October 1998 draft of a contribution to Salvador Barberà, Peter J. Hammond and Christian Seidl (eds.) *Handbook of Utility Theory, Vol. 2* (in preparation for Kluwer Academic Publishers).

22 INTERPERSONALLY COMPARABLE UTILITY

Peter J. Hammond

Abstract: This chapter supplements the earlier reviews in Hammond (1991a) and Suzumura (1996) by concentrating on four issues. The first is that in welfare economics interpersonal comparisons are only needed to go beyond Pareto efficiency or Pareto improvements. The second concerns the need for interpersonal comparisons in social choice theory, to escape Arrow's impossibility theorem. The third issue is how to revise Arrow's independence of irrelevant alternatives condition so that interpersonal comparisons can be accommodated. Finally, and most important, the chapter presents a form of utilitarianism in which interpersonal comparisons can be interpreted as ethical preferences for different personal characteristics.

22.1 INTRODUCTION AND OUTLINE

Over many years, interpersonal comparisons of utility have had a significant role to play in economics. Utility began as a concept which Frances Hutcheson, Cesare Beccaria, Jeremy Bentham, John Stuart Mill, and Henry Sidgwick sought to use as a basis for a general ethical theory that is simple yet profound. Classical utilitarian theory relied on interpersonal comparisons because it required that there be a common unit in which one can measure each person's pleasure or happiness before adding to arrive at a measure of total happiness. According to Bentham, one should then proceed to subtract each person's pain or misery, also measured in the same common unit, in order to arrive at a measure of total utility.

For economists, the notion of utility later became much more sophisticated. Consumer or demand theory had been based on a notion of utility, and the requirement that the marginal utilities of spending wealth on different commodities should be equalized. Following the ideas pioneered by Hicks, Allen, and Samuelson, a revised demand theory was built on the foundation of a binary preference relation, perhaps revealed by the consumer's own behaviour. In positive economics this meant that utility became an ordinal rather than a cardinal concept. And also that one then lacked a common unit with which one could measure and compare different individuals' utilities. Robbins (1938) then felt justified in making his widely cited claim that interpersonal comparisons of utility are unscientific.

Welfare economic theory, however, and the related discipline of social choice theory, have retained their links to ethical theory. In fact, without ethical content, both theories would become empty shells, as Little (1957, pp. 79–80) for one has pointed out. For this reason, interpersonal comparisons continue to play a significant role in both these theories.

This chapter will not attempt to survey the large literature on interpersonal comparisons. One reason for this is that there is no reason to repeat what has already appeared in Hammond (1991a) or Suzumura (1996). Instead, I would like to focus attention on four specific questions which arise in connection with interpersonal comparisons.

Of these four questions, the first is why economists need these particular value judgements that Robbins deemed unscientific. In fact, what would remain of welfare economics and of social choice theory if one refused to make any interpersonal comparisons at all?

The second question relates to the first, because it asks what can be done with interpersonal comparisons. Section 2 begins by arguing that much of the theory of welfare economics avoids such comparisons. It also points out how, in welfare economics, they can be used to determine what weights to place on different individuals' gains and losses. In social choice theory, however, as discussed in Section 3, Kenneth Arrow's famous "dictatorship" theorem concerns the impossibility of making reasonable social choices without interpersonal comparisons. Section 3 also illustrates how interpersonal comparisons allow many possible escapes from Arrow's theorem, depending upon whether one can make comparisons of utility levels or of utility units.

As Arrow himself pointed out, interpersonal comparisons are effectively excluded by the independence of irrelevant alternatives (IIA) condition which he imposed on any social welfare function. The third question is whether IIA can be modified in a way that can be satisfied when social choice does depend on interpersonal comparisons. Section 4 summarizes some of the answers that were provided more formally in Hammond (1991b).

The fourth and last question may well strike the reader as being the most important. To the extent that interpersonal comparisons are unavoidable, how can they be made, and what meaning can they be given? Before answering this question, Section 5 first motivates a form of utilitarianism in which interpersonal comparisons play a crucial role. This motivation is based on the expected utility model described in Chapter 5. Then, following Hammond (1991a), Section 6 argues that the only really coherent answer to the fourth question is that interpersonal comparisons have to be seen as revealed by the choice of persons — or better, by the ethical choice of population size and of the distribution of personal characteristics within the population.

Section 7 contains some concluding remarks.

22.2 WELFARE ECONOMICS

22.2.1 Pareto Efficiency

Welfare economics is an enormous subject, touching every branch of the discipline. Here, I shall not attempt to summarize more than a few of the most crucial results, while explaining which of them rely on interpersonal comparisons.

Modern welfare economics, like modern social choice theory, begins with an article by Kenneth Arrow. In 1951, he presented the two fundamental theorems of the subject. Of course, there were antecedents in well known classic works by Enrico Barone, Vilfredo Pareto, Oskar Lange, Abba Lerner, Paul Samuelson, Maurice Allais, and others. But these earlier authors limited themselves to incomplete and local results based on the differential calculus. Whereas Arrow's analysis was global, exploiting the notion of convexity and the separating hyperplane theorem.

According to the first of these two theorems, each Walrasian equilibrium allocation is Pareto efficient, at least if consumers' preferences are locally non-satiated. According to the second theorem, any Pareto efficient allocation not on the boundary of the attainable set is a Walrasian equilibrium, provided that preferences satisfy appropriate convexity and continuity assumptions. For present purposes, it is enough to recall that these two theorems relate the set of Pareto efficient allocations to the set of Walrasian equilibria with lump-sum transfers. To describe these two sets, there is evidently no need for interpersonal comparisons. Such comparisons serve only to choose among the elements of each set, which is really a social choice problem anyway.

In each of the world's contemporary national economies, there remain many imperfections which prevent the Pareto efficient allocation of resources. For example, there are public goods, external effects, distortionary taxes of the kind needed to finance public goods and to institute measures that alleviate poverty, etc. These invevitable imperfections limit the relevance to practical economics of the two fundamental efficiency theorems. In fact, these theorems are too idealistic because they characterize allocations which are perfect — or at least perfectly efficient.

22.2.2 Pareto Improvements

For this reason, the results concerning the gains from free trade and free exchange might appear to be much more useful. Most economists think of these as belonging to the field of international economics. But there is a general third theorem of welfare economics concerning not only the gains from international trade, but also the gains from market integration, from profit maximization by a firm, from free competition between firms, from replacing a distortionary tax with lump-sum taxes raising the same revenue, and from technical progress that enhances the efficiency of production. All are really instances of one general theorem, as pointed out in Hammond and Sempere (1995).

The third theorem shows that, if a new market is opened, or if existing markets are made more efficient, there is a potential Pareto improvement in the sense described originally by Barone (1908), though more commonly ascribed to Kaldor (1939) and Hicks — see the articles the latter published during the years 1939–1946 that are reprinted in Part II of Hicks (1981).[1] That is, even if some people who initially lose because of adverse relative price movements caused by the new markets or the increase in efficiency, they can always be compensated so that everybody gains in the end. Thus, an actual Pareto improvement becomes possible. But in this connection, one is always looking for a Pareto improvement, in which everybody gains and nobody loses. In this way, the need for interpersonal comparisons has still been avoided.

22.2.3 Private Information

These three classical theorems all rely on the assumption that lump-sum redistribution is possible without limit. Yet in reality we lack the information needed to arrange such redistribution in a suitable manner. As Vickrey (1945) and Mirrlees (1971) understood very well, it is impossible to have ideal lump-sum taxes based on workers' inherent abilities. These abilities cannot be observed. Instead, one sees only the incomes which workers can earn by deploying their abilities. So, instead of an ideal tax on inherent ability, one is forced to substitute a distortionary tax on income.

A worker's inherent ability is merely one kind of private information. There are many other kinds — for example, a consumer's preferences and endowments, or a producer's true technology and associated cost function. Each piece of private information creates its own "incentive" constraint. Roger Guesnerie (1995) and I have independently analysed economies with very many agents who possess some private information. We showed how lump-sum transfers must generally be independent of private information. And how incentives are preserved only by what public finance economists generally regard as "distortionary" taxes that depend on individual transactions, as well as on the distribution of privately known personal characteristics in the population. Then the two theorems linking Pareto efficient allocations to perfect markets lose virtually all their relevance. The usual Pareto frontier becomes replaced by a "second-best" Pareto frontier, which recognizes incentive constraints as well as the usual requirements of physical feasibility. Further discussion and references can be found in Hammond (1990).

Guesnerie and I have also considered what would remain possible if individuals could manipulate not only by concealing or misrepresenting their private information, but if they could also combine in small groups with other individuals in order to exchange goods on the side, in a hidden economy beyond the control of the fiscal authorities. These extra manipulations imply that one can only have linear pricing for each good whose transactions cannot be observed

[1]For an assessment of Barone's earlier contribution, see Chipman and Moore (1978).

by the authorities. In this way, extra constraints arise and one is forced down to a "third-best" Pareto frontier. However, in the absence of externalities or public goods, all three frontiers contain whatever allocations would result from a policy of total laisser faire, without any attempts to redistribute wealth in order to move around the first-best frontier. See also Blackorby and Donaldson (1988), as well as Hammond (1997).

We still lack simple or intuitive mathematical characterizations of the constrained Pareto frontiers. There are no fundamental theorems like the two proved by Arrow. Nevertheless, it is evident that any such constrained Pareto frontier can be described without any need to make interpersonal comparisons. Both the Pareto criterion and the relevant incentive constraints can be described by making use of information only about individual preference orderings. Only the ethical social choice of a point or subset of the frontier requires interpersonal comparisons.

The third theorem is much less modified than the first two when one takes account of private information and the resulting incentive constraints. As shown in Hammond and Sempere (1995), Pareto improvements can still be ensured if the tax on each commodity is varied in a way that freezes the after-tax prices (and wages) faced by all consumers; this still allows prices faced by producers to vary in order to clear markets. In addition, after-tax dividends paid by firms to consumers should be frozen. But even this theorem concerns potential Pareto gains, and so still avoids any need for interpersonal comparisons.

