

Similarity as the Basis of Social Welfare Judgments

By Kenneth J. Arrow

January 26, 1949

The present note seeks to clarify and amplify the positive implications of the results presented in an earlier paper of the present author's.<sup>1</sup>

1. Restatement of the Main Theorem

In Possibility, certain conditions are laid down on the formation of social welfare functions. In particular, it is required that the social preference scales be formed from individual preference scales and that the social decision between two alternatives be independent of the desires of individuals involving any alternatives other than the given two.<sup>2</sup> These conditions together serve to exclude interpersonal comparisons of social utility either by some form of direct measurement or by comparison with other alternative social states (compensation principle).<sup>3</sup> Therefore the main result of Possibility can be restated as follows:

If we exclude the possibility of interpersonal comparisons of utility, then the only methods of passing from individual tastes to social preferences

---

1. K.J. Arrow, The Possibility of a Universal Social Welfare Function, Project Rand, 1948, RAD(I)-289, referred to as Possibility.

2. Possibility, Conditions 1 and 3, pp. 9-10

3. The arguments against a meaningful interpersonal comparison of utilities are old. Even Bentham had his doubts. "Tis in vain to talk of adding quantities which after the addition will continue distinct as they were before, one man's happiness will never be another man's happiness; a gain to one man is no gain to another: you might as well pretend to add 20 apples to 20 pears." (Quoted in W.C. Mitchell, "Bentham's Felicific Calculus," The Backward Art of Spending Money and Other Essays (New York, 1937), p. 184, reprinted from Political Science Quarterly, XXXIII (1918), pp. 161-183). I have presented some arguments on this subject in Possibility, pp. 9-10, and in Second Thoughts on Social Welfare Indices, c.c. 237, p. 11.

which will be satisfactory and which will work for all individual preference scales are either conventional or dictatorial.<sup>4</sup>

The word, "satisfactory," in the above statement means that the social welfare function is to reflect individuals' desires positively (Possibility, Condition 2, p. 10) and that the resultant social tastes shall be represented by a preference scale having the usual properties ascribed to individual preference scales (Possibility, Condition 1, p. 9), and Axioms I-II, p. 4).

If we wish to make social welfare judgments which depend on all individual tastes, i.e., are not conventional or dictatorial, then we must relax some of the conditions imposed. In the present note it will continue to be maintained that there is no meaningful interpersonal comparison of utilities and that the conditions wrapped up in the word, "satisfactory," are to be accepted. The only condition that remains to eliminate is that the method of forming a social preference scale work properly for all possible individual preference scales. That is, it will now be supposed that it is known in advance that the individual preference scales  $R_1, \dots, R_n$  for social actions satisfy certain conditions, and it is required to find a social welfare function which will be satisfactory for all individual preference scales satisfying those restrictions.

## 2. Individualistic Restrictions

One important possibility is to impose on the individual preference scales two conditions which in fact have almost invariably been assumed in works on welfare economics: (1) each individual's comparison of two alternative social states depends only on the commodities that he

---

4. Possibility, Theorem 2, p. 14. The above formulation owes a great deal to comments by T.C. Koopmans, Cowles Commission for Research in Economics; F. Modigliani, University of Illinois; H. Simon, Illinois Institute of Technology; and T.W. Schultz, University of Chicago. See F. Modigliani, Contributions to the Discussion of K.J. Arrow's Paper on "The Possibility of a Universal Social Welfare Function," See C.C. 236, pp. 4-5. For the terms, "conventional and dictatorial see Possibility, Def. 8-0 pp. 10-11.

received (and labor that he expends) in the two states, i.e., he is indifferent as between any two social states in which his own consumption-leisure-saving situations are the same or at least indifferent to him;<sup>5</sup> (2) in comparing two personal situations in one of which he receives at least as much of each commodity (including leisure and saving as commodities) and more of at least one commodity than in the other, the individual will prefer the first situation. However, it can be shown that in a world of more than one commodity, these restrictions do not suffice to remove the paradox.

