

Second Thoughts on Social Welfare Indices

by Kenneth J. Arrow

December 10, 1948

This note is designed to make more concrete some of the abstract concepts and conditions appearing in an earlier paper by the present author,¹ and to correct and amplify some of the discussion there.

I. SOME GENERAL CONSIDERATIONS ON WELFARE

1. Idealist Philosophy and Social Welfare.

The comments on social philosophy in Section 1 of "Possibility" are in some places misleading or incorrect. The identification of idealism with anti-democratic viewpoints and nominalism with democracy, though quite common, is not in fact compatible with the actual history of political theory.² Many idealist philosophies have had democratic implications for politics, e.g., Rousseau, Kant, and T.H. Green. At least, some of these philosophers distinguished between the noumenal will, the desire for the rational good inherent in each individual, and the phenomenal will, the desires as actually revealed after corruption by the environment.³ All individuals have the same noumenal will, but their phenomenal wills may vary because of external factors. From this viewpoint, the attainment of social welfare is best accomplished by creating an environment in which phenomenal and noumenal wills most nearly

1. The Possibility of a Universal Social Welfare Function, RAD(L)-289, Project Rand, referred to below as Possibility.

2. For this point, I am indebted to D. Easton of the Dept. of Political Science, University of Chicago.

3. Utilitarianism and welfare economics have usually based their official doctrines on the phenomenal will, either denying the existence of the noumenal will or feeling that it is irrelevant. However, it will be noted below that at least one distinction of this type has received recognition, though it usually enters as an exception to the general principle.

coincide. John Milton⁴ and Rousseau, among many others, have held that free interchange of ideas is the surest way of arriving at the social good.⁵

This model has much in common with the statistical problem of pooling the opinions of a group of experts to arrive at a best forecast; here, individuals are considered as experts at detecting the noumenal will.⁶ An anti-democrat, such as Plato, might argue that only for a minority is there enough resemblance between the two wills to make them useful experts in this sense, while a pro-democrat might argue that all men have natively an equal portion of the light; all are children of God. The analogy to the problem of pooling experts' opinions is, of course, incomplete; for, in the social welfare problem, the very method of pooling, i.e., of social decision, may affect the degree of expertness of individuals. E.g., marking off a certain group, such as army officers, to hold power, may, because of the greater leisure and better living conditions which follow, enable them to make better decisions, even though there was no native superiority to the remainder; or, it may cause them to lose contact with the daily problems of the ordinary man, and so lower their ability to make "good" decisions in certain contexts.

4. "All opinions, yea errors, known, real and collated, are of main service and assistance towards the speedy attainment of what is truest." "Areopagitica," p. 690, Complete Poetry and Selected Prose of John Milton, Modern Library, New York, 1942.

5. Another pro-democratic argument from the same viewpoint is that the very act of establishing a dictator or élite to decide on the social good will lead to a distortion of the phenomenal from the noumenal will for the ruling group. "Power always corrupts; and absolute power corrupts absolutely." (Lord Acton)

6. This analogy was pointed out to me by O. Helmer, Project Rand. For the development and application of the statistical methods used in this problem, see M.A. Girshick and A. Skogstad, A Pilot Study in Predictions, MW-258, and A. Kaplan, M.A. Girshick, and A. Skogstad, The Prediction of Social and Technological Events, RAD (L)-303, both of Project Rand.

2. Knowledge and the Meaning of Social Alternatives.

If the assumption that the noumenal will is the same for all is dropped, there may still be a distinction between the two wills. Thus Dewey distinguishes between ultimate ends and ends in view for individuals, the latter being means for achieving the former, which are, however, usually unconscious. The relation between the two is not unique, however; for a given act of ultimate ends, there may be differing sets of ends in view, depending on the greater or lesser knowledge of the individual as to the best means of achieving his ultimate ends (and, indeed, as to what his ultimate ends are). The reality of this effect of knowledge in economic life is well evidenced by the fact that the price ratio between two chemically indistinguishable brands of aspirin may exceed ten to one; here the hierarchy of ends in view (the phenomenal will) leads to differing preferences for different brands of aspirin, though they are in fact equally efficacious in achieving the ultimate end (noumenal will) of mitigating headaches. Though welfare economics in its formal aspects typically takes individuals' overt behavior as unanalyzable, few economists would oppose laws against false advertising.

