INDEPENDENCE OF IRRELEVANT INTERPERSONAL COMPARISONS

Peter J. Hammond, Department of Economics
European University Institute, Badia Fiesolana,
50016 S. Domenico di Fiesole (FI), Italy;
and Stanford University, CA 94305–6072, U.S.A.

December 1989; revised August 1990.

To appear in *Social Choice and Welfare*

ABSTRACT

Arrow's independence of irrelevant alternatives (IIA) condition makes social choice depend only on personal rather than interpersonal comparisons of relevant social states, and so leads to dictatorship. Instead, a new "independence of irrelevant interpersonal comparisons" (IIIC) condition allows anonymous Paretian social welfare functionals such as maximin and Sen's "leximin," even with an unrestricted preference domain. But when probability mixtures of social states are considered, even IIIC may not allow escape from Arrow's impossibility theorem for individuals' (ex-ante) expected utilities. Modifying IIIC to permit dependence on interpersonal comparisons of relevant probability mixtures allows Vickrey-Harsanyi utilitarianism.

Journal of Economic Literature classfication: 025

IRRELEVANT INTERPERSONAL COMPARISONS

> *Thus, if we wish to go beyond the comparisons that are possible using only the* [Pareto] *principle of the new welfare economics, the issue is not whether we can do so without making interpersonal comparisons of satisfactions. It is rather, what sorts of interpersonal comparisons are we willing to make. Unless the comparisons allowed by Arrow's Condition 3* [independence of irrelevant alternatives] *could be shown to have some ethical priority, there seems to be no reason for confining consideration to this group.*
>
> — Hildreth (1953, p. 91)

1. Introduction

It is becoming widely acknowledged that interpersonal comparisons of utility are likely to provide the only ethically satisfactory escape from Arrow's (1950, 1963, 1983) impossibility theorem. Their incorporation into social choice theory, however, requires some appropriate reformulation of Arrow's independence of irrelevant alternatives (IIA) condition.

An Arrow social welfare function (ASWF) makes the social ordering depend upon individual preference orderings. Sen (1970a, 1970b, 1977, 1982) showed how interpersonal comparisons of utility could be included in formal social choice theory by making the social ordering depend upon individuals' utility functions via a social welfare functional (SWFL). Unlike a profile of preference orderings, a profile of utility functions can incorporate various kinds of interpersonal comparisons, such as comparisons of utility levels, or of utility differences, or both. This became clear in the ensuing work by d'Aspremont and Gevers (1977), Sen (1977), Roberts (1980b), Blackorby, Donaldson and Weymark (1984), d'Aspremont (1985) and others.

This body of work has left two major issues unresolved, or at last unsatisfactorily resolved. The first is the interpretation of the interpersonal comparisons themselves. This has left room for much confusion in the literature on interpersonal comparisons, despite the significant contributions of Harsanyi (1955, 1976, 1977)

and many others. Elsewhere (Hammond, 1990) I have discussed the idea that interpersonal comparisons represent the relative ethical desirability of individuals with different personal types or characteristics, as well as of social states. In fact, I postulate the existence of a fundamental preference ordering (cf. Harsanyi, 1955, Tinbergen, 1957, Rawls, 1959, 1971, and Kolm, 1972) on the space of all *personal consequences* — which are pairs consisting of social states and personal characteristics. This postulate leads to interpersonal comparisons of utility levels, as explained further in Section 2 below. When probability mixtures of personal consequences are also considered and the maximization of the expected value of some von Neumann-Morgenstern social welfare function is assumed, the same postulate also leads to interpersonal comparisons of utility units, as explained in Section 8 below.

Arrow's IIA requires the social choice from any set of social states Z to depend only on individuals' preferences restricted to Z, and so excludes interpersonal comparisons. So the second issue is how to amend Arrow's IIA condition in order to allow interpersonal comparisons and, in particular, how to amend it without having it lose all its natural appeal, which elsewhere I have even sought to reinforce (Hammond, 1977, 1986). For SWFL's, a natural adaptation of IIA is to make the social ordering over any set depend only on individuals' utility functions restricted to that set, as in d'Aspremont and Gevers (1977), Sen (1977), Roberts (1980b), etc. One may call this condition "independence of irrelevant utilities" (IIU), since the "relevant utilities" are those of the "relevant alternatives" in Arrow's sense. The IIU condition, however, rather obviously allows dependence on irrelevant alternatives once interpersonal comparisons are based on individuals' own preferences for different personal characteristics. For, when personal characteristics and preferences are fixed, IIA requires us to ignore all the information which preferences for personal characteristics provide, and so interpersonal comparisons cannot affect social choice after all.

This observation prompts consideration of what really is a "relevant alternative" — a question that, so far as I know, has really only been taken up by

Strasnick (1977) in the past. In particular, are relevant alternatives social states, as in Arrow's work, or are they perhaps personal consequences, as might be thought appropriate for ethical decisions? More specifically, what are relevant preferences or comparisons? Are they individual preferences or personal comparisons over relevant social states, as they were for Arrow? Or are they fundamental preferences or interpersonal comparisons over relevant personal consequences?

This paper will consider the implications of allowing some particular interpersonal comparisons to become relevant. Section 2 sets out a basic framework for social choice theory which differs somewhat from the standard framework. Ethical views will be allowed to vary, so that the interpersonal comparisons which they generate can change. Section 3 considers the form which Arrow social welfare functions take within this new framework. Then Section 4 reconsiders Arrow's IIA condition when ethical views are allowed to vary and formulates a slightly more powerful "independence of irrelevant personal comparisons" (IIPC) condition. This leads to a strengthened form of Arrow's impossibility theorem in which not even a dictator's identity can depend on variable ethical views.

Section 5 formulates the first new independence condition which allows interpersonal comparisons to influence the social ordering. It is called "independence of irrelevant interpersonal comparisons" (IIIC). The same Section 5 also sets out an associated independence condition for a fixed society. Section 6 relates this to the independence condition for "generalized social welfare functions" of Hammond (1976). If individuals' welfare orderings can be represented by utility functions, then this reduces to the IIU condition discussed above.

Generalized social welfare functions make no attempt to allow for the special structure of expected utility maximizing preferences for random consequences. Accordingly, Section 7 extends the framework of Sections 2, 3, 4, 5 and 6 in order to allow risk. In fact Section 7 allows general preferences for risky consequences, and claims that nothing new emerges. In Section 8, however, it is assumed that both social and individual welfare orderings have expected utility representations. Under the usual Pareto condition, there is then a social welfare function which is a

non-negatively weighted sum of individual welfare functions (cf. Harsanyi, 1955).

In this framework of expected utility maximization, it is natural to extend both the IIPC condition of Section 4 and the IIIC condition of Section 5 to sets of risky consequences. Yet, under the assumption of a sufficiently rich domain of both personal characteristics and ethical views, Section 9 shows that even IIIC only allows dictatorial preferences. Accordingly, Sections 10 and 11 formulate two "independence of irrelevant (inter)personal comparisons of mixtures" conditions (IIPCM and IIICM). In Section 10 it is shown how IIICM *does* allow Vickrey-Harsanyi utilitarianism. In fact, it allows the maximization of an unweighted sum of fundamental (expected) utilities. This is the classical utilitarian criterion, somewhat reinterpreted, which I have sought to defend in Hammond (1987, 1988, 1990). Section 11 shows how interpersonal comparisons are essential. Just cardinalizing utilities is not enough because, without interpersonal comparisons, IIPCM is equivalent to IIPC. This accords with the earlier impossibility results for non-comparable cardinal utilities due to Sen (1970a) and Osborne (1976).

Section 12 is a brief concluding summary.

2. Social States, Personal Consequences and Fundamental Preferences

Let the *membership* of a society — the fixed finite set of individuals — be M, with at least two members. Each member $i \in M$ has a *personal characteristic* θ_i in the set Θ of all possible personal characteristics. Each $\theta \in \Theta$ determines everything relevant to evaluating the welfare of all θ-persons — their preferences, needs, etc. A *society* is a profile $\theta^M = \langle \theta_i \rangle_{i \in M}$ of personal characteristics, one for each individual $i \in M$.

