CHAPTER VIII

NOTES ON THE THEORY OF SOCIAL CHOICE, 1963

When the first edition of this book was prepared, the relevant literature was summarized in four pages. Since 1951 there has been a considerably greater volume of discussion. In preparing a second edition, I felt that the most useful procedure would be to append to my previous discussion a series of reflections inspired by the recent discussions. Despite their high quality, I do not find it obligatory to prepare a revision of the text itself. It is not the results that have been significantly affected [1] but their interpretation and the relation to other contemporary work. There is no attempt here at exhausting the relevant references.

Let me begin by calling attention to some excellent expositions of the theory of social choice which have appeared in the last decade. A remarkably clear exposition at an elementary level has been given by M. Barbut.[2] A number of interesting papers on the foundations of welfare economics has appeared in a special issue of *Économie Appliquée*.[3] I particularly call attention to a remarkable exposition of the theory of collective choice and the general problem of aggregation due to G.-Th. Guilbaud.[4]

A recent paper by W. Vickrey [5] has reviewed the proof of the main theorem of this book under somewhat different assumptions and traced out some possible implications and further lines of development in the foundations of the theory. By restating the axioms in a somewhat different form, he has achieved an extremely simple exposition which brings out the lines of the argument very clearly.

Perhaps the most complete and up-to-date summary of the problem of

[1] There are, however, some revisions that must be made in precise mathematical statements, as Blau (see Section II.4) has shown.

[2] "Quelques aspects mathématiques de la décision rationnelle," *Les Temps Modernes*, Vol. 15, October, 1959, pp. 725–45, trans. as "Does the Majority Ever Rule?" *Portfolio and Art News Annual*, No. 4, 1961, pp. 79–83, 161–68.

[3] Volume 5, October–December, 1952.

[4] "Les théories de l'intérêt général et la problème logique de l'agrégation," *ibid.*, pp. 501–84.

[5] "Utility, Strategy, and Social Decision Rules," *Quarterly Journal of Economics*, Vol. 74, November, 1960, pp. 507–35.

aggregation of individual choices into collective ones, with particular emphasis on political aspects, has been carried out by W. Riker.[6] The economic literature has been ably reviewed by J. Rothenberg.[7]

Finally, it should be noted that the pioneer work of D. Black (see pp. 75–80 of the text) has recently been collected and systematically expounded in a book.[8]

I. Historical Remarks

I must confess to a certain want of diligence in tracking down the historical origins of the theories of social choice. When I first studied the problem and developed the contradictions in the majority rule system, I was sure that this was no original discovery, although I had no explicit reference, and sought to express this knowledge by referring to the "well-known 'paradox of voting'" (text, p. 2). When the basic ideas of the book were first read as a paper to the Econometric Society in December, 1948, Professor C. P. Wright of the University of New Brunswick called my attention to the work of E. J. Nanson.[9] Nanson, in discussing a proposal of his for a method of election, refers without great emphasis to the possibility of intransitivity arising from majority choice (pp. 213–214) for which he gives no previous reference. It is true, however, that the tone of his remarks does not suggest that this possibility is a discovery of his own, although it is rather difficult to be sure.

However, Guilbaud [10] notes that the paradox was known and developed by the Marquis de Condorcet in the eighteenth century,[11] and refers to the paradox therefore as the Condorcet effect. This development was part of Condorcet's great interest in methods of election and essentially, therefore, in the theory of social choice. His work, in turn,

[6] "Voting and the Summation of Preferences: An Interpretive Bibliographic Review of Selected Developments During the Last Decade," *American Political Science Review*, Vol. 55, December, 1961, pp. 900–11.

[7] J. Rothenberg, *The Measurement of Social Welfare*, Englewood Cliffs, New Jersey: Prentice-Hall, 1961.

[8] *The Theory of Committees and Elections*, Cambridge, U. K.: Cambridge University Press, 1958.

[9] "Methods of Election," *Transactions and Proceedings of the Royal Society of Victoria*, Vol. 19, 1882, pp. 197–240.

[10] *Op. cit.*, pp. 513–15.

[11] Most especially in his *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris, 1785. A thorough study of Condorcet's thought appears in G.-G. Granger, *La Mathématique Social du Marquis de Condorcet*, Paris: Presses Universitaires de France, 1956. The work of Condorcet on voting is mostly analyzed in Chapter 3, especially pp. 94–129, an extensive summary of Condorcet's *Essai*.

seems to have been inspired by an earlier paper, that of Jean-Charles de Borda.[12]

Black [13] has given a history of the theory of social choice, starting with the work of Borda and including that of Condorcet, Laplace, Nanson, Galton, and most especially C. L. Dodgson (Lewis Carroll). In regard to the last, he has uncovered some previously unpublished pamphlets in which Dodgson cryptically, although with great acumen, analyzed problems of elections and particularly what he called "cyclical majorities." Both Dodgson's work and Black's comments on it and on the circumstances of its origin are extremely worthwhile.

Black's excellent history makes superfluous any need for recapitulation here. However, there are a few comments that may be of some interest. Borda's starting point is the fact that plurality voting among several candidates could easily lead to a very unreasonable choice. His major point—and this, I believe, has been decisive in all subsequent work—is that the entire ordering of the individual voters among alternative candidates is needed for social decision. The method Borda proposes is the rank-order method as defined on p. 27 of the present text. This, as Borda observes, gives equal weight to the differences between adjacent candidates as well as to different voters. The first raises thus the problem of the measurability of utility, the second that of interpersonal comparisons. He justifies the first step by an argument essentially based on ignorance. If a voter ranks B between A and C, then we have as much reason to suppose that the difference in intensity between A and B is greater than the difference in intensity between B and C as to suppose that it is less. The adding up of different individual votes is justified on the grounds of equality of voters. These themes have continued to recur. The argument of L. Goodman and H. Markowitz [14] may be regarded as giving in effect an axiomatic justification of Borda's position.

As Granger and Black both observed, Condorcet has really two different approaches. In the one most in line with subsequent developments, as well as with Borda's work, the chief contribution has been what might be termed the *Condorcet criterion*, that a candidate who receives a majority as against each other candidate should be elected.

[12] "Mémoire sur les élections au scrutin," *Mémoires de l'Académie Royale des Sciences*, 1781, pp. 657–65. The paper was presented in 1770. For an English translation and incisive comments, see A. de Grazia, "Mathematical Derivation of an Election System," *Isis*, Vol. 44, June, 1953, pp. 42–51.

[13] *Op. cit.* in footnote 8, Part II.

[14] "Social Welfare Functions Based on Individual Rankings," *American Journal of Sociology*, Vol. 58, November, 1952, pp. 257–62.

This implicitly accepts the view of what I have termed the independence of irrelevant alternatives (see text, pp. 26–28). It was in this context that Condorcet discovered that pairwise majority comparisons might lead to intransitivity and hence to an indeterminacy in the social choice. Condorcet did propose some methods for dealing with the general case, but these have not been found clear by subsequent writers.

The second approach is closely related to the theory of juries which Condorcet and others were studying. Here the implication is rather that the voters are judges of some truth rather than expressing their own preferences. This position is essentially a stochastic version of an idealistic position and has been discussed previously (see text, pp. 85–86).

The work of Nanson and Dodgson deserves some mention, although we must refer to Black's work and the original sources for fuller treatment. Both criticize various well-known methods of voting, mainly in terms of failing to satisfy the Condorcet criterion. Most proposed methods do not even accomplish this much, not even the rank-order method of Borda. Nanson suggests the following procedure based on having the orderings of every voter for all candidates: Rank the candidates according to the rank-order method. Then eliminate all those whose total votes is less than the average. With the remaining candidates form the rank-orders again, considering only those candidates, and repeat the process until one candidate is selected. Nanson shows that if there exists a candidate who has a majority against any other, he will be selected. Of course, the paradox of voting cannot be eliminated by this technique.