22.2.4 Measures of Individual Gain and Loss

So far, I have argued that the major theorems of Paretian welfare economics do not rely on interpersonal comparisons. But these major theorems cannot be applied easily to real issues of economic policy, such as how to provide affordable medical services, or lower unemployment, or reduce poverty, or provide more adequate housing, while avoiding excessive taxes or risks of high inflation. According to the familiar old proverb, "It is an ill wind that blows nobody any good." This applies even in economics. For example, a deep recession brings a lot of business for accountants and others who are responsible for winding up bankrupt firms. The reverse is: "It is a good wind that blows nobody any ill" — in other words, it is difficult to find a true Pareto improvement. In practice, real economic policy choices make some people better off, others worse off. The choice between policies then does require interpersonal comparisons.

Still, a great deal can be learned about the effects of economic policy choices even without interpersonal comparisons. This is because any economic policy reform or decision can be regarded as having effects on each separate individual. So one should be able to calculate or estimate each individual's net benefit from any policy decision. In principle, it is usually possible even to construct a money metric measure of net benefit. This is done by finding what increase or decrease in wealth would have exactly the same effect on the individual's welfare as the policy decision being contemplated, provided that private good prices and public good quantities remained fixed at their *status quo* values. It is *not* done, except possibly very inaccurately, by calculating consumer surplus based on the area under an uncompensated Marshallian demand curve. For details, see Hammond (1994) or Becht (1995), amongst others. The measure that results is closely related to Hicks' equivalent variation. It tells us how much each particular individual gains or loses from a policy change, which is immensely valuable information. Yet the construction of different individuals' measures of net benefit does not require any interpersonal comparisons.

At this stage, many economists of the Chicago school, following Harberger (1971), succumb to the temptation of just adding different individuals' monetary measures. "A dollar is a dollar," they might say, regardless of how deserving is the recipient. Implicitly, they attach equal value to the extra dollar a rich man will spend on a slightly better bottle of wine and to the dollar a poor woman needs to spend on life-saving medicine for her child. Of course, any such judgement is a value judgement, even an interpersonal comparison, which lacks scientific foundation. Thus, the "surplus economists" who just add monetary measures, often of consumer surplus rather than individual welfare, make their own value judgements and their own interpersonal comparisons. Moreover, their comparisons not only lack scientific content, but most people also find them totally unacceptable from an ethical point of view. Surely it is better to avoid interpersonal comparisons altogether rather than make them in such a biased way.

Many economists, including even Harberger (1978) himself (though very reluctantly), have suggested multiplying each individual's monetary measure of gain by a suitable "welfare weight" in order to arrive at a suitable welfare-weighted total measure of benefit for society as a whole. The ratios of these welfare weights evidently represent the (constant) marginal rates of substitution between the wealth levels of the corresponding individuals in a social welfare function. These ratios reflect interpersonal comparisons between the supposed ethical worth of marginal monetary gains occurring to different individuals, even if one follows the Chicago school in equating all the welfare weights to 1. Such welfare-weighted sums can be used to identify directions in which small enough policy changes are deemed beneficial for society as a whole.

Many economists have advocated considering welfare-weighted sums even for changes that are not small. Yet policies having a significant impact on the distribution of real wealth are also likely to change the marginal rates of substitution which lie behind the different relative welfare weights. So one needs to be more careful. This is an issue I have discussed at greater length in Hammond (1994). Of course, interpersonal comparisons will play an inevitable role in determining any suitable measure of social welfare.

To summarize this section, as long as welfare economics concerns itself only with (constrained) Pareto efficient allocations, or with (potential) Pareto improvement, there is no need for interpersonal comparisons. Even without such comparisons, one can still describe the Pareto frontier, with or without constraints of various kinds, and also look for Pareto improvements. Moreover, it is possible to construct measures of net monetary gain for each separate individual. As discussed in Hammond (1990), such individual measures already provide very useful information; much more is provided by the joint statistical distribution of these measures and of other relevant personal characteristics, such as education, family circumstances, age, or family background. In principle, this joint distribution can and should be estimated by the best possible econometric techniques. It does not depend on any interpersonal comparisons. Its interpretation depends on only one ethical value judgement — namely, the judgement that information about different individuals' behaviour can determine how those individuals' measures of benefit should be estimated. That is a serious value judgement, but one which is indispensable for the neo-classical theory of welfare economics. Without this judgement, one would have to consider issues such as how much paternalism is desirable.

In the end, much welfare analysis is possible without interpersonal comparisons. They play a role only in choosing among different Pareto efficient allocations. Or more generally, in deciding whether to institute a reform which benefits one set of individuals but harms another. Or when one wants to construct a measure of social welfare. These considerations lead us to the theory of social choice, to which I now turn.

22.3 SOCIAL CHOICE

22.3.1 Arrow's Dictatorship Theorem

Like modern welfare economics, modern social choice theory starts with a 1951 publication by Kenneth Arrow — in this case, the first edition of *Social Choice and Individual Values*, based on his Ph.D. thesis submitted to Columbia University. This and the earlier article (Arrow, 1950) presented his famous "impossibility" theorem. Though this result is well known, I will cover it briefly in order to introduce some terminology which will be useful later, and also to offer a slightly different interpretation.

Because Arrow deliberately sought to avoid interpersonal comparisons, he defined a social welfare function on a domain of individual preference profiles. Let X be the universal set of social states defined so that society is required to have one social state from some feasible subset of X. Let N be a finite set of n individuals. Let $\mathcal{R}(X)$ denote the set of all logically possible (complete and transitive) preference orderings on X. Let $\mathcal{R}^N(X)$ denote the Cartesian product of n different copies of the set $\mathcal{R}(X)$. Each $R^N = \langle R_i \rangle_{i \in N} \in \mathcal{R}^N(X)$ is a *preference profile* consisting of one preference ordering $R_i \in \mathcal{R}(X)$ for each individual $i \in N$. Then an *Arrow social welfare function* (ASWF) is a mapping $f : D \to \mathcal{R}(X)$ defined on a domain $D \subset \mathcal{R}^N(X)$ of preference profiles, whose value is some social welfare preference ordering on X.

Arrow imposed four conditions on such ASWFs. Of these, the first three are:

(U) *Unrestricted domain*: The domain D on which f is defined is equal to the whole Cartesian product set $\mathcal{R}^N(X)$.

(P) *Weak Pareto*: Let P_i ($i \in N$) and P denote the strict preference relations derived from the profile $R^N = \langle R_i \rangle_{i \in N}$ and from the corresponding social ordering $R = f(R^N)$, respectively. Then, for any pair $x, y \in X$, it must be true that $x \; P \; y$ whenever $x \; P_i \; y$ for all $i \in N$.

(I) *Independence of irrelevant personal comparisons* (usually called "independence of irrelevant alternatives"): Let A be any non-empty subset of X. Given any two preference relations $Q, Q' \in \mathcal{R}(X)$, write $Q =_A Q'$ to indicate that Q and Q' coincide on the set A — i.e., that $aQb \iff aQ'b$, for all $a, b \in A$. Let $\mathcal{R}^N$ and $\tilde{\mathcal{R}}^N$ be two preference profiles, with corresponding social orderings $R = f(R^N)$ and $\tilde{R} = f(\tilde{R}^N)$. Then it is required that $R =_A \tilde{R}$ whenever $R_i =_A \tilde{R}_i$ for all $i \in N$.

Under these three conditions, and provided that X contains at least three different social states, Arrow proved that there must be a *dictator* $d \in N$ who, given any $x, y \in X$, has the power to ensure that xPy whenever $x \; P_d \; y$.

Arrow's fourth condition was that, provided $\#N > 1$, there should be no such dictator. This explains why it is impossible to find any ASWF satisfying his four conditions. But obviously Arrow's result can be re-cast as a *dictatorship theorem*, claiming that (U), (P) and (I) together imply the existence of a dictator. In fact, the true conclusion of Arrow's theorem is that one cannot avoid interpersonal comparisons, because it is necessary to choose the dictator!

If X contains only two alternative social states, a non-dictational ASWF satisfying (U), (I) and (P) is majority rule, weighted or not. But then the choice of weights requires interpersonal comparisons — e.g., one vote per citizen over 18 years of age, but no votes for non-citizens or for children under 18.

22.3.2 Consequentialism and Ordinality

Faced with such dictatorship when $\#X \geq 3$, it is tempting to revert to an incomplete preference relation, such as that which emerges from the unanimity or *Pareto rule*, requiring that $x \; P \; y$ whenever $x \; R_i \; y$ for all $i \in N$ and also $x \; P_h \; y$ for some $h \in N$. But such incomplete preference relations create other problems. Suppose, for example, that we apply the Pareto rule when there are at least four different social states $a, b, c, d \in X$, and when $N = \{ i, j \}$ consists of exactly two individuals. Suppose too that $b \; P_i \; c \; P_i \; a \; P_i \; d$ and $a \; P_j \; d \; P_j \; b \; P_j \; c$. The left half of Figure 1 shows the *utility possibility set*, for any pair of individual utility functions that represents these preferences. Evidently, a and b are Pareto efficient, whereas c and d are inefficient. The Pareto rule therefore selects a and b.

Consider now the decision tree T shown in the right half of Figure 1. At the initial node n_0 of T, the Pareto rule allows the choice of either n_1 or n_2, anticipating a continuation from n_1 to a or from n_2 to b. Suppose that the tie between n_1 and n_2 is broken by allowing individual j to choose. Then she is likely to choose n_1, anticipating her favourite social state a. But at n_1 it is no longer possible to choose b, so c becomes Pareto efficient. Moreover, if it becomes i's turn to choose, he will choose c over a, so the final outcome will be

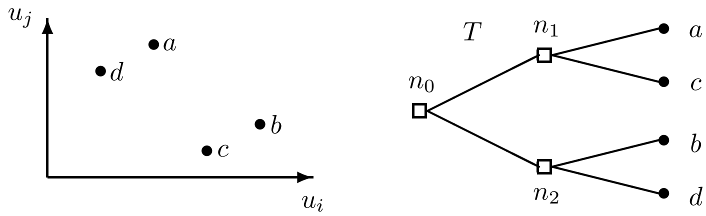

Figure 22.1 The utility possibility set and the decision tree T

c, which is Pareto inefficient. Alternatively, if society's first move is to n_2, the social state a becomes infeasible, so d becomes Pareto efficient, and would be chosen by i in preference to b. In order to avoid inefficient outcomes such as c or d, at the nodes n_1 or n_2 it is necessary to remember that these choices are inefficient, and so to be avoided.

