and x belong to $C(S')$ for every two-element subset S' of S containing x.

Proof: Lemmas 3 and 4.

Lemma 6. If S contains the single element x, then x belongs to $C(S)$.

Proof: I' and II'.

Lemma 7. If $C(S)$ is a rational choice function and R is defined by Definition 3, then R is a weak ordering relation.

Proof: It is to be shown that R so defined satisfies I and II.

(1) The set $C([x,y])$ contains either x or y by I' and II'; hence, I follows from Definition 3.

(2) Suppose x R y and y R z. Let S be the set containing the three elements x,y, and z. If $[x,y]$ did not intersect $C(S)$, $C(S)$ would contain the single element z, by I' and II'. Then $[y,z]$ would intersect $C(S)$, so that y would belong to $C(S)$ by Lemma 2 and Definition 3. This is a contradiction, so that $[x,y]$ intersects $C(S)$. Hence, x belongs to $C(S)$ by Lemma 2, and therefore to $C([x,z])$, by Lemma 2. Hence, II is satisfied.

Lemma 8. If R is a weak ordering relation, and C(S) is defined by Definition 4, then C(S) is a rational choice function.

Proof: It is to be shown that C(S) satisfied $I'-IV'$.

(1) I' follows immediately from Definition 4.

(2) Let n be the number of elements in S. For n=1, II' follows from (1) of Lemma 1. Suppose II' holds for n. Let S contain n+1 elements, S_1 be an n-element subset. For some x in S_1, x R y for all y in S_1 by the induction hypothesis. Let z be the single element in S but not in S_1. If x R z, then II' holds. If not, then z R x, by I, and therefore, z R y for all y in S_1 by II. As z R z by Lemma 1, z R y for all y in S, so that z belongs to C(S).

(3) Under the hypotheses of III', x belongs to $C(S')$ for every two-element subset of S_c containing x, among others, so that x R y for all y in S, verifying III'.

(4) Let x belong to $S' \wedge C(S)$. As x belongs to C(S), x R y for all y in S and in particular in S'. As x belongs to S', x belongs to $C(S')$ by definition. Hence, every element of $S' \wedge C(S)$ belongs to $C(S')$.

Now let x be any element of $C(S')$. By hypothesis, there is an element y belonging to $S' \wedge C(S)$. As y belongs to S', x R y; as y belongs to C(S), y R z for all z in S. Hence, x R z for all z in S by II, so that x belongs to C(S) and hence to $S' \wedge C(S)$.

Lemma 9. If C(S) is a rational choice function, R is defined by Definition 3, and $C'(S)$ is defined by Definition 4, then $C(S) = C'(S)$ for all S.

Proof: The element x belongs to $C'(S)$ if and only if x R y for all y in S and therefore if and only if x belongs to $C(S')$ for every two-element subset S' of S containing x. Lemma 9 then follows from Lemma 5.

Lemma 10. If R is a weak ordering relation, C(S) is defined by Definition 4, and R' is defined by Definition 3, then x R y if and only if x R' y.

Proof: By definition, x R' y if and only if x belongs to $C([x,y])$ and therefore if and only if x R x and x R y. As x R x always holds, Lemma 10 is proved.

THEOREM 1. Definitions 3 and 4 establish a one-one correspondence between weak ordering relations and rational choice functions.

Proof: Lemmas 7 - 10.

Theorem 1 permits us to use indifferently the language of rational choice functions and that of weak ordering relations.

For use in subsequent sections, two other types of ordering relations will be defined here.

Definition 5. R is said to be a partial weak ordering relation if

(1) x R x for all x, and

(2) x R y and y R z imply x R z.

In terms of preference scales, partial weak ordering relations permit us to consider the case where no choice at all, not even the choice of indifference, can be made between two possible decisions. Partial weak ordering relations have also been referred to as quasi-ordering relations.[11] A lemma which relates partial weak ordering relations to weak ordering relations will be useful later on.

Lemma 11. If R is a partial weak ordering relation on a space X and S a subset of X such that for all x,y in S, neither x R y nor y R x, and if there exists a weak ordering relation T' on X such that x R y implies x T' y, then for every weak ordering relation T'' on S, there is a weak ordering relation T on X such that x R y implies x T y and x T'' y implies x T y for all x, y in S.