Unlike most social choice theory, it will be convenient here to think of ethical views as being variable. The reason is that social choice rules will be considered which are allowed to depend upon variable interpersonal comparisons. Thus e, e' will denote possible ethical views or *ethics*, in the form of parameters affecting both social and individual preference orderings, including interpersonal comparisons, and E will denote the domain of all possible ethics.

Arrow (1963) takes both the society θ^M and the ethic e as given, not influenced by the social choice. There is a set X of at least three *social states*. As explained in the introduction, I shall allow choices to affect personal characteristics in order to have interpersonal comparisons (or, more exactly, comparisons between personal characteristics rather than between persons). So the choice space of conceivable consequences (cf. Hammond, 1986, 1987) is expanded from X to the product space $X \times \Theta^M$ whose members are pairs (x, θ^M) consisting of social states $x \in X$ and of societies $\theta^M \in \Theta^M$.

For any set S of possible social or individual states, let $\mathcal{R}(S)$ denote the set of all logically possible preference orderings over S. It will then be assumed that there is a mapping $\rho : E \to \mathcal{R}(X \times \Theta^M)$ which, for every possible ethic $e \in E$, determines the *social welfare ordering* $R = \rho(e)$ on $X \times \Theta^M$ as a function of e.

Of course, the preference ordering $\rho(e)$ is meant to be able to determine on its own the appropriate mode of behaviour in any ethical decision problem. The task of social choice theory, however, is to construct such an ordering from information about individual preferences or welfare orderings, as well as from interpersonal comparisons. In this sense, social choice theory is a study of how to construct individualistic ethical decision rules.

Interpersonal comparisons are ethical preferences regarding the personal characteristic θ as well as the social state x. Thus I shall assume that there is a mapping $\tilde{\rho} : E \to \mathcal{R}(X \times \Theta)$ which, for every possible ethic $e \in E$, determines the *fundamental (interpersonal) welfare ordering* $\tilde{R} = \tilde{\rho}(e)$ on the space $X \times \Theta$ of personal consequences — i.e., on combinations (x, θ) of social states together with personal characteristics. Indeed, when $\tilde{R} = \tilde{\rho}(e)$, one can interpret the strict preference statement $(x, \theta)\ \tilde{P}\ (y, \eta)$ as signifying that, according to the ethic e, it is preferable for society to have a θ-person in social state x rather than an η-person in social state y.

Following Harsanyi (1955) and Kolm (1972), I shall consider only a single ethic $e \in E$ which applies to all persons, and is independent of i. This is in contrast to Suppes (1966). The problem of reconciling the conflicting ethical views of different

members of a society will not be considered in this paper.

Finally, the social welfare ordering $\rho(e)$ is meant to be based upon individual welfare. It should therefore be true that there is a *social welfare function with ordinal interpersonal comparisons* (or SWF) $\phi : \mathcal{R}(X \times \Theta) \to \mathcal{R}(X \times \Theta^M)$ which determines the social welfare ordering $\rho(e) \in \mathcal{R}(X \times \Theta^M)$ as a function of the fundamental welfare ordering $\tilde{\rho}(e) \in \mathcal{R}(X \times \Theta)$. For this to be true for every ethic $e \in E$ requires that $\rho(e) \equiv \phi[\tilde{\rho}(e)]$. The following commutative function diagram, in which ι denotes the identity mapping, illustrates this.

$$\begin{array}{ccc} \mathcal{R}(X \times \Theta) & \xrightarrow{\iota} & \mathcal{R}(X \times \Theta) \\ \uparrow \tilde{\rho} & & \downarrow \phi \\ E & \xrightarrow{\rho} & \mathcal{R}(X \times \Theta^M) \end{array}$$

In most of what follows, I shall also assume that the (weak) *Pareto condition* (P) is satisfied, in the sense that when $\tilde{R} = \tilde{\rho}(e)$ is the fundamental ordering, and when $R = \rho(e) = \phi(\tilde{R})$ is the social ordering, then

$$(x, \theta_i)\ \tilde{P}\ (y, \eta_i) \quad (\text{all } i \in M) \Longrightarrow (x, \theta^M)\ P\ (y, \eta^M).$$

Thus, if every individual $i \in M$ is made better off as a result of changing both the social state and personal characteristics, then that change is to be preferred.

3. Arrow Social Welfare Functions

For each ethic $e \in E$ and each fixed society $\theta^M \in \Theta^M$, the social welfare ordering $R = \rho(e)$ on $X \times \Theta^M$ induces an obvious ordering $R(\theta^M, e)$ on the space of social states X which is given by

$$x\ R(\theta^M, e)\ y \iff (x, \theta^M)\ \rho(e)\ (y, \theta^M).$$

Arrow's original social choice problem can be regarded as how to determine the social ordering $R(\theta^M, e)$ on X as a function of θ^M — especially, as a function of the "individual values" or preferences in each society θ^M. Thus, in a society

θ^M, and given both the ethic $e \in E$ and the fundamental interpersonal ordering $\tilde{R} = \tilde{\rho}(e)$ on $X \times \Theta$, for each individual $i \in M$ there is an obvious induced *individual welfare ordering* $R_i(\theta_i, e)$ on X, depending on the personal characteristic θ_i, which is given by

$$x \; R_i(\theta_i, e) \; y \iff (x, \theta_i) \; \tilde{\rho}(e) \; (y, \theta_i).$$

Actually, for any $\theta \in \Theta$ and $e \in E$, the ordering $R_i(\theta, e)$ is independent of i. But the notation $R_i(\theta_i, e)$ seems clearer than just $R(\theta_i, e)$, since it emphasizes that $R_i(\theta_i, e)$ is the welfare ordering for individual i.

Recall from section 2 that the characteristic θ_i determines, amongst other things, individual i's own preference ordering. Notice, however, that each individual's welfare ordering $R_i(\theta_i, e)$ also depends in general upon the ethic e. In neoclassical welfare economics, it is usual to postulate "consumer sovereignty", whereby $R_i(\theta_i, e)$ is the preference ordering that determines i's behaviour when his type is θ_i, and so is unaffected by e. Here, the ethic is allowed to affect $R_i(\theta_i, e)$ because the ordering represents the ethical concept of individual welfare rather than the positive concept of individual behaviour. An individual's welfare ordering may change either because the personal characteristic (including the preference ordering) changes, or because the ethic changes. Indeed, $R_i(\theta_i, e)$ may not be the preference ordering revealed by an individual's behaviour because there may not even be any consistent revealed preference ordering at all. It may also differ because some revealed preferences are judged to be unethical — e.g., excessive intolerance for others. And it may differ because some revealed preferences are based on misinformation. In economic contexts, e helps determine any deviations from consumer sovereignty. For a recent extended discussion of some reasons for distinguishing individuals' welfare orderings from their preferences, as well as a suggestion for how to do so, see Haslett (1990).

Let $\mathcal{R}_i(X)$ denote the set of all logically possible preference orderings on X which individual i could have; evidently $\mathcal{R}_i(X) = \mathcal{R}(X)$. Then let $\mathcal{R}^M(X) = \prod_{i \in M} \mathcal{R}_i(X)$. An *Arrow social welfare function* (or ASWF) is then defined as a mapping $f : \mathcal{R}^M(X) \to \mathcal{R}(X)$ from the domain $\mathcal{R}^M(X)$ of all logically possible

preference profiles $R^M = \langle R_i \rangle_{i \in M}$ over X to the set $\mathcal{R}(X)$ of possible social preference orderings on X. In the framework being used here, such an ASWF exists if and only if the following function diagram commutes for every $\theta^M \in \Theta^M$ and for some f which must be independent of both θ^M and e:

$$\begin{array}{ccccc} \mathcal{R}(X \times \Theta) & \xrightarrow{\iota} & \mathcal{R}(X \times \Theta) & \xrightarrow{\langle R_i(\theta_i, e) \rangle_{i \in M}} & \mathcal{R}^M(X) \\ \uparrow \tilde{\rho} & & \downarrow \phi & & \downarrow f \\ E & \xrightarrow{\rho} & \mathcal{R}(X \times \Theta^M) & \xrightarrow{R(\theta^M, e)} & \mathcal{R}(X) \end{array}$$

Note that, if an ASWF exists and the above Pareto rule is satisfied, then for any fixed society $\theta^M \in \Theta^M$, one must have

$$x \; \tilde{P}_i(\theta_i, e) \; y \quad (\text{all } i \in M) \Longrightarrow x \; P(\theta^M, e) \; y.$$

This is just the usual form of the weak Pareto condition.