Thus, when the Pareto rule is applied within a decision tree and each subtree, it makes the set of possible consequences of behaviour depend on the structure of the decision tree. This violates the consequentialist principle of rational decision-making (as explained in Chapter 5), requiring that acts in any decision tree or subtree be chosen for their (good) consequences, and for no other reason such as the tree structure. The implication of this principle is that decisions must maximize a (complete and transitive) preference ordering.[2] From now on, it will be assumed throughout that any social decision does indeed maximize a preference or social welfare ordering.

22.3.3 Consequentialism and Independence

As another possible escape from dictatorship, one might consider instead an ASWF that violates independence condition (I). A prominent example is the *Borda rule*, defined as follows under the supposition that X is finite. Given any profile $R^N \in \mathcal{R}^N(X)$, define for each $i \in N$ the "Borda utility function" by $B_i(x) := \#\{ y \in X \mid x \, P_i \, y \}$ for each $x \in X$. Note that $x \, R_i \, y$ iff $B_i(x) \geq B_i(y)$, so this *is* a utility function that represents R_i. Then define the *Borda count* by $B(x) := \sum_{i \in N} B_i(x)$ for all $x \in X$. Finally, define $R = f(R^N)$ as the social ordering which satisfies $x \, R \, y$ iff $B(x) \geq B(y)$. It is easy to see that this defines an ASWF satisfying conditions (U) and (P). Also, there is no dictator. Condition (I), however, is violated, as will be shown shortly. Worse, the Borda rule also violates the consequentialist requirement that decisions should have

[2]There is some relationship here to Arrow's (1963, p. 120) own concept of "path independence". But previous formalizations of this concept by Plott (1973) and others do not always imply the existence of a preference ordering. However, see Campbell (1978) for a different justification of ordinality, which is discussed further in by Bandyopadhyay (1988). Other important work using axioms describing behaviour in decision trees is due to Arthur Burks (1977). Indeed, had I known of it earlier, I would have been glad to acknowledge Burks' important contribution in my chapters 5 and 6 prepared for Volume 1 of this *Handbook*. I am most grateful to Peter Wakker for bringing his work to my attention.

consequences that are independent of the structure of the decision tree that society faces.

X	a	b	c	d	e
b_I	4	3	2	1	0
b_{II}	2	1	0	4	3
B	6	4	2	5	3

$X(n_1)$	a	d	e
b'_I	2	1	0
b'_{II}	0	2	1
B'	2	3	1

Table 22.1 The Borda counts in the tree T and in the subtree $T(n_1)$

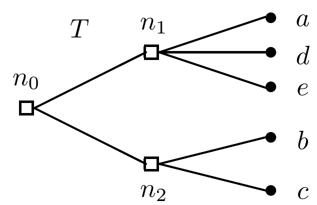

Figure 22.2 A decision tree T illustrating the Borda rule

To substantiate these claims, suppose that a, b, c, d, e are five different social states in X, and that $N = \{i, j\}$ consists of two individuals. Suppose that $a\ P_i\ b\ P_i\ c\ P_i\ d\ P_i\ e$ and that $d\ P_j\ e\ P_j\ a\ P_j\ b\ P_j\ c$. Then the Borda utility functions and Borda count are given in the left part of Table 1. Thus, a is the optimal choice from $\{a, b, c, d, e\}$. In the decision tree T illustrated in Figure 2, it is optimal to move first from n_0 to n_1.

However, suppose that the Borda rule is applied once again to the subtree $T(n_1)$ after reaching node n_1. Now b and c are no longer relevant alternatives. The new feasible set is $X(n_1)$. The Borda utility functions and Borda count become revised as indicated in the right part of Table 1. So now the optimal choice is d rather than a. The outcome of applying the Borda rule in the decision tree is d. Yet, if the decision tree only had one decision node, forcing an immediate choice of one social state from the set $\{a, b, c, d, e\}$, the result would be a. Once again, consequentialism is violated because decisions have consequences that depend on the tree structure. In Hammond (1977), it was proved that in fact consequentialism requires (I) to be satisfied.

One concludes that rational social decision-making is impossible without some form of interpersonal comparisons, even if these only serve to choose a dictator.

22.3.4 Social Welfare Functionals

So what form of rational social decision-making is possible *with* interpersonal comparisons? This question was a major preoccupation during the 1970s. Following a preliminary idea due to Suppes (1966), later Sen (1970) formulated the general concept of a *social welfare functional*, whose domain consists of profiles of utility functions rather than preference orderings. That is, the social ordering R is a function of the form $F(\langle U_i \rangle_{i \in N})$, where each U_i is a utility function mapping X to the real line $\mathbb{R}$. Moreover, these utility function profiles could be constructed with the help of suitable interpersonal comparisons. Such comparisons, or the lack of them, determine which profiles $\langle U_i' \rangle_{i \in N}$ of transformed individual utility functions should be regarded as *equivalent* to $\langle U_i \rangle_{i \in N}$, because they preserve all the relevant information contained in the utility functions. When two profiles $\langle U_i \rangle_{i \in N}$ and $\langle U_i' \rangle_{i \in N}$ are equivalent in this sense, the *invariance principle* requires that $F(\langle U_i \rangle_{i \in N}) = F(\langle U_i' \rangle_{i \in N})$.

In order to illustrate the possibilities somewhat, suppose that the two utility function profiles $\langle U_i \rangle_{i \in N}$ and $\langle U_i' \rangle_{i \in N}$ are equivalent iff there exist real constants α and β, with $\beta > 0$, such that $U_i'(x) = \alpha + \beta U_i(x)$ for all $i \in N$ and all $x \in X$. Note that such transformations preserve interpersonal comparisons of *utility levels* of the form $U_i(x) > U_j(y)$, as well as comparisons of *utility differences* of the form $U_i(x) - U_i(y) > U_j(y) - U_j(x)$. That is

$$\begin{aligned} & U_i(x) > U_j(y) \iff U_i'(x) > U_j'(y) \\ \text{and} \quad & U_i(x) - U_i(y) > U_j(y) - U_j(x) \iff U_i'(x) - U_i'(y) > U_j'(y) - U_j'(x) \end{aligned}$$

Now let $v_k(x)$ denote the kth smallest individual utility level in each social state $x \in X$ — i.e., $v_k(x)$ must be the unique real number satisfying

$$\#\{\, i \in N \mid U_i(x) < v_k(x) \,\} < k \leq \#\{\, i \in N \mid U_i(x) \leq v_k(x) \,\}$$

Then a whole class of SWFLs which are invariant under the transformations specified above are those given by

$$x \; R \; y \iff \sum\nolimits_{i=1}^{n} r_k \, v_k(x) \geq \sum\nolimits_{i=1}^{n} r_k \, v_k(y)$$

for any collection r_k ($k = 1$ to n) of real constants. These constants should be positive, or at least non-negative, if the SWFL is to satisfy the Pareto rule (P). One special case of some importance arise when $r_1 = 1$ and $r_k = 0$ for all $k > 1$. This gives the "Rawlsian" maximin rule, with

$$x \; R \; y \iff \min_i \{U_i(x)\} \geq \min_i \{U_i(y)\}$$

A second special case occurs when $r_k = 1$ for all k. This gives the "utilitarian" SWFL, with

$$x \; R \; y \iff \sum\nolimits_{i=1}^{n} U_i(x) \geq \sum\nolimits_{i=1}^{n} U_i(y)$$

But there are many other possibilities, of course. There are also different possible degrees of interpersonal comparability. For a discussion of the various possibilities, see Roberts (1980b), Blackorby, Donaldson and Weymark (1984, 1990), and d'Aspremont (1985). Certainly, explicitly introducing interpersonal comparisons allows the unpalatable conclusion of Arrow's theorem to be avoided.

22.4 INDEPENDENCE OF IRRELEVANT INTERPERSONAL COMPARISONS

Introducing interpersonal comparisons, however, produces SWFLs which violate Arrow's independence condition (I). This brings us to our third question: whether some appealing modification of condition (I) might be satisfied by a suitable class of SWFLs.

Once again, I shall merely indicate some of the possibilities by concentrating on two particularly important examples — namely, the "Rawlsian" maximin and utilitarian SWFLs that were presented in the previous section.

22.4.1 Interpersonal Comparisons of Utility Levels and Maximin

The maximin SWFL evidently requires interpersonal comparisons of utility levels. Equivalently, there should exist an *interpersonal ordering* $\tilde{R}$ on the Cartesian product space $X \times N$ whose members are pairs (x, i) consisting of a social state $x \in X$ combined with a individual $i \in N$. A preference statement such as $(x, i)\ \tilde{R}\ (y, j)$ should be interpreted as indicating that it is no worse for society to have individual i be in social state x than it is to have individual j be in social state y. The interpersonal ordering is similar in spirit to the notion of "extended sympathy" discussed by Arrow (1963) — see also Arrow (1977). Two other early discussions of such level comparisons occur in Suppes (1966) and Sen (1970).

Following Hammond (1976), define a *generalized social welfare function* (or GSWF) as a mapping $g : \mathcal{R}(X \times N) \to \mathcal{R}(X)$ that determines the social ordering on X as a function $R = g(\tilde{R})$ of the interpersonal ordering $\tilde{R}$ on $X \times N$. This is a generalization of an Arrow social welfare function insofar as any ASWF can be used to generate a particular GSWF. For, given the interpersonal ordering $\tilde{R}$ on $X \times N$, one can first define for each $i \in N$ the individual ordering $R_i(\tilde{R})$ on X by $x\ R_i(\tilde{R})\ y \iff (x, i)\ \tilde{R}\ (y, i)$ (all $x, y \in X$). Then, given the ASWF f, one can go on to define the value of the induced GSWF by $g(\tilde{R}) := f(\langle R_i(\tilde{R})\rangle_{i \in N})$.

On the other hand, not every GSWF corresponds to an ASWF. Indeed, one that clearly does not is the *maximin* GSWF defined by

$$x\ R\ y \iff \exists i \in N; \forall j \in N : (y, j)\ \tilde{R}\ (y, i) \quad \text{and} \quad (x, j)\ \tilde{R}\ (y, i)$$

In other words, i is the worst off person in state y, and in state x nobody is worse off than i is in state y. Note that, if the *interpersonal utility function* $\tilde{U}(x, i)$ represents $\tilde{R}$ on $X \times N$ in the sense that $\tilde{U}(x, i) \geq \tilde{U}(y, j) \iff (x, i)\ \tilde{R}\ (y, j)$, then

$$x\ R\ y \iff \min_{i \in N} \{\tilde{U}(x, i)\} \geq \min_{i \in N} \{\tilde{U}(y, i)\}$$

Hence, provided that $\tilde{R}$ can be represented by an interpersonal utility function on $X \times N$, this GSWF is identical to the maximin rule considered in Section 3.4.