Proof: Define x T y as follows: if x,y in S, say x T y if and only if x T'' y; if neither x nor y in S, say x T y if and only if x T' y; if x in S and y not in S, say x T y; if x not in S and y in S, say not y T x. As T' is a weak ordering relation on X and therefore on the complement of S, and T'' is a weak ordering on S, it follows easily that T is a weak ordering relation. Further, if x R Y, then neither x nor y are in S and therefore x T' y implies x T y; but x T' y holds by assumption.

---

[11] G. Birkhoff, Lattice Theory, New York, (1940), p.7

<u>Definition 6.</u> R is said to be a **strong ordering relation** if

    (1) For all x, not x R x;

    (2) For all x $\neq$ y, either x R Y or y R x;

    (3) x R y and y R z imply x R z.

A strong ordering relation is a natural generalization of the relation "less than" for real numbers. The following lemma states an obvious property of strong ordering relations which will be useful later. First, we shall define the ternary relation, "betweenness".

<u>Definition 7.</u> If R is a strong ordering relation, define B(x,y,z) to mean x R y and y R z, or z R y and y R x.

<u>Lemma 12.</u> If x,y,z are distinct, then exactly one of the following holds: B(x,y,z), B(y,x,z), B(y,z,x).

3. The Aggregation of Preference Relations

It will now be supposed that there are a number of individuals, each of whom has a preference relation in a given space of alternatives. Let i stand for an individual, and R_i for his preference relation, which is assumed to be a weak ordering relation. The letter V, possibly with subscripts or superscripts, will denote a set of individuals; the letter S, possibly similarly modified, will denote a set of alternatives.

The problem of social welfare is to form a function of the individual preference patterns such that the values of the function are themselves weak ordering relations. This may be expressed in the following condition, letting R be the social preference scale considered as a function of $R_1,\ldots,R_n$, where n is the number of individuals.

<u>Condition 1.</u> For all $R_1,\ldots,R_n$, R is a weak ordering relation. (Universality of Social ordering).

It is also natural to insist that the preferences of individuals be reflected affirmatively in the social preference, i.e., if two sets of individual preference patterns are the same except that one alternative is higher on the preference scale

of some individuals in the second case than in first, then the alternative in question

is not rated lower by society in the second case than in the first. To state this con-

dition precisely, let P be the preference relation corresponding to R in accordance

with Definition 1; similarly, we will let P_i be the preference relation corresponding

to the weak ordering R_i.

Condition 2. Let R and R' be the social orderings corresponding to the sets of

individual orderings $R_1,\ldots,R_n$, and $R_1',\ldots,R_n'$, respectively, and let P and P' be the

corresponding preference relations. Suppose that x P y, and that, for all i, x' R_i y'

if and only if x' R' y' for all x', y' not equal to x, and that x R_i y' implies

x R_i' y' and x P_i y' implies x P_i' y' for all y'. Then x P' y. (Monotonicity)

If C(S) is the rational choice function corresponding to R in accordance with

Theorem 1, it must be interpreted as the choice which society would make if the space

of alternatives were restricted to S. This being so, C(S) should be independent of

the very existence of alternatives outside of S, and therefore should depend only on

the individual preference scales within S.

Condition 3. If, for all i and all x and y in S, x R_i y if and only if x R_i' y,

then C(S) = C'(S), where C(S) is the social rational choice function derived from the

individual preference scales $R_1,\ldots,R_n$, and C'(S) is the social rational choice function

derived from the individual preference scales $R_1',\ldots,R_n'$. (Independence of irrelevant

alternatives)

There are social welfare functions satisfying Conditions 1-3; the main theorem of

this essay is to show that these functions fall into one of the two classes given by

the following definitions.

Definition 8. A social welfare function is said to be conventional if there exits

a pair x,y of distinct alternatives such that x R y independently of $R_1,\ldots,R_n$.

Definition 9. A social welfare function is said to be dictatorial if there exists an

individual i such that for all x and y, x P y whenever x P_i y regardless of the

preferences of all individuals other than i.

Condition 4. The social welfare function is not to be conventional.

Condition 5. The social welfare function is not to be dictatorial.