Dodgson accepts fully the Condorcet criterion, so much so that he comes to the conclusion that if a cyclical majority persists, then there should be "no election." [15] This position is rather curious. Indeed, Dodgson elsewhere shrewdly points out that "no election," if it is an allowable alternative, should be considered on a par with all the candidates.[16] This seems quite inconsistent with the previous statement. It does point, however, to an important empirical truth, especially about legislative matters rather than the choice of candidates: The status quo does have a built-in edge over all alternative proposals.

Rather interestingly, though, in a discussion of alternative proposals for the case of cyclical majorities,[17] he points out the unreasonableness of their conclusions by reference to another criterion; namely, he points out that the candidate who would have a majority over all others with the least number of interchanges of orderings on the individual prefer-

[15] Black, *op. cit.*, p. 230.
[16] Black, *op. cit.*, p. 232.
[17] Black, *op. cit.*, pp. 227–30.

ence scales is not chosen. This suggests that one could elevate this to a principle for election when we do not admit the possibility of "no election," or do not wish to give that alternative any preference. In principle, and especially with computing machines, the criterion itself provides a method of choosing candidates. I do not know whether there is any simple way of characterizing this principle. It does coincide with Nanson's method for the case of three candidates.

II. The Formal Statement of the Conditions and a New Exposition of the Proof

This section consists of some diverse observations on the formal aspects of the theory. First, I show that some of the conditions can be replaced by the Pareto principle (that a unanimity of individual preferences implies a social preference). Since the Pareto principle is universally accepted, the new set of conditions will be easier to compare with other formulations of the problem of social choice. Second, I give what appears to be the simplest formulation of the proof that the conditions are inconsistent. Third, I state some other, stronger, conditions leading to majority rule of some type; these developments are due to May and Murakami. Finally, I comment on an error in the original statement of the theorem, which was discovered by Blau.

1. *The Pareto Principle and the Conditions for a Social Welfare Function*

The Pareto principle was originally given in the text (p. 36) as a form of the compensation principle. We give it here in a slightly weaker form (involving only strict preferences).

Condition P: *If $x\ P_i y$ for all i, then $x\ P\ y$.*

(In words, if every individual prefers x to y, then so does society.)

To meet an important objection raised by Blau,[18] which will be discussed in paragraph 4, I will also at this stage replace Conditions 1 and 2 by the stronger

Condition 1′: *All logically possible orderings of the alternative social states are admissible.*

Condition 2′: *For a given pair of alternatives, x and y, let the individual preferences be given.* (By Condition 3, these suffice to determine the

[18] J. H. Blau, "The Existence of Social Welfare Functions," *Econometrica*, Vol. 25, April, 1957, pp. 302–13.

social ordering.) *Suppose that x is then raised in some or all of the individual preferences. Then if x was originally socially preferred to y, it remains socially preferred to y after the change.*

Condition 1′ is unnecessarily strong, but it will be useful in permitting a simpler exposition of the proof. Condition 2′ is equivalent to Condition 2 in the presence of Conditions 1′ and 3.

We note here

THEOREM 1: *Condition P is deducible from Conditions 2′, 3, and 4 (Positive Association of Social and Individual Values, Independence of Irrelevant Alternatives, and Citizens' Sovereignty).*

Indeed, Condition P is identical with Consequence 3 (text, p. 54). Consequence 1 (whose proof involves a fallacy pointed out by Blau) is identical with Condition 2′. A careful reading of the proofs of Consequences 2 and 3 (pp. 52–54) shows then that their proofs depend only on Conditions 2′, 3, and 4.

In the next paragraph, the inconsistency of Conditions 1′, 3, P, and 5 will be shown. Since Condition P follows from Conditions 2′, 3, and 4, the inconsistency of Conditions 1′, 2′, 3, 4, and 5 will follow. However, the new list of conditions will be easier to compare with other formulations of the social choice problem; this comparison is made in Section III.1.[19]

2. *A Proof of the General Possibility Theorem* [20]

In this paragraph, we demonstrate

THEOREM 2: *Conditions* 1′, 3, P, *and* 5 *are inconsistent.*[21]

Hence the substitution of the Pareto principle for Conditions 2 and 4 does not affect the inconsistency shown in the text. Furthermore,

[19] The foregoing does not mean that Condition 2 or 2′ (Positive Association of Social and Individual Values) can be disregarded in the general development of the theory of social choice. It can be replaced by the Pareto principle for the purpose of demonstrating the inconsistency of the original set of conditions; but if one of the other conditions is dropped in an effort to find a consistent set, then Condition 2 or 2′ is certainly stronger than the Pareto condition (if we continue to maintain Condition 4, the absence of imposed social decisions). For example, a version of Condition 2′ is used by May in constructing a justification for majority rule; see paragraph 3.

[20] This proof originally appeared in my paper, "Le principe de rationalité dans les décisions collectives," *Économie Appliquée*, Vol. 5, October, 1952, pp. 469–84, with, however, some confusing misprints.

[21] Blau states a corresponding theorem (*op. cit.*, p. 309) but adds Condition 2′ to the list of conditions. It appears to be superfluous; the inconsistency holds without assuming Condition 2′ as the subsequent proof will show.

since Condition P follows from Conditions 2, 3, and 4, as shown in the preceding paragraph, this proof also constitutes a proof of the theorem shown in the text (except for the substitution of the stronger Condition 1′ for Condition 1). This proof, and indeed any proof of which I am aware, does not differ essentially from that given in the text, but I believe the general lines of the proof and the role played by the different conditions are a little more evident.

We use a slightly different definition of a decisive set than that in the text (Definition 10, p. 52). A set of individuals V is decisive for x against y if x is socially chosen when every individual in V prefers x to y and every individual not in V prefers y to x.

The proof falls into two parts. It is first shown that if an individual is decisive for some pair of alternatives, then he is a dictator, which is excluded by Condition 5; the impossibility theorem itself then follows very easily from the first result and the Pareto principle.

That an individual decisive for some pair of alternatives must be a dictator will be shown to follow from the assumptions of Collective Rationality (Condition 1′ together with the definition of a social welfare function), the Pareto principle, and the Independence of Irrelevant Alternatives (Condition 3). We distinguish one individual, called I, and introduce the following notations for statements about the social welfare function, or constitution:

(1) $x \bar{D} y$ *means that* x *is socially preferred to* y *whenever* I *prefers* x *to* y, *regardless of the orderings of other individuals;*

(2) $x D y$ *means that* x *is socially preferred to* y *if individual* I *prefers* x *to* y *and all other individuals have the opposite preference.*

This notation is only legitimate because of Condition 3, which assures us that the choice between x and y depends only on the preferences of all individuals concerning those two alternatives. If Condition 2 were assumed, the two concepts would be equivalent (see Consequence 2, p. 53), but we are now assuming the Pareto principle instead.

Note that the statement, $x \bar{D} y$, implies $x D y$ and that $x D y$ is the same as the assertion that I is a decisive set for x against y.

Suppose then that $x D y$ holds for some x and y. We will first suppose that there are only three alternatives altogether. Let the third alternative be z. Suppose I orders the alternatives, x, y, z, in descending order, whereas all other individuals prefer y to both x and z, but may have any preferences as between the last two. Then I prefers x to y, whereas all others prefer y to x; from (2), this means that $x P y$. All individuals prefer y to z; by Condition P, $y P z$. Then by transitivity, $x P z$; but then this holds whenever $x P_I z$, regardless of the orderings of other

individuals. In symbols,

(3) $$x\ D\ y \text{ implies } x\ \overline{D}\ z.$$

Again suppose $x\ D\ y$, but now suppose that I orders the alternatives, z, x, y, whereas all other individuals prefer both z and y to x. By a similar argument, $x\ P\ y$ and $z\ P\ x$, so that $z\ P\ y$.

