# CARDINAL WELFARE, INDIVIDUALISTIC ETHICS, AND INTERPERSONAL COMPARISONS OF UTILITY[1]

JOHN C. HARSANYI
University of Queensland

## I

THE naïve concept of social welfare as a sum of intuitively measurable and comparable individual cardinal utilities has been found unable to withstand the methodological criticism of the Pareto school. Professor Bergson[2] has therefore recommended its replacement by the more general concept of a social welfare function, defined as an arbitrary mathematical function of economic (and other social) variables, of a form freely chosen according to one's personal ethical (or political) value judgments. Of course, in this terminology everybody will have a social welfare function of his own, different from that of everybody else, except to the extent to which different individuals' value judgments happen to coincide with one another. Actually, owing to the prevalence of individualistic value judgments in our society, it has been generally agreed that a social welfare function should be an increasing function of the utilities of individuals: if a certain situation, $X$, is preferred by an individual to another situation, $Y$, and if none of the other individuals prefers $Y$ to $X$, then $X$ should be regarded as socially preferable to $Y$. But no other restriction is to be imposed on the mathematical form of a social welfare function.

Recently, however, Professor Fleming[3] has shown that if one accepts one further fairly weak and plausible ethical postulate, one finds one's social welfare function to be at once restricted to a rather narrow class of mathematical functions so as to be expressible (after appropriate monotone transformation of the social welfare and individual utility indexes if necessary) as the weighted sum of the individuals' utilities. This does not mean, of course, a return to the doctrine that the existence of an additive cardinal utility function is intuitively self-evident. The existence of such a function becomes, rather, the consequence of the ethical postulates adopted and is wholly dependent on these postulates. Still, Fleming's results do in a sense involve an unexpected revival of some views of the pre-Pareto period.

In this paper I propose, first of all, to examine the precise ethical meaning of Fleming's crucial postulate and to show that it expresses an *individualistic* value judgment going definitely beyond the generally adopted individualistic postu-

[1] I am indebted to my colleagues at the University of Queensland, Messrs. R. W. Lane and G. Price, for helpful comments. Of course, the responsibility for shortcomings of this paper and for the opinions expressed in it is entirely mine.

[2] A. Bergson (Burk), "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, LII (February, 1938), 310–34, and "Socialist Economics," in *A Survey of Contemporary Economics*, ed. H. S. Ellis (Philadelphia, 1949), esp. pp. 412–20.

[3] M. Fleming, "A Cardinal Concept of Welfare," *Quarterly Journal of Economics*, LXVI (August, 1952), 366–84. For a different approach to the same problem see L. Goodman and H. Markovitz, "Social Welfare Functions Based on Individual Rankings," *American Journal of Sociology*, Vol. LVIII (November, 1952).

late mentioned earlier, though it represents, as I shall argue, a value judgment perfectly acceptable according to common ethical standards (Sec. II). I shall also attempt to show that, if both social and individual preferences are assumed to satisfy the von Neumann–Morgenstern-Marschak axioms about choices between uncertain prospects, even a much weaker ethical postulate than Fleming's suffices to establish an additive cardinal social welfare function (Sec. III). In effect, it will be submitted that a mere logical analysis of what we mean by value judgments concerning social welfare and by social welfare functions leads, without any additional ethical postulates, to a social welfare function of this mathematical form (Sec. IV). Finally, I shall turn to the problem of interpersonal comparisons of utility, which gains new interest by the revival of an additive cardinal welfare concept, and shall examine what logical basis, if any, there is for such comparisons (Sec. V).

## II

Fleming expresses his ethical postulates in terms of two alternative conceptual frameworks: one in terms of an "*ideal utilitarianism*" of G. E. Moore's type, the other in terms of a *preference* terminology more familiar to economists. Though he evidently sets greater store by the first approach, I shall adopt the second, which seems to be freer of unnecessary metaphysical commitments. I have also taken the liberty of rephrasing his postulates to some extent.

*Postulate A* (*asymmetry of social preference*).—If "from a social standpoint"[4] situation $X$ is preferred to situation $Y$, then $Y$ is not preferred to $X$.

*Postulate B* (*transitivity of social preference*).—If from a social standpoint $X$ is preferred to $Y$, and $Y$ to $Z$, then $X$ is preferred to Z.

*Postulate C* (*transitivity of social indifference*).—If from a social standpoint neither of $X$ and $Y$ is preferred to the other, and again neither of $Y$ and $Z$ is preferred to the other, then likewise neither of $X$ and $Z$ is preferred to the other.

These three postulates are meant to insure that "social preference" establishes a *complete ordering* among the possible social situations, from which the existence of a social welfare function (at least of an ordinal type) at once follows. (Actually, two postulates would have sufficed if, in the postulates, "weak" preference, which does not exclude the possibility of indifference, had been used instead of "strong" preference.)