Condition 4 seems to be very sweeping, since it denies the possibility of any decision's being removed from popular control. However, all that is really needed is the following condition:

Condition 4'. There is a set S containing at least three alternatives such that R is neither conventional nor dictatorial on S.

To guard against trivialities, the following condition is imposed:

Condition 6. There are at least three alternatives.

It will be shown that Conditions 1-6 are inconsistent. In what follows, V_0 will be the null set of individuals, V' a set containing a single individual, and V'' the set of all individuals. A number of consequences will be drawn from the conditions, leading to a contradiction.

Condition 3 implies that in considering $C(S)$, we can disregard all preferences among alternatives not in S. Also, R, and therefore $C(S)$, is completely defined by considering only two-element sets of alternatives. This shows that measurability of individual utility is irrelevant to the ordering of social utilities.

Definition 10. V is said to be decisive for x against y if x P y whenever x P_i y for all i in V.

Note that the definition of decisive set is defined by the process of forming the social preference scale from individual preference scales and does not depend on the actual individual preference scales.

Consequence 1. V is decisive for x against y if and only if x P y whenever x P_i y for all i in V and Y P_i x for all i not in V.

Proof: Necessity follows directly from Definition 10.

Sufficiency: In the set $[x,y]$, let R_i be defined as follows: $x P_i y$ for all i in V, $y P_i x$ for all i not in V. By hypothesis, $x P y$. Let $R_1', \ldots, R_n'$ be any other individual preference scales in $[x,y]$ such that $x P_i y$ for all i in V. If i belongs to V, R_i is identical with R_i' in $[x,y]$, so that the conditions of Condition 2 are satisfied; if i does not belong to V, the conditions of Condition 2 are satisfied vacuously. Therefore, $x P' y$, so that V is decisive for x against y.

Consequence 2. For every x and y, there is a decisive set for x against y.

Proof: If we interchange x and y in Definition 8, it follows from Condition 4 that there exists a set of individual preference relations $R_1, \ldots, R_n$ such that not $y R x$, and therefore such that $x P y$. Let V be the set of individuals such that $x R_i y$. Let $R_1', \ldots, R_n'$ be defined in $[x,y]$ as follows: $x P_i' y$ for i in V, $y P_i' x$ for all i not in V. As $x P_i' y$ implies $x R_i' y$, by Lemma 1, it follows from Condition 2 that $x P' y$. By Condition 3 and Consequence 1, V is decisive for against y.

Consequence 3. For every x and y, V'' is decisive for x against y.

Proof: Let V be the decisive set for x against y guaranteed by Consequence 2. If $x P_i y$ for all i in V'', then in particular $x P_i y$ for all i in V, and therefore by Definition 10, $x P y$. Hence, V'' is decisive by Consequence 1 and Condition 3.

Consequence 4. If V' is decisive for either x against y or y against z, then V' is decisive for x against z, where x, y, z are distinct alternatives.

Proof: (1) Suppose V' is decisive for x against y. Give the individual in V' the number 1. Suppose $x P_1 y$ and $y P_1 z$, $y P_i z$ and $z P_i x$ for all $i \neq 1$. Then, by Definition 10, $x P y$. For all i, $y P_i z$, so that $y P z$, by Consequence 3. Hence, $x P z$ by Condition 1 and Lemma 1, while $x P_1 z$, $z P_i x$ for $i \neq 1$. By Consequence 1 and Condition 3, V' is decisive for x against z.

(2) Suppose V' is decisive for y against z. Again suppose that $x P_1 y$ and $y P_1 z$; let $z P_i x$ and $x P_i y$ for $i \neq 1$. Then, again, $x P y$ and $y P z$, so that $x P z$, while $x P_1 z$ and $z P_i x$ for $i \neq 1$. Hence, V' is again decisive for x against z.

<u>Consequence 5</u>. There exists no i and no x and y such that $x \; P \; y$ whenever $x \; P_i \; y$ regardless of the preference scales of all other individuals.

Proof: Suppose there were such an i and an x and y. Let V' be the set consisting of the sole individual i. V' is decisive for x against y by Definition 10. By Consequence 4, V' is decisive for x against any y' and for any x' against y. By repeated application of Consequence 4, it follows that V' is decisive for every x' against every $y' \neq x'$. But this contradicts Condition 5.