(4) $$x\ D\ y \text{ implies } z\ \overline{D}\ y.$$

Interchanging y and z in (4) yields

(5) $$x\ D\ z \text{ implies } y\ \overline{D}\ z.$$

Replacing x by y, y by z, and z by x in (3) yields

(6) $$y\ D\ z \text{ implies } y\ \overline{D}\ x.$$

Since $x\ \overline{D}\ z$ implies $x\ D\ z$, and $y\ \overline{D}\ z$ implies $y\ D\ z$, we can, by chaining the implications (3), (5), and (6), deduce

(7) $$x\ D\ y \text{ implies } y\ \overline{D}\ x.$$

If we interchange x and y in (3), (4), and (7), we arrive at the respective implications

$$y\ D\ x \text{ implies } y\ \overline{D}\ z,$$

$$y\ D\ x \text{ implies } z\ \overline{D}\ x,$$

$$y\ D\ x \text{ implies } x\ \overline{D}\ y,$$

and these can each be chained with the implication (7) to yield

(8) $$x\ D\ y \text{ implies } y\ \overline{D}\ z,\ z\ \overline{D}\ x, \quad \text{and} \quad x\ \overline{D}\ y.$$

Implications (3), (4), (7), and (8) together can be summarized as saying

(9) *If $x\ D\ y$, then $u\ \overline{D}\ v$ holds for every ordered pair u, v from the three alternatives x, y, and z;*

i.e., individual I is a dictator for the three alternatives.

Because of Condition 1′, we can extend this result to any number of alternatives by an argument due to Blau.[22] Suppose $a\ D\ b$ holds, and let x and y be any pair of alternatives. If x and y are the same as a and b, either in the same or in the reverse order, we add a third alternative c to a and b; then we can apply (9) to the triple a, b, c and deduce $x\ \overline{D}\ y$ by letting $u = x$, $v = y$. If exactly one of x and y is distinct from a and b, add it to a and b to form a triple to which again (9) is applicable. Finally, if both x and y are distinct from a and b, two steps are needed.

[22] *Op. cit.*, p. 310.

First, add x to a and b, and deduce from (9) that $a \bar{D} x$ and therefore $a D x$. Then, again applying (9) to the triple a, x, y, we find that $x \bar{D} y$. Thus, $a D b$ for some a and b implies that $x \bar{D} y$ for all x and y, i.e., individual I is a dictator. From the Condition of Nondictatorship (Condition 5), it can be concluded that

(10) *$x D y$ cannot hold for any individual I and any pair x, y.*

The remainder of the proof is now an appropriate adaptation of the paradox of voting. By Condition P, there is at least one decisive set for any ordered pair, x, y, namely, the set of all individuals. Among all sets of individuals which are decisive for some pairwise choice, pick one such that no other is smaller; by (10) it must contain at least two individuals. Let V be the chosen set, and let the ordered pair for which it is decisive be x, y. Divide V into two parts, V_1, which contains only a single individual, and V_2, which contains all the rest. Let V_3 be the set of individuals not in V. Consider now the case where the preference order of V_1 is x, y, z, that of all members of V_2 is z, x, y, and that of all members of V_3 is y, z, x.

Since V is decisive for x against y, and all members of V prefer x to y while all others have the opposite preference, $x P y$. On the other hand, it is impossible that society prefers z to y since that would require that V_2 be decisive on this issue; this is impossible since V_2 has fewer members than V, which, by construction, has as few members as a decisive set can have. Hence, $y R z$, and, since $x P y$, society must prefer x to z. But then the single member of V_1 would be decisive, and we have shown that to be impossible.

Thus the contradiction is established.

3. *The Principles of Equality and Neutrality*

Since the conditions given in the text or in the preceding paragraph are inconsistent, there is not much point in strengthening any of them unless another is dropped. One approach most consistent with electoral practice has been to weaken or omit the condition of Collective Rationality. Guilbaud [23] weakens the condition to require simply avoiding contradictions on two successive decisions (rather than three, when the full force of transitivity comes into play). He then conducts the bulk of his analysis under the assumption that the decision rule between any two alternatives be the same. This assumption has later been termed "neutrality," and may be defined formally as follows.

[23] *Op. cit.*, fn. 4, pp. 555–72.

CONDITION OF NEUTRALITY: *Let $T(x)$ be a one-one transformation of the set of alternatives into itself which preserves all individual orderings. Let the environment S be transformed into the environment S' by the transformation T. Then the social choice from S, $C(S)$, is transformed by T into the social choice, $C(S')$, from the environment S'.*

If we confine ourselves to choices among pairs of alternatives, neutrality says essentially that the decisive sets for one pair of alternatives are the same as those for another.

Under the assumption of neutrality, Guilbaud shows that his weaker form of the Collective Rationality condition requires that the decision rules are completely specified by the decisive sets, with the conditions that enlarging a decisive set leads to another decisive set and that the set of all individuals is decisive. Guilbaud then seeks to drop the neutrality assumption and argues that one can extend the previous result by introducing fictitious individuals whose preferences are specified in advance (p. 569). The same results then apply except that the decisive sets may contain these ideal voters.[24]

The principle of neutrality is not intuitively basic, although one expects it to hold in a wide variety of ordinary decisions; practice always insists on special decision rules (special majorities, for example) for particular decisions. K. O. May[25] has combined neutrality with a second and more significant condition.

CONDITION OF EQUALITY: *A permutation of the individual orderings among individuals leaves all social choices unchanged.*

May adds to these conditions a requirement that the social choice be well defined and that a slightly stronger version of the Positive Association of Social and Individual Values holds. For the case of preference (pair-wise choice) which is all that he considers,[26] he shows that major-

[24] The argument is very terse, and I am unable to determine if it is thoroughly correct.

[25] "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision," *Econometrica*, Vol. 20, October, 1952, pp. 680–84.

[26] He remarks (*ibid.*, p. 680) that, "Since it follows that the pattern of group choice may be built up if we know the group preferences for each pair of alternatives, the problem [of determining group choices from all sets] reduces to the case of two alternatives." This, however, would only be correct if transitivity were also assumed. Otherwise, there is no necessary connection between choices from two-member sets and choices from larger sets. If there are more than two alternatives, then it is easy to see that many methods of choice satisfy all of May's conditions, for example, both plurality voting and rank-roder summation. A complete characterization of all social decision processes satisfying May's conditions when the number of alternatives is any finite number does not appear to be easy to achieve.

ity voting is the only method of social decision satisfying his conditions.[27]

The Condition of Equality is clearly highly appealing; it is also closely connected with Kant's categorical imperative.

Murakami[28] generalized May's results, making use of some theorems on logical functions. The conditions that social pair-wise choice be well defined and that Condition 2′ (see paragraph 1) hold imply that social preference is determined by a generalized form of majority voting. More specifically, there is prescribed a sequence of steps, in each of which decision is made by majority choice, but the outcome of one step may be used as a ballot in the next, and also some ballots with outcomes prescribed in advance may be added to the ballots of real individuals. If, in addition, neutrality is required, then the fixed ballots can be eliminated.[29]

4. *An Error in the Statement of the General Possibility Theorem*

Blau has pointed out that the General Possibility Theorem is incorrectly stated.[30] The difficulty in the proof arises in attempting to reduce the general problem to the case of three alternatives; specifically, Conditions 2 and 5 were implicitly assumed to hold for the three-member subset mentioned in Condition 1; however, the conditions could hold for a set of alternatives and not for a subset. If Conditions 1 and 2 are replaced by Conditions 1′ and 2′, Blau shows that the difficulty is avoided, as we have seen in paragraph 4. However, Condition 1′ is a strong condition, as the discussion in the text (p. 24) shows.

Murakami[31] suggested as an alternative to the strengthening of Condition 1 a weakening of the definition of a dictator (and thereby a strengthening of the Condition of Nondictatorship). If we introduce

[27] In a subsequent note, May showed that his four conditions were completely independent in the sense that one could find a social decision procedure which satisfied any subset of the conditions and failed to satisfy the others. See "A Note on the Complete Independence of the Conditions for Simple Majority Decision," *Econometrica*, Vol. 21, January, 1953, pp. 172–73. It follows that a possibly useful way to classify social decision procedures is to enumerate which of May's conditions it satisfies.

[28] Y. Murakami, "Some Logical Properties of Arrowian Social Welfare Function," *The Journal of Economic Behavior*, Vol. 1, No. 1, April, 1961, pp. 77–84.