The maximin GSWF satisfies obvious extensions to the domain $\mathcal{R}(X \times N)$ of interpersonal orderings of Arrow's conditions of unrestricted domain, Pareto, and non-dictatorship. Indeed, it even satisfies the *anonymity condition* requiring that $g(\tilde{R}) = g(\tilde{R}')$ whenever $\tilde{R}$ and $\tilde{R}'$ are two interpersonal orderings on $X \times N$ which are related in the sense that, for some permutation $\sigma : N \to N$ of the individuals in N, one has

$$(x, i)\ \tilde{R}'\ (y, j) \iff (x, \sigma(i))\ \tilde{R}\ (y, \sigma(j))$$

Maximin does not satisfy the strict Pareto condition (P*) requiring that $x\ P\ y$ if $(x, i)\ \tilde{R}\ (x, i)$ for all $i \in N$ and there exists $j \in N$ such that $(x, j)\ \tilde{P}\ (x, j)$. However, the maximin GSWF can be made to satisfy (P*) by extending it lexicographically to the *leximin* GSWF. This GSWF is easier to specify after first defining a *ranking* $r_i(x)$ in each social state $x \in X$ as any mapping from N to $\{1, 2, \ldots, n\}$ such that $r_i(x) \geq r_j(x) \iff (x, i)\ \tilde{R}\ (x, j)$ for all $i, j \in N$. Ties can be broken arbitrarily. Let $i_r(x)$ denote the unique rth ranked individual in state x. Then the maximin GSWF satisfies $x\ R\ y \iff (x, i_1(x))\ \tilde{R}\ (y, i_1(y))$, because any individual who is given the rank 1 in a particular social state must have lower utility than anybody else in the same social state. The lexicographic extension of this rule is specified by

$$\begin{aligned} x\ P\ y \iff \exists r \in \{1, 2, \ldots, n\} : \quad & (x, i_k(x))\ \tilde{I}\ (y, i_k(y)) \quad (k = 1, 2, \ldots, r-1) \\ \text{and} \quad & (x, i_r(x))\ \tilde{P}\ (y, i_r(y)) \end{aligned}$$

Obviously, this definition implies that

$$x\ I\ y \iff (x, i_k(x))\ \tilde{I}\ (y, i_k(y))\ (k = 1, 2, \ldots, n)$$

Neither maximin nor leximin satisfies Arrow's independence condition (I), however. To see this, consider any non-empty $A \subset X$ and any two interpersonal orderings $\tilde{R}$ and $\tilde{R}'$ on $X \times N$. Then it is not true that $R_i(\tilde{R}) =_A R_i(\tilde{R}')$ (all $i \in N$) implies $g(\tilde{R}) =_A g(\tilde{R}')$; the social ordering of the elements of A depends on interpersonal comparisons of utility levels as well as on the profile $\langle R_i(\tilde{R}) \rangle_{i \in N}$ of induced individual orderings restricted to A.

Instead of Arrow's independence of irrelevant personal comparisons, the maximin and leximin GSWFs both satisfy a less demanding condition, which I like to call *independence of irrelevant interpersonal comparisons* (or IIIC). This requires that, if $\emptyset \neq A \subset X$ and $\tilde{R} =_{A \times N} \tilde{R}'$, then $g(\tilde{R}) =_A g(\tilde{R}')$. Condition (IIIC) is weaker than (I) because $\tilde{R} =_{A \times N} \tilde{R}'$ implies that $R_i(\tilde{R}) =_A R_i(\tilde{R}')$ (all $i \in N$), but the converse is not true.

Apart from leximin, there are many other GSWFs which also satisfy conditions (U), (IIIC), (P*) and anonymity. One other possible SWFL, for example, is the "leximax" rule defined by

$$\begin{aligned} x\ P\ y \iff \exists r \in \{1, 2, \ldots, n\} : \quad & (x, i_k(x))\ \tilde{I}\ (y, i_k(y)) \quad (k = r+1, \ldots, n) \\ \text{and} \quad & (x, i_r(x))\ \tilde{P}\ (y, i_r(y)) \end{aligned}$$

As shown by Roberts (1980a, b), all the other possible rules satisfying these four conditions involve a lexicographic hierarchy of "dictatiorial positions". Of all these SWFLs, only leximin satisfies the additional equity axiom formulated in Hammond (1976) — see also Hammond (1979). The main conclusion, however, is that the maximin and leximin SWFLs do satisfy independence condition (IIIC), even though they generally do not satisfy condition (I).

22.4.2 Interpersonal Comparisons of Utility Differences and Utilitarianism

Chapter 5 discusses axioms that are sufficient to imply that behaviour in risky decision trees should maximize the expected value of a von Neumann–Morgenstern utility function. One might argue that higher normative standards should apply to social than to individual decision-making. So these earlier axioms seem no less applicable to social decision-making than they are to individual behaviour.

Following the notation of Chapter 5, let $\Delta(X)$ denote the set of simple probability distributions on the domain X, which now consists of social states. That is, each member $\lambda \in \Delta(X)$ is a mapping $\lambda : X \to [0,1]$ for which there is a finite *support* $F \subset X$ such that $\lambda(x) > 0 \iff x \in F$, and also $\sum_{x \in F} \lambda(x) = 1$. Given any $\lambda \in \Delta(X)$ and any real-valued function f on X, denote the *expected value* of f w.r.t. λ by $\mathbb{E}_\lambda f(x) := \sum_{x \in F} \lambda(x)\, f(x)$. Assume now that, because the axioms of Chapter 5 are satisfied, or for any other reason, there is a *von Neumann–Morgenstern* (or NM) *Bergson social welfare function* $w : X \to \mathbb{R}$ whose expected value *represents* the social ordering R on $\Delta(X)$ in the sense that, whenever $\lambda, \mu \in \Delta(X)$, then $\lambda \; R \; \mu \iff \mathbb{E}_\lambda w(x) \geq \mathbb{E}_\mu w(x)$. Assume too that there is an NM *interpersonal welfare function* $v : X \times N \to \mathbb{R}$ whose expected value represents the interpersonal ordering $\tilde{R}$ on $\Delta(X \times N)$ in the sense that, whenever $\tilde{\lambda}, \tilde{\mu} \in \Delta(X \times N)$, then $\tilde{\lambda} \; \tilde{R} \; \tilde{\mu} \iff \mathbb{E}_{\tilde{\lambda}} v(x,i) \geq \mathbb{E}_{\tilde{\mu}} v(x,i)$.

Now assume that the Pareto condition (P) is replaced by the *Pareto indifference condition* (P^0) requiring that, whenever $\lambda, \mu \in \Delta(X)$ satisfy $\mathbb{E}_\lambda v(x,i) = \mathbb{E}_\mu v(x,i)$ for all $i \in N$, then $\mathbb{E}_\lambda w(x) = \mathbb{E}_\mu w(x)$. Under this assumption and some additional domain conditions, Harsanyi (1955) showed that there must exist constant "welfare weights" ω_i $(i \in N)$ such that $w(x) \equiv \sum_{i \in N} \omega_i \, v(x,i)$ on X. That is, one must have a weighted utilitarian Bergson social welfare function. Of course, if condition (P) is supplemented by (P^0) rather than replaced by it, then the welfare weights ω_i must be non-negative, for all $i \in N$, and at least one ω_j must be positive. For other proofs showing that Harsanyi's result is valid even without additional domain conditions, see Border (1985), Coulhon and Mongin (1989), Broome (1990), and also Hammond (1992). A similar result also appears later in Section 5.4 of this chapter.

In this framework it is hardly surprising that Arrow's condition (I) forces interpersonal comparisons to be ignored. Then Arrow's impossibility theorem implies that there must be a dictator $d \in N$ such that $\omega_d > 0$ and $\omega_i = 0$ for all $i \in N \setminus \{d\}$. What initially may be surprising, however, is that in the present framework involving the social choice of risky consequences or consequence lotteries in $\Delta(X)$, condition (IIIC) has exactly the same strong and unacceptable implication, provided the domain of possible utility functions is sufficiently rich. A formal result can be found in Hammond (1991b, Section 9).

The basic explanation is that (IIIC) requires the social ordering R restricted to any finite set $A \subset X$ to remain invariant under any non-linear strictly increasing transformation of the function $v(x,i)$. For this to be true when R is represented by $\sum_{i \in N} \omega_i \, v(x,i)$ on the set A, generally the sum must collapse to the single term $\omega_d \, v(x,d)$ for some $d \in N$.

The obvious remedy is to weaken the independence condition still further. The new condition, called *independence of irrelevant interpersonal comparisons of mixtures* (or IIICM), requires that, if $\emptyset \neq A \subset X$ and the two interpersonal orderings $\tilde{R}$ and $\tilde{R}'$ satisfy $\tilde{R} =_{\Delta(A \times N)} \tilde{R}'$, then the two associated social orderings R and R' should satisfy $R =_A R'$.

For any non-empty $A \subset X$, the fact that $\tilde{R}$ is represented by $\mathbb{E}v(x,i)$ on $\Delta(A \times N)$ implies that $v(x,i)$ is determined uniquely on the set $A \times N$ up to a positive affine transformation — as discussed in Chapter 5, for instance. This obviously implies that the function which maps each $x \in A$ to $\sum_{i \in N} \omega_i \, v(x,i)$ is determined uniquely up to a positive affine transformation — in particular, the social ordering R is determined uniquely on the set A. Hence, unlike (IIIC), condition (IIICM) is weak enough to be satisfied when R is represented by $\sum_{i \in N} \omega_i \, v(x,i)$ with $\omega_i \neq 0$ for at least two different individuals $i \in N$. There is no need for a dictatorship or any other restriction on the constants ω_i $(i \in N)$, except the obvious requirement that all should be positive if the strong Pareto condition (P*) holds.[3] In particular, utilitarianism — whether weighted or unweighted — satisfies independence condition (IIICM). It even satisfies the formally stronger condition requiring that $R =_{\Delta(A)} R'$ whenever $\tilde{R} =_{\Delta(A \times N)} \tilde{R}'$.