In this connection, it must be pointed out that the alternatives, among which social preference is to be defined in Possibility (see p. 4), may be interpreted in (at least) two ways: (1) each alternative is a vector whose components are values of the various particular decisions actually made by the government, such as tax rates, expenditures, anti-monopoly policy, and price policies of socialized enterprises; (2) each alternative is a complete description of the state of every individual throughout the future. For convenience, we will refer to the first interpretation of the concept, "alternative," as the concept, "social decision," to the second as the concept, "social end."

Social ends either are themselves ultimate ends or at least completely determine the ultimate ends; the relation between social ends and ultimate ends is only unknown to the extent that the ultimate ends are unknown. On the other hand, it is also true that social decisions determine, wholly or partially, social ends; but the relation here is a matter of the empirical laws of the social sciences. In the present state of these disciplines, it is too much to expect the relation to be well-known. Indeed, there seem to be intrinsic limits to knowing all the consequences of any decision; K. Gödel has shown that there is no method which will determine whether any given statement is true or false in a finite length of time,⁷ while the quantum physicists have shown that there are intrinsic limits to the possibility of measurement. Therefore, the relation between the preference scales concerning social ends and those concerning social decisions is afflicted with uncertainty.

One of the great advantages of abstract postulational methods is the fact that the same system may be given several different interpretations, permitting a considerable saving of time. In the present case, the argument given in Possibility goes through whether the variables x, y, \dots , are assumed to refer to social decisions or social ends. It might therefore be argued that the problem is equally acute in either case. However, Dewey would hold that ultimate ends arise out of biological needs and so are, in part at least, objective. Thus, preference scales for social ends, while not identical from individual to individual, are likely to be more similar than individual preference scales for social decisions. It may be that the biological basis of

7. K. Gödel, "Über formal unentscheidbare sätze der Principia Mathematica und verwandter systeme I," Monatshefte für Mathematik und Physik, 38 (1938), 173-198. See also A.M. Turing, "On Computable Numbers, with an Application to the Entscheidungs problem," Proceedings of the London Mathematical Society, Series 2, 42 (1937), 230-265. For a simpler proof of Gödel's results, see E.L. Post, "Recursively Enumerable Sets of Positive Integers," Bulletin of the American Mathematical Society, 50 (1944), 284-316.

ultimate ends limits preferences about them sufficiently so that a social welfare function can be formed; then the social ordering of social decisions should be based on the social ordering of social ends plus the use of scientific and statistical methods to limit the amount of ignorance in passing from decisions to ends and to limit the effects of the remaining ignorance.⁸

It may also be remarked that, in terms of the analysis in the present section, the doctrine of "enlightened self-interest" would be justified in the event that it was assumed that all individuals had the same ultimate ends for the community. In that case, different opinions on social issues arise from lack of knowledge and can be removed by discovering the truth and letting it be widely known. In our present pessimistic age, even this seems like a very difficult problem, not to be dismissed as lightly as it was by our more exuberant predecessors of the last century.

8. The fact of uncertainty as to the relation between decisions and ends has important implications for the controversy as to the relative efficiency of socialism and capitalism, particularly in a dynamic economy. It may well be argued that centralized planning will necessarily reduce this uncertainty since more facts are available to the central planners. It is true that the effect of a given error may be greater under centralization; but this will not be so if a rational method of planning against uncertainty is adopted. On this point, see A. Wald, "Foundations of a General Theory of Sequential Decision Functions," Econometrica, 15 (1947), 279-313, and earlier papers cited there; and K.J. Arrow, Planning under Uncertainty: A Preliminary Report, C.C. 3.1. The relation of uncertainty to the optimum form of the economic system has been ignored in most of the recent formal controversy, though it has received a great deal of recognition on a less formal level, probably because of the lack of a well-developed theory of uncertainty. Thus, the discussion by J.R. Hicks (Value and Capital, 2nd ed., Oxford, 1946, p. 135), while stimulating, is vitiated by his very limited certainty-equivalent theory of uncertainty, the inadequacies of which have been well shown by A.G. Hart, "Risks, Uncertainty, and the Unprofitability of Compounding Probabilities," in O. Lange, F. McIntyre, and T.O. Yntema, Studies in Mathematical Economics and Econometrics, Chicago, 1942, pp. 110-118.