[29] Strictly speaking, this last result has been proved only for the case where no individual is indifferent between any two alternatives.

[30] *Op. cit.* in fn. 18.

[31] Y. Murakami, "A Note on the General Possibility Theorem of the Social Welfare Function," *Econometrica*, Vol. 29, April, 1961, pp. 244–46.

CONDITION 5′: *Among the triples of alternatives satisfying Condition* 1, *there is at least one on which no individual is a dictator.*

Then the arguments of paragraph 2, with slight modification, show

THEOREM 3. *Conditions* 1, 3, *P, and* 5′ *are inconsistent.*

Condition 5′ is a very reasonable extension of the notion of a dictator; to deny it means roughly to say that for every choice on which there can be real disagreement there is a dictator.

III. WHAT IS THE PROBLEM OF SOCIAL CHOICE?

A long series of distinguished critics have argued, in one form or another, that the problem of social choice has been incorrectly posed in this book.[32] I will now argue that these criticisms are based on misunderstandings of my position and indeed of the full implications of the critics' own positive views. Upon close examination, all implicitly accept the essential formulation stated here: The social choice from any given environment is an aggregation of individual preferences. The true grounds for disagreement are the conditions which it is reasonable to impose on the aggregation procedure, and even here it is possible to show that the limits of disagreement are not as wide as might be supposed from some of the more intemperate statements made.

1. *Welfare Judgments*

Bergson's deservedly classic paper of 1938 [33] set the pattern in which most of the subsequent discussion, and in particular my own, has been carried on. It is a refined form of classical utilitarianism, but one which at least faces the problem of commensurating the utilities of different individuals.

[32] See I. M. D. Little, "Social Choice and Individual Values," *Journal of Political Economy*, Vol. 60, October, 1952, pp. 422–32, and "L'avantage collectif," *Économie Appliquée*, Vol. 5, October–December, 1952, pp. 455–68; A. Bergson, "On the Concept of Social Welfare," *Quarterly Journal of Economics*, Vol. 68, May, 1954, pp. 233–52; M. C. Kemp, "Arrow's General Possibility Theorem," *Review of Economic Studies*, Vol. 21, 1953–54, pp. 240–43; J. M. Buchanan, "Individual Choice in Voting and the Market," *Journal of Political Economy*, Vol. 62, August, 1954, pp. 334–43; E. J. Mishan, "An Investigation into Some Alleged Contradictions in Welfare Economics," *Economic Journal*, Vol. 68, September, 1957, pp. 445–54; G. Tullock, Appendix 2 to J. M. Buchanan and G. Tullock, *The Calculus of Consent*, Ann Arbor: The University of Michigan Press, 1962, especially pp. 331–34.

[33] "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, Vol. 52, February, 1938, pp. 310–34.

At the risk of boring the reader with the obvious, let me restate the logic of the welfare judgment, as crystallized in Bergson's formulation. Any social decision has consequences for the individual members of the society. It is first assumed that for each individual there is a way of evaluating these consequences; we do not at the moment specify whether this evaluation is cardinal or ordinal. But so far there are as many evaluations as there are individuals in the society. Classical utilitarianism implicitly and Bergson explicitly call for a second-order evaluation, which I for the moment call a welfare judgment. This is an evaluation of the consequences to all individuals which is based on the evaluations for individuals; specifically, if in each individual valuation, two sets of consequences are indifferent, then the welfare judgment as between the two must also be one of indifference.

The individual values are thus the raw material out of which the welfare judgment is manufactured. As good economists, let us look more closely at the technology of the transformation process and the sensitivity of the output to variations in the inputs.

The collection of welfare judgments on all possible environments is determined by the individual valuation schedules, or individual orderings of social states, to use the terminology of this book. This point is accepted and even urged by Bergson and by Little, although their emphasis sounds very different. They argue that the social welfare function (in Bergson's sense) takes individuals' tastes as given. Given a set of individual orderings, the Bergson procedure is first to associate to each ordering a corresponding utility indicator and then to prescribe a function of the individual utilities whose values serve as an indicator of social welfare. But presumably such a social welfare function could be drawn up for any given set of individual orderings; if we substitute a new set of individual orderings, we have, in Bergson's terminology, a new social welfare function, i.e., a new formula for determining all possible welfare judgments.

This formulation is in no wise different from mine; to each given set of individual orderings is associated a social choice function (see Definition 4, p. 23 of text [24]). It would perhaps have been better for me to use a different term from "social welfare function" for the process of determining a social ordering or choice function from individual orderings, although the difference between Bergson's definition and my own was

[24] The definition cited associates a social ordering with each set of individual orderings; it might have been better to speak more generally of a social choice function and leave for a special assumption the question whether the social choice function should be an ordering. The desirability of the transitivity condition has been criticized and needs discussion; see Section V.

pretty carefully spelled out (p. 23 of text). I will therefore now use the term "constitution," as suggested by Kemp and Asimakopulos.[35] The difference, however, is largely terminological; to have a social welfare function in Bergson's sense, there must be a constitution.[36]

The real point of the Bergson-Little criticisms, then, is not the formulation of the social choice problem as the choice of a constitutuion, but the acceptability of my particular set of conditions. These conditions are of two kinds: those that relate to the social choice function produced by any given set of individual orderings, and those that relate to the way social choices vary with respect to changes in the individual orderings. It is the latter set which are rejected by Bergson and Little. This point was made very clearly by Little: "If tastes change, we may expect a new ordering of all the conceivable states; but we do not require the difference between the new and the old ordering should bear any particular relation to the changes of taste which have occurred. We have, so to speak, a new world and a new order; and we do not demand correspondence between the change in the world and the change in the order." [37]

Just how much weight this point can bear is seen better from the revised set of conditions given in Section II.1 than from the original set. It turns out that there is only one condition which in fact requires a correspondence of the type that Little objects to, the Independence of Irrelevant Alternatives. The area of fundamental disagreement thus narrows down to this one assumption, to be discussed further in Section IV. It should be noted that of the other conditions, two, Collective Rationality and the Pareto principle, are satisfied by Bergson's social welfare function, and the third, Nondictatorship, hardly raises any fundamental questions.

I therefore conclude that the formulation of the social choice problem as the selection of a constitution is not only compatible with the views

[35] M. C. Kemp and A. Asimakopulos, "A Note on 'Social Welfare Functions' and Cardinal Utility," *Canadian Journal of Economics and Political Science*, Vol. 18, May, 1952, pp. 195-200.

[36] A similar failure to appreciate the inescapability of the logical need for a constitution is found in the criticism of M. C. Kemp, *op. cit.*, footnote 32. He argues that the choice of a decision procedure cannot be made except in terms of a particular choice situation, since choice procedures will be evaluated in terms of their consequences. But this argument in no way invalidates the need for a constitution; it says in fact that for each choice situation (or environment), there is an appropriate way of combining individual orderings to arrive at a social decision. This is exactly equivalent to saying that we can combine individual orderings to arrive at a system of welfare judgments sufficient to arrive at a social choice from any environment.

[37] "Social Choice and Individual Values," *op. cit.* in footnote 32, pp. 423–24.

of such critics as Bergson, Little, and Kemp, but in fact is a logical corollary of their positive position.

2. *The Social Decision Process*

Little [38] has argued cogently that a rule for social decision-making is not the same as a welfare judgment. A welfare judgment requires that some one person is judge; a rule for arriving at social decisions may be agreed upon for reasons of convenience and necessity without its outcomes being treated as evaluations by anyone in particular.

This distinction is well taken. I would consider that it is indeed a social decision process with which I am concerned and not, strictly speaking, a welfare judgment by any individual. That said, however, I am bound to add that in my view a social decision process serves as a proper explication for the intuitive idea of social welfare. The classical problems of formulating the social good are indeed of the metaphysical variety which modern positivism finds meaningless; but the underlying issue is real. My own viewpoint towards this and other ethical problems coincides with that expressed by Popper: "Not a few doctrines which are metaphysical, and thus certainly philosophical, can be interpreted as hypostatizations of methodological rules." [39] All the writers from Bergson on agree on avoiding the notion of a social good not defined in terms of the values of individuals. But where Bergson seeks to locate social values in welfare judgments by individuals, I prefer to locate them in the actions taken by society through its rules for making social decisions. This position is a natural extension of the ordinalist view of values; just as it identifies values and choices for the individual, so I regard social values as meaning nothing more than social choices.