22.5 EXPECTED SOCIAL WELFARE

22.5.1 Social and Personal Consequences

The objectively expected utility functions of Chapter 5, and the arguments that were used to justify them, will now be applied to social decision problems. The result will be a form of utilitarianism that allows interpersonal comparisons to be interpreted as preferences for different personal characteristics.

First, given any $i \in N$, write X_i for a copy of the set X whose members x_i are i's *personalized social states*. As in the theory of public goods (Foley, 1970, p. 70; Milleron, 1972 etc.), it helps to imagine that we could somehow choose different social states $x_i \neq x_j$ for individuals i and j whenever they are different members of N, even though this may well be impossible in practice. Think how many social conflicts could be avoided if only we were each allowed to choose our own favourite social state! But the requirement that $x_i = x_j$ for all $i, j \in N$ can be imposed on the decision problem at a later stage.

[3] See Weymark (1991, 1993, 1995) for discussion of this and other similar sign restrictions on the welfare weights.

In addition to social states in the conventional sense, it will be convenient to consider also for each $i \in N$ a space of *personal characteristics* $\theta_i \in \Theta_i$. Such characteristics determine *i's* preferences, interests, talents, and everything else (apart from the social state) which is ethically relevant in determining the welfare of individual i. In Section 5.5, θ_i will even indicate whether or not individual i ever comes into existence.

For each individual $i \in N$, a *personal consequence* is a pair $z_i = (x_i, \theta_i)$ in the Cartesian product set $Z_i := X_i \times \Theta_i$ of personalized social states x_i and personal characteristics θ_i. Then, in a society whose membership N is fixed, a typical *social consequence* consists of a profile $z^N = (z_i)_{i \in N} \in Z^N := \prod_{i \in N} Z_i$ of such personal consequences — one for each individual member of society (both actual and potential). The consequence domain $Y = Z^N$ will consist of all such social consequences, with typical member $y = z^N$.

The theory of expected utility that was expounded and motivated in Chapter 5 can now be applied to the class of all decision problems with consequences in Z^N. The implication is the existence of a unique cardinal equivalence class of von Neumann–Morgenstern *social welfare functions* $w(y) \equiv w(z^N)$, defined on the space of social consequences, whose expected value should be maximized in every (finite) social decision problem. The only difference is that the consequence domain consists of social consequences. What is most important, however, is the idea that each personal consequence $z_i \in Z_i$ captures everything of ethical relevance to individual i — by definition, nothing else, including no other individual's personal consequence, can possibly be relevant to i's welfare.

22.5.2 Individualistic Consequentialism

A general random social consequence is some joint probability distribution $\lambda \in \Delta(Z^N)$ over the product space Z^N of different individuals' personal consequences. Such personal consequences could be correlated between different individuals, or they could be independent. The extent of this correlation should be of no consequence to any individual, however. For, provided that everything relevant to individual $i \in N$ really has been incorporated in each personal consequence $z_i \in Z_i$, all that really matters to i is the marginal distribution $\lambda_i \in \Delta(Z_i)$ of i's own consequences. This leads to the *individualistic consequentialism* hypothesis requiring any two lotteries $\lambda, \mu \in \Delta(Z^N)$ to be regarded as equivalent random consequences whenever, for every individual $i \in N$, the marginal distributions $\lambda_i = \mu_i \in \Delta(Z_i)$ of i's consequences are precisely the same. This means in particular that

$$\lambda_i = \mu_i \text{ (all } i \in N) \implies \mathbb{E}_\lambda\, w(z^N) = \mathbb{E}_\mu\, w(z^N)$$

— i.e., λ and μ must be indifferent according to the relevant expected utility criterion whenever the personal marginal distributions are all equal.

Succinctly stated, individual consequentialism amounts to requiring that only each individual's probability distribution of personal consequences be relevant when evaluating any social probability distribution. There is no reason to take account of any possible correlation between different individuals' personal consequences.

22.5.3 Individual Welfarism

Consider any decision problem having the special property that there is only one individual $i \in N$ whose distribution of personal consequences is affected by any feasible decision. Hence, there must be a profile $\bar{\lambda}_{-i} \in \prod_{h \in N \setminus \{i\}} \Delta(Z_h)$ of fixed lotteries $\bar{\lambda}_h \in \Delta(Z_h)$ $(h \in N \setminus \{i\})$ for all other individuals, as well as a set $F_i \subset \Delta(Z_i)$ of feasible lotteries over i's personal consequences, such that the feasible set of lotteries is $F_i \times \{\, \bar{\lambda}_{-i} \,\} \subset \Delta(Z^N)$. A decision problem with this property will be called *individualistic*.

The second individualistic axiom which I shall use is *individual welfarism*. This requires that for each $i \in N$ there is a unique cardinal equivalence class of *individual welfare functions* $w_i(z_i)$ with the property that, in any individualistic decision problem having $F_i \times \{\, \bar{\lambda}_{-i} \,\} \subset \Delta(Z^N)$ as the feasible set of lotteries, the social decision should maximize the expected value $\mathbb{E}_{\lambda_i} w_i(z_i)$ of w_i w.r.t. λ_i over the set $F_i \subset \Delta(Z_i)$ of feasible probability distributions over i's personal consequences. In particular, the social decision should be independent of $\bar{\lambda}_{-i}$.

This last independence property is the key hypothesis here. The motivation is that, if only consequences to i are affected by any decision, the fixed consequences to all other individuals are ethically irrelevant — assuming, as I do, that everything relevant to ethical decision making is already included in the consequences, and that only (distributions over) personal consequences matter.

Thus, whenever there is "no choice" in the personal consequences of all other individuals, the social objective becomes identical to the only affected individual's welfare objective. Note especially that individual welfarism poses no restrictions on what is allowed to count as part of a personal consequence and so to affect each individual's welfare. All it says is that, in "one person situations," social welfare is effectively identified with that one person's individual welfare.

22.5.4 Utilitarianism

Individual welfarism has a much more powerful implication, however, when it is combined with individualistic consequentialism as defined in Section 5.2. To see this, define the expected utility functions $U : \Delta(Z^N) \to \mathbb{R}$ and $U_i : \Delta(Z^i) \to \mathbb{R}$ by $U(\lambda^N) := \mathbb{E}_{\lambda^N} w(z^N)$ and $U_i(\lambda_i) := \mathbb{E}_{\lambda_i} w_i(z_i)$ $(i \in N)$ respectively. Now fix any profile $\bar{\lambda}^N \in \Delta(Z^N)$. Let n denote the number of individuals in the set N. Following an argument due to Fishburn (1970, p. 176) which was also used in the proof of Lemma 4.4 in Chapter 6, note that for all $\lambda^N \in \Delta(Z^N)$ one has

$$\sum\nolimits_{i \in N} \frac{1}{n} \, (\lambda_i, \bar{\lambda}_{-i}) = \frac{n-1}{n} \, \bar{\lambda}^N + \frac{1}{n} \, \lambda^N$$

Because the function U must preserve probability mixtures, it follows that

$$\frac{1}{n} \sum\nolimits_{i \in N} U(\lambda_i, \bar{\lambda}_{-i}) = \frac{n-1}{n} U(\bar{\lambda}^N) + \frac{1}{n} U(\lambda^N)$$

Therefore

$$U(\lambda^N) = \sum\nolimits_{i \in N} U(\lambda_i, \bar{\lambda}_{-i}) - (n-1) U(\bar{\lambda}^N)$$

But individual welfarism implies that $U(\lambda_i, \bar{\lambda}_{-i})$ and $U_i(\lambda_i)$ must be cardinally equivalent functions of λ_i. So, for each $i \in N$, there exist constants $\delta_i > 0$ and $\bar{\alpha}_i$ such that

$$U(\lambda_i, \bar{\lambda}_{-i}) = \bar{\alpha}_i + \delta_i U_i(\lambda_i)$$

for all $\lambda_i \in \Delta(Z^i)$. Therefore

$$U(\lambda^N) = \sum\nolimits_{i \in N} [\bar{\alpha}_i + \delta_i U_i(\lambda_i)] - (n-1) U(\bar{\lambda}^N) = \alpha + \sum\nolimits_{i \in N} \delta_i U_i(\lambda_i)$$

where $\alpha := \sum_{i \in N} \bar{\alpha}_i - (n-1) U(\bar{\lambda}^N)$. Hence, there must exist an additive constant α and a set of positive multiplicative constants δ_i $(i \in N)$ such that

$$w(z^N) \equiv \alpha + \sum\nolimits_{i \in N} \delta_i w_i(z_i)$$

Then, however, since the individual and social welfare functions are only unique up to a cardinal equivalence class, for each $i \in N$ we can replace the individual welfare function $w_i(z_i)$ by the cardinally equivalent function $\tilde{w}(z_i) := \delta_i w_i(z_i)$, and the social welfare function $w(z^N)$ by the cardinally equivalent function $\tilde{w}(z^N) := w(z^N) - \alpha$. The result is that

$$\tilde{w}(z^N) = w(z^N) - \alpha = \sum\nolimits_{i \in N} \delta_i w_i(z_i) = \sum\nolimits_{i \in N} \tilde{w}_i(z_i)$$

This takes us back to the simple addition of individual "utilities," once these have all been suitably normalized. Because of this possible normalization, I shall assume in future that

$$w(z^N) \equiv \sum\nolimits_{i \in N} w_i(z_i).$$

Note, however, that these utility functions are by no means the same as those in other more traditional versions of utilitarianism. They are merely representations of appropriate ethical social decisions in individualistic decision problems, without any necessary relationship to classical or other concepts of utility such as happiness, pleasure, absence of pain, preference satisfaction, etc. Indeed, the functions should probably be thought of more as indicators of individual value rather than as any measure of individual utility or even welfare. This is a major difference from Harsanyi's (1955) utilitarian theory. On the other hand, the additive structure of that theory is preserved.

22.5.5 Non-Existence

So far the set of individuals N has been treated as fixed. Yet many ethical issues surround decisions affecting the size of future generations, as well as the precise characteristics of those individuals who will come into existence. That is, both the number and the composition of the set N are of great ethical significance. Thus, it would seem that N itself should be treated as variable consequence along with z^N, as indeed it was in Hammond (1988). For some of the most recent work on the ethics of variable population, see Blackorby, Bossert and Donaldson (1995, 1996, 1997, 1998) and several other articles by the same authors.