3. Social Welfare and the Problem of Aggregation.

Since, under any interpretation, an alternative is a vector with many components, the construction of a social welfare function is an aggregation problem in a double sense; aggregation over individuals and aggregation over the different components of the vector. The latter problem arises already for a single individual. R. Bellman, in a discussion of Possibility,⁹ suggests that the attempt to set up an ordering of decisions involves a fallacy of one-dimensionalization. By way of analogy, he considers the measurement of intelligence. Originally, intelligence was measured as an arbitrarily weighted sum of scores on various tests of specific traits. Actually, these specific test scores turn out to have virtually zero correlation, so that we must say that intelligence is really a vector with incommensurable components. (However, since there is presumably some tendency for the scores in allied skills to be more highly correlated than in widely differing skills, the elements of the vector are in some sense a continuous function of the traits to which they refer, so that, Bellman suggests, the incommensurability is not absolute.) Analogously, social and individual decisions and actions are multi-dimensional vectors among some of whose components, at least, there appears to be little commensurability.

Bellman's analogy can be confronted with another which will clarify the situation. Consider an employer faced with the problem of deciding which of two applicants to hire for a job. The abilities of each individual are represented by components of a vector; let us suppose that neither individual is better at every skill than the other. The abilities are incommensurable; how does the employer choose? Clearly, he introduces ^ameasure not intrinsic to the individuals; which ability vector will contribute most to the profits of the firm?

9. R. Bellman, On the Concept of Utility and Decision-Making, D(L)-319, Project Rand, 1948.

For our purposes, however, we must ask a more difficult question: what is it that enables an individual to make the choices he does in daily consumption of widely disparate items. We can, of course, simply say that people do make decisions and hence Bellman's argument cannot be correct; but there still remains the question, why? Dewey's theory of biological needs seems to be one answer; if extended to include psychological needs, it would undoubtedly be acceptable to Freud and his followers. Thus, Bellman's question may be answered with the old proverb, "Man is the measure of all things."

This answer suggests that much information can be obtained for the theory of demand by going behind the preference schedules to the ultimate ends served, taking into account, of course, the problem of knowledge discussed in Section 2. Informal remarks on this question are, indeed, found in the writings of most economists,¹⁰ but the whole problem can hardly yet be said to have entered the main body of economic discussion, either qualitative or statistical. An analytic tool which may be useful here is the study of the functional structure of the variables entering into the utility function suggested by W. Leontief.¹¹

II. SOME PROPOSED FORMS FOR THE SOCIAL WELFARE FUNCTION

1. The Bergson Social Welfare Function.

The reference to Bergson's work on page 2 of Possibility is somewhat misleading. Bergson starts off by postulating an economic welfare function whose arguments describe completely the economic state of all individuals;

10. E.g., A. Marshall, Principles of Economics, 8th Edition, New York, 1948, pp. 86-91. The most systematic discussion of at least one part of the theory of demand from the viewpoint of deeper ends is still: T. Veblen, The Theory of the Leisure Class, Chicago, 1898. An excellent informal discussion of the problems discussed here is contained in W.C. Mitchell, "The Backward Art of Spending Money," American Economic Review, II (1912), 269-281, reprinted in W.C. Mitchell, The Backward Art of Spending Money and Other Essays, New York and London, 1937, pp. 3-19.

11. W. Leontief, "Introduction to a Theory of the Internal Structure of Functional Relationships," Econometrica, 15(1947), 361-373; for the application to demand theory, see particularly 370-373.

i.e., he imposes Condition 1 in the terminology of Possibility. At a later point in the analysis, however, he introduces the Fundamental Value Propositions of Individual Preference, which correspond essentially to Conditions 2 and 4.¹² In this special case it is remarked that the social welfare function depends only on the preference scales of the individuals.