In fact, the Bergson formulation cannot be kept distinct from the interpretation of social welfare in terms of social decision processes. In the first place, the argument of paragraph 1 shows that the Bergson social welfare function is necessarily a constitution, that is, a potential social decision process; the body of welfare judgments made by a single individual are determined, in effect, by the social decision process which the individual would have society adopt if he could. In the second place, the location of welfare judgments in any individual, while logically possible, does not appear to be very interesting. "Social welfare" is related to social policy in any sensible interpretation; the welfare judgments formed by any single individual are unconnected with action and therefore sterile. Bergson recognizes the possible difficulty in his 1954

[38] *Ibid.*, p. 427, pp. 430–32.

[39] K. Popper, *The Logic of Scientific Discovery*, New York: Basic Books, 1959, p. 55.

paper; I quote the passage at length since it displays the issue so well.

"I have been assuming that the concern of welfare economics is to counsel individual citizens generally. If a public official is counseled, it is on the same basis as any other citizen. In every instance reference is made to some ethical values which are appropriate for the counseling of the individual in question. In all this I believe I am only expressing the intent of welfare writings generally; or if this is not the intent, I think it should be. But some may be inclined nevertheless to a different conception, which allows still another interpretation of Arrow's theorem. *According to this view the problem is to counsel not citizens generally but public officials* [emphasis added]. Furthermore, the values to be taken as data are not those which might guide the official if he were a private citizen. The official is envisaged instead as more or less neutral ethically. His one aim in life is to implement the values of other citizens as given by some rule of collective decision-making. Arrow's theorem apparently contributes to this sort of welfare economics the negative finding that no consistent social ordering could be found to serve as a criterion of social welfare in the counseling of the official in question." [40]

I need only add that my interpretation of the social choice problem agrees fully with that given by Bergson beginning with the italicized statement.

Where Bergson, Little, and I seek in varying ways to explicate the notion of social welfare in operational terms, Buchanan's positivism is more extreme.[41] Choice is only individual; the very concept of social welfare is inadmissible, and my use of the term "collective rationality" (by which I meant that social choices corresponding to any given set of individual orderings were so interrelated as to satisfy the definition of an ordering) was strongly attacked on the grounds that only individuals can be rational.[42] Nevertheless, Buchanan and Tullock [43] do put great

[40] "On the Concept of Social Welfare," p. 242.

[41] "Individual Choice . . . ," *op. cit.* in fn. 32.

[42] I generally take the view that a definition is a more or less useful convention, no more. But Buchanan and Little both take very seriously the exact words used in the definiendum. Thus, Buchanan does not so much say that collective rationality, as I define it, is an unsatisfactory condition for a constitution as flatly deny that there can be anything called collective rationality. No doubt, words have penumbras of meaning which may not be easy to eliminate in the reader's mind; but this type of confusion seems secondary when explicit definitions are set forth. Buchanan and Little at some points substitute verbal quibblings for genuine argument.

[43] *The Calculus of Consent*, *op. cit.* in fn. 32, Chapter 6. Since the subsequent references I will make to this work are largely negative, I should note that they do not pertain to the work's essential contribution which is, in my opinion, of major importance.

stress on the selection of a constitution as the central step in developing a social choice mechanism.

3. *Welfare Economics, Compensation and Log-Rolling*

Unlike the criticisms discussed in the last two paragraphs, which raise important questions of meaning, I feel that those I am about to discuss represent elementary confusions or word play. Little, Bergson, and Mishan all agree that my theorem is not part of welfare economics. Thus, "Arrow's work has no relevance to the traditional theory of welfare economics, which culminates in the Bergson-Samuelson formulation" (Little); "The theorem has little or no bearing on welfare economics" (Bergson).[44] The most obvious remark to be made is that one can hardly think of a less interesting question about my theorem than whether it falls on one side or another of an arbitrary boundary separating intellectual provinces. Since Little and Bergson (and also Samuelson, according to Little) accept the Bergson social welfare function as part of welfare economics, the arguments of the last two paragraphs show that any attempt to divide welfare economics in their sense from the theory of social choice must be artificial. At the very least, welfare economics, no matter how defined, has something to do with the public adoption of economic policy, and it is hard to see how any study of the formation of social decisions can have "no relevance to" or "no bearing on" welfare economics.

A more natural division between welfare economics and the theory of social choice, if there is any point in locating one, is provided by defining the former as the implications of Pareto optimality. I take this to be Mishan's meaning in reading my work out of the honorific domain of welfare economics. In the terms used in this book, instead of seeking a social ordering, we confine ourselves to the unanimity quasi-ordering.

Of course, as has been repeated over and over in the literature, this "new welfare economics" says nothing about choices among Pareto-optimal alternatives.[45] The purpose of the soical welfare function was precisely to extend the unanimity quasi-ordering to a full social ordering.

[44] Little, "Social Choice and Individual Values," p. 425; Bergson, "On the Concept of Social Welfare," p. 243.

[45] This point is made with admirable clarity by R. G. Davis, "Comment on Arrow and the 'New Welfare Economics,'" *Economic Journal*, Vol. 68, December, 1958, pp. 834–35, in a comment on Mishan, *op. cit.* in footnote 32. In effect, Mishan accepted Davis's argument; see "Arrow and the 'New Welfare Economics': A Restatement," *ibid.*, Vol. 68, September, 1958, pp. 595–97. For some odd editorial reason, Mishan's rejoinder to Davis appeared before Davis's note.

The compensation principle, based on the new welfare economics, is not an adequate substitute for a social ordering, as has been shown by Scitovsky, Baumol, and others.[46] However, it has been again used by Buchanan and Tullock in their construction. They very correctly argue that trading of votes on different issues (log-rolling) is essentially equivalent to the payment of compensation, but they fail to recognize the ambiguity of the compensation principle as a social decision process. Tullock goes so far as to state that "in processes in which votes are traded . . . the particular type of irrationality described by Arrow is impossible." [47] He seems to be under the impression that I am describing a procedure for deciding separate issues, without any regard to complementarities and substitutions among them. This is a simple misreading; I am concerned, as are the writers on the compensation principle, with choices among social states; a social state is a whole bundle of issues, and I presupposed that all possible combinations of decisions on the separate issues are considered as alternative social states. That this included log-rolling seemed to me so obvious as not to be worth spelling out. The paradox of social choice cannot be so easily exorcised.[48]

IV. The Independence of Irrelevant Alternatives and Interpersonal Comparisons of Intensity

1. *Ordinalism, Observability, and the Independence of Irrelevant Alternatives*

The essential point of the modern insistence on ordinal utility is the application of Leibniz's principle of the identity of indiscernibles. Only observable differences can be used as a basis for explanation. In the field of consumers' demand theory, the ordinalist position turned out to create no problems; cardinal utility had no explanatory power above and beyond ordinal.

[46] T. Scitovsky, "A Note on Welfare Propositions in Economics," *Review of Economic Studies*, Vol. 9, November, 1941, pp. 77–88; W. J. Baumol, "Community Indifference," *ibid.*, Vol. 14, No. 1, 1946–7, pp. 44–48; Chapter IV of the text.

[47] *Op. cit.* in fn. 32, p. 332.

[48] Dahl also notes that an election, in which many issues enter, cannot be interpreted as indicating a majority on any specific issue; he therefore speaks of "minorities" rule as opposed to either majority or minority rule; see *A Preface to Democratic Theory*, Chicago: The University of Chicago Press, 1956, pp. 127–32. However, the significant question is not the existence of a majority on each issue but the existence of a majority on the bundle of issues represented by the candidate over any other attainable bundle.