A simpler alternative to the arguments in these papers, however, is to treat "non-existence" for any individual $i \in N$ as a particular personal characteristic $\theta_i^0 \in \Theta_i$ which i could have, and then to define N as the set of all potential rather than actual individuals. In this way, N is partitioned into the two sets $N^* := \{ i \in N \mid \theta_i \neq \theta_i^0 \}$ of actual individuals who do come into existence, and $N^0 := \{ i \in N \mid \theta_i = \theta_i^0 \}$ of individuals whose potential existence remains unrealized. Actually, not much generality is lost by doing this, for the following reason. Assuming that only a finite number of individuals can ever be born before the world comes to an end (as seems quite reasonable, despite economists' models of steady state growth, etc.), one can regard each identifier $i \in N$ as just an integer used to number all the individuals who come into existence, more or less in the temporal order of their birth. Everything that is really relevant about an individual i, including date of birth, can be included in i's personal characteristic θ_i. Accordingly, every individual who is ever born certainly gets numbered. Also, unless all the maximum possible number of individuals does actually come into existence, there will be "unused" numbers which refer to potential rather than actual individuals.

For those individuals $i \in N^0$ who never come into existence, the concept of individual welfare hardly makes any sense. In decision-theoretic terms, this means that non-existent individuals are not affected by social decisions — all social decisions are the same to them (except for decisions giving rise to a positive probability of their coming into existence, of course). Consider now, for any $i \in N$, an individualistic decision problem whose feasible set F_i has the property that $\lambda_i \in F_i$ only if $\lambda_i(X_i \times \{\theta_i^0\}) = 1$ — i.e., the probability of i not existing is always 1, no matter what decision is taken. Since all consequences in F_i are the same to this certainly non-existent individual, this suggests that all social decisions with consequences in F_i are equally ethically appropriate from the point of view of individual i alone. This motivates the assumption that, for some constant w_i^0, individual i's welfare function $w_i(z_i)$ should satisfy $w_i(x_i, \theta_i^0) = w_i^0$ for all $x_i \in X_i$. Thus, w_i^0 can be regarded as the constant "welfare of non-existence," which is entirely independent of the social state or any aspect of any social consequence in which i never exists.[4]

After making this assumption, one more useful normalization of individuals' welfare functions is possible. Replace each $w_i(z_i)$ by the function

$$\tilde{w}_i(z_i) := w_i(z_i) - w_i^0$$

This function is cardinally equivalent because a constant has merely been subtracted. Then, of course, $\tilde{w}_i(x_i, \theta_i^0) = 0$ for all $x_i \in X_i$, and so $\tilde{w}_i(z_i) = 0$ whenever $i \in N^0$. Similarly, replace $w(z^N) \equiv \sum_{i \in N} w_i(z_i)$ by the cardinally equivalent function

$$\tilde{w}(z^N) := w(z^N) - \sum_{i \in N} w_i^0$$

Then, however,

$$\tilde{w}(z^N) \equiv \sum_{i \in N} [w_i(z_i) - w_i^0] \equiv \sum_{i \in N} \tilde{w}_i(z_i) \equiv \sum_{i \in N^*} \tilde{w}_i(z_i)$$

where N^* is the set of individuals who ever come into existence. So only individuals in the set N^* need be considered when adding all individuals' welfare levels.

Once again, it will be assumed from now on that this normalization has been carried out. Because $N^* = \{ i \in N \mid \theta_i \neq \theta_i^0 \}$, it follows that

$$w(x^N, \theta^N) = w(z^N) \equiv \sum_{i \in N^*} w_i(z_i) = \sum_{i \in N^*} w_i(x_i, \theta_i)$$

Maximizing this social objective is formally identical to classical utilitarianism. But as already pointed out in Section 5.4, the resemblance is only formal because the individual welfare functions $w_i(z_i)$ mean something quite different. In particular, the zero level of this function is, by its very construction, just the minimum level of individual welfare at which it is ethically appropriate to cause the individual to come into existence.[5] This does much to dilute the strength of Parfit's (1984) "repugnant conclusion," which is that classical utilitarianism recommends creating very many extra individuals who are barely able to live above a subsistence level set so low that anyone who was forced to live below it would prefer not to have been born at all. Here we can escape the repugnant

[4] Blackorby *et al.*, in the works cited previously, prefer to call w_i^0 the *critical level* of i's utility; for them, a life has zero utility, by definition, when the individual is no better or worse off than by never having been born.

[5] A similar construction is used by Dasgupta (1993, ch. 13), who also provides a much more thorough philosophical discussion. The zero level in his approach, as well as in that outlined above, corresponds to the "critical level" considered by Blackorby *et al.* In this connection, Blackorby, Bossert and Donaldson (1998, p. 17) are mistaken when they claim that my approach uses "individual 'preferences' that cover states in which the person does not exist" — although this may not be entirely clear from the paper Hammond (1988) which they cite. In fact, the approach advocated here uses ethical *social* preferences throughout, even for decisions affecting only one individual, and this treatment of non-existence is one of the important ways in which social preferences differ from individual preferences.

conclusion because there is nothing to prevent the ethical values embodied in the normalized individual welfare function $w_i(z_i)$ from making w_i positive only if individual i would actually be quite well off if allowed to come into existence. The fact that the personal consequence z_i makes individual i glad to be alive is not by itself sufficient to make $w_i(z_i)$ positive, though many might argue that is a necessary condition.

Note too that having $w_i(z_i)$ positive would only be a sufficient condition on its own for wanting i to exist if i's existence could somehow be brought about without interfering with anybody else. Yet children cannot exist without having (or having had) parents. So the personal benefits (or costs) to i of coming into existence have to be weighed against any costs and benefits to other individuals, especially i's parents, etc. Some further discussion of such issues occurs in Hammond (1988).

22.6 INTERPRETING INTERPERSONAL COMPARISONS

22.6.1 Comparing Persons

The fourth and last question posed in the introduction concerned the way in which one can make interpersonal comparisons, as well as the significance that can be attached to such comparisons. As discussed in Hammond (1991a), a limitation of the SWFL approach to social choice theory was its failure to answer such questions satisfactorily. As Myerson (1985, pp. 238–9) points out, one should avoid interpreting the comparison $(x, i)\ \tilde{P}\ (y, j)$ as meaning that i prefers state x to being j in state y, or that j in state y would prefer being i in state x, because most such comparisons do not correspond to any choice that persons i or j could ever be forced to make. Nor, in fact, could anybody else be forced to make this choice.[6]

Some other interpretations of interpersonal comparisons can be given decision theoretic significance, however. For instance, even a dictatorial Arrow social welfare function requires interpersonal comparisons because the dictator has to be chosen. Similarly, interpersonal comparisons are involved in electing the president of a nation or of a university, and also in choosing which of one's friends should receive a spare opera ticket (Harsanyi, 1987).

Of course, the decisions in these examples involve comparisons of individuals; it may not be so obvious that they involve interpersonal comparisons of utility. But somebody who votes in a national presidential election is surely expressing an opinion concerning the relative utilities to the nation of the different candidates as potential presidents. Also, it is reasonable that the choice of which friend is to receive a spare opera ticket should maximize the donor's utility function over the set of possible recipients. Naturally, the utility in-

[6] Myerson goes so far as to claim that "Interpersonal comparisons of utility cannot be given decision theoretic significance." This seems an unjustified extension of the valid weaker claim that the standard interpretation of interpersonal comparisons set out in this paragraph lends them no practical decision theoretic significance.

crease that each different presidential candidate would derive personally from being elected is only of marginal relevance to most voters. Similarly, the person who gives away the spare opera ticket may not always choose the friend who is likely to derive the most enjoyment or personal benefit from the performance; a particularly close or deserving friend might be chosen instead, if giving the ticket to that friend would generate more utility for the donor.

Thus, the interpersonal comparisons that are relevant to decisions involving the choice of a person are comparisons of the utilities of different people to the chooser, rather than comparisons of different people's own utilities. This important but little appreciated difference should be borne in mind throughout the ensuing discussion.

22.6.2 Revealed Interpersonal Comparisons

The social welfare functional (SWFL) approach to social choice theory, which was discussed in Section 3.4, never made explicit the interpersonal comparisons on which it was based. It is now my duty to explain how the particular version of utilitarianism expounded in Section 5 can remedy this defect.

In fact, as pointed out in Hammond (1991a), there are interpersonal comparisons embodied in the social welfare function $w(z^N) = \sum_{i \in N^*} w_i(z_i)$. These interpersonal comparisons also meet Myerson's (1985) criticism because they can be given decision-theoretic significance. For the level comparison $w_h(z_h) > w_i(z_i)$ means that society is better off creating individual h with personal consequence z_h rather than individual i with personal consequence z_i. And the difference comparison $w_h(z_h) - w_h(z'_h) < w_i(z'_i) - w_i(z_i)$, which is of course equivalent to $w_h(z_h) + w_i(z_i) < w_h(z'_h) + w_i(z'_i)$, really does mean that moving h from z_h to z'_h and i from z_i to z'_i produces a benefit to society (if nobody else is affected). If there is a loss to h, this must be outweighed by the gain to i. Alternatively, if there is a loss to i, this must be outweighed by the gain to h.

Indeed, even welfare ratios acquire meaning. For $w_h(z_h)/w_i(z_i)$ can be regarded as the marginal rate of substitution between individuals like h facing personal consequence z_h and individuals like i facing personal consequence z_i. If this ratio is greater than 1, for instance, then society could gain by creating more individuals like h and fewer like i. And if $w_h(z_h)/w_i(z_i) = 10$, this means that society should be indifferent between creating 10 individuals like i and one individual like h. Thus, the claim that a Brahmin has 10 times the utility (or welfare) of an Untouchable does have meaning, even if most of us would regard the kinds of decision implied by such a claim as highly unethical and obnoxious.[7]

[7]Robbins (1938, p. 636) attributes to Sir Henry Maine a story of a Brahmin who, upon meeting a Benthamite, was moved to say: "I am ten times as capable of happiness as that untouchable over there." See also Sen (1973, pp. 81–2). The Brahmin's statement appears quite obnoxious, but actually is not immediately relevant to any social decision, except insofar as it was addressed to a Benthamite. After all, the statement is about the capacity for

So we have a "cardinal ratio scale" measure of individual welfare, with "cardinal full comparability" of both welfare levels and differences, as well as a clearly defined zero level of welfare. Yet, according to the theory expounded above, of all the SWFLs considered by Roberts (1980b) which have this property, only the simple sum is ethically appropriate. The social welfare functional is no longer left indeterminate, therefore, as usually happens in the SWFL approach to social choice theory.