2. The Summation of Utilities.

The statement on page 3 of Possibility that "The only concrete form that has been proposed for Bergson's social welfare function is the compensation principle developed by Hotelling," refers only to social welfare functions satisfying the Fundamental Value Propositions of Individual Preference. Even then, it does not seem entirely accurate. Bergson himself discusses the older social welfare function, the sum of individual utilities (see Possibility, page 1), and though he argues against it, he does not find that it contradicts the value propositions referred to above.¹³ Yet the sum of utilities would seem to depend not only on the preference scales of individuals but also on the particular indicators employed. The only way that I can see of making the sum of utilities depend only on the indifference loci is the following: Since to each preference scale, there corresponds an infinite number of utility indicators, set up an arbitrary rule which assigns to each preference scale one of its utility indicators; then the sum of the particular utility indicators chosen by the rule is a function of the individual preference scales which can be used to establish a social ordering.

Obviously, this formation of the sum of utilities will lead to different decisions in a given situation with different choices of the rule. For any

12. A. Bergson (Burk), "A Reformulation of Certain Aspects of Welfare Economics," Quarterly Journal of Economics, LII (1938), 318-320. For the Conditions, see Possibility, 10-11.

13. Bergson, Op.cit., pp. 324, 327-8.

rule, Condition 1 (Possibility, page 9) is satisfied. However, Conditions 2 and 3 essentially prescribe that for a given situation, the choice made shall vary in a particular way with certain variations in the preference scales of the individuals. This being so, it is clear that for the sum of utilities to satisfy Conditions 2 and 3, it would be necessary for the rule to be stringently limited; in fact, Theorem 2 (Possibility, page 14) guarantees that the only rules consistent with Conditions 2 and 3, if any, are of a trivial nature. I have not been able to construct a special proof of this fact for the sum of utilities which is essentially different from the proof of the general theorem.

3. The Compensation Principle.

There are two variants of the compensation principle: (1) only compensated changes in economic structure shall be undertaken; (2) a change in economic structure is to be undertaken if the algebraic sum of the money payments which would be made by individuals if they were to be exactly as well off in the second situation is positive. In the present notation, the wealth and income of each individual are components of the various alternatives. Therefore, the first variant of the compensation principle in the present terminology appears as the rule saying that one alternative is preferred to another if everybody is at least as well off in the first situation as in the second and at least one person is better off. This ordering is discussed briefly in Possibility, pp. 18-19. Pragmatically, it may be objected to as an undue sanctification of the status quo; from the present viewpoint, it suffers the disadvantage of being a partial ordering and therefore cannot answer the question as to what choice to make out of a given set of alternatives, except in very special cases.

F. Modigliani has suggested that in changing from one state to another which is better for all concerned, the difficulties arising from partial ordering

can be avoided by making a conventional decision as to the distribution of the surplus. Thus, if there are two individuals, we may rule that a change from the present situation will only be permitted if the richer individual would be exactly as well off as now while the poorer individual is better off. This still yields only a partial ordering but it may be held that this partial ordering suffices for all cases which will actually arise, since retention of the status quo should always be included in sets S which are considered. This matter requires further investigation.

The difficulties with the second variant of the compensation principle which arise even in the case of a comparison between two alternatives have been pointed out by T. Scitovsky.¹⁴ It may be noted that the real force of Scitovsky's argument is not felt until at least three alternatives are considered, for there are various ways of eliminating the paradox for two alternatives. Thus, we may say that change is not permitted unless compensation is actually paid; the objections noted above to this procedure only apply when more than two alternatives are concerned. Or we may say that one situation is better than another only if the compensation principle says that it is better when viewed in both ways; otherwise, there is indifference. This last statement may be referred to as the Scitovsky form of the compensation principle. It is easy to show by means of an example that indifference defined by the Scitovsky form of the compensation principle is not transitive.