It is the great merit of Bergson's 1938 paper [49] to have carried the same principle into the analysis of social welfare. The social welfare function was to depend only on indifference maps; in other words, welfare judgments were to be based only on interpersonally observable behavior.

The Condition of Independence of Irrelevant Alternatives extends the requirement of observability one step farther. Given the set of alternatives available for society to choose among, it could be expected that, ideally, one could observe all preferences among the available alternatives, but there would be no way to observe preferences among alternatives not feasible for society.

The austerity imposed by this condition is perhaps stricter than necessary; in many situations, we do have information on preferences for nonfeasible alternatives. It can be argued that, when available, this information should be used in social choice, and some possibilities in this direction will be briefly commented on in the following paragraphs. But clearly, social decision processes which are independent of irrelevant alternatives have a strong practical advantage. After all, every known electoral system satisfies this condition.

It may be worth noting more explicitly than in the text that the market mechanism also operates independently of irrelevant alternatives. If we alter the utility functions of individuals with respect to allocations which are socially infeasible, we do not alter the competitive equilibrium. Indeed, the decentralization of knowledge which is such a virtue of the market mechanism is incompatible with the use of utility comparisons among irrelevant alternatives in arriving at resource allocations.[50]

This paragraph may be concluded by summarizing some conclusions of Guilbaud and Inada, which bring out the underlying meaning of Independence of Irrelevant Alternatives in somewhat different ways. Guilbaud [51] argues that a social welfare function based on individual utilities which are meaningful only up to monotonic transformations

[49] "A Reformulation . . .," pp. 318–20; see also P. A. Samuelson, *Foundations of Economic Analysis*, Cambridge, Massachusetts: Harvard University Press, 1947, p. 228.

[50] Since the market mechanism does satisfy the Condition of Independence of Irrelevant Alternatives, it must violate another condition, which is clearly that of Collective Rationality (Condition 1 or 1'). This violation is precisely the well-known intersection of community indifference curves. Samuelson's social indifference curves (see P. A. Samuelson, "Social Indifference Curves," *Quarterly Journal of Economics*, Vol. 70, February, 1956, pp. 1–22) satisfy the condition of Collective Rationality but violate that of Independence of Irrelevant Alternatives; the income redistribution associated with each change in social production possibilities requires a high degree of centralization of knowledge about individual utility functions.

[51] *Op. cit.*, pp. 576–84

must be what is mathematically termed a functional, not a function in the ordinary sense. That is, each social choice must depend on the entire individual ordering. Inada's [52] result is somewhat similar; if the social welfare function is to be determined from individual orderings, then the marginal social rates of substitution among commodity allocations to individuals cannot be determined exclusively from the individuals' marginal rates of substitution at that point; even local social choices must depend upon individual preferences in the large.[53]

2. *The Impossibility of Purely Ethical Comparisons*

Bergson, in both his papers, takes the position that interpersonal comparisons are purely ethical in nature; this is just another way of saying that only individual indifference maps (and not cardinal utilities) are used in forming welfare judgments. However, especially in his second paper, his exposition brings out the difficulties involved.[54] He comes at points very close to asserting a meaningful empirical interpersonal comparison: "The ultimate criterion would be the perfectly plausible one of the comparative degree to which wants of different orders are satisfied for different individuals. . . . The individual members of the community are all supposed to order social states on the ethical premise that distribution should be according to need." The meaningfulness of "wants of different orders" or of "need" seems to depend on an empirical comparison. But Bergson then veers away: "It might even be necessary to pair by separate ethical premises all the indifference curves of each household with all those of every other one." Later, although not necessarily rejecting the existence of meaningful empirical interpersonal comparisons, he reemphasizes that purely ethical comparisons are adequate for the formation of social welfare judgments: "The criterion must be ethical in character. This does not by itself rule out empirical comparability, but it means that even with this supposition one must establish

[52] K. Inada, "On the Economic Welfare Function," Technical Report No. 97, Contract Nonr-225(50) for the Office of Naval Research, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California, July 13, 1961.

[53] This result presupposes that the social welfare function is defined for all possible utility functions of individuals subject only to the usual conditions of quasi-concavity and monotonicity. If it is known *a priori* that the utility functions of all individuals are purely individualistic (see Chapter VI) and fall into some very restricted class, such as displaying constant marginal utility for some one commodity, then social welfare functions can be drawn up for which social marginal rates of substitution are completely determined by individual rates at the same point.

[54] "On the Concept of Social Welfare," pp. 244–45.

why the criterion is ethically compelling . . . If one can advance the Utilitarian criterion with empirical comparability then it should also be possible to do so without it." [55]

It seems that the last sentence should almost be inverted. If there is no empirical way of comparing two states (say, indifference curves of two different individuals), there can be no ethical way of distinguishing them. Value judgments may equate empirically distinguishable phenomena, but they cannot differentiate empirically indistinguishable states. The pairing of indifference curves referred to above requires that there be some operational meaning, if only an ideal one, to the comparison.

3. *Interpersonal Comparisons and Enlarged Indifference Maps*

If empirically meaningful interpersonal comparisons have to be based on indifference maps, as we have argued, then the Independence of Irrelevant Alternatives must be violated. The information which enables us to assert that individual A prefers x to y more strongly than B prefers y to x must be based on comparisons by A and B of x and y not only to each other but also to other alternatives.

In fact, the indifference map that must be used will probably contain many more dimensions than might ordinarily be considered. Verbal or other expressive behavior might possibly be added to the economists' more usual comparison of bundles. One cannot help being influenced by such psychological work as that of S. S. Stevens,[56] who could find a cardinal scale for such a phenomenon as pitch by asking subjects which of two reference notes a given note was closer to. The resemblance to (better, identity with) a well-known line of thought, stemming from Pareto, which justifies cardinal utility on the basis of ordinal comparisons of *changes* from one commodity vector to another, is obvious.[57]

The strategy of deriving empirically and then ethically meaningful interpersonal comparisons is apparently to proceed as follows. (1) The preference system of the individual is explored with respect to many irrelevant alternatives, indeed well beyond the regions that are relevant for individual choice. (2) This wider realm of choices can, of course, be

[55] *Ibid.*, pp. 250–51.

[56] S. S. Stevens, "The Psychophysics of Sensory Function," *The American Scientist*, Vol. 48, June, 1961, pp. 226–53; see especially the discussion of alternative scales on pp. 231–33. The method mentioned in the text is referred to by Stevens as a "category scale"; the "magnitude scale," which he prefers, is even more of a break with the typical views of economists on the possibilities of observing behavior.

[57] The classic and still best statement of this position is that of R. Frisch, "Sur un problème d'économie pure," *Norsk Matematisk Forenings Skrifter*, Serie I, No. 16, 1926, pp. 1–40.

represented by utility functions; although these are in principle unique only up to monotonic transformations, in fact a set of utility functions sufficiently restricted to define a cardinal utility appear most natural (as in the use of cardinal utility in explaining risk-taking). (3) Finally, one of the utility differences is widely accepted to have an interpersonal ethical significance and can be used to arrive at the ethical interpersonal equation of utility differences which are not directly comparable.

A simple illustration of the preceding is the suggestion of Dahl [58] that voting and other forms of political activity may be taken as an interpersonal measure of preference intensity. It is assumed, of course, that voting is per se a disutility. Let x and y be two alternative social states, let v be the act of voting, and let (x, v) be the combination of the social state x and the act of voting. If an individual prefers (x, v) to y, then his preference for x over y must be at least sufficiently great to cover the disutility of voting. We can find a family of utility indicators in the joint space of social states and the act of voting. Suppose in particular we can find one utility indicator for which the social state and the act of voting are independent, i.e., the utilities of the two are additive. Then to say that (x, v) is preferred to y is to say that

$$U(x) - c(v) > U(y),$$

where U is the utility ascribed to a social state, and $c(v)$ is the disutility ascribed to voting. Equivalently, we can say that

$$U(x) - U(y) > c(v).$$

If finally we make the ethical assumption that the disutility of voting is equivalent for all individuals, then the utility differences between x and y for all those voting for x can be assumed at least equal to the disutility of voting and vice versa.