*Postulate D* (*positive relation of social preferences to individual preferences*).—If a given individual $i$ prefers situation $X$ to situation $Y$, and none of the other individuals prefers $Y$ to $X$, then $X$ is preferred to $Y$ from a social standpoint.

As already mentioned Postulate D expresses a generally accepted individualistic value judgment.

Finally, Fleming's Postulate E states essentially that on issues on which two individuals' interests (preferences) conflict, all other individuals' interests being unaffected, social preferences should depend exclusively on comparing the relative social importance of the interests at stake of each of the two individuals concerned. In other words, it requires that

[4] Of course, when I speak of preferences "from a social standpoint," often abbreviated to "social" preferences and the like, I always mean preferences based on a given individual's value judgments concerning "social welfare." The foregoing postulates are meant to impose restrictions on *any* individual's value judgements of this kind, and thus represent, as it were, value judgments of the second order, that is, value judgments concerning value judgments. Later I shall discuss the concept of "preferences from a social standpoint" at some length and introduce the distinctive term "ethical preferences" to describe them (in Sec. IV). But at this stage I do not want to prejudge the issue by using this terminology.

the distribution of utilities between each pair of individuals should be judged separately on its own merits, independently of how utilities (or income) are distributed among the other members of the community.

*Postulate E (independent evaluation of the utility distribution[5] between each pair of individuals).*—(1) There are at least three individuals. (2) Suppose that individual $i$ is indifferent between situations $X$ and $X'$ and also between situations $Y$ and $Y'$, but prefers situations $X$ and $X'$ to situations $Y$ and $Y'$. Suppose, further, that individual $j$ is also indifferent between $X$ and $X'$ and between $Y$ and $Y'$, but (unlike individual $i$) prefers $Y$ and $Y'$ to $X$ and $X'$. Suppose also that all other individuals are indifferent between $X$ and $Y$, and likewise between $X'$ and $Y'$.[6] Then social preferences should always go in the same way between $X$ and $Y$ as they do between $X'$ and $Y'$ (that is, if from a social standpoint $X$ is preferred to $Y$, then $X'$ should also be preferred to $Y'$; if from a social standpoint $X$ and $Y$ are regarded as indifferent, the same should be true of $X'$ and $Y'$; and if from a social standpoint $Y$ is preferred to $X$, then $Y'$ should also be preferred to $X'$).

Postulate E is a natural extension of the individualistic value judgment expressed by Postulate D. Postulate D already implies that if the choice between two situations $X$ and $Y$ happens to affect the interests of the individuals $i$ and $j$ only, without affecting the interests of anybody else, social choice must depend exclusively on $i$'s and $j$'s interests—provided that $i$'s and $j$'s interests *agree* in this matter. Postulate E now adds that in the assumed case social choice must depend exclusively on $i$'s and $j$'s interests (and on weighing these two interests one against the other in terms of a consistent ethical standard), even if $i$'s and $j$'s interests are in *conflict*. Thus both postulates make social choice dependent solely on the *individual* interests directly affected.[7] They leave no room for the separate interests of a superindividual state or of impersonal cultural values[8] (except for the ideals of equity incorporated in the ethical postulates themselves).

At first sight, Postulate E may look inconsistent with the widespread habit of judging the "fairness" or "unfairness" of the distribution of income between two individuals, not only on the basis of these two people's personal conditions and needs, but also on the basis of comparing

[5] The more general term "utility distribution" is used instead of the term "income distribution," since the utility enjoyed by each individual will, in general, depend not only on his own income but also, owing to external economies and diseconomies of consumption, on other people's incomes.

[6] It is not assumed, however, that the other individuals are (like $i$ and $j$) indifferent between $X$ and $X'$ and between $Y$ and $Y'$. In effect, were this restrictive assumption inserted into Postulate E, this latter would completely lose the status of an independent postulate and would become a mere corollary of Postulate D.

[7] In view of consumers' notorious "irrationality," some people may feel that these postulates go too far in accepting the consumers' sovereignty doctrine. These people may reinterpret the terms in the postulates referring to individual preferences as denoting, not certain individuals' actual preferences, but rather their "true" preferences, that is, the preferences they *would* manifest under "ideal conditions," in possession of perfect information, and acting with perfect logic and care. With some ingenuity it should not be too difficult to give even some sort of "operational" meaning to these ideal conditions, or to some approximation of them, acceptable for practical purposes. (Or, alternatively, these terms may be reinterpreted as referring even to the preferences that these individuals *ought* to exhibit in terms of a given ethical standard. The latter interpretation would, of course, deprive the postulates of most of their individualistic meaning.)