Notice in particular how an ASWF can exist even when there is a fundamental welfare preference ordering $\tilde{R} \in \mathcal{R}(X \times \Theta)$; as Arrow (1963) explains, he originally wished to avoid interpersonal comparisons, and so an ASWF ignores all the information which $\tilde{\rho}(e)$ provides except the derived individual orderings $R_i(\theta_i, e)$ $(i \in M)$ on X (cf. Sen, 1977). In this sense, any ASWF is independent of all interpersonal comparisons, relevant or not.

4. Arrow's Independence of Irrelevant Personal Comparisons

The condition which Arrow called "independence of irrelevant alternatives" (IIA) restricts even further the dependence of $R(\theta^M, e)$ upon θ^M, and makes social choice independent of irrelevant personal comparisons. To express this and related conditions formally, it is convenient to introduce some further notation. First, given any binary relation Q on a set A, and given any subset B of A, let Q_B denote the restriction of Q to B, which is defined for all pairs $a, b \in A$ by

$$a \; Q_B \; b \iff a, b \in B \text{ and } a \; Q \; b.$$

Next, define the equivalence relation $=_B$ between the two relations Q, Q' on A by

$$Q =_B Q' \iff Q_B = Q'_B$$

whenever B is a subset of A. One can read $Q =_B Q'$ as, "Q is equal on B to Q'."

Specifically, suppose that Z is any subset of X; then Z is a (possible) set of *relevant alternatives*. Given Z, all other members of X are *irrelevant alternatives*. According to IIA, when Z is the set of "relevant alternatives", two societies $\theta^M, \eta^M \in \Theta^M$ are regarded as equivalent if the individuals' welfare orderings among the members of Z are identical. According to the IIPC condition to be defined below, variations in the ethic e which leave these welfare orderings unchanged are also irrelevant to the social welfare orderings of all pairs in Z. Thus only personal welfare comparisons concerning pairs within Z are relevant; all other personal comparisons (and all interpersonal comparisons) are deemed irrelevant.

Formally, the ASWF $f : \mathcal{R}^M(X) \to \mathcal{R}(X)$ is said to satisfy *independence of irrelevant personal comparisons* (IIPC) provided that, for all pairs of societies $\theta^M, \eta^M \in \Theta^M$, all pairs of ethics $e, e' \in E$, and all subsets Z of X, one has

$$R_i(\theta_i, e) =_Z R_i(\eta_i, e') \quad (\text{all } i \in M) \Longrightarrow R(\theta^M, e) =_Z R(\eta^M, e').$$

So indeed individuals' personal comparisons of relevant alternatives (in Z) suffice to determine the social ordering on Z, no matter what the ethic may be. And the social ordering on Z is unaffected by changes either in personal characteristics or in the ethic, as long as the individual welfare orderings of the relevant alternatives in Z are unaffected.

Once again, the following commutative function diagram may help to explain this IIPC condition. It presumes the existence of an ASWF $f : \mathcal{R}^M(X) \to \mathcal{R}(X)$. The condition IIPC requires that, for every subset $Z \subset X$, there must be a corresponding *restricted* ASWF $f_Z : \mathcal{R}^M(Z) \to \mathcal{R}(Z)$ relating the domains $\mathcal{R}^M(Z)$ and $\mathcal{R}(Z)$ consisting of the restrictions to Z of possible individual preference profiles and possible social orderings respectively.

$$\begin{array}{ccc} \mathcal{R}^M(X) & \xrightarrow{R_Z^M} & \mathcal{R}^M(Z) \\ \Big\downarrow f & & \Big\downarrow f_Z \\ \mathcal{R}(X) & \xrightarrow{R_Z} & \mathcal{R}(Z) \end{array}$$

Here, of course, R_Z is used to denote the mapping whose value is the restriction to the set Z of the ordering $R \in \mathcal{R}(X)$, and $R_Z^M = \langle R_{iZ} \rangle_{i \in M}$ is used to denote the mapping whose value is the profile of the restrictions R_{iZ} of the individual welfare orderings R_i.

Say that the ASWF $f : \mathcal{R}^M(X) \to \mathcal{R}(X)$ satisfies the *unrestricted domain* condition (U) provided that, for each ethic $e \in E$, each individual $i \in M$, and every logically possible preference ordering $\bar{R} \in \mathcal{R}(X)$, there exists at least one personal characteristic $\theta_i \in \Theta$ such that $R_i(\theta_i, e) = \bar{R}$. Then we have:

THEOREM (ARROW'S "IMPOSSIBILITY" THEOREM). *The three conditions (U), (IIPC) and (P) together imply the existence of a dictator d, whose identity is independent of e, such that for every society θ^M, every ethic e, and every pair of social states $x, y \in X$, one has*

$$x \; P_d(\theta_d, e) \; y \Longrightarrow x \; P(\theta^M, e) \; y.$$

The only novel feature of the above definition and result is the introduction of variations in the ethic e which are separate from variations in "individual values." This extra feature actually requires that the above version of Arrow's theorem be given a slightly new proof, since the standard proof only shows the existence of a dictator $d(e)$ whose identity depends on e. Nevertheless, it is trivial to combine (IIPC), as enunciated above, with the unrestricted domain assumption (U) in order to show that d must in fact be independent of e.

Note how, in the absence of the Pareto criterion, the results of Hansson (1969) and Wilson (1972) imply that, when both (U) and (IIPC) are satisfied, then there are three possibilities: either (i) there is a dictator as above; or (ii) there is an "anti-dictator"; or (iii) there is universal social indifference between all pairs of social states in X.

5. Independence of Irrelevant Interpersonal Comparisons

As remarked above, and as Arrow himself fully intended, of course, the construction of an ASWF specifically excludes interpersonal comparisons. This is because preferences regarding variations in θ are entirely ignored when the society θ^M is fixed. Moreover, when IIPC is imposed for any subset Z of X, only individuals' personal comparisons regarding pairs of alternatives in Z are treated as relevant.

It is important to realize that IIPC forces us to ignore many preferences for relevant personal consequences, even when the society θ^M is fixed. To support this claim, consider first the set

$$\Theta(\theta^M) := \{\, \eta \in \Theta \mid \exists i \in M :\ \theta_i = \eta \,\}$$

of all the personal characteristics that some individual in the fixed society θ^M possesses. Let Z be the set of relevant social states. Then IIPC treats as relevant precisely the preferences between those pairs $(x,\theta),(y,\eta) \in Z \times \Theta(\theta^M)$ with the property that $\theta = \eta$ — i.e., all personal comparisons. So *all* the personal consequences in the set $Z \times \Theta(\theta^M)$ are relevant for *some* comparisons. But IIPC specifically excludes interpersonal comparisons of pairs with different characteristics θ_i, θ_j for different individuals $i, j \in M$. Independence of irrelevant interpersonal comparisons (IIIC), on the other hand, will treat the entire restricted fundamental interpersonal preference relation $\tilde{R}_{Z\times\Theta(\theta^M)}$ as relevant. Indeed, it will say that the social ordering of relevant social consequences depends only on the ethical views which govern the fundamental ordering of the corresponding relevant personal consequences.

First, corresponding to any subset Y of "relevant" social consequences in $X \times \Theta^M$, define each individual i's corresponding set of "relevant personal consequences" as

$$Y_i := \{\, (x,\eta) \in X \times \Theta \mid \exists (x,\theta^M) \in Y :\ \theta_i = \eta \,\}.$$

Then define the set of all relevant personal consequences as

$$\tilde{Y} := \bigcup\nolimits_{i\in M} Y_i = \{\, (x,\eta) \in X \times \Theta \mid \exists (x,\theta^M) \in Y :\ \eta \in \Theta(\theta^M) \,\}.$$

Now say that *independence of irrelevant interpersonal comparisons* (IIIC) is satisfied provided that, for every pair of ethics e, e' and every subset Y of relevant social consequences in $X \times \Theta^M$, one has

$$\tilde{\rho}(e) =_{\tilde{Y}} \tilde{\rho}(e') \Longrightarrow \rho(e) =_Y \rho(e').$$

Thus, as the ethic varies, the ordering of social consequences depends only on the ordering of relevant personal consequences. Notice how important it is to let the ethic vary; if $e = e'$ then both the hypothesis and the implication above become tautologies and so IIIC would have no force whatsoever if ethical views were fixed.