Unfortunately, most of the links in this chain are weak. In the first place, all that can be established about a utility difference is an inequality; we know that the utility difference of one who votes for x is at least the disutility of voting, but it might be a good deal more. Hence, counting the votes is not sufficient to find the sum of the utilities. Nor is this problem peculiar to this example; it is intrinsic in the mechanism of political choice.

Second, the ethical assumption of interpersonal equality of the disutility of voting is not one which we would really insist on; the disutility of voting might be expected to differ among individuals depending on wealth or intelligence or health.

[58] *Op. cit.*, pp. 134–5.

Third, even with all the strong assumptions that have been made, the problems raised by not requiring the Independence of Irrelevant Alternatives occur. If the voting actually takes place, then the disutility of voting has been incurred and has partially wiped out for the winners their gain in the victory of x; at the same time those losers who felt strongly enough to vote for y have not only suffered the social acceptance of x but also the disutility of voting. Thus if x is chosen as the result of voting, it could easily happen that the sum of utilities of the entire process, including the act of voting, be less than if y were imposed without voting.

Of course, the best solution of all, according to this logic, would be to impose x without voting. This indeed would be the true meaning of using irrelevant alternatives as a measure of preference intensities; it is the willingness to incur the sacrifice of voting, not the sacrifice itself, which is the measure of intensity. But how can this be made operational?

Basically, I can only leave the matter with this conundrum. But in the next two paragraphs two particular classes of interpersonal comparisons based on enlarged indifference maps will be discussed.

4. *Extended Sympathy* [59]

One type of interpersonal comparison to be found in the most ancient ethical writings has yet to receive significant expression and formalization in political and economic contexts. It is exemplified, in perhaps an extreme form, by an inscription supposedly found in an English graveyard.

> Here lies Martin Engelbrodde,
> Ha'e mercy on my soul, Lord God,
> As I would do were I Lord God,
> And Thou wert Martin Engelbrodde.

People seem prepared to make comparisons of the form: State x is better (or worse) for me than state y is for you.[60] This is certainly one way of approaching the notion of an appropriate income distribution; if I am richer than you, I may find it easy to make the judgment that it is better for you to have the marginal dollar than for me.

[59] I am indebted to H. K. Zassenhaus for stressing to me the importance of this notion for welfare judgments.

[60] A more formal presentation of this notion is given by my colleague, P. Suppes, "Two Formal Models for Moral Principles," Technical Report No. 15, Office of Naval Research Contract Nonr 225(17), Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California, November 1, 1957, pp. 17–18.

The ordinalist would ask what possible meaning the comparison could have to anyone; a comparison should represent at least a conceivable choice among alternative actions. Interpersonal comparisons of the extended sympathy type can be put in operational form; the judgment takes the form: It is better (in my judgment) to be myself in state x than to be you in state y.

In this form, the characteristics that define an individual are included in the comparison. In effect, these characteristics are put on a par with the items usually regarded as constituting an individual's wealth. The possession of tools would ordinarily be regarded as part of the social state; why not the possession of the skills to use the tools and the intelligence which lies behind those skills? Individuals, in appraising each other's states of well-being, consider not only material possessions but also find themselves "desiring this man's scope and that man's art." [61]

The principle of extended sympathy as a basis for interpersonal comparisons seems basic to many of the welfare judgments made in ordinary practice. But it is not easy to see how to construct a theory of social choice from this principle.

5. *The Ability to Discriminate*

A recurrent approach to interpersonal comparisons has been the use of just noticeable utility differences as interpersonally valid units.[62] The clearest discussion, which will be analyzed here, is that of Goodman and Markowitz. They argue that no individual can make indefinitely fine comparisons of alternatives. Hence, it may be supposed that each individual has only a finite number of levels of discrimination; a change from one level to the next represents the minimum difference which is discernible to an individual. Goodman and Markowitz then make the basic ethical assumption that the significance of a change from one discretion level to the next is the same for all individuals and independent of the level from which the change is made.

[61] For an interesting discussion of the moral implications of the position that many attributes of the individual are similar in nature to external possessions, see V. C. Walsh, *Scarcity and Evil*, Englewood Cliffs, New Jersey: Prentice-Hall, 1961. I am indebted for this reference to R. A. Mundell.

[62] See the discussion of Borda in Part I, p. 99; F. Y. Edgeworth, *Mathematical Psychics*, London: C. Kegan and Paul, 1881, pp. 7–8; W. E. Armstrong, "Utility and the Theory of Welfare," *Oxford Economics Papers*, New Series, Vol. 3, October, 1951, pp. 259–71; and L. Goodman and H. Markowitz, *op. cit.*, fn. 14. For excellent analyses of Armstrong's and Goodman-Markowitz's views, see Rothenberg (*op. cit.* in fn. 7), Chapters 7 and 8.

The consequence of this assumption (in conjunction with other, more usual, conditions, Collective Rationality, the Pareto principle, and Equality) is that social choice is made according to a sum of individual utilities, where the utility of any individual for any social state is the number of discrimination levels below the level in which the individual places the given social state.

Two problems are raised by this procedure: one is the operational meaning of the discrimination levels and the other is the ethical desirability of using discrimination levels as an interpersonal measure. In regard to the first, it must be asked what experiments could one conduct, at least ideally, which would define the discrimination level to be attached to any social alternative by any individual? Actually, Goodman and Markowitz do not give a clear account. They distinguish [63] between knowing the true discrimination level l_{ij} of the alternative j for the individual i and the ranking a_{ij} of that alternative among a fixed set of alternatives. They suggest that only the latter are observable and should or could be used. The implied experiment can be stated as follows: There is a subset S of all logically possible social states such that every environment (set of actually attainable social states) which is at all likely to occur will be a subset of S. Each individual then draws his indifference map within S; by the assumption of finite discriminatory ability, he will simply divide S into a finite number of subsets within each of which he is completely indifferent. These subsets are then assigned ranks, which are the utilities.

The ranks can, however, easily depend on the choice of S. Suppose, for example, that a new commodity becomes available but is prohibitively expensive. If S is expanded by adding distributions of commodities including the new one, it may contain elements more desirable than any of the old ones, as well as some less desirable than some old ones. The additional components of the social state vector will increase the possibility of discrimination, so that it is to be expected that the enlargement of S will introduce new discrimination levels whose ranks lie between some of the old ones. Then the perceived ranks, a_{ij}, of the original alternatives will be altered by the introduction of new alternatives which may not be technologically feasible.

This objection is, of course, simply another illustration of the argument for the principle of Independence of Irrelevant Alternatives.

The ethical desirability of basing welfare judgments on discrimination levels can be examined best by means of an illustration. Since the fundamental function of any theory of social welfare is to supply criteria for income distribution, let us assume the existence of a single commodity

[63] *Op. cit.*, pp. 5–6.

(income) to be distributed between two individuals; the total quantity of the commodity will be taken as 1. Let x be the amount given to individual 1, $1 - x$ the amount given to individual 2. According to Goodman and Markowitz, x is chosen to maximize $U_1(x) + U_2(1 - x)$, where the functions U_i represent the discrimination levels of the two individuals.