[8] These postulates do not exclude, however, the possibility that such consideration may influence the relative weights given to different individuals' utilities within the additive social welfare function. Even by means of additional postulates, this could be excluded only to the extent to which the comparison of individual utilities can be put on an objective basis independent of individual value judgments (see Sec. V).

their incomes with the incomes of the other members of their respective social groups. Thus people's judgments on the income distribution between a given worker and his employer will also depend on the current earnings of other similar workers and employers. But the conflict with Postulate E is more apparent than real. In a society with important external economies and diseconomies of consumption, where the utility of a given income depends not only on its absolute size but also on its relation to other people's incomes, it is not inconsistent with Postulate E that, in judging the income distribution between two individuals, other people's incomes should also be taken into account. An income distribution between a given worker and a given employer, which in the original situation seemed perfectly "fair" in terms of a given ethical standard, may require adjustment in the worker's favor, once wages have generally gone up, since the worsening of this worker's position relative to that of his fellows must have reduced him to a lower level of utility.

Postulate E requires that the distribution of *utility* between two individuals (once the utility levels of the two individuals are given) should always be judged independently of how utility and income are distributed among other members of the society. In the absence of external economies and diseconomies of consumption, this would necessarily also mean judging the distribution of *income* between two individuals independently of the incomes of others. In the presence of such economies and diseconomies, however, when the utility level of any person depends not only on his own income but also on other persons' incomes, it is not inconsistent with Postulate E that our value judgment on the distribution of income between two individuals should be influenced by the income distribution in the rest of the society—in so far as the income distribution in the rest of the society affects the utility levels of these two individuals themselves and consequently the distribution of utility between them. Postulate E demands only that, once these effects have been allowed for, the distribution of income in the rest of the society must not have any further influence on our value judgment.

## III

In accordance with prevalent usage in welfare economics, Fleming's postulates refer to social or individual preferences between *sure prospects* only. However, it seems desirable to have both sorts of preferences defined for choices between *uncertain prospects* as well. More often than not, we have to choose in practice between social policies that promise given definite results only with larger or smaller probabilities. On the other hand, if we subscribe to some sort of individualistic ethics, we should like to make social attitude toward uncertainty somehow dependent on individual attitudes toward it (at least if the latter do not manifest too patent and too great an inconsistency and irrationality).

Since we admit the possibility of external economies and diseconomies of consumption, both social and individual prospects will, in general, specify the amounts of different commodities consumed and the stocks of different goods held by all individuals at different future dates (up to the time horizon adopted), together with their respective probabilities.

As the von Neumann–Morgenstern axioms[9] or the Marschak postulates[10]

[9] See J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (2d ed.; Princeton, 1947), pp. 641 ff.

[10] J. Marschak, "Rational Behavior, Uncertain Prospects, and Measurable Utility," *Econometrica*,

equivalent to them (which latter I shall adopt) are essential requirements for rational behavior, it is natural enough to demand that both social and individual preferences[11] should satisfy them. This gives us:

*Postulate a.*—Social preferences satisfy Marschak's Postulates I, II, III′, and IV.

*Postulate b.*—Individual preferences satisfy the same four postulates.

In addition, we need a postulate to secure the dependence of social preferences on individual preferences:

*Postulate c.*—If two prospects $P$ and $Q$ are indifferent from the standpoint of every individual, they are also indifferent from a social standpoint.

Postulate $c$ once more represents, of course, an individualistic value judgment—though a very weak one, comparable to Fleming's Postulate D rather than to his Postulate E.

I propose to show that Postulate $c$ suffices to establish that the cardinal social welfare function defined by Postulate $a$ can be obtained as a weighted sum of the cardinal individual utility functions defined by Postulate $b$ (on the understanding that the zero point of the social welfare function is appropriately chosen).

*Theorem I.*—There exists a social welfare function such that its actuarial value is maximized by choices conformable to the social preferences given. This social welfare function is unique up to linear transformation.

*Theorem II.*—For each individual there exists a utility function such that its actuarial value is maximized by choices conformable to the individual's preferences. This utility function is unique up to linear transformation.

Both theorems follow from Marschak's argument.

Let $W$ denote a social welfare function satisfying Theorem I and $U_i$ denote a utility function of the $i$'th individual, satisfying Theorem II. Moreover, let $W$ be chosen so that $W = 0$ if for all the $n$ individuals $U_1 = U_2 = \ldots = U_n = 0$.

*Theorem III.*—$W$ is a single-valued function of $U_1, U_2, \ldots, U_n$. This follows, in view of Theorems I and II, from Postulate $c$.

*Theorem IV.*—$W$ is a homogeneous function of the first order of $U_1, U_2, \ldots, U_n$.

*Proof.*—We want to show that, if the individual utilities $U_1 = u_1$; $U_2 = u_2$; $\ldots$; $U_n = u_n$ correspond to the social welfare $W = w$, then the individual utilities $U_1 = k \cdot u_1$; $U_2 = k \cdot u_2$; $\ldots$; $U_n = k \cdot u_n$ correspond to the social welfare $W = k \cdot w$.