The condition IIIC requires that, for every $Y \subset X \times \Theta^M$, the following function diagram must commute:

$$\begin{array}{ccc} \mathcal{R}(X \times \Theta) & \xrightarrow{\tilde{R}_{\tilde{Y}}} & \mathcal{R}(\tilde{Y}) \\ \Big\downarrow \phi & & \Big\downarrow \phi_Y \\ \mathcal{R}(X \times \Theta^M) & \xrightarrow{R_Y} & \mathcal{R}(Y) \end{array}$$

Indeed, this diagram shows clearly how IIIC is precisely that property of an SWF $\phi : \mathcal{R}(X \times \Theta) \to \mathcal{R}(X \times \Theta^M)$ which corresponds to property IIPC of an ASWF.

Take the special case when the society θ^M is fixed, so that the set of relevant social consequences is $Y := Z \times \{\theta^M\}$ for some subset Z of X, and the set of relevant personal consequences is $\tilde{Y} := Z \times \Theta(\theta^M)$. Then IIIC implies that

$$\tilde{\rho}(e) =_{Z \times \Theta(\theta^M)} \tilde{\rho}(e') \Longrightarrow \rho(e) =_{Z \times \{\theta^M\}} \rho(e').$$

From this and the earlier definition of the orderings $R(\theta^M, e), R(\theta^M, e') \in \mathcal{R}(X)$, it follows that IIIC reduces to the following condition, called IIIC(θ^M):

$$\tilde{\rho}(e) =_{Z \times \Theta(\theta^M)} \tilde{\rho}(e') \Longrightarrow R(\theta^M, e) =_Z R(\theta^M, e').$$

6. Generalized Social Welfare Functions

So far, personal consequences have been defined as pairs $(x, \theta) \in X \times \Theta$. An equivalent description for any *given* society $\theta^M \in \Theta^M$ is simply the set $X \times M$ of pairs (x, i), because then any θ^M and (x, i) together determine the personal consequence (x, θ_i). This idea relates to the concept of a *generalized social welfare function* (GSWF), defined as a mapping $g : \mathcal{R}(X \times M) \to \mathcal{R}(X)$, as in Hammond (1976) and Roberts (1980a).

Given the society θ^M, the ethical views e, and the resulting fundamental ordering $\tilde{\rho}(e)$ on $X \times \Theta$, define the induced *interpersonal ordering* $\hat{R}(\theta^M, e)$ on $X \times M$ by

$$(x, i)\ \hat{R}(\theta^M, e)\ (y, j) \iff (x, \theta_i)\ \tilde{\rho}(e)\ (y, \theta_j).$$

Then the social welfare ordering $\rho(e)$ on X corresponds to such a GSWF if and only if, for all pairs of ethics $e, e' \in E$ and all pairs of societies $\theta^M, \eta^M \in \Theta^M$, one has

$$\hat{R}(\theta^M, e) =_{X \times M} \hat{R}(\eta^M, e') \Longrightarrow R(\theta^M, e) =_X R(\eta^M, e').$$

When X is replaced by any subset Z of X in the above implication, one has a GSWF which satisfies a condition which was called "independence of irrelevant alternatives" in Hammond (1976); it is a significant weakening of IIPC, however, and so I now prefer to say that *generalized independence of irrelevant alternatives* (GIIA) is satisfied provided that, for all $e, e' \in E$, all $\theta^M, \eta^M \in \Theta^M$, and all subsets Z of X, one has

$$\hat{R}(\theta^M, e) =_{Z \times M} \hat{R}(\eta^M, e') \Longrightarrow R(\theta^M, e) =_Z R(\eta^M, e').$$

This implies the existence of a restricted GSWF $g_Z : \mathcal{R}(Z \times M) \to \mathcal{R}(Z)$. Both the existence of a GSWF and the condition GIIA are illustrated by the following commutative diagram, in which some obvious notation has been introduced in order to describe certain natural mappings:

$$\begin{array}{ccccc}
\mathcal{R}(X\times\Theta) & \xrightarrow{\hat{R}(\theta^M,e)} & \mathcal{R}(X\times M) & \xrightarrow{R_{Z\times M}} & \mathcal{R}(Z\times M)\\
\downarrow\phi & & \downarrow g & & \downarrow g_Z\\
\mathcal{R}(X\times\Theta^M) & \xrightarrow{R(\theta^M,e)} & \mathcal{R}(X) & \xrightarrow{R_Z} & \mathcal{R}(Z)
\end{array}$$

The following result shows how GIIA is a natural independence condition for GSWF's because it is almost equivalent to IIIC(θ^M) for the corresponding SWF $\phi:\mathcal{R}(X\times\Theta)\to\mathcal{R}(X\times\Theta^M)$.

THEOREM. *(1) GIIA implies IIIC(θ^M) for every θ^M. (2) Conversely, suppose that the domain E of possible ethics e places no restriction at all on the domain of possible fundamental orderings $\tilde{R}=\tilde{\rho}(e)$ on $X\times\Theta$, so that the range $\tilde{\rho}(E)$ of the mapping $\tilde{\rho}:E\to\mathcal{R}(X\times\Theta)$ must include the whole of $\mathcal{R}(X\times\Theta)$. In this case, if the social ordering $R=\rho(\theta^M,e)$ in each fixed society $\theta^M\in\Theta^M$ can be derived from the fundamental ordering $\tilde{R}=\tilde{\rho}(e)$ through a GSWF $g:\mathcal{R}(X\times M)\to\mathcal{R}(X)$, and if IIIC($\theta^M$) is satisfied, then so is GIIA.*

PROOF: (1) Suppose that $Z\subset X$, that $\theta^M\in\Theta^M$, and that $\tilde{\rho}(e)=_{Z\times\Theta(\theta^M)}\tilde{\rho}(e')$. Then, for all pairs $(x,i),(y,j)\in Z\times M$, one has

$$\begin{aligned}
&(x,i)\ \hat{R}(\theta^M,e)\ (y,j)\iff(x,\theta_i)\ \tilde{\rho}(e)\ (y,\theta_j)\\
\iff&(x,\theta_i)\ \tilde{\rho}(e')\ (y,\theta_j)\iff(x,i)\ \hat{R}(\theta^M,e')\ (y,j)
\end{aligned}$$

using the definitions of $\hat{R}(\theta^M,e),\hat{R}(\theta^M,e')$ and the above hypothesis. Therefore $\hat{R}(\theta^M,e)=_{Z\times M}\hat{R}(\theta^M,e')$. From GIIA it follows that $R(\theta^M,e)=_Z R(\theta^M,e')$. So IIIC(θ^M) has been verified.

(2) Conversely, suppose that $\hat{R}(\theta^M,e)=_{Z\times M}\hat{R}(\eta^M,e')$. Then, for all $x,y\in Z$ and all $i,j\in M$, one has

$$\begin{aligned}
&(x,\theta_i)\ \tilde{\rho}(e)\ (y,\theta_j)\iff(x,i)\ \hat{R}(\theta^M,e)\ (y,j)\\
\iff&(x,i)\ \hat{R}(\eta^M,e')\ (y,j)\iff(x,\eta_i)\ \tilde{\rho}(e')\ (y,\eta_j)
\end{aligned}$$

using the definitions of $\hat{R}(\theta^M,e),\hat{R}(\eta^M,e')$ and the above hypothesis.

By the unrestricted domain hypothesis that $\tilde{\rho}(E) = \mathcal{R}(X \times \Theta)$, one can now certainly find an ethic $\bar{e} \in E$ for which the associated fundamental ordering $\tilde{\rho}(\bar{e})$ on $X \times \Theta$ satisfies

$$(x, \theta_i)\ \tilde{\rho}(\bar{e})\ (y, \theta_j) \iff (x, \eta_i)\ \tilde{\rho}(e')\ (y, \eta_j)$$

for all $x, y \in X$ and all $i, j \in M$. But this construction implies that $\hat{R}(\theta^M, \bar{e}) = \hat{R}(\eta^M, e')$. Then, because of the hypothesis that a GSWF exists, it must be true that $R(\theta^M, \bar{e}) = R(\eta^M, e')$ and so that $R(\theta^M, \bar{e}) =_Z R(\eta^M, e')$ in particular.