For suppose that among the possible alternatives were three, none of which gave any individual at least as much of both commodities as any other. For example, suppose that there are two individuals and a total of 10 units of each of two commodities. Consider three alternative distributions described by the following table:

| Alternative | Individual 1 |             | Individual 2 |             |
|-------------|--------------|-------------|--------------|-------------|
|             | Commodity 1  | Commodity 2 | Commodity 1  | Commodity 2 |
| 1           | 5            | 1           | 5            | 9           |
| 2           | 4            | 2           | 6            | 8           |
| 3           | 3            | 3           | 7            | 7           |

The restrictions imposed do not tell us anything about the way either individual orders these alternatives. All preferences are permitted, so that we are essentially back in the original situation of unrestricted choice where the paradox holds.<sup>6</sup>

In fact, the same example shows that even further restrictions that have sometimes been used will not suffice. Thus, in the individualistic

---

5. See e.g., P.A. Samuelson, Foundations of Economic Analysis, (Harvard, 1947) pp 222-224, A. Bergson (Burk), "A Reformulation of Certain Aspects of Welfare Economics," pp. 318-320, Quarterly Journal of Economics, L II (1928), O. Lange, "Foundations of Welfare Economics", Econometrica, Vol. 10, (1942), p. 216.

6. This result is a special case of Possibility, Theorem 3, p. 16

case, it is occasionally further assumed, at least as a basis for welfare statements, that all men have the same preferences in regard to their individual situations. Given any ordering of the various social alternatives by the two individuals, we can construct a preference scale for the six individual situations involved by assuming the three available to individual 2 to be ranked according to his preferences and to be all superior to the three individual situations available to individual 1, the last being ranked among themselves in accordance with individual 1's tastes. If each allocation of each commodity in alternative 2 is reduced by .5, it is even possible to embed this last preference scale for the individual situations into an indifference map with convex indifference curves, so even assuming convexity does not impose enough restriction on individual preference scales to permit a satisfactory social welfare function.

The results of this section suggest strongly that the difficulties in forming a social welfare function arise from the differing social attitudes which follow from the individualistic hypothesis, especially in the case of similar tastes for individual consumption. It follows that the possibility of social welfare judgments rests upon similarity of attitudes towards social alternatives.

### 3. Group Choice in the Theory of Games.

In the present development of the theory of multi-person games, an important role is played by the assumption that each possible coalition ranks all possible alternative strategies according to the sum of the payments to its members.<sup>7</sup> It may be asked why this group choice function does not run into the paradox discussed here. It is true that it is assumed in the theory of games that each individual is playing for his own interests;<sup>8</sup>

---

7. J. Von Neumann and O. Morgenstern, Theory of Games and Economic Behavior, Second Edition, p. 264. The above remarks is implicit in definitions of effective set and domination.

8. op. cit., pp. 8-9.

but it has already been pointed out (see Section 2) that the individualistic hypothesis is insufficient to yield a satisfactory social welfare function.

In the theory of games, there is really a one-commodity world, ignoring for the moment the question of chance events. In such a situation, if we make assumptions (1) and (2) on page 2, there is only one preference scale for social alternative possible to any one individual; he ranks social alternatives according to the amount of money he gets under each. In such a situation, the individual preference scales are not variables; Conditions 2 and 3 in Possibility become irrelevant, since they relate to the variation in the social preference scale corresponding to certain specified types of changes in the individual preference scales. In the present case, any social preference scale will satisfy the conditions laid down in Possibility,<sup>9</sup> so no contradiction can arise.

The situation in the theory of games is somewhat more complicated. If there are a number of different possible money payments, any probability distribution over this range is also the outcome of a possible social alternative, and each individual ranks not money payments but probability distributions of money payments. If no other restriction were imposed, the probabilities of different outcomes act like different commodities, and the situation of Section 2 reappears. However, Von Neumann and Morgenstern assume for the purposes of the theory that individuals rank probability distributions of income solely according to the expected value of the return; this ranking is implicit in the assumption of a transferable utility.<sup>10</sup> Again, therefore, the ranking by an individual of alternative social states is prescribed in advance, and the situation is that of the preceding paragraph.

---

9. Pp. 9-10

10. Von Neumann and Morgenstern, op. cit., pp. 604, 629n.

4. Complete and Partial Unanimity.

Suppose that we do not assume in advance the shape of the preferences of any one individual, but we do assume that all individuals have the same preferences for social alternatives. This implies a social-minded attitude and also a homogenous society. If we consider the preferences in question to refer not to expressed preferences but to the preferences which would be expressed if the corruptions of the environment were removed, the assumption of unanimity is the idealist view of political philosophy.<sup>11</sup> In this case, the obvious way of defining the social welfare function is to choose some one individual and then say that the social preference scale shall be the same as his. This satisfies all the conditions set forth in Possibility (pp. 9-11) except the condition that the social welfare function not be dictatorial. Under the assumptions of this section, since it makes no difference who is dictator, the condition of non-dictatorship loses its intrinsic desirability.