This will be shown first for the case where $0 \leqq k \leqq 1$. Suppose that prospect $O$ represents $U_1 = U_2 = \ldots = U_n = 0$

---

XVIII (1950), 111–41, esp. 116–21. Marschak's postulates can be summarized as follows. *Postulate I (complete ordering)*: The relation of preference establishes a complete ordering among all prospects. *Postulate II* (continuity): If prospect $P$ is preferred to prospect $R$, while prospect $Q$ has an intermediate position between them (being preferred to $R$ but less preferred than $P$), then there exists a mixture of $P$ and $R$, with appropriate probabilities, such as to be exactly indifferent to $Q$. *Postulate III′ (sufficient number of nonindifferent prospects)*: There are at least four mutually nonindifferent prospects. *Postulate IV (equivalence of mixture of equivalent prospects)*: If prospects $Q$ and $Q'$ are indifferent, then, for any prospect $P$, a given mixture of $P$ and $Q$ is indifferent to a similar mixture of $P$ and $Q'$, (that is, to a mixture of $P$ and $Q'$ which has the same probabilities for the corresponding constituent prospects).

Postulate I is needed to establish the existence of even an *ordinal* utility (or welfare) function, while the other three postulates are required to establish the existence of a *cardinal* utility (or welfare) function. But, as Postulates II and III are almost trivial, Postulate IV may be regarded as being decisive for cardinality as against mere ordinality.

[11] There are reasons to believe that, in actuality, individual preferences between uncertain prospects do not always satisfy these postulates of rational behavior (for example, owing to a certain "love of danger"; see Marschak, *op. cit.*, pp. 137–41). In this case we may fall back again upon the preferences each individual *would* manifest under "ideal conditions" (see n. 5).

for the different individuals and consequently represents $W = 0$ for society, while prospect $P$ represents $U_1 = u_1$; $U_2 = u_2$; . . . ; $U_n = u_n$ for the former and $W = w$ for the latter. Moreover, let $Q$ be the mixed prospect of obtaining either prospect $O$ (with the probability $1 - p$) or prospect $P$ (with the probability $p$). Then, obviously, $Q$ will represent $U_1 = p \cdot u_1$; $U_2 = p \cdot u_2$; . . . ; $U_n = p \cdot u_n$ for the individuals and $W = p \cdot w$ for society. Now, if we write $k = p$, a comparison between the values of the variables belonging to prospect $P$ and those belonging to prospect $Q$ will, in view of Theorem III, establish the desired result for the case where $0 \leqq k \leqq 1$ ($p$, being a probability, cannot be $<0$ or $>1$).

Next let us consider the case where $k < 0$. Let us choose prospect $R$ so that prospect $O$ becomes equivalent to the mixed prospect of obtaining either prospect $R$ (with the probability $p$) or prospect $P$ (with the probability $1 - p$). A little calculation will show that in this case prospect $R$ will represent $U_1 = (1 - 1/p) \cdot u_1$; $U_2 = (1 - 1/p) \cdot u_2$; . . . ; $U_n = (1 - 1/p) \cdot u_n$ for the different individuals and $W = (1 - 1/p) \cdot w$ for society. If we now write $k = 1 - 1/p$, a comparison between the variables belonging to $R$ and those belonging to $P$ will establish the desired result for the case $k < 0$ (by an appropriate choice of the probability $p$, we can make $k$ equal to any negative number).

Finally, the case where $k > 1$ can be taken care of by finding a prospect $S$ such that prospect $P$ becomes equivalent to the mixed prospect of obtaining either $S$ (with a probability $p$) or $O$ (with a probability $1 - p$). Then this prospect $S$ will be connected with the values $U_1 = 1/p \cdot u_1$; $U_2 = 1/p \cdot u_2$; . . . ; $U_n = 1/p \cdot u_n$ and $W = 1/p \cdot w$. If we now write $k = 1/p$ we obtain the desired result for the case where $k > 1$ (by an appropriate choice of $p$ we can make $k$ equal to any number $>1$).

*Theorem V.*—$W$ is a weighted sum of the individual utilities, of the form

$$W = \Sigma a_i \cdot U_i ,$$

where $a_i$ stands for the value that $W$ takes when $U_i = 1$ and $U_j = 0$ for all $j \neq i$.

*Proof.*—Let $S_i$ be a prospect representing the utility $U_i$ to the $i$th individual and the utility zero to all other individuals. Then, according to Theorem IV, for $S_i$ we have $W = a_i \cdot U_i$.