In addition, the above construction also ensures that

$$(x, \theta_i)\ \tilde{\rho}(\bar{e})\ (y, \theta_j) \iff (x, \theta_i)\ \tilde{\rho}(e)\ (y, \theta_j)$$

for all $x, y \in Z$ and all $i, j \in M$, because both are satisfied if and only if $(x, \eta_i)\ \tilde{\rho}(e')\ (y, \eta_j)$. Therefore $\tilde{\rho}(\bar{e}) =_{Z \times \Theta(\theta^M)} \tilde{\rho}(e)$, and so IIIC$(\theta^M)$ implies that $R(\theta^M, \bar{e}) =_Z R(\theta^M, e)$.

The conclusions of the last two paragraphs imply that $R(\theta^M, e) =_Z R(\eta^M, e')$ whenever $\hat{R}(\theta^M, e) =_{Z \times M} \hat{R}(\eta^M, e')$, as required for GIIA to be valid. ∎

7. Random Consequences

Up to now attention has been restricted to preferences in the absence of risk. Vickrey (1945, 1960) and Harsanyi (1955) derived their particular form of utilitarianism from the need to consider risky consequences. And one obviously wants a theory of social choice that is applicable to decisions in the face of risk. Thus I shall consider the set $\Delta(X \times \Theta^M)$ of simple probability measures on $X \times \Theta^M$ — i.e., discrete probability distributions which each attach probability one to some finite subset of $X \times \Theta^M$. Considering instead the space $\mathcal{M}(X \times \Theta^M, \mathcal{A})$ of general probability measures on some σ-algebra $\mathcal{A}$ over $X \times \Theta^M$ would merely add technical complications.

The set $\Delta(X \times \Theta^M)$ is a mixture space, in the sense of Herstein and Milnor (1953). That is, any probability mixture of a finite set of elements of $\Delta(X \times \Theta^M)$ is itself a member of $\Delta(X \times \Theta^M)$. For every ethic $e \in E$, the social welfare ordering $\rho(e)$ is now defined on this mixture space, and the fundamental welfare ordering

$\tilde{\rho}(e)$ is defined on the mixture space $\Delta(X \times \Theta)$. The SWF becomes a mapping $\phi : \mathcal{R}(\Delta(X \times \Theta)) \to \mathcal{R}(\Delta(X \times \Theta^M))$, of course.

Without any further structure on preferences, there is little to add to the previous analysis — all that happens is that $X \times \Theta^M$, $X \times \Theta$ and their various subsets get replaced by (appropriate subsets of) $\Delta(X \times \Theta^M)$ or $\Delta(X \times \Theta)$. For example, given any simple probability measure $\lambda \in \Delta(X \times \Theta^M)$ and any individual $i \in M$, define $\lambda_i := \text{marg}_{X \times \Theta_i} \lambda \in \Delta(X \times \Theta)$ as the marginal probability distribution which is induced by λ on i's personal consequences (x, θ_i). Thus

$$\lambda_i(x, \theta_i) := \sum\nolimits_{\theta_{-i} \in \Theta_{-i}} \lambda(x, \theta_i, \theta_{-i})$$

where Θ_{-i} denotes the Cartesian product space $\prod_{j \in M \setminus \{i\}} \Theta_j$ of profiles of other individuals' characteristics, with typical member θ_{-i}. Then the Pareto condition (P) takes the form:

$$\lambda_i \ \tilde{P}(e) \ \mu_i \quad (\text{all } i \in M) \Longrightarrow \lambda \ P(e) \ \mu.$$

The three main independence conditions IIPC, IIIC and IIIC(θ^M) which have been introduced so far can now be fairly easily extended from subsets of X to subsets of $\Delta(X)$ in order to accommodate such random consequences. The three resulting *extended* independence conditions will be called EIIPC, EIIIC and EIIIC(θ^M) respectively. Their precise definitions are given in the appendix. In fact, though, IIIC(θ^M) — which turns out to be the weakest of all the four conditions IIIC, IIIC(θ^M), EIIIC and EIIIC(θ^M) that admit interpersonal comparisons — will be strong enough to establish the impossibility theorem of Section 8. Accordingly the extended conditions EIIIC and EIIIC(θ^M) play no role at all in the following analysis. Nor does EIIPC.

8. Harsanyi's Expected Utilitarianism

Following Herstein and Milnor's (1953) familiar axioms, I shall now assume that, for each ethic $e \in E$, the corresponding social welfare ordering $R = \rho(e)$ on $\Delta(X \times \Theta^M)$ satisfies the two conditions:

(i) *Probability Independence*. If $\lambda, \mu, \nu \in \Delta(X \times \Theta^M)$ and $0 < \alpha \leq 1$, then

$$\alpha\,\lambda + (1-\alpha)\,\nu \; R \; \alpha\,\mu + (1-\alpha)\,\nu \iff \lambda \; R \; \mu.$$

(ii) *Weak Continuity*. For any $\lambda, \mu, \nu \in \Delta(X \times \Theta^M)$, the two sets

$$\{\,\alpha \in [0,1] \mid \alpha\,\lambda + (1-\alpha)\,\mu \; R \; \nu\,\} \quad \text{and} \quad \{\,\alpha \in [0,1] \mid \nu \; R \; \alpha\,\lambda + (1-\alpha)\,\mu\,\}$$

are always closed subsets of the real line interval $[0,1]$.

These two conditions imply that for every e there exists a unique cardinal equivalence class of *von Neumann-Morgenstern social welfare functions* (NMSWF's) $w(\cdot,\cdot;e) : X \times \Theta^M \to \Re$ such that, for every pair $\lambda, \mu \in \Delta(X \times \Theta^M)$, one has

$$\lambda \; \rho(e) \; \mu \iff \mathbb{E}_\lambda\, w(x, \theta^M; e) \geq \mathbb{E}_\mu\, w(x, \theta^M; e).$$

Here $\mathbb{E}_\lambda$ denotes expected value with respect to λ, and $\mathbb{E}_\mu$ similarly.

Next, the space of personal consequences is similarly extended to the mixture space $\Delta(X \times \Theta)$ of simple probability measures on $X \times \Theta$, and the fundamental welfare ordering $\tilde{R} = \tilde{\rho}(e)$ on this space is also assumed to satisfy the Herstein and Milnor axioms for every $e \in E$. So there is a cardinal equivalence class of *fundamental* von Neumann-Morgenstern welfare functions (NMWF's) $v(\cdot,\cdot;e) : X \times \Theta \to \Re$ such that, for every pair $\lambda, \mu \in \Delta(X \times \Theta)$, one has

$$\lambda \; \tilde{\rho}(e) \; \mu \iff \mathbb{E}_\lambda\, v(x, \theta; e) \geq \mathbb{E}_\mu\, v(x, \theta; e).$$

For every e, the NMWF v is unique up to a cardinal equivalence class of functions which result from increasing linear transformations of the form

$$\tilde{v}(x, \theta; e) \equiv \alpha(e) + \beta(e)\, v(x, \theta; e)$$

where both $\alpha(e)$ and $\beta(e)$ are independent of θ, with $\beta(e)$ positive. Thus $v(\cdot, \cdot; e)$ embodies interpersonal comparisons of both utility levels and utility units — what Roberts (1980b) calls "cardinal full comparability."