Of course, this extra determinacy of the functional form comes at a price, since now all the indeterminacy has been displaced into the individual welfare function. In Sen's version of social choice theory with interpersonal comparisons, as well as in Harsanyi's version of utilitarianism, one could argue that each individual's utility measure had some objective reality which was independent of any particular ethical values; here instead the relevant ethical values, as revealed by judgements of what are the right decisions in particular individualistic decision problems, have been used to construct the relevant individual welfare measure.

22.6.3 Assessment

The theory set out in this section and the previous one gives rise to a unique cardinal equivalence class of "von Neumann–Morgenstern social welfare functions." These are defined so that it is their expected values which society should maximize over the feasible set of random social consequences. The social consequences to be considered amount to interpersonal profiles of personal consequences, with each personal consequence summarizing everything which is ethically relevant to the corresponding individual. There should be a welfare function for society as a whole, and also individual welfare functions determining what decisions are ethically acceptable when there is only one individual whose personal consequences can be affected by the decision. The relevant individual's welfare function should be cardinally equivalent to the social welfare function in all such "one person situations." This implies that the social welfare function can be expressed as the sum of suitably normalized individual welfare functions whose expected values it is right to maximize in one person situations where only one individual's personal consequence lottery is affected by any decision that could be made. Thus, the theory entails a form of classical utilitarianism which is close in spirit to that of Harsanyi (1955), even though the interpretation of individual utility or welfare functions is quite different.

Finally, on the assumption that decisions affecting individuals who never come into existence regardless of what decision is made do not matter, and so can be made arbitrarily, it was shown that individuals' welfare levels could be normalized to zero for individuals who never exist. Then it is enough to sum the welfare levels of all individuals who do come into existence, and even to use

---

happiness rather than about any ethical measure of individual welfare; though Benthamites confuse the two, by definition, there is no reason for the rest of us to do so.

this criterion to determine who should come into existence — exactly as with classical utilitarianism. The most obvious defects of that ethical theory are avoided, however, by interpreting individual welfare as a purely ethical concept representing what behaviour *ought* to maximize in decision trees affecting the personal consequences of only one individual. Indeed, it is probably more appropriate to give the name " ethical value functions" to the resulting measures of individual welfare.

In particular, if an individual's measure of welfare or ethical value is negative, this indicates that society's ethical value would have been higher had that individual never come into existence; this is quite consistent with the individual feeling better off personally than if he or she had never been born. So, just like the critical levels used by Blackorby *et al.*, this approach is also able to escape the most damaging features of Parfit's (1987) repugnant conclusion.

22.7 CONCLUDING REMARKS

Sen (1970) pioneered the social welfare functional approach to social choice theory, which many of us followed during the ensuing decade. This approach allows the social welfare ordering to depend on broader information than the profiles of individual orderings that form the domain of an Arrow social welfare function. In particular, a social welfare functional could accommodate interpersonal comparisons of utility. This approach was very useful in alerting us to the possibility that the iron grip of Arrow's "dictatorship theorem" could be relaxed provided that one not only admitted interpersonal comparisons, but also replaced Arrow's independence of irrelevant alternatives condition by some form of "independence of irrelevant interpersonal comparisons," as in Hammond (1991).

In retrospect, however, the social welfare functional approach can now be seen as having several quite serious defects. One is the failure to explain how interpersonal comparisons of utility are to be interpreted. A second is its failure to provide an unambiguous procedure for embodying interpersonal comparisons in the social welfare functional that generates the social welfare ordering.

In fact, utility functions are typically determined only up to an equivalence class that contains many functions representing the same preferences. For this reason, interpersonal comparisons of utility, thought of as comparisons of different individuals' utilities, make little sense until put into a more appropriate framework. One more suitable alternative is the interpersonal ordering $\tilde{R}$ on $X \times N$ considered in Section 4 which, when represented by a single interpersonal utility function $\tilde{U}$ on $X \times N$, gives meaning to comparisons between utility levels $\tilde{U}(x, i)$ and $\tilde{U}(y, j)$ for any pair (x, i) and (y, j) in $X \times N$. A second alternative considers the interpersonal ordering $\tilde{R}$ on the set $\Delta(X \times N)$ of simple lotteries over $X \times N$ which, if it can be represented by the the expected value of each von Neumann–Morgenstern utility function in a cardinal equivalence class, gives meaning to comparisons between utility differences, and even to ratios of utility differences. In this way, the fundamental concept becomes the interpersonal ordering that is represented by an interpersonal utility function which happens to give meaning to interpersonal comparisons of utility, rather than starting out with different individuals' utility functions which one then tries to compare.

Though this direct use of an interpersonal ordering seems a definite improvement, it still leaves us with the question of what this ordering is meant to represent, and how it and the interpersonal comparisons it implies should be reflected in social preferences. Rather than face these questions directly, Sections 5 and 6 are intended to lay out the details of a comprehensive ethical decision theory, based on consequentialist principles of the kind that were discussed in Chapters 5 and 6 of Volume 1. This leads to a form of utilitarianism requiring the maximization of the expected sum of individual utility functions that all lie within one common cardinal equivalence class. Indeed, by considering what variations might be desirable in the set of individuals who come into existence, these utilities can be given a meaningful zero level and determined up to a common ratio scale.

The utility functions constructed in this way, however, reflect each person's relative ethical value, rather than their utility in the sense that is generally understood by most social choice and decision theorists. Individual preferences and individual values are important considerations in determining the measure of an individual's ethical value, but do not determine it; other considerations that are relevant to ethical decision-making also have to be included, and sometimes even allowed to dominate.

This inevitable dependence on ethical values can be seen as the price that has to be paid for having individual value functions which, by construction, represent appropriate ethical preferences. As a by-product, the theory gives interpersonal comparisons the relatively simple and concrete interpretation discussed in Section 6. It also tells us exactly how to embody such comparisons into our social objective. In this way, the theory remedies some otherwise quite serious shortcomings in the social welfare functional approach.

Finally, this chapter has considered only the case when there is a single interpersonal ordering, or when ethical decisions are made by some kind of benevolent ethical dictator. It does not consider what is implied by the divergence of ethical opinions that seems inevitable in any real human society, notwithstanding the arguments of Harsanyi (1955) and others. Indeed, suppose that all individuals subscribe to the theory set out in Sections 5 and 6, but have divergent ethical values. Then they will have different cardinal equivalence classes of "von Neumann–Morgenstern ethical value function" whose expected value they think it is right to maximize. Now, this is exactly the setting for Sen's (1970, Theorem 8*2, pp. 129–30) cardinal extension of Arrow's impossibility theorem. Moreover, as shown by Bordes, Hammond and Le Breton (1997), there is little hope of escaping the need to dictate the ethical values by restricting the domain of admissible profiles of different individuals' opinions concerning what the ethical value function should be. Unless, that is, one admits interpersonal comparisons of ethical values in a way that allows some weighted average of different individuals' versions of the ethical value function to be constructed. But then it seems that the different weights would have to be determined dictatorially.

Ostensibly, the original purpose of introducing interpersonal comparability into social choice theory was to by-pass the Arrow theorem and so allow different individuals' welfare measures to be aggregated into a social welfare measure without the need for dictatorship. In the end, this enterprise has proved less successful than might have been hoped. It is true that one can give positive weight to many different individuals' welfare measures, and so avoid the crudest form of dictatorship. But as discussed by Suzumura (1983, 1996), by Roberts (1995, 1996, 1997), and by Nagahisa and Suga (1998), faced with diversity of ethical opinion, dictatorship of ethical values appears inevitable if one is going to have a (complete) social welfare ordering satisfying some form of Pareto criterion. Giving this important topic the attention it deserves, however, would take us too far away from the main topic of this chapter.

Moreover, according to the theory propounded in Sections 5 and 6 of this chapter, the relevant concept of individual utility or welfare has undergone drastic change, so that it also depends no less on what may have to be a dictated concept of ethical personal value than it does on an individual's own preferences. Although the new theory is designed to have more secure ethical foundations, its link to empirical reality remains more tenuous than ever. Probably that was only to be expected.

Acknowledgments

Much of this chapter is based on two lectures delivered at the Institut International de Paris La Défense during June, 1995. My thanks to Serge Kolm for inviting me to give those lectures, and for the interesting discussion with him and other members of the audience that ensued. Section 5 is adapted from parts of Hammond (1996). The final version of this chapter was prepared during a visit to the Institute of Economic Research, Hitotsubashi University, sponsored by the Japanese Ministry of Education. I should like to express my gratitude to them, and to my host, Kotaro Suzumura.

References

Arrow, K.J. (1950) "A Difficulty in the Concept of Social Welfare," *Journal of Political Economy* **58**, 328–346; reprinted in Arrow (1983), ch. 1, pp. 1–29.

Arrow, K.J. (1951, 1963) *Social Choice and Individual Values* (2nd. edn.) (New Haven: Yale University Press).

Arrow, K.J. (1977) "Extended Sympathy and the Possibility of Social Choice," *American Economic Review (Papers and Proceedings)* **67**, 219–25; reprinted in Arrow (1983), ch. 11, pp. 147–161.

Arrow, K.J. (1983) *Collected Papers of Kenneth J. Arrow, Vol. 1: Social Choice and Justice* (Cambridge, Mass.: The Belknap Press of Harvard University Press).

Bandyopadhyay, T. (1988) "Revealed Preference Theory, Ordering and the Axiom of Sequential Path Independence" *Review of Economic Studies*, **55**, 343–351.

Barone, E. (1908) "Il ministero della produzione nello stato colletivista," *Giornale degli Economisti* **37**, 267–293 and 391–414; translated as "The Ministry of Production in the Collectivist State" in F.A. Hayek (ed.) (1935), *Collectivist Economic Planning* (London: Routledge & Kegan Paul), pp. 310–334; reprinted in P. Newman (ed.), *Readings in Mathematical Economics, Vol. I: Value Theory* (Baltimore: The Johns Hopkins Press).

Becht, M. (1995) "The Theory and Estimation of Individual and Social Welfare Measures," *Journal of Economic Surveys* **9**, 53–87.

Blackorby, C., W. Bossert and D. Donaldson (1995) "Intertemporal Population Ethics: Critical-Level Utilitarian Principles," *Econometrica*, **63**, 1303–20.