It will now be shown that Conditions 1-6 lead to a contradiction. Let x', y', z' be any three distinct alternatives, as guaranteed by Condition 6. For every ordered pair of these, there is at least one decisive set by Consequence 2. Of all such decisive sets, consider the smallest; if this is not unique, select any one of the smallest sets; and designate that set by V_1. Let it be decisive for x against y, and denote by z the third of the alternatives x', y', z'. Let the number of individuals in V_1 be k.

Number the individuals in $V_1, 1, \ldots, k$, and number the remaining individuals $k+1, \ldots, n$. Let V' contain the sole individual 1, V_2 individuals $2, \ldots, k$, and V_3 individuals $k+1, \ldots, n$. Let the preference scales of the various individuals in the set of the three alternatives x, y, z be defined as follows:

$$\text{i in } V': \quad x \; P_i \; y, \; y \; P_i \; z;$$

$$\text{i in } V_2: \quad z \; P_i \; x, \; x \; P_i \; y;$$

$$\text{i in } V_3: \quad y \; P_i \; z, \; z \; P_i \; x.$$

As $x \; P_i \; y$ for all i in V_1, $x \; P \; y$ by definition of decisive set. For all i in V_2, $z \; P_i \; y$, while $y \; P_i \; z$ for all i not in V_2. If $z \; P \; y$, then V_2 would be decisive for z against y by Consequence 1; but V_2 contains only $k-1$ elements, while the smallest decisive set contains k elements by construction. Therefore, not $z \; P \; y$, and hence $y \; R \; z$, by Lemma 1. As $x \; P \; y$ and $y \; R \; z$, $x \; R \; z$ by Lemma 1 and Condition 1. But $x \; P_i \; z$ for i in V', while $z \; P_i \; x$ for all i not in V', contradicting Consequence 5.

THEOREM 2. If the number of alternatives exceeds two, every social welfare function satisfying Conditions 1-3 is either conventional or dictatorial.

This theorem suggests the desirability of relaxing one of the conditions. From the very meaning of a social decision process, it is hard to see how Conditions 2 or 3 can be weakened. Two specific social welfare functions which violate Condition 3 will be examined, and it will be clear that in these cases, at any rate, the functions are undesirable. The remaining sections will be concerned with weakening of the various other conditions.

4. Conventional and Dictatorial Social Welfare Functions

It is obvious that conventional and dictatorial social welfare functions satisfying Conditions 1-3 exist. For a conventional social welfare function, let R be any weak ordering of the alternatives independent of $R_1,\ldots,R_n$. Then R satisfies Conditions 1-3 and in fact is not dictatorial.

For a dictatorial social welfare function, let R coincide with R_1. Then R satisfies Conditions 1-3 and in addition is not conventional.

We could not restrict dictatorial social welfare functions to the case where x R y if and only if x R_i y, if we wished to preserve Theorem 2. For a counter-example, let R be defined as follows: x P y if and only if for some i, x I_j y for j < i, x P_i y; this relation R satisfies Conditions 1-4 and is dictatorial under Definition 9 but not under the proposed redefinition.

5. The Number of Alternatives

If the number of alternatives is zero or one, the problem is meaningless. If the number of alternatives is two, there are methods for which no difficulty can arise with transitivity. We might, for example, use majority voting as the social welfare function, interpreting a tie as social indifference. This viewpoint is essentially the basis of the Anglo-American two-party system.

6. Independence of Irrelevant Alternatives[12]

One example of a social welfare function violating Condition 3 is the following: Assume, following von Neumann and Morgenstern[13], that for each individual utility is measurable up to linear transformations; and assume further that there is a maximum and a minimum utility for each individual. Then the utility for each individual can be defined uniquely by letting the maximum utility for each individual be one and the minimum zero. Then order social preferences by the sum of the individual utilities.

Suppose there are three alternatives and three individuals. Let two of the individuals have the utility 1 for x, .9 for y, and 0 for z; and let the third individual have the utility 1 for y, .5 for x, and 0 for z. On the above criterion, y would be chosen as against x. Clearly, z is a very undesirable alternative, since each individual regards it as worst. If z were blotted out of existence, it should not make any difference to the final outcome; yet, doing so would cause the first two individuals to have utility 1 for x and 0 for y, while the third individual will have utility 0 for x and 1 for y, so that now x is preferred to y. This is clearly unsatisfactory.