Let $T$ be the mixed prospect of obtaining either $S_1$ or $S_2$ or . . . $S_n$, each with probability $1/n$. Then $T$ will represent the individual utilities $U_1/n$, $U_2/n$, . . . , $U_n/n$ and the social welfare

$$W = \frac{1}{n} \cdot \Sigma a_i \cdot U_i .$$

In view of Theorem IV, this directly implies that if the individual utility functions take the values $U_1$, $U_2$, . . . , $U_n$, respectively, the social welfare function has the value

$$W = \Sigma a_i \cdot U_i ,$$

as desired.[12]

## IV

In the pre-Pareto conceptual framework, the distinction between social welfare and individual utilities was free of ambiguity. Individual utilities were assumed to be directly given by introspection, and social welfare was simply their sum. In the modern approach, however, the distinction is far less clear. On the one hand, our social welfare concept has

[12] If we want a formal guaranty that no individual's utility can be given a negative weight in the social welfare function, we must add one more postulate (for instance, Postulate D of Sec. II).

come logically nearer to an individual utility concept. Social welfare is no longer regarded as an objective quantity, the same for all, by necessity. Rather, each individual is supposed to have a social welfare function of his own, expressing his own individual values—in the same way as each individual has a utility function of his own, expressing his own individual taste. On the other hand, our individual utility concept has come logically nearer to a social welfare concept. Owing to a greater awareness of the importance of external economies and diseconomies of consumption in our society, each individual's utility function is now regarded as dependent not only on this particular individual's economic (and noneconomic) conditions but also on the economic (and other) conditions of all other individuals in the community—in the same way as a social welfare function is dependent on the personal conditions of all individuals.

At the same time, we cannot allow the distinction between an individual's social welfare function and his utility function to be blurred if we want (as most of us do, I think) to uphold the principle that a social welfare function ought to be based not on the utility function (subjective preferences) of *one* particular individual only (namely, the individual whose value judgments are expressed in this welfare function), but rather on the utility functions (subjective preferences) of *all* individuals, representing a kind of "fair compromise" among them.[13] Even if both an individual's social welfare function and his utility function in a sense express his own individual preferences, they must express preferences of different sorts: the former must express what this individual prefers (or, rather, would prefer) on the basis of impersonal social considerations alone, and the latter must express what he actually prefers, whether on the basis of his personal interests or on any other basis. The former may be called his "ethical" preferences, the latter his "subjective" preferences. Only his "subjective" preferences (which define his utility function) will express his preferences in the full sense of the word as they actually are, showing an egoistic attitude in the case of an egoist and an altruistic attitude in the case of an altruist. His "ethical" preferences (which define his social welfare function) will, on the other hand, express what can in only a qualified sense be called his "preferences": they will, by definition, express what he prefers only in those possibly rare moments when he forces a special impartial and impersonal attitude upon himself.[14]

In effect, the ethical postulates pro-

[13] This principle is essentially identical with Professor Arrow's "nondictatorship" postulate in his *Social Choice and Individual Values* (New York, 1951), p. 30 (see also n. 12).

[14] Mr. Little's objection to Arrow's nondictatorship postulate (see Little's review article in the *Journal of Political Economy*, LX [October, 1952], esp. 426–31) loses its force, once the distinction between "ethical" and "subjective" preferences is noted. It does, then, make sense that an individual should morally *disapprove* (in terms of his "ethical" preferences) of an unequal income distribution which benefits him financially, and should still *prefer* it (in terms of his "subjective" preferences) to a more egalitarian one or should even *fight* for it—behavior morally regrettable but certainly not logically inconceivable.

Arrow's distinction between an individual's "tastes" (which order social situations only according to their effects on his own consumption)and his "values" (which take account also of external economies and diseconomies of consumption and of ethical considerations, in ordering social situations) does not meet the difficulty, since it does not explain how an individual can without inconsistency accept a social welfare function conflicting with his own "values." This can be understood only if his social welfare functions represents preferences of another sort than his "values" do. (Of course, in my terminology Arrow's "values" fall in the class of "subjective" preferences and not in the class of "ethical" preferences, as is easily seen from the way in which he defines them.)

posed in Sections II and III—namely, Postulates D, E, and *c*—can be regarded as simply an implicit definition of what sort of "impartial" or "impersonal" attitude is required to underlie "ethical" preferences: these postulates essentially serve to exclude nonethical subjective preferences from social welfare functions. But this aim may also be secured more directly by explicitly defining the impartial and impersonal attitude demanded.

I have argued elsewhere[15] that an individual's preferences satisfy this requirement of impersonality if they indicate what social situation he would choose if he did not know what his personal position would be in the new situation choosen (and in any of its alternatives) but rather had an equal *chance* of obtaining any of the social positions[16] existing in this situation, from the highest down to the lowest. Of course, it is immaterial whether this individual does not in fact know how his choice would affect his personal interests or merely disregards this knowledge for a moment when he is making his choice. As I have tried to show,[17] in either case an impersonal choice (preference) of this kind can in a technical sense be regarded as a choice between "uncertain" prospects.

This implies, however, without any additional ethical postulates that an individual's impersonal preferences, if they are rational, must satisfy Marschak's axioms and consequently must define a cardinal social welfare function equal to the arithmetical mean[18] of the utilities of all individuals in the society (since the arithmetical mean of all individual utilities gives the actuarial value of his uncertain prospect, defined by an equal probability of being put in the place of any individual in the situation chosen).