Assume that the Pareto condition (P) of Section 6 is supplemented by the Pareto indifference condition (P^0)

$$\lambda_i \; \tilde{I} \; \mu_i \quad (\text{all} \;\; i \in M) \Longrightarrow \lambda \; I \; \mu,$$

where $\tilde{I}$ and I are the indifference relations generated respectively by the fundamental and social welfare orderings $\tilde{\rho}(e)$ and $\rho(e)$. Then, as Harsanyi (1955) originally argued, for every ethic $e \in E$, the corresponding NMSWF $w(\cdot, \cdot; e)$ — whose expected value represents the social ordering $\rho(e)$ on the set $\Delta(X \times \Theta^M)$ — must be a non-negatively weighted sum

$$w(x, \theta^M; e) \equiv \sum\nolimits_{i \in M} \omega_i(e) \, v(x, \theta_i; e)$$

of the values $v(x, \theta_i; e)$ of the fundamental NMWF whose expectation represents the fundamental welfare ordering $\tilde{\rho}(e) \in \mathcal{R}(\Delta(X \times \Theta))$. Actually, Harsanyi did not consider variations in θ^M, but this makes no difference to the conclusion — see Hammond (1987). Notice too that the "welfare weights" $\omega_i(e)$ are allowed to reflect varying ethical views, as is the fundamental NMWF $v(\cdot, \cdot; e)$. Also the Pareto condition (P) implies that $\sum_{i \in M} \omega_i(e) > 0$.

9. Another Impossibility Theorem

Suppose that the welfare weighted sum $\sum_{i \in M} \omega_i(e) \, v(x, \theta_i; e)$ does not give rise to a dictatorship. In other words, suppose that there is no fixed individual $d \in M$, independent of e, for whom $\omega_i(e) > 0$ only if $i = d$. Then there must exist at least two different individuals $j, k \in M$ and two (possibly coincident) ethics $e, e' \in E$ for which $\omega_j(e)$ and $\omega_k(e')$ are both positive.

Consider now what happens when Z consists of just the pair x, y of sure social states, and when the society θ^M is also certain. Say that the domain is

sufficiently rich provided that there exist two more ethics $e_1, e_2 \in E$, three personal characteristics $\bar{\theta}, \theta', \theta'' \in \Theta$, and a large real number $K > 1$, which together satisfy:

(i) $\omega_j(e_1) > 0$ and $\omega_k(e_2) > 0$;

(ii) for both the fundamental welfare orderings $\tilde{R} = \tilde{\rho}(e_1)$ and $\tilde{R} = \tilde{\rho}(e_2)$, one has

$$(x, \bar{\theta})\ \tilde{I}\ (y, \bar{\theta})\ \tilde{P}\ (x, \theta')\ \tilde{P}\ (y, \theta')\ \tilde{P}\ (y, \theta'')\ \tilde{P}\ (x, \theta'');$$

(iii) for all $z \in \{ x, y \}$ and all $\theta \in \{ \bar{\theta}, \theta', \theta'' \}$ one has

$$v(z, \theta; e_2) \equiv \min \{ v(z, \theta; e_1) - v(y, \theta''; e_1),\ K[v(z, \theta; e_1) - v(y, \theta''; e_1)] \};$$

(iv) $\omega_j(e_1)\,[v(x, \theta'; e_1) - v(y, \theta'; e_1)] > \omega_k(e_1)\,[v(y, \theta''; e_1) - v(x, \theta''; e_1)]$;

(v) $K > \dfrac{\omega_j(e_2)\,[v(x, \theta'; e_2) - v(y, \theta'; e_2)]}{\omega_k(e_2)\,[v(y, \theta''; e_1) - v(x, \theta''; e_1)]}$.

The idea here is that the ethic should be able to vary sufficiently from both e and e' in order to satisfy (ii), (iii), (iv) and (v), while the welfare weights still satisfy (i). In particular, this requires some independence in the possible variations of the welfare weights and of the fundamental NMWF which is defined on $X \times \Theta$.

Let θ^M be the society with $\theta_j = \theta'$, $\theta_k = \theta''$, and $\theta_i = \bar{\theta}$ otherwise. By (ii) one has $v(x, \theta'; e_1) > v(y, \theta'; e_1)] > v(y, \theta''; e_1)$ and also $v(x, \theta''; e_1) < v(y, \theta''; e_1)$. Therefore (iii) implies that, for $z \in \{ x, y \}$, one has

$$\begin{aligned} v(z, \theta'; e_2) &= v(z, \theta'; e_1) - v(y, \theta''; e_1); \\ v(z, \theta''; e_2) &= K\,[v(z, \theta''; e_1) - v(y, \theta''; e_1)]. \end{aligned}$$

Thus

$$\begin{aligned} v(x, \theta'; e_2) - v(y, \theta'; e_2) &= v(x, \theta'; e_1) - v(y, \theta'; e_1) > 0; \\ v(x, \theta''; e_2) - v(y, \theta''; e_2) &= K\,[v(x, \theta''; e_1) - v(y, \theta''; e_1)] < 0. \end{aligned}$$

Then, however, (v) implies that

$$\begin{aligned} \omega_j(e_2)\,[v(x, \theta'; e_2) - v(y, \theta'; e_2)] &= \omega_j(e_2)\,[v(x, \theta'; e_1) - v(y, \theta'; e_1)] \\ &< \omega_k(e_2)\,K\,[v(y, \theta''; e_1) - v(x, \theta''; e_1)] \\ &= \omega_k(e_2)\,[v(y, \theta''; e_2) - v(x, \theta''; e_2)]. \end{aligned}$$

Next, note how (ii) implies that, for $q = 1$ and 2,

$$w(x, \theta^M; e_q) - w(y, \theta^M; e_q) = \omega_j(e_q)\,[v(x, \theta'; e_q) - v(y, \theta'; e_q)] \\ + \omega_k(e_q)\,[v(x, \theta''; e_q) - v(y, \theta''; e_q)]$$

because all the other terms $\sum_{i \in M \setminus \{\, j,k \,\}} \omega_i(e_q)\,[v(x, \bar{\theta}; e_q) - v(y, \bar{\theta}; e_q)]$ of the sum are zero. Now, when $q = 1$ this expression is positive, by (iv), but when $q = 2$ it is negative, as an implication of (v). So $x\ P(\theta^M, e_1)\ y$ and $y\ P(\theta^M, e_2)\ x$. This violates the condition IIIC(θ^M) because, when $Y = \{\, x, y \,\} \times \{\theta^M\}$ and therefore $\tilde{Y} = \{\, x, y \,\} \times \{\, \bar{\theta}, \theta', \theta'' \,\}$, it follows that $R(\theta^M, e_1) \neq_Y R(\theta^M, e_2)$ even though (ii) implies that $\tilde{\rho}(e_1) =_{\tilde{Y}} \tilde{\rho}(e_2)$.

The above argument confirms that, at least when the domain E of possible ethics allows sufficient independent variations of the welfare weights $\omega_i(e)$ ($i \in M$) and of the fundamental NMWF $v(\cdot, \cdot; e)$, then nonlinear transformations of the form which appears in (iii) above will only preserve the ordering generated by $\mathbb{E} \sum_{i \in M} \omega_i(e)\, v(x, \theta_i; e)$ when there is a dictator. Thus:

PROPOSITION. *Suppose that IIIC(θ^M) is satisfied, for every society $\theta^M \in \Theta^M$, and that the set of ethics E gives rise to a sufficiently rich domain of possible fundamental interpersonal orderings $\tilde{\rho}(e) \in \mathcal{R}(\Delta(X \times \Theta))$. Suppose too that the expected utility hypotheses and the two Pareto conditions (P) and (P^0) are satisfied. Then there exists a dictator d whose identity is independent of e.*

This result says that only a dictatorship succeeds in making the following function diagram commute for every society $\theta^M \in \Theta^M$ and subset $Z \subset X$. The diagram uses the notation $\mathcal{R}^*(\Delta(S))$ to denote the set of all preference orderings on $\Delta(S)$ which can be represented by the expected value of any member of a unique cardinal equivalence class of von Neumann-Morgenstern utility functions. In addition, $[w]$ and $[v]$ each denote such an equivalence class representing respectively the social welfare ordering $\rho(e) \in \mathcal{R}^*(\Delta(X \times \Theta^M))$ and the fundamental welfare ordering $\tilde{\rho}(e) \in \mathcal{R}^*(\Delta(X \times \Theta))$.