Another social welfare function, applicable to a finite number of alternatives, which violates Condition 3 is the rank-order method of voting: Let each individual rank the alternatives and then weigh each choice, the higher weight going to the more preferred choice. The socially chosen alternative is that with the highest weighted sum of votes.

Let there be three individuals and four alternatives, x, y, z, and w. Let individuals 1 and 2 rank then in the order x, y, z, w, while individual 3 rank then in the order z, w, x, y. Let the weights for first, second, third and fourth choice be 4, 3, 2, 1 respectively. Then x is chosen. Alternative y is always inferior to x,

---

[12] The examples in this section were suggested in discussions with A. Kaplan and with G. E. Forsythe, National Bureau of Standards.

[13] See footnote 4.

so that its deletion should not affect the choice of x. But if y is deleted, and the first, second, and third choices are weighted 4,3,2 or 3,2,1 respectively, we now find that x and z are now tied.

7. Universality of Social Ordering.

Condition 1 might be weakened in two ways: (1) require that R be a weak ordering only for some restricted range of possible individual preference patterns $R_1,\ldots,R_n$; (2) require only that R be a partial weak ordering.

The first weakening is quite realistic. For example, under individualistic assumptions, any social decision which gives individual i more of each commodity will be preferred by him, and it would be pointless to require that R even be defined for $R_1,\ldots R_n$ not consistent with this condition. Such restrictions are of the nature of weak partial orderings. Condition 1 might be replaced by the following condition.

Condition 1'. Let $Q_1\ldots,Q_n$ be a specified set of partial weak orderings. Then for every $R_1,\ldots,R_n$ such that x Q_i y implies x R_i y, R is a weak ordering relation.

Suppose that $Q_1,\ldots,Q_n$ are such that there exists a set S containing at least three alternatives such that for all i and all distinct alternatives x, y in S, neither x Q_i y nor y Q_i x. An example of this would be a decision to distribute several commodities among the individuals. Consider three possible distributions, such that no one gives more of each commodity to any one individual than any other distribution. In this case, Lemma 11 applies for each i, if we substitute Q_i for R and R_i for T'. It follows then $Q_1,\ldots,Q_n$ impose no restriction on the ordering within S, so that the whole analysis leading up to Theorem 2 is applicable.

THEOREM 3. If $Q_1,\ldots,Q_n$ are weak partial orderings for which there exists a set S contain at least three alternatives such that for all i and all distinct alternatives x, y in S, neither x Q_i y nor y Q_i x, then every social welfare function satisfying Conditions 1', 2, and 3 is either conventional or dictatorial.

A more radical restriction upon the range of possible individual preference patterns has been proposed by D. Black.[14] Black's result for a finite number of alternatives is here generalized to any space of alternatives.

A. There exists a strong ordering relation S such that for each i, x R_i y and B(x,y,z) together imply y P_i z.

Here, B(x,y,z) is defined by Definition 7. Loosely, this condition means that the space of alternatives can be represented linearly in such a way that each individual has a most preferred position and the preference for other decisions decreases as the alternative moves farther away from the most preferred position in either direction. This is the same idea as the left-to-right ordering of political parties in Continental politics.

Let N(x,y) be the number of individuals for whom x R_i y. Then define a social welfare function, the method of majority decision, by the requirement that x R y means N(x,y) $\geq$ N(y,x).

Lemma 13. If R is the majority decision method, and if, for all i, x R_i y implies z P_i w, then x R y implies z R w.

Proof: Suppose the hypothesis and the condition x R Y both hold. From the hypothesis, N(z,w) $\geq$ N(x,y); since x R y, N(x,y) $\geq$ N(y,x). From the hypothesis, w R_i z implies y P_i x for each i, using Lemma 1, so that N(w,z) $\leq$ N(y,x). Hence, N(z,w) $\geq$ N(w,z), so that z R w.

Condition 1''. For all R_1,...,R_n satisfying A, R is a weak ordering relation.