Blackorby, C., W. Bossert and D. Donaldson (1996) "Intertemporally Consistent Population Ethics: Classical Utilitarian Principles," in K.J. Arrow, A.K. Sen and K. Suzumura (eds.) *Social Choice Re-examined, Vol. 2*, IEA Conference Volume, no. 117 (London: Macmillan), ch. 13, pp. 137–162.

Blackorby, C., W. Bossert and D. Donaldson (1997) "Birth-Date Dependent Population Ethics: Critical-Level Principles" *Journal of Economic Theory* **77**, 260–284.

Blackorby, C., W. Bossert and D. Donaldson (1998) "Uncertainty and Critical-Level Population Principles," *Journal of Population Economics*, **11**, 1–20.

Blackorby, C. and D. Donaldson (1988) "Cash versus Kind, Self-Selection, and Efficient Transfers," *American Economic Review* **78**, 691–700.

Blackorby, C., D. Donaldson and J. Weymark (1984) "Social Choice with Interpersonal Utility Comparisons: A Diagrammatic Introduction," *International Economic Review*, **25**, 327–356.

Blackorby, C., D. Donaldson and J. Weymark (1990) "A Welfaristic Proof of Arrow's Theorem," *Recherches Economiques de Louvain*, **56**, 259–286.

Border, K.C. (1985) "More on Harsanyi's Utilititarian Cardinal Welfare Theorem," *Social Choice and Welfare* **1**, 279–281.

Bordes, G., P.J. Hammond and M. Le Breton (1997) "Social Welfare Functionals on Restricted Domains and in Economic Environments," Stanford University, Department of Economics Working Paper No. 97-023.

Broome, J. (1990) "Bolker–Jeffrey Expected Utility Theory and Axiomatic Utilitarianism," *Review of Economic Studies* **57**, 477–502.

Burks, A. (1977) *Chance, Cause, Reason: An Inquiry into the Nature of Scientific Evidence* (Chicago, University of Chicago Press).

Campbell, D.E. (1978) "Realization of Choice Functions," *Econometrica*, **46**, 171–180.

Chipman, J.S. and J.C. Moore (1978) "The New Welfare Economics, 1939–1974," *International Economic Review* **19**, 547–584.

Coulhon, T. and P. Mongin (1989) "Social Choice Theory in the Case of Von Neumann–Morgenstern Utilities," *Social Choice and Welfare* **6**, 175–187.

Dasgupta, P. (1993) *An Enquiry into Well-Being and Destitution* (Oxford: Clarendon Press).

D'Aspremont, C. (1985) "Axioms for Social Welfare Orderings," in L. Hurwicz, D. Schmeidler and H. Sonnenschein (eds.) *Social Goals and Social Organization* (Cambridge: Cambridge University Press), ch. 2, pp. 19–76.

Fishburn, P.C. (1970) *Utility Theory for Decision Making* (New York: John Wiley).

Foley, D.K. (1970) "Lindahl's Solution and the Core of an Economy with Public Goods," *Econometrica*, **38**, 66–72.

Hammond, P.J. (1976) "Equity, Arrow's Conditions, and Rawls' Difference Principle," *Econometrica*, **44**, 793–804.

Hammond, P.J. (1977) "Dynamic Restrictions on Metastatic Choice," *Economica*, **44**, 337–350.

Hammond, P.J. (1979) "Equity in Two Person Situations: Some Consequences," *Econometrica*, **47**, 1127–35.

Hammond, P.J. (1988) "Consequentialist Demographic Norms and Parenting Rights," *Social Choice and Welfare*, **5**, 127–145.

Hammond, P.J. (1990) "Theoretical Progress in Public Economics: A Provocative Assessment," *Oxford Economic Papers*, **42**, 6–33; also in P.J.N. Sinclair and M.D.E. Slater (eds.) *Taxation, Private Information and Capital* (Oxford: Clarendon Press, 1991).

Hammond, P.J. (1991a) "Interpersonal Comparisons of Utility: Why and How They Are and Should Be Made," in J. Elster and J.E. Roemer (eds.), *Interpersonal Comparisons of Well-Being* (Cambridge: Cambridge University Press), ch. 7, pp. 200–254.

Hammond, P.J. (1991b) "Independence of Irrelevant Interpersonal Comparisons," *Social Choice and Welfare*, **8**, 1–19.

Hammond, P.J. (1992) "Harsanyi's Utilitarian Theorem: A Simpler Proof and Some Ethical Connotations," in R. Selten (ed.) *Rational Interaction: Essays in Honor of John Harsanyi* (Berlin: Springer-Verlag, 1992), pp. 305–319.

Hammond, P.J. (1994) "Money Metric Measures of Individual and Social Welfare Allowing for Environmental Externalities," in W. Eichhorn (ed.) *Models and Measurement of Welfare and Inequality* (Berlin: Springer-Verlag) pp. 694–724.

Hammond, P.J. (1996) "Consequentialist Decision Theory and Utilitarian Ethics," in F. Farina, F. Hahn, and S. Vannucci (eds.) *Ethics, Rationality, and Economic Behaviour* (Oxford: Clarendon Press, 1996), pp. 92–118.

Hammond, P.J. (1997) "Multilaterally Strategy-Proof Mechanisms in Random Aumann–Hildenbrand Macroeconomies," Stanford University Department of Economics Working Paper No. 97-022; revised version to appear in M. Wooders (ed.) *Topics in Game Theory and Mathematical Economics* (American Mathematical Society).

Hammond, P.J. and J. Sempere (1995) "Limits to the Potential Gains from Economic Integration and Other Supply Side Policies," *Economic Journal* **105**, 1180–1204.

Harberger, A.C. (1971) "Three Basic Postulates for Applied Welfare Analysis," *Journal of Economic Literature* **9**, 785–797.

Harberger, A.C. (1978) "On the Use of Distributional Weights in Social Cost Benefit Analysis," *Journal of Political Economy* **86**, S87–S120.

Harsanyi, J.C. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* **63**, 309–321; reprinted in Phelps (1973) and Harsanyi (1976).

Harsanyi, J.C. (1976) *Essays on Ethics, Social Behavior, and Scientific Explanation* (Dordrecht: D. Reidel).

Harsanyi, J.C. (1987) "Interpersonal Utility Comparison," in J. Eatwell, M. Milgate and P. Newman (eds.) *The New Palgrave Dictionary of Economics* (London: Macmillan).

Little, I.M.D. (1957) *A Critique of Welfare Economics* (Oxford: Oxford University Press).

Milleron, J.C. (1972) "Theory of Value with Public Goods: A Survey Article," *Journal of Economic Theory* **5**, 419–477.

Mirrlees, J.A. (1971) "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies* **38**: 175–208.

Myerson, R. (1985) "Bayesian Equilibrium and Incentive Compatibility: An Introduction," in L. Hurwicz, D. Schmeidler and H. Sonnenschein (eds.) *Social Goals and Social Organization* (Cambridge: Cambridge University Press), ch. 8, pp. 229–259.

Nagahisa, R. and K. Suga (1998) "Impossibility Theorems with Interpersonally Comparable Welfare Levels: The 'Extended Sympathy' Approach Reconsidered" *Japanese Economic Review*, **49**, 171–190.

Parfit, D. (1987) *Reasons and Persons* (Oxford: Oxford University Press).

Phelps, E.S. (ed.) (1973) *Economic Justice* (Harmondsworth: Penguin Books).

Plott, C.R. (1973) "Path Independence, Rationality, and Social Choice," *Econometrica*, **41**, 1075–1091.

Robbins, L. (1938) "Interpersonal Comparisons of Utility: A Comment," *Economic Journal*, **48**, 635–641.

Roberts, K.W.S. (1980a) "Possibility Theorems with Interpersonally Comparable Welfare Levels," *Review of Economic Studies* **47**, 409–420.

Roberts, K.W.S. (1980b) "Interpersonal Comparability and Social Choice Theory," *Review of Economic Studies* **47**, 421–439.

Roberts, K.W.S. (1995) "Valued Opinions or Opinionated Values: The Double Aggregation Problem" in K. Basu, P.K. Pattanaik, and K. Suzumura (eds.) *Choice, Welfare and Development: A Festschrift in Honour of Amartya K. Sen* (Oxford: Oxford University Press, 1995), ch. 8, pp. 141–165.

Roberts, K.W.S. (1996) "Interpersonal Comparisons of the Extended Sympathy Type and the Possibility of Social Choice: Discussion" in K.J. Arrow, A.K. Sen and K. Suzumura (eds.) *Social Choice Re-examined, Vol. 2* IEA Conference Volume, no. 117 (London: Macmillan), pp. 230-236.

Roberts, K.W.S. (1997) "Objective Interpersonal Comparisons of Utility" *Social Choice and Welfare*, **14**, 79–96.

Sen, A.K. (1970) *Collective Choice and Social Welfare* (San Francisco: Holden-Day).

Sen, A.K. (1973) *On Economic Inequality* (Oxford: Clarendon Press).

Suppes, P. (1966) "Some Formal Models of Grading Principles," *Synthese* **6**: 284–306.

Suzumura, K. (1983) "Resolving Conflicting Views of Justice in Social Choice," in P.K. Pattanaik and M. Salles (eds.) *Social Choice and Welfare* (North-Holland, 1983), ch. 8, pp. 125–149.

Suzumura, K. (1996) "Interpersonal Comparisons of the Extended Sympathy Type and the Possibility of Social Choice," in K.J. Arrow, A.K. Sen and K. Suzumura (eds.) *Social Choice Re-examined, Vol. 2* IEA Conference Volume, no. 117 (London: Macmillan), ch. 15, pp. 202–229.

Vickrey, W.S. (1945) "Measuring Marginal Utility by the Reactions to Risk," *Econometrica* **13**: 319–333.

Weymark, J.A. (1991) "A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism," in J. Elster and J.E. Roemer (eds.), *Interpersonal Comparisons of Well-Being* (Cambridge: Cambridge University Press), ch. 8, pp. 255–320.

Weymark, J.A. (1993) "Harsanyi's Social Aggregation Theorem and the Weak Pareto Principle," *Social Choice and Welfare*, **10**, 209–221.

Weymark, J.A. (1995) "Further Remarks on Harsanyi's Social Aggregation Theorem and the Weak Pareto Principle," *Social Choice and Welfare*, **12**, 87–92.