CENTER FOR OPERATIONS RESEARCH & ECONOMETRICS

UNIVERSITE CATHOLIQUE DE LOUVAIN

Discussion Paper

8013

ON JOHN HARSANYI'S DEFENCES OF UTILITARIANISM

by

Charles Blackorby, David Donaldson

and

John A. Weymark

April 1980

CORE, 34 VOIE DU ROMAN PAYS, 1348 LOUVAIN-LA-NEUVE, BELGIUM, 010/41.81.81

ON JOHN HARSANYI'S DEFENCES OF UTILITARIANISM

by

Charles Blackorby*

David Donaldson*

and

John A. Weymark**

Abstract

Harsanyi has advanced three arguments in defence of utilitarianism. One defence is based on the separability of the social-welfare function with respect to individual utilities, the second defence requires social and individual preferences to satisfy the expected utility hypothesis, while the third defence considers the choices that would be made if there were an equal chance of obtaining any position in society. We develop a framework in which these three defences may be compared. We demonstrate that with complete welfare information each of these proposals results in a different ordering of the social alternatives, and that none of these is classical utilitarianism.

* University of British Columbia

** CORE and Duke University

Weymark's stay at CORE has been made possible by the financial support of the Institut des Sciences Economiques of the Université Catholique de Louvain. We would like to thank Claude d'Aspremont for his detailed comments on a preliminary draft.

1. INTRODUCTION

There have been many attempts to justify particular social-evaluation rules. Of these rules, Rawls' difference principle (maximin or its lexicographic counterpart) and utilitarianism have received the most attention. John Harsanyi has attempted three important arguments for utilitarianism [1953, 1955, 1976, 1977]. In this paper, we develop a framework in which these three defences may be compared. We are able to produce three specific models from our general framework by employing different sets of assumptions.

In spite of a quarter century of discussion of Harsanyi's work, there has yet to emerge a consensus about the precise specification of each of Harsanyi's models. Thus, while each of our models is similar to Harsanyi's own, we do not claim to have replicated his models exactly. In deriving each of our models from our general framework, we try to make clear exactly which assumptions are critical for each model. Further, we discuss the relationships among the models and their relationships with the utilitarian rule.

To define the utilitarian rule, we assume the existence of a "Bentham" (extended) utility function, B , whose image $B(x,i)$ is the utility that person i gets in state x . This function is assumed to be numerically significant, and allows complete interpersonal comparability. In this notation, the utilitarian rule ranks state x at least as good as state y if and only if $\sum_i B(x,i) \geqslant \sum_i B(y,i)$. Thus, the Bentham utility function provides an objective standard with which to rank alternative social states. The existence of objectively valid interpersonal comparisons of utility is not essential for our results; an alternative subjectivist approach is briefly considered in our concluding remarks.

In each model an agent is endowed with an extended utility function and in the first and second models an agent is also endowed with a social-evaluation functional. An *extended utility function* maps social stations - being person i in state x - into the real numbers. By considering the restriction of an extended utility function to situations involving only the evaluating agent, we obtain the ordinary notion of a utility function. A *social-evaluation functional* maps extended utility functions into orderings of social states. An extended utility function can be thought of as representing an agent's regular preferences while a social-evaluation functional represents his moral preferences.

We employ two types of restrictions on the social-evaluation functionals. The first restriction captures the notion that an agent may have difficulties in making precise interpersonal utility comparisons because of inadequate information. The set of extended utility functions is partitioned into information sets; members of the same set cannot be distinguished from one another using the available information. Our "information invariance" assumption requires social-evaluation functionals to map all elements of the same information set into the same social ordering.[1] For example, full numerical comparability means that each information set is a singleton and, hence, no restriction is placed on the social-evaluation functional. On the other hand, cardinal comparability means that each information set con-

[1] See Sen [1977b; 1979].

tains all those extended utility functions which are positive affine transformations of each other.

Our second restriction relates the actual extended utility function used to form the social evaluation to the Bentham utility function. We shall be assuming that each agent attempts to employ an extended utility function which incorporates objectively valid interpersonal utility comparisons. However, given the information available, it may not be possible to implement this objective completely. Consequently, we require each agent to choose an extended utility function from the same information set as the Bentham utility function. In the case of numerical comparability, this requires each agent to use B as his extended utility function while in the case of cardinal comparability this means that each agent's extended utility function can be written as a positive affine transformation of the Bentham utility function.

We assume that the social-evaluation functionals are of the welfarist type (satisfying Pareto indifference, independence, and unlimited domain). Because of this, we are able to replace each social-evaluation functional with a *social-welfare-function*, a real-valued function defined on the individuals' utility values alone.

The first model (Section 4) is based on a separability condition on the social-welfare function. It requires that, if a number of individuals are indifferent between two states, the social ordering will depend on the utilities of the remaining individuals, *independently* of the utility levels of the first group.