THEOREM 4. The method of majority decision is a social welfare function satisfying Conditions 1'' and 2-5 for any number of alternatives, provided the number of individuals is odd.

Proof: If R is the method of majority decision, it clearly satisfied Conditions 2-5.

14
    D. Black, "On the Rationals of Group Decision-Making," Journal of Political Economy, LVI (1948), pp. 23-34. "The Decisions of a Committee Using a Special Majority", Econometrica, 16 (1948), pp. 245-261.

It is only necessary to show that Condition 1″ is satisfied, i.e., that R satisfies I and II when $R_1,\ldots,R_n$ are restricted by A. The condition of connexity is clearly satisfied; it only remains to verify transitivity.

If any two of x,y,z are equal, then it is trivial that x R y and y R z imply x R z. Suppose x,y,z distinct, and suppose x R y and y R z hold. By Lemma 12, there are three possibilities: B(x,y,z), B(y,x,z), and B(y,z,x).

(1) B(x,y,z): From A, $x\ R_i\ y$ implies $y\ P_i\ z$ and therefore $x\ P_i\ z$. By Lemma 13, x R z follows from x R y.

(2) B(y,x,z): Suppose $y\ R_i\ z$ but not $x\ P_i\ z$. Then $z\ R_i\ x$, and therefore $y\ R_i\ x$. Then, from A, $x\ P_i\ z$, which is a contradiction. Therefore $y\ R_i\ z$ implies $x\ P_i\ z$, so that x R z follows from y R z by Lemma 13.

(3) B(y,z,x): Suppose $y\ R_i\ z$. Then $z\ P_i\ x$, by A, and $y\ P_i\ x$. Let N' be the total number of individuals for whom $y\ P_i\ x$, and N the total number of individuals. Then $N(x,y) = N - N'$; also, $N(y,x) \geq N'$, so that from x R y it follows that $N - N' \geq N'$, or $N' \leq N/2$. But since $y\ R_i\ z$ implies $y\ P_i\ x$, $N' \geq N(y,z)$. As $y\ R_i\ z$ or $z\ R_i\ y$ for all i, $N(y,z) + N(z,y) \geq N$. As y R z by assumption, $N(y,z) \geq N(z,y)$, so that $N(y,z) \geq N/2$. Therefore, $N' \geq N/2$, or $N' = N/2$. This contradicts the assumption that N is odd; hence, case (3) cannot arise.

The condition that the number of individuals be odd is essential. Suppose there were two individuals, one of whom preferred x to y and y to z, while the other preferred y to z and z to x. These preference scales satisfy A if the ordering x,y,z is taken as the basic strong ordering. Then majority decision yields that x is indifferent to y and y is preferred to z, but x is indifferent and not preferred to z.

The second weakening of Condition 1, namely, that R be only a partial weak ordering, has, in fact, been the approach of such positive work as exists in welfare economics, as mentioned in Section 1. Social preference is defined by saying that x R y means $x\ R_i\ y$ for all i. This relation clearly satisfies all conditions if Condition 1 is

modified to require only that R be a partial weak ordering. It may be noted that most of the substantive theorems derived also assume that there is at least one commodity such that each individual will prefer more of it to less, all other things being equal. That is, $R_1,\ldots,R_n$ are restricted to be consistent with certain preassigned partial weak orderings.

8. Game Problems in Social Welfare Theory

In the preceding analysis, it has been presupposed that the individual preference patterns were known. However, if some social welfare function is established, and if the individual preference patterns are to be obtained by some form of questionnaire, such as voting, there is always the danger of false answers to take advantage of the machinery in the manner of a game. For example, under plurality voting, individuals do not vote for hopeless minority candidates even though they may prefer them. To insure proper social welfare, the rules of the electoral game must be so devised as to insure that expressed preferences coincide with actual preferences. This problem is allied to the problem of games of fair division.[15]

9. Acknowledgments.

I wish to express my thanks to the following individuals at Project RAND:
A. Kaplan, University of California at Los Angeles, and J. W. Youngs, University of Indiana, for guidance in formulating the problem; D. Blackwell, Howard University, and O. Helmer for other helpful discussions.

KJA/hm

---

[15] H. Steinhaus, "The Problem of Fair Division", Abstract, *Econometrica*, 16 (1948) pp. 101-104.