Example: Suppose there are two individuals and two commodities; let (x, y, z, w) mean that individual 1 has amounts x and y of commodities 1 and 2 respectively, while individual 2 has amounts z and w of commodities 1 and 2,

14. T. Scitovsky, "A Note on Welfare Propositions in Economics," Review of Economic Studies, IX (1941), 77-88. See Possibility, p. 3.

respectively. Let state 1 be defined by (3, 3, 2, 4), state 2 by (4, 2, 2, 4), and state 3 by (4, 3, 3, 2). Further suppose that individual 1 prefers commodity bundles (3, 5), (5, 1), (3, 4), and (4, 2) in that order, while individual 2's ordering is (3, 3), (4, 2), (3, 2), and (2, 4). Then it can be shown that the Scitovsky form of the compensation principles makes states 1 and 2 indifferent and also states 2 and 3 indifferent, while state 3 is preferred to state 1 and indeed by both parties.

Any use of a compensation principle in which compensation is not actually paid involves the use of tastes for non-realizable alternatives as a measure of intensity of wants. But the essential difficulty is that they measure not the intensity of desire for the social state by itself but only the relation of that intensity to the intensity of desire for wealth (or whatever commodity the compensation is paid in.) Thus, in comparing two states, one of which is not competitive, the result of the compensation principle may not be invariant under changes in the commodity in which compensation is carried out. Even in comparing two competitive situations, this may be true if the changes involved are large, since only for infinitesimal changes are the marginal rates of substitution equal. I believe this argument takes on additional force when it is realized that any given individual only consumes a few of all possible commodities; if any commodity is used to measure compensation which is not consumed by some individual, the marginal rate of substitution between that commodity and wealth will not be the same for that individual (in general) as for others.

In a vaguely intuitive way, it seems to me that the fact that alternatives, as defined here, already include all of the relevant elements precludes using compensating variations in any one component as an effective way of measuring intensities. I believe this is essentially the same point as that made in the last paragraph, but I am not sure.

It might seem that the Scitovsky form of the compensation principle could at least lead to a partial ordering which would cover more cases than the principle of actually paying compensation. The work of Hicks and Kuznets discussed in the next section makes this appear likely.

4. The Valuation of the Social Income.

The problem of evaluating the real social income is one step more complicated than the classical welfare economics in that it requires a judgment as to the social preference between two states to be made not on the basis of a knowledge of the indifference maps of all individuals but solely on the basis of observed price-quantity data in the two situations and possibly in other situations, but in any case only in a finite number of positions. Even for one individual this is not always possible.¹⁵

An important series of papers by Hicks and Kuznets¹⁶ deals with the problem of at least setting up a partial ordering among the observed positions of the economy. Hicks says that state B is better than state A if there is some redistribution among people of the total quantities of all commodities produced under B which would make each individual better off than he was under A. As a sufficient (though not necessary) condition that B be better than A, he advances the following criterion: Compute Laspeyre's (base-year weight) and Paasche's (current-year weight) price indices for situation B with situation A as base. Deflate the money national income in situation B by both indices;

15. P.A. Samuelson, Foundations of Economic Analysis, Cambridge, Mass., 1947, pp. 146-156.

16. J.R. Hicks, "The Valuation of the Social Income," Economica, VII (1940), N.S., 105-124; idem, "The Valuation of the Social Income - A Comment on Professor Kuznets' Reflections," ibid., XV (1948), N.S., 163-172; S. Kuznets, "On the Valuation of Social Income - Reflections on Professor Hicks' Article," ibid., XV (1948), N.S., 1-16, 118-131.

if both deflated figures exceed the national income in A, B will be better than A. Kuznets agrees that if there are no changes in population and if every commodity is consumed by every individual, then Hicks's criterion will in fact meet the base reversal test; indeed, it will actually be transitive and so establish a partial ordering.¹⁷ I have not had the opportunity to see how rigorous the proof is.

As a final remark, I would suggest that, just as in the case of a single individual,¹⁸ it is likely that a more inclusive partial ordering can be obtained if in the comparison of any two states use is made of price-quantity observation in other states (which will give a better picture of the indifference maps of the individuals under the assumption of constant wants).

17. Kuznets, Op.cit., 2-6, 124-131.

18. Samuelson, op.cit., pp. 151-153.