This rather trivial result reinforces the suggestion at the end of Section 2 that like attitudes towards social alternatives (not like tastes for individual consumption) are needed for the formation of social judgments. Some values which might give rise to such similarity of social attitudes are the desires for freedom, for national power, and for equality;<sup>12</sup> by their very nature, likeness in individual tastes leads to likeness in desires for

---

11. K. J. Arrow, "Second Thoughts on Social Welfare Indices," C.C. 237, pp. 1-5.

12. These are here considered as ends in themselves; they also have an instrumental significance in furthering or hindering other ends, and in individual who favors, e.g., equality as an end in itself may nevertheless favor a certain amount of inequality in order to increase total output.

social alternatives. Somewhat less direct in its social implication is the desire for prolongation of life, which we may take to be one of the most universal of all human motives. This desire is essentially individualistic, extending to only a few individuals at most; but since the means for achieving increased longevity are in such large part social, there is a strong factor making for like attitudes on special issues. Differences may still arise due to imperfect knowledge.<sup>13</sup> Of like nature are the various types of collective consumption.

---

13.. F. G. Dickinson, American Medical Association, has suggested orally that the prolongation of life could itself be used as a social welfare guide. This view is especially attractive since the greater part of social, and particularly economic, activity is devoted to that end, broadly construed. In contrast to such other basic motives, such as sex and prestige, the desire for longevity is socializing rather than divisive, although if one considers the world as one unit, the relation between population and food supply is probably such that strong elements of conflict remain. However, life cannot be taken as a sole objective, since, for most human beings, there are specific situations in which human beings are willing to give up their lives in the pursuit of other values, whether these be the aggrandizement of the political unit to which one owes obedience ("Dulce et decorum est pro patria mori" - Horace) or the desire for freedom ("It is better to die on your feet than to live on your knees" - Dolores Ibarruri). From a more practical viewpoint, longevity is probably too insensitive to short-run economic adjustments to serve as a meaningful guide, especially in view of the great uncertainty existing as to the factors making for prolongation of life.

The very measurement of length of life for a whole society involves most of the ambiguities already found in economic welfare analysis. Thus, in evaluating the relative importance of different causes of death, simple death rates have been objected to on the grounds that they ignore age at death which should be considered since it determines the loss to society occasioned by the death. Therefore, new measures are proposed which seek, at least to a rough approximation, to measure the economic loss to society occasioned by the deaths due to each cause. See F. G. Dickinson and E. L. Welker, What is the Leading Cause of Death? Two New Measures, Bulletin 64, Bureau of Medical Economic Research, American Medical Association, Chicago, 1948. Of course, carried to its logical conclusion, this course involves all the difficulties already encountered in ascribing a meaning to marginal social productivity.

However the correct mathematical generalization of the unanimity condition is not easy to see. The most obvious generalization is to assume that all individuals are unanimous about some choices but not necessarily about others. That is, among all possible ordered pairs  $(x,y)$  of social alternatives, there are some for which it is known that all individuals feel alike. These common feelings can be expressed by saying that there is a partial ordering  $Q$  of all social alternatives such that the preference scales of all individuals are compatible with this same partial ordering  $Q$ .<sup>14</sup> If there are three alternatives such that  $Q$  tells us nothing about the comparison of any two, i.e., if there are three alternatives such that we do not assume unanimity of agreement as to the choice between any two, then clearly the assumption of a partial unanimity does not exclude any of the difficulties encountered from arising in the social ordering of the three alternatives in question.<sup>15</sup>

#### 5. The Decision Process as a Value.<sup>16</sup>

Up to now, no attempt has been made to find guidance by considering the components of the vector which defines the social state. One especially interesting analysis of this sort considers that among the variables which taken together define the social state, one is the very process by the society makes its choice. This is especially important if the mechanism of

---

14. For the definition of a partial ordering, see Possibility, page 8.

15. This result can be rigorously deduced from Theorem 3 of Possibility, page 16. First suppose that not merely is it known that certain choices will be unanimous but in fact it is known in advance what the choice will be. That is, suppose  $Q$  is known in advance. If we let  $S$  be the set of three alternatives mentioned in the text, then Theorem 3 applies with  $Q_1, \dots, Q_n$  all equal to  $Q$ . If now we merely postulate unanimity of those choices without specifying how the choices will come out, we have clearly imposed less restraints on the preference scales of individuals and therefore a fortiori it remains valid that the only possible social welfare functions are conventional or dictatorial.

16. The approach in this section owes much to suggestions of M. Friedman, University of Chicago, and P.J. Bjerve, Norwegian Ministry of Trade.



choice itself has a value to the individuals in the society. E.g., an individual may have a positive preference for achieving a given distribution through the free market mechanism over achieving the same distribution through rationing by government. If the decision process is interpreted broadly to include the whole socio-psychological climate in which social decisions are made, the reality and importance of such preferences, as opposed to preferences about the distribution of goods, is obvious.