Each individual could receive from 0 to 1 unit, but because of his finite discriminatory ability this interval is divided into a finite number of segments, within each of which the individual cannot discriminate. Suppose then individual 1 divides the interval into m equal parts, within each of which he is indifferent; for definiteness, suppose that each part includes its lower limit and excludes its upper. The amount 1 is in a class by itself. The utility function for individual 1 is then

$$U_1(x) = [mx],$$

where $[y]$ means the largest integer not exceeding y. Similarly, if individual 2 divides the interval from 0 to 1 into n parts, his utility function is

$$U_2(1 - x) = [n(1 - x)].$$

Suppose now the $n > m$, i.e., individual 2 is capable of finer discrimination than individual 1. Then the Goodman-Markowitz criterion yields the remarkable conclusion that the entire income should go to individual 2. For suppose that $x > 0$ at the optimum. Then $U_1(x) \geq 0$. If $U_1(x) > 0$, let

$$U_1(x) = r.$$

Then

$$mx \geq [mx] = U_1(x) = r,$$

so that

$$U_2(1 - x) = [n(1 - x)] \leq \left[n - r\left(\frac{n}{m}\right)\right].$$

Since $n > m$ and $r > 0$,

$$n - r\left(\frac{n}{m}\right) < n - r,$$

and since the right-hand side is an integer,

$$U_2(1 - x) \leq \left[n - r\left(\frac{n}{m}\right)\right] < n - r = n - U_1(x),$$

or

$$U_1(x) + U_2(1 - x) < n.$$

On the other hand, if $x = 0$, $1 - x = 1$, we have

$$U_1(0) + U_2(1) = [0] + [n] = n > U_1(x) + U_2(1 - x),$$

which contradicts the presumed optimality of $x > 0$, provided $U_1(x) > 0$. If $x > 0$, but $U_1(x) = 0$, note that

$$U_1(x) + U_2(1 - x) = 0 + [n(1 - x)] < n = U_1(0) + U_2(1),$$

so that the same contradiction holds.[64]

That a slight difference in sensitivity should lead to complete inequality hardly seems ethically reasonable.[65]

V. Collective Rationality

The condition that welfare judgments should take the form of an ordering has been much less disputed than others, although Buchanan (see Section III.3) argues that rationality is a property of individuals only and there is no reason to attribute rationality to society. There have been other incidental comments (e.g., by Kemp [66]) that a social decision process might well sacrifice transitivity if necessary to satisfy other conditions.

The two properties which characterize an ordering are connectedness and transitivity (see p. 13 of text). Connectedness, when understood, can hardly be denied; it simply requires that some social choice be made from any environment. Abstention from a decision cannot exist; some social state will prevail.

[64] This result is obviously closely related to the discussion of J. von Neumann and O. Morgenstern of games with discrete utility scales; there too the less discerning player is completely discriminated against. See *Theory of Games and Economic Behavior*, second edition, Princeton, New Jersey: Princeton University Press, 1947, pp. 614–16.

[65] Examples such as the above may be more convincing than they should be, because of hidden empirical content. If in fact we never run across two individuals related as above, then our intuitive rejection of the completely unequal income distribution may be incorrect; our intuition is guided by situations we have encountered, and it could be that if we were to meet two individuals, one of whom is more sensitive than the other, no matter what their income levels are, we would accept the Goodman-Markowitz solution. Goodman and Markowitz have suggested to me orally that in fact discrimination levels would become closer together at lower levels of income (this is one interpretation of diminishing marginal utility of income); thus, the sum of utilities would be maximized at some point short of complete inequality. Since in realistic cases, it could be argued, the Goodman-Markowitz social welfare function does not lead to blatant absurdity, we should not use the evidence of the example in the text. Actually, it would not be difficult to construct examples similar to the one in the text for which, however, increasing ability to discriminate at lower levels of income is sufficient to prevent complete inequality, but nevertheless a moderate difference in the number of levels of discrimination leads to a very great inequality of income.

[66] *Op. cit.* in fn. 32, pp. 242–43.

However, it cannot be denied that there is an important sense in which one may speak, although loosely, of the absence of decision. In any historically given situation there is a social state which has a preferred status in social choice in that it will be adopted in the absence of a specific decision to the contrary. Politically, the status quo has this property, as is frequently all too obvious. If one considers the entire distribution of goods, the preferred alternative is slightly different. Suppose that social state x obtains when the environment is S. Now suppose there is an innovation, so that the environment expands. In a basically free enterprise economy there will be an adaptation which is controlled by property laws and the workings of the system. The resulting social state, say x^1, which in general differs from x, will automatically prevail unless there is a specific legislative decision to alter it.

There is a fundamental divergence of opinion, which has not been fully recognized, on whether or not social choices should be historically conditioned, or, equivalently, whether or not inaction is an alternative different from other alternatives.[67] The social welfare function approach, whether in Bergson's version or in mine, and "populistic democracy," as Dahl [68] terms it, both imply that the social choice at any moment is determined by the range of alternative social states available (given the preferences of individuals); there is no special role given to one alternative because it happens to be identical to or derived from a historically given one. In the case of choice between two alternatives, this point of view tends toward majority rule.[69]

Buchanan and Tullock, on the other hand, distinguish between retaining and changing the status quo most clearly in the following quotation: [70] "We must sharply differentiate between two kinds of decisions: (1) the positive decision that authorizes action for the social group, and (2) the negative decision that effectively blocks action proposed by another group. If a group is empowered to make decisions resulting in positive action by/for the whole group, we shall say that this group effectively 'rules' for the decisions in question. It does not seem meaningful to say

[67] The sharpest confrontation of opposing views is found in the controversy between A. Downs, "In Defense of Majority Voting," *Journal of Political Economy*, Vol. 69, April, 1961, pp. 192–99, and G. Tullock, "Reply to a Traditionalist," *ibid.*, pp. 200–03.

[68] See Dahl, *op. cit.*, Chapter 2, especially p. 41.

[69] The refusal to consider any alternative as being preferred on historical grounds removes one reason for nonneutrality (see Section II.3) and hence, according to May's argument, implies majority rule if his other, less controversial, conditions are accepted. However, there may be other reasons for rejecting neutrality in specific situations, so that qualified majorities may still be called for.

[70] *Op. cit.*, pp. 258–59.

that the power to block action constitutes effective 'rule.' . . . The power of blocking action is not what we normally mean, or should mean, when we speak of 'majority rule' or 'minority rule.'" The asymmetry between action and inaction is closely related to their support of unanimity as the ideal criterion of choice; [71] under such a rule, the status quo is a highly privileged alternative.

It is against this background that the importance of the transitivity condition becomes clear. Those familiar with the integrability controversy in the field of consumer's demand theory [72] will observe that the basic problem is the same: the independence of the final choice from the path to it. Transitivity will insure this independence; from any environment, there will be a chosen alternative, and, in the absence of a deadlock, no place for the historically given alternative to be chosen by default.

That an intransitive social choice mechanism may as a matter of observed fact produce decisions that are clearly unsatisfactory has been brought out in different ways by Riker [73] and by Dahl. [74] Riker's emphasis is on the possibility that legislative rules may lead to choice of a proposal opposed by a majority, Dahl's rather on the possibility that the rules lead to a deadlock and therefore a socially undesired inaction. The notion of a "democratic paralysis," a failure to act due not to a desire for inaction but an inability to agree on the proper action, seems to me to deserve much further empirical, as well as theoretical, study.

Collective rationality in the social choice mechanism is not then merely an illegitimate transfer from the individual to society, but an important attribute of a genuinely democratic system capable of full adaptation to varying environments.

[71] In the absence of costs of decision-making; it is only these costs which, in their theory, explain the acceptance of less-than-unanimous decisions.

[72] See footnote 5, p. 13 of text. Since the writing of the text, several important papers have been written in this area; see H. Houthakker, "Revealed Preference and the Utility Function," *Economica*, New Series, Vol. 17, May, 1950, pp. 159–74; H. Uzawa, "Preference and Rational Choice in the Theory of Consumption," Chapter 9 in K. J. Arrow, S. Karlin, and P. Suppes (eds.), *Mathematical Methods in the Social Sciences, 1959*, Stanford, California: Stanford University Press, 1960. Ville's paper has been translated by P. K. Newman as "The Existence-Conditions of a Total Utility Function," *Review of Economic Studies*, Vol. 19, No. 2, 1951–52, pp. 123–128. An excellent historical survey has been given by P. A. Samuelson, "The Problem of Integrability in Utility Theory," *Economica*, Vol. 17, November, 1950, pp. 355–85.

[73] W. H. Riker, "The Paradox of Voting and Congressional Rules for Voting on Amendments," *American Political Science Review*, Vol. 52, June, 1958, pp. 349–66.

[74] *Op. cit.*, pp. 39–41.