More exactly, if the former individual has any objective criterion for comparing his fellows' utilities with one another and with his own (see Sec. V), his social welfare function will represent the unweighted mean of these utilities, while in the absence of such an objective criterion it will, in general, represent their weighted mean, with arbitrary weights depending only on his personal value judgments. In the former case social welfare will in a sense be an objective quantity, whereas in the latter case it will contain an important subjective element; but even in this latter case it will be something very different from the utility function of the individual concerned.[19]

## V

There is no doubt about the fact that people do make, or at least attempt to make, interpersonal comparisons of utility, both in the sense of comparing different persons' total satisfaction and in the

[15] See my "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking," *Journal of Political Economy*, LXI (October, 1953), 434–35.

[16] Or, rather, if he had an equal chance of being "put in the place of" any individual member of the society, with regard not only to his objective social (and economic) conditions, but also to his subjective attitudes and tastes. In other words, he ought to judge the utility of another individual's position not in terms of his own attitudes and tastes but rather in terms of the attitudes and tastes of the individual actually holding this position.

[17] *Op. cit.*

[18] Obviously, the (unweighted or weighted) *mean* of the individual utilities defines the same social welfare function as their *sum* (weighted by the same relative weights), except for an irrelevant proportionality constant.

[19] The concept of ethical preferences used in this section implies, of course, an ethical theory different from the now prevalent subjective attitude theory, since it makes a person's ethical judgments the expression, not of his subjective attitudes in general, but rather of certain special unbiased impersonal attitudes only. I shall set out the philosophic case for this ethical theory in a forthcoming publication. (For a similar view, see J. N. Findlay, "The Justification of Attitudes," *Mind*, N.S., LXIII [April, 1954], 145–61.)

sense of comparing increments or decrements in different persons' satisfaction.[20] The problem is only what logical basis, if any, there is for such comparisons.

In general, we have two indicators of the utility that *other* people attach to different situations: their preferences as revealed by their actual choices, and their (verbal or nonverbal) expressions of satisfaction or dissatisfaction in each situation. But while the use of these indicators for comparing the utilities that a *given* person ascribes to different situations is relatively free of difficulty, their use for comparing the utility that *different* persons ascribe to each situation entails a special problem. In actual fact, this problem has two rather different aspects, one purely metaphysical and one psychological, which have not, however, always been sufficiently kept apart.

The *metaphysical* problem would be present even if we tried to compare the utilities enjoyed by different persons with identical preferences and with identical expressive reactions to any situation. Even in this case, it would not be inconceivable that such persons should have different susceptibilities to satisfaction and should attach different utilities to identical situations, for, in principle, identical preferences may well correspond to different absolute levels of utility (as long as the ordinal properties of all persons' utility functions are the same[21]), and identical expressive reactions may well indicate different mental states with different people. At the same time, under these conditions this logical possibility of different susceptibilities to satisfaction would hardly be more than a metaphysical curiosity. If two objects or human beings show similar behavior in *all* their relevant aspects open to observation, the assumption of some unobservable hidden difference between them must be regarded as a completely gratuitous hypothesis and one contrary to sound scientific method.[22] (This principle may be called the "principle of unwarranted differentiation." In the last analysis, it is on the basis of this principle that we ascribe mental states to other human beings at all: the denial of this principle would at once lead us to solipsism.[23] Thus in the case of persons with similar preferences and expressive reactions we are fully entitled to assume that they derive the same utilities from similar situations.

In the real world, of course, different people's preferences and their expressive reactions to similar situations may be rather different, and this does represent a very real difficulty in comparing the utilities enjoyed by different people—a difficulty in addition to the metaphysical difficulty just discussed and independent of it. I shall refer to it as the *psychological* difficulty, since it is essentially a question of how psychological differences between people in the widest sense (for example,

[20] See I. M. D. Little, *A Critique of Welfare Economics* (Oxford, 1950), chap. iv. I have nothing to add to Little's conclusion on the *possibility* of interpersonal comparisons of utility. I only want to supplement his argument by an analysis of the *logical basis* of such comparisons. I shall deal with the problem of comparisions between total utilities only, neglecting the problem of comparisons between differences in utility, since the social welfare functions discussed in the previous sections contain only total utilities of individuals.

[21] Even identical preferences among uncertain prospects (satisfying the Marschak axioms) are compatible with different absolute levels of utility, since they do not uniquely determine the zero points and the scales of the corresponding cardinal utility functions.

[22] By making a somewhat free use of Professor Carnap's distinction, we may say that the assumption of different susceptibilities of satisfaction in this case, even though it would not be against the canons of *deductive* logic, would most definitely be against the canons of *inductive* logic.

[23] See Little, *A Critique of Welfare Economics*, pp. 56–57.

differences in consumption habits, cultural background, social status, and sex and other biological conditions, as well as purely psychological differences, inborn or acquired) affect the satisfaction that people derive from each situation. The problem in general takes the following form. If one individual prefers situation $X$ to situation $Y$, while another prefers $Y$ to $X$, is this so because the former individual attaches a *higher* utility to situation $X$, or because he attaches a *lower* utility to situation $Y$, than does the latter—or is this perhaps the result of both these factors at the same time? And, again, if in a given situation one individual gives more forcible signs of satisfaction or dissatisfaction than another, is this so because the former feels more intense satisfaction or dissatisfaction, or only because he is inclined to give stronger expression to his feelings?