From a logical point of view, some care has to be taken in defining the decision process, since the choice of decision process in any given case is effected by a decision process. There is no deep circularity here, however. If  $x$  is the vector describing a possible social state, let  $x_1$  be the components of that vector which are not decision processes; let  $x_2$  be the process of deciding among the alternative possible  $x_1$ 's; in general, let  $x_n$  be the process of deciding among the alternative possible  $x_{n-1}$ 's. We may refer to  $x_1$  as the first - order decision,  $x_2$  as a second-order decision, etc.; then an  $n^{\text{th}}$ -order decision is a process of making an  $(n-1)^{\text{st}}$ -order decision. Any particular social state is described in its entirety by a vector of the form  $(x_1, x_2, \dots, x_n, \dots)$ . In describing the United States government, we might say that  $x_1$  is a proposed bill or, more precisely, the proposed bill taken into conjunction with all the legislation now on the books;  $x_2$  is the process by which bills are enacted into law by Congress and the President;  $x_3$  is the process of choosing a Congress and President, set down by the Constitution; and  $x_4$  is the process of constitutional amendment.

Suppose that for some value of  $n$ , there is one possible  $x_n$  which is so strongly desired by all individuals that they prefer any social state which involves accepting that particular  $x_n$  to any which does not. For

example, the belief in democracy may be so strong that any decision on the distribution of goods arrived at democratically may be preferred to such a decision arrived at in other ways, even though all individuals might have preferred the second distribution of goods to the first if it had been arrived at democratically. In such a case, again, our social welfare problem may be regarded as solved, since the unanimous agreement on the decision process resolves the conflicts as to the decisions themselves.

Some such valuation as the above seems to be implicit in every stable political structure. However, there is a certain empirical element in practice; individuals prefer certain political structures over others not only because of their liking for the structure as such but also because they have some idea of the preference patterns of the other individuals in the society and feel that on the whole they can expect the particular structure in question, taken in conjunction with the expected behavior of other individuals under that structure, to yield decisions on current matters which will usually be acceptable to themselves. Thus, we may expect that social welfare judgments can usually be made when there is both a widespread agreement on the decision process and a widespread agreement on the desirability of everyday decisions. Indeed, the sufficiency of the former alone, as implied in the preceding paragraph, would require that individuals ascribe an incommensurably greater value to the process than to the decisions reached under it, a proposition which hardly seems like a credible representation of the psychology of most individuals in a social situation.

#### 6. An Example of Black's Postulate.

D. Black has analyzed the possibility of forming a social welfare function under a special set of restrictions on  $R_1, \dots, R_n$ . Let  $U_1, \dots, U_n$  be utility indicators for  $R_1, \dots, R_n$  are such that the alternative social states can be represented by a one-dimensional variable in such a way that the graphs of

$U_1, \dots, U_n$  have a single peak. Under these assumptions, Black shows that one can construct a social welfare function by saying that society prefers  $x$  to  $y$  if a majority of the individuals prefer  $x$  to  $y$ . This social welfare function has all the properties desired.

It is possible to find economic examples in which Black's postulate may be expected to be satisfied. Thus, suppose that for reasons of technological efficiency, it is important that all workers work the same number of hours, and it is desired to fix the number of hours to be worked. If we assume that wages are to be paid in accordance with marginal productivity, the real wage rate is a known decreasing function of the number of hours selected; hence, each social alternative is completely specified by a single number, the number of hours to be worked. For each individual, the relation between wages and hours worked defines an income-leisure transformation curve. Under individualistic assumptions, we may assume that individuals rank different numbers of hours to be worked by considering the corresponding points on the income-leisure transformation curve and comparing the income-leisure indifference curves which pass through them. We may reasonably suppose that there is one point on the transformation curve for which the individual's utility is maximized and that the individual's utility decreases as the number of hours worked varies in either direction from the optimum. Then Black's postulate is satisfied, and we could find one number of hours worked which would command a majority vote over any other.<sup>17</sup>

The fact that Black's restrictions on the individual preference scales suffice to permit a social welfare function casts new light on what is meant by similarity of social attitudes. In Black's case, individuals can have

---

17. Actually, as Black points out, it is not actually necessary to consider all pairwise comparisons in order to find the optimum. It suffices, under his assumptions, to look at the first choices and find the median. See "The Decisions of a Committee", p. 250, *Econometrica*, 16(1948) pp. 250

varied first choices; but they must have a fundamentally similar attitude towards the classification of the alternatives, since they all order the alternatives in the same way.