$$\begin{array}{ccccc}
\mathcal{R}^*(\Delta(X\times\Theta)) & \xrightarrow{\iota} & \mathcal{R}^*(\Delta(X\times\Theta)) & \xrightarrow{[v]_{Z\times\Theta(\theta^M)}} & \mathcal{R}(Z\times\Theta(\theta^M)) \\
\uparrow{\scriptstyle [v]} & & \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \phi_{Z\times\{\theta^M\}}} \\
E & \xrightarrow{[w]} & \mathcal{R}^*(\Delta(X\times\Theta^M)) & \xrightarrow{[w]_{Z\times\{\theta^M\}}} & \mathcal{R}(Z\times\{\theta^M\})
\end{array}$$

10. Independence of Irrelevant Interpersonal Comparisons of Mixtures

The following weakening of IIIC proves useful in avoiding the dictatorship to which it leads when consequences are risky. Say that *independence of irrelevant interpersonal comparisons of mixtures* (IIICM) is satisfied provided that, for every subset Y of social consequences $(x,\theta^M)\in X\times\Theta^M$ with an assocated set $\tilde{Y}$ of personal consequences $(x,\theta)\in X\times\Theta$, and for each pair of ethics $e,e'\in E$, one has

$$\tilde{\rho}(e)=_{\Delta(\tilde{Y})}\tilde{\rho}(e')\Longrightarrow\rho(e)=_Y\tilde{\rho}(e').$$

Like our previous independence conditions, this one can also be illustrated with a commutative function diagram, as follows:

$$\begin{array}{ccccc}
\mathcal{R}^*(\Delta(X\times\Theta)) & \xrightarrow{\iota} & \mathcal{R}^*(\Delta(X\times\Theta)) & \xrightarrow{[v]_{\tilde{Y}}} & \mathcal{R}^*(\Delta(\tilde{Y})) \\
\uparrow{\scriptstyle [v]} & & \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \phi_Y} \\
E & \xrightarrow{[w]} & \mathcal{R}^*(\Delta(X\times\Theta^M)) & \xrightarrow{[w]_Y} & \mathcal{R}(Y)
\end{array}$$

Thus preferences for risky relevant personal consequences count even though Y contains only sure social consequences. This may seem peculiar, but appears to be the only way of extending rational Paretian social choice to cover risky consequences without having a dictatorship.

Notice that IIICM is indeed satisfied when $\rho(e)$ is represented on $X\times\Theta^M$ by

$$w(x,\theta^M;e)\equiv\sum\nolimits_{i\in M}\omega_i\,v(x,\theta_i;e)$$

as in Section 8, except that now ω_i is independent of e. For when $\tilde{Y}$ has at least two members and $\tilde{\rho}(e)=_{\Delta(\tilde{Y})}\tilde{\rho}(e')$, it follows that the two NMWF's $v(x,\theta;e)$ and

$v(x, \theta; e')$ are cardinally equivalent on $\tilde{Y}$. That is, there exists an additive constant α and a positive multiplicative constant β such that

$$v(x, \theta; e') \equiv \alpha + \beta\, v(x, \theta; e)$$

for all $(x, \theta) \in \tilde{Y}$. Thus

$$\begin{aligned} w(x, \theta^M; e') &\equiv \sum\nolimits_{i \in M} \omega_i\, v(x, \theta_i; e') \equiv \alpha \sum\nolimits_{i \in M} \omega_i + \beta \sum\nolimits_{i \in M} \omega_i\, v(x, \theta_i; e) \\ &\equiv \alpha \sum\nolimits_{i \in M} \omega_i + \beta\, w(x, \theta^M; e) \end{aligned}$$

for all $(x, \theta^M) \in Y$. Therefore $w(\cdot, \cdot; e)$ and $w(\cdot, \cdot; e')$ are cardinally equivalent on Y, and so not only represent identical preferences on this set, but also have expected values representing identical preferences on the mixture set $\Delta(Y)$. A similar argument then shows that when IIICM is extended to random consequences, as IIIC was in Section 7, the corresponding extended condition EIIICM is also satisfied.

Of course, one can also insist upon anonymity, so that $\omega_i = \omega$ (all $i \in M$) where $\omega > 0$. Then, after a harmless normalization, the expected value of the sum

$$w(x, \theta^M; e) \equiv \sum\nolimits_{i \in M} v(x, \theta_i; e)$$

represents $\rho(e)$ on $\Delta(X \times \Theta^M)$.

11. Independence of Irrelevant Personal Comparisons of Mixtures

So IIICM does allow an escape from the dictatorship of Arrow's impossibility theorem to a form of expected utilitarianism along the lines pioneered by Vickrey and Harsanyi. IIICM differs from IIPC in allowing both interpersonal comparisons and preferences for risky consequences to become relevant. Section 9 showed how just introducing interpersonal comparability through the IIIC condition is not enough by itself to avoid dictatorship — preferences for risky consequences must be considered as well. The question still remains whether allowing just dependence on preferences for risky consequences, without interpersonal comparisons, could

do the same. In fact it does not, which should come as no suprise in view of Sen's (1970a) version of Arrow's theorem for cardinal utilities (see also Osborne, 1976).

Actually, an even stronger statement is possible. Even when interpersonal comparisons are not relevant, say that "independence of irrelevant personal comparisons of mixtures" (or IIPCM) is satisfied provided that, for every set $Z \subset X$ of sure social states, every pair of societies $\theta^M, \eta^M \in \theta^M$, and every pair of ethics $e, e' \in E$, it is true that

$$R_i(\theta_i, e) =_{\Delta(Z)} R_i(\eta_i, e') \text{ (all } i \in M) \Longrightarrow R(\theta^M, e) =_Z R(\eta^M, e').$$

Thus the hypothesis of IIPCM is stronger than that of IIPC because it requires the preference profiles to be identical for all lotteries in the mixture set $\Delta(Z)$, rather than just for all sure social states in Z. Accordingly, IIPCM is superficially a less restrictive condition than IIPC. Yet in fact it turns out to be equivalent when preferences for risky consequences can be represented by the expected value of a unique cardinal equivalence class of utility functions.

To show this equivalence, suppose that IIPCM is satisfied, and then consider any pair set $Z = \{ x, y \} \subset X$ of sure social states. Now, if $x \; P_i(\theta_i, e) \; y$ for any individual $i \in M$, characteristic $\theta_i \in \Theta$, and ethic $e \in E$, it must be true that, for any pair of lotteries $\lambda\, x + (1-\lambda)\, y$ and $\mu\, x + (1-\mu)\, y$ in $\Delta(Z)$ (where $\lambda, \mu \in [0, 1]$), one has

$$\lambda\, x + (1-\lambda)\, y \; P_i(\theta_i, e) \; \mu\, x + (1-\mu)\, y \iff \lambda > \mu.$$

On the other hand, if $y \; P_i(\theta_i, e) \; x$, then it must be true that

$$\lambda\, x + (1-\lambda)\, y \; P_i(\theta_i, e) \; \mu\, x + (1-\mu)\, y \iff \lambda < \mu$$

for any such pair. While if $x \; I_i(\theta_i, e) \; y$, then for all such pairs it must be true that $\lambda\, x + (1-\lambda)\, y \; I_i(\theta_i, e) \; \mu\, x + (1-\mu)\, y$, no matter what values λ and μ may have in the interval $[0, 1]$. Thus the preference ordering $R_i(\theta_i, e)$ on the mixture set $\Delta(Z)$ is completely determined by preferences over just the two extreme points in the pair set Z.

Given the two profiles of characteristics $\theta^M, \eta^M \in \theta^M$, and the pair of ethics $e, e' \in E$, suppose it is true that $R_i(\theta_i, e) =_Z R_i(\eta_i, e')$ for all $i \in M$, as in the hypothesis of IIPC for the pair set Z. Then the argument of the previous paragraph shows that $R_i(\theta_i, e) =_{\Delta(Z)} R_i(\eta_i, e')$ for all $i \in M$, which is the hypothesis of IIPCM for this same pair set Z. So IIPCM implies that $R(\theta^M, e) =_Z R(\eta^M, e')$. Thus, for an arbitrary pair set $Z \subset X$, it has been shown that

$$R_i(\theta_i, e) =_Z R_i(\eta_i, e') \text{ (all } i \in M) \Longrightarrow R(\theta^M, e) =_Z R(\eta^M, e'),$$

or that IIPC is true for this pair set.