This psychological difficulty is accessible to direct empirical solution to the extent to which these psychological differences between people are capable of change, and it is therefore possible for some individuals to make direct comparisons between the satisfactions open to one human type and those open to another.[24] Of course, many psychological variables are not capable of change or are capable of change only in some directions but not in others. For instance, a number of inborn mental or biological characteristics cannot be changed at all, and, though the cultural patterns and attitudes of an individual born and educated in one social group can be considerably changed by transplanting him to another, usually they cannot be completely assimilated to the cultural patterns and attitudes of the second group. Thus it may easily happen that, if we want to compare the satisfactions of two different classes of human beings, we cannot find any individual whose personal experiences would cover the satisfactions of both these classes.

Interpersonal comparisons of utility made in everyday life seem, however, to be based on a different principle (which is, of course, seldom formulated explicitly). If two individuals have opposite preferences between two situations, we usually try to find out the psychological differences responsible for this disagreement and, on the basis of our general knowledge of human psychology, try to judge to what extent these psychological differences are likely to increase or decrease their satisfaction derived from each situation. For example, if one individual is ready at a given wage rate to supply more labor than another, we tend in general to explain this mainly by his having a lower disutility for labor if his physique is much more robust than that of the other individual and if there is no ascertainable difference between the two individuals' economic needs; we tend to explain it mainly by his having a higher utility for income (consumption goods) if the two individuals' physiques are similar and if the former evidently has much greater economic needs (for example, a larger family to support).

Undoubtedly, both these methods of tackling what we have called the "psychological difficulty" are subject to rather large margins of error.[25] In general, the greater the psychological, biological, cultural, and social differences between two

[24] On the reliability of comparisons between the utility of different situations before a change in one's "taste" (taken in the broadest sense) and after it, see the first two sections of my "Welfare Economics of Variable Tastes," *Review of Economic Studies*, XXI, (1953–54), 204–8.

[25] Though perhaps it would not be too difficult to reduce these margins quite considerably (for example, by using appropriate statistical techniques), should there be a need for more precise results.

people, the greater the margin of error attached to comparisons between their utility.

Particular uncertainty is connected with the second method, since it depends on our general knowledge of psychological laws, which is still in a largely unsatisfactory state.[26] What is more, all our knowledge about the psychological laws of satisfaction is ultimately derived from observing how changes in different (psychological and other) variables affect the satisfactions an individual obtains from various situations. We therefore have no direct empirical evidence on how people's satisfactions are affected by the variables that, for any particular individual, are *not* capable of change. Thus we can, in general, judge the influence of these "unchangeable" variables only on the basis of the correlations found between these and the "changeable" variables, whose influence we can observe directly. For instance, let us take sex as an example of "unchangeable" variables (disregarding the few instances of sex change) and abstractive ability as an example of "changeable" variables. We tend to assume that the average man finds greater satisfaction than the average woman does in solving mathematical puzzles *because*, allegedly, men in general have greater abstractive ability than women. But this reasoning depends on the implicit assumption that differences in the "unchangeable" variables, if unaccompanied by differences in the "changeable" variables, are in themselves immaterial. For example, we must assume that men and women equal in abstractive ability (and the other relevant characteristics) would tend to find the same satisfaction in working on mathematical problems.

Of course, the assumption that the "unchangeable" variables in themselves have no influence is *ex hypothesi* not open to direct empirical check. It can be justified only by the a priori principle that, when one variable is alleged to have a certain influence on another, the burden of proof lies on those who claim the existence of such an influence.[27] Thus the second method of interpersonal utility comparison rests in an important sense on empirical evidence more indirect[28] than that underlying the first method. On the other hand, the second method has the advantage of also being applicable in those cases where no one individual can possibly have wide enough personal experience to make direct utility comparisons in terms of the first method.

In any case, it should now be sufficiently clear that interpersonal compari-

[26] Going back to our example, for instance, the disutility of labor and the utility of income are unlikely to be actually independent variables (as I have tacitly assumed), though it may not always be clear in which way their mutual influence actually goes. In any case, income is enjoyed in a different way, depending on the ease with which it has been earned, and labor is put up with in a different spirit, depending on the strength of one's need for additional income.

[27] This principle may be called the "principle of unwarranted correlation" and is again a principle of inductive logic, closely related to the principle of unwarranted differentiation referred to earlier.

[28] There is also another reason for which conclusions dependent on the principle of unwarranted correlation have somewhat less cogency than conclusions dependent only on the principle of unwarranted differentiation. The former principle refers to the case where two individuals differ in a certain variable $X$ (in our example, in sex) but where there is no special evidence that they differ also in a certain other variable $Y$ (in susceptibility to satisfaction). The latter principle, on the other hand, refers to the case where there is no ascertainable difference at all between the two individuals in any observable variable whatever, not even in $X$ (in sex). Now, though the assumption that these two individuals differ in $Y$ (in susceptibility to satisfaction) would be a gratuitous hypothesis in either case, obviously it would be a less unnatural hypothesis in the first case (where there is some observed difference between the two individuals) than in the second case (where there is none).

sons of utility are not value judgments based on some ethical or political postulates but rather are factual propositions based on certain principles of inductive logic.

At the same time, Professor Robbins[29] is clearly right when he maintains that propositions which purport to be interpersonal comparisons of utility often contain a purely *conventional* element based on ethical or political value judgments. For instance, the assumption that different individuals have the same susceptibility to satisfaction often expresses only the egalitarian value judgment that all individuals should be treated equally rather than a belief in a factual psychological equality between them. Or, again, different people's total satisfaction is often compared on the tacit understanding that the gratification of wants regarded as "immoral" in terms of a certain ethical standard shall not count. But in order to avoid confusion, such propositions based on ethical or political restrictive postulates must be clearly distinguished from interpersonal comparisons of utility without a conventional element of this kind.

It must also be admitted that the use of conventional postulates based on personal value judgments may sometimes be due not to our free choice but rather to our lack of the factual information needed to give our interpersonal utility comparisons a more objective basis. In effect, if we do not know anything about the relative urgency of different persons' economic needs and still have to make a decision, we can hardly avoid acting on the basis of personal guesses more or less dependent on our own value judgments.

On the other hand, if the information needed is available, individualistic ethics consistently requires the use, in the social welfare function, of individual utilities not subjected to restrictive postulates. The imposition of restrictive ethical or political conventions on the individual utility functions would necessarily qualify our individualism, since it would decrease the dependence of our social welfare function on the actual preferences and actual susceptibilities to satisfaction, of the individual members of the society, putting in its place a dependence on our own ethical or political value judgments (see nn. 5 and 6).

To sum up, the more complete our factual information and the more completely individualistic our ethics, the more the different individuals' social welfare functions will converge toward the same objective quantity, namely, the unweighted sum (or rather the unweighted arithmetic mean) of all individual utilities. This follows both from (either of two alternative sets of) ethical postulates based on commonly accepted individualistic ethical value judgments and from the mere logical analysis of the concept of a social welfare function. The latter interpretation also removes certain difficulties connected with the concept of a social welfare function, which have been brought out by Little's criticism of certain of Arrow's conclusions.

Of course, the practical need for reaching decisions on public policy will require us to formulate social welfare functions—explicitly or implicitly—even if we lack the factual information needed for placing interpersonal comparisons of utility on an objective basis. But even in this case, granting the proposed ethical postulates (or the proposed interpretation of

[29] See L. Robbins, "Robertson on Utility and Scope," *Economica*, N.S., XX (1953), 99–111, esp. 109; see also his *An Essay on the Nature and Significance of Economic Science* (2d ed.; London, 1948), chap. vi; and his "Interpersonal Comparisons of Utility," *Economic Journal*, XLIII (December, 1938), 635–41.

the concept of a social welfare function), our social welfare function must take the form of a weighted sum (weighted mean) of all individual utility functions, with more or less arbitrary weights chosen according to our own value judgments.

There is here an interesting analogy with the theory of statistical decisions (and, in general, the theory of choosing among alternative hypotheses). In the same way as in the latter, it has been shown[30] that a rational man (whose choices satisfy certain simple postulates of rationality) must act *as if* he ascribed numerical subjective probabilities to all alternative hypotheses, even if his factual information is insufficient to do this on an objective basis—so in welfare economics we have also found that a rational man (whose choices satisfy certain simple postulates of rationality and impartiality) must likewise act *as if* he made quantitative interpersonal comparisons of utility, even if his factual information is insufficient to do this on an objective basis.

Thus if we accept individualistic ethics and set public policy the task of satisfying the preferences of the individual members of the society (deciding between conflicting preferences of different individuals according to certain standards of impartial equity), our social welfare function will always tend to take the form of a sum (or mean) of individual utilities; but whether the weights given to these individual utilities have an objective basis or not will depend wholly on the extent of our factual (psychological) information.

[30] See Marschak's discussion of what he calls "Ramsey's norm," in his paper on "Probability in the Social Sciences," in *Mathematical Thinking in the Social Sciences*, ed. P. F. Lazarsfeld (Glencoe, Ill., 1954), Sec. I, esp. pp. 179–87; also reprinted as No. 82 of "Cowles Commission Papers" (N.S.).

For a survey of earlier literature see K. J. Arrow, Alternative Approaches to the Theory of Choice in Risk-taking Situations," *Econometrica*, XIX (October, 1951), 404–37, esp. 431–32, and the references there quoted.