To complete the argument requires noting that, if IIPC holds for every pair set, then it must be true for every subset $Z \subset X$. But this is well known and easy to show (see, for instance, d'Aspremont and Gevers, 1977).

Merely cardinalizing utilities, therefore, does nothing at all to weaken the IIPC condition. Interpersonal comparisons must also be allowed, as in the IIICM condition of Section 10.

12. Concluding Summary

Arrow's "independence of irrelevant alternatives" (IIA) condition — or, more precisely, the new but closely related "independence of irrelevant personal comparisons" (IIPC) condition of Section 4, which was designed to allow variable ethical views — has been modified in order to allow interpersonal comparisons to affect the social ordering. Section 5 presented a weakened "independence of irrelevant interpersonal comparisons" (IIIC) condition which recognized that both IIA and IIPC exclude interpersonal comparisons of relevant personal consequences.

In Section 9 IIIC was shown to lead rather easily to a dictatorship when random consequences are considered. Accordingly a specific weakening to allow consideration of preferences for random personal consequences was introduced in Section 10. This new "independence of irrelevant interpersonal comparisons of mixtures" (IIICM) condition was then shown to be consistent with Harsanyi's fundamental utilitarianism (cf. Hammond, 1987). Later, in Section 11, it was

shown how the corresponding weakening for IIPC (without interpersonal comparisons being allowed to count) would make no difference. Indeed, in retrospect, this need to consider probability mixtures should not be too surprising, since that is how utilities are cardinalized. On its own, IIIC allows only ordinal utilities, and so at most only ordinal level comparability of different individuals' utilities. IIICM, on the other hand, enables random social states to be ordered according to an expected utility maximizing criterion.

ACKNOWLEDGEMENTS

An earlier version of this paper, with the title ™Independence of Irrelevant Personal Consequences," was prepared with research support from the U.S. National Science Foundation under Grant No. SES 83±02460 to Stanford University, and presented at the 5th World Congress of the Econometric Society in Boston in August 1985. The current version is a minor revision of the paper which was presented to the Deutsche Forschungsgemeinschaft Conference on ™Distributive Justice" in Bonn-Bad Godesberg in August 1989. Besides the above institutions, I am also much indebted to Kenneth Arrow and to Alessandro Petretto for raising questions which this paper attempts to answer, and to Salvador Barberà, other participants in the Hoover Seminar on Collective Choice in 1985, and some anonymous referees for prompting great improvements both in the main concept and in the exposition.

APPENDIX. The Three Extended Independence Conditions

The following three definitions of the extended independence conditions EIIPC, EIIIC, and EIIIC(θ^M) were promised in Section 7 of the paper.

First, say that EIIPC (or *extended independence of irrelevant personal comparisons*) is satisfied provided that, for all subsets $Z \subset \Delta(X)$, all societies $\theta^M \in \Theta^M$, and all pairs of ethics $e, e' \in E$, one has

$$R_i(\theta_i, e) =_Z R_i(\eta_i, e') \quad (\text{all } i \in M) \Longrightarrow R(\theta^M, e) =_Z R(\eta^M, e').$$

Before the next two conditions EIIIC and EIIIC(θ^M) can be described properly, a little more notation is needed. Given any set $\Lambda \subset \Delta(X \times \Theta^M)$ of random

social consequences, define for each individual $i \in M$ the corresponding set

$$\Lambda_i := \{\, \lambda_i \in \Delta(X \times \Theta) \mid \exists \lambda \in \Lambda :\ \lambda_i = \mathrm{marg}_{X \times \Theta_i}\, \lambda \,\}$$

of marginal distributions on i's personal consequences. Then let $\tilde{\Lambda} := \cup_{i \in M} \Lambda_i$ be the set of marginal distributions over all individuals' personal consequences.

Now say that *extended independence of irrelevant interpersonal comparisons* (EIIIC) is satisfied provided that, for all such subsets $\Lambda \subset \Delta(X \times \Theta^M)$ and all pairs of ethics $e, e' \in E$, one has

$$\tilde{\rho}(e) =_{\tilde{\Lambda}} \tilde{\rho}(e') \Longrightarrow \rho(e) =_{\Lambda} \rho(e').$$

Finally, for any fixed society $\theta^M \in \Theta^M$, one obviously says that EIIIC(θ^M) is satisfied provided that

$$\tilde{\rho}(e) =_{\Xi \times \{\theta^M\}} \tilde{\rho}(e') \Longrightarrow R(\theta^M, e) =_{\Xi} R(\theta^M, e')$$

for all subsets $\Xi \subset \Delta(X)$ and all pairs of ethics $e, e' \in E$. Evidently EIIIC implies that EIIIC(θ^M) is true for all fixed societies $\theta^M \in \Theta^M$.

REFERENCES

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

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

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

C. d'Aspremont (1985), "Axioms for Social Welfare Orderings," ch. 1, pp. 19–76 of L. Hurwicz, D. Schmeidler and H. Sonnenschein (eds.), *Social Goals and Organization* (Cambridge: Cambridge University Press).

C. d'Aspremont and L. Gevers (1977), "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies*, 44: 199–209.

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

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

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

P.J. Hammond (1986), "Consequentialist Social Norms for Public Decisions," ch. 1, pp. 3–27 of W.P. Heller, R.M. Starr and D.A. Starrett (eds.) *Social Choice and Public Decision Making: Essays in Honor of Kenneth J. Arrow, Vol. I* (Cambridge: Cambridge University Press).

P.J. Hammond (1987), "On Reconciling Arrow's Theory of Social Choice with Harsanyi's Fundamental Utilitarianism," ch. 4, pp. 179–222 of G.R. Feiwel (ed.) *Arrow and the Foundations of the Theory of Economic Policy* (London: Macmillan).

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

P.J. Hammond (1990), "Interpersonal Comparisons of Utility: Why and How They Are and Should Be Made," European University Institute, Working Paper No. ECO 90/3; to appear in J. Elster and J. Roemer (eds.), Proceedings of the Sloan Conference on "Interpersonal Comparability of Welfare" (New York: Cambridge University Press).

B. Hansson (1969), "Group Preferences," *Econometrica*, 37: 50–54.

J.C. Harsanyi (1953), "Cardinal Utility in Welfare Economics and the Theory of Risk-Taking," *Journal of Political Economy*, 61: 434–435.

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

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

J.C. Harsanyi (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations* (Cambridge: Cambridge University Press).

D.W. Haslett (1990), "What is Utility?" *Economics and Philosophy*, 6: 65–94.

I.N. Herstein and J. Milnor (1953), "An Axiomatic Approach to Measurable Utility," *Econometrica*, 21: 291–297.

C. Hildreth (1953), "Alternative Conditions for Social Orderings," *Econometrica*, 21: 81–94.

S.-C. Kolm (1972), *Justice et Equité* (Paris: Editions du Centre National de la Recherche Scientifique).

D.K. Osborne (1976), "Irrelevant Alternatives and Social Welfare," *Econometrica*, 44: 1001–1015.

J. Rawls (1959), "Justice as Fairness," *Philosophical Review*, 67: 164–94.

J. Rawls (1971) *A Theory of Justice* (Cambridge, Mass.: Harvard University Press).

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

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

A.K. Sen (1970a), *Collective Choice and Social Welfare* (San Francisco: Holden Day and London: Oliver and Boyd).

A.K. Sen (1970b), "Interpersonal Aggregation and Partial Comparability," *Econometrica*, 38: 393–409; reprinted in Sen (1982).

A.K. Sen (1977), "On Weights and Measures: Informational Constraints in Social Welfare Analysis," *Econometrica*, 45: 1539–1572; reprinted in Sen (1982).

A.K. Sen (1982), *Choice, Welfare and Measurement* (Oxford: Basil Blackwell).

S. Strasnick (1977), "Ordinality and the Spirit of the Justified Dictator," *Social Research*, 44: 668–690.

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

J. Tinbergen (1957), "Welfare Economics and Income Distribution," *American Economic Review (Papers and Proceedings)*, 47: 490–503.

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

W.S. Vickrey (1960), "Utility, Strategy and Social Decision Rules," *Quarterly Journal of Economics*, 74: 507–535.

R. Wilson (1972), "Social Choice Theory without the Pareto Principle," *Journal of Economic Theory*, 5: 478–487.