In the case of numerical comparability and the assumption of symmetry in the Bentham utilities, then, for the kth evaluator, x is at least as good as y if and only if $\sum_i \varphi^k(B(x,i)) \geqslant \sum_i \varphi^k(B(y,i))$ where φ^k is an arbitrary increasing and continuous function. (See the Corollary to Theorem 4). Thus, these orderings of the social alternatives are not the utilitarian rule, even if each individual has the same moral preferences ($\varphi^k = \varphi$ for all k). In order to generate the utilitarian result, we must assume that each social-evaluation functional satisfies information invariance with cardinal comparability. That is, if individual utilities are subjected to a common positive affine transformation, the social ordering is unchanged. This latter result is not very satisfying, since it depends critically on an informational restriction that must be the same across all (moral) agents.

The second model (Section 5) extends the discussion to uncertain events. The social-welfare functions and the individual utility functions satisfy the expected utility hypothesis, and we assume that there is agreement about the (subjective) probabilities for each event. If the principle of acceptance is applied to only events which occur for certain, then with numerical comparability and symmetry, agent k ranks state x for certain at least as good as state y for certain if and only if $\sum_i g^k(B(x,i)) \geqslant \sum_i g^k(B(y,i))$ where g^k is an arbitrary increasing and continuous function. The function g^k may be different from φ^k (above). Extending the application of the acceptance principle to uncertain events, $g^k = g$ for all k. Again, we do *not* obtain the utilitarian rule. Further, since there is no reason that $g = \varphi$, the results of the previous section and this one do not agree with each other. However, as in Section 4, if we relax numerical comparability to cardinal comparability, it is possible to obtain the utilitarian rule in this model as well.

The third model (Section 6) is somewhat different from the first two. An individual is deprived of the knowledge of who he is going to be in society. If he has an equal chance of being anyone in his society, we may interpret his ranking of society's alternatives as a sort of moral ranking (even though it may be based on self-interest). With numerical comparability, the ranking again takes on the same functional form as in the second model. Therefore, the ranking of social states is not the utilitarian ranking. However, as was the case with the earlier models, the utilitarian rule can be obtained with cardinal comparability.

We are thus led to the conclusion that with complete welfare information, that is, when it is possible to make numerically-significant interpersonal utility comparisons, none of our models result in the utilitarian rule nor do all three models result in the same social-evaluation rule. However, it is possible to generate the utilitarian rule in each model, but only by making an information assumption which seems quite arbitrary and which is certainly not empirically accurate.

There can be no doubt that Harsanyi has make one of the most significant contributions to welfare economics; his equal-probability model, in particular, represents a great intellectual advance. Our results do not, of course, detract from these contributions nor do they provide an argument against utilitarianism; our argument merely makes clear that this type of extended sympathy argument does not provide a complete rationale for utilitarianism. This will not surprise many readers; for example, Sen [1979, pp. 60-63] argues that an independent axiomatization of utilitarianism is needed in addition to Harsanyi's work. We hope that the present paper provides the axiomatization Sen is asking for, and also hope that the role of the alternative sets of assumptions employed and the interrelationships between the three models is now more clearly understood than before.

2. SOCIAL-EVALUATION FUNCTIONALS

Let $S = \{x,y,z,\ldots\}$ be a set of social alternatives with at least three elements. In Section 5, we further assume that S is a connected subset of $\mathbb{R}^s$, Euclidean s-space. In comparing our three models we require S to be the same set, and thus be connected; however, the results in Sections 4 and 6 are valid for more general sets. Society consists of n $(n \geqslant 3)$ individuals with $N = \{1,2,3,\ldots,n\}$. A *social evaluation* is an ordering of the elements of S. This evaluation is made on the basis of an ordering, assumed representable by a real-valued function, of the elements of $S \times N$. In other words, comparisons of "being person i in state x " with "being person j in state y" are used to form the social evaluation of alternative social states.

Formally, a *social-evaluation functional* $F : \mathcal{U} \to \mathcal{R}$ is a mapping from the set $\mathcal{U}$ of admissible real-valued extended utility functions[2] on $S \times N$ to the set $\mathcal{R}$ of orderings of S. Elements of $S \times N$ are called *social stations* with typical element denoted (x,i) while elements of $\mathcal{R}$ are social

[2] Extended utility functions are defined on $S \times N$ while (personal) utility functions are defined on S.

evaluations. For $U \in \mathcal{U}$, $U(x,i)$ is interpreted to be the utility which the evaluator attributes to person i in state x. We define $R_U := F(U)$.

This framework is extremely flexible, allowing for a wide variety of assumptions concerning the measurability of some particular individual's utility function $U(\cdot,i)$ and the commensurability of different individuals' utilities. Different commensurability and measurability assumptions are obtained by partitioning $\mathcal{U}$ into information sets and requiring all extended utility functions in the same information set to be mapped by F into the same ordering of S. For example, in the Arrow [1951] framework no interpersonal comparisons of utility are possible (commensurability assumption) and all individual utility functions $U(\cdot,i)$ are assumed to be ordinal (measurability assumption). Thus if U' can be obtained from U by, possibly independent, increasing transformations of the $U(\cdot,i)$, then $F(U) = F(U')$. It is of utmost importance to note that there is an inverse relationship between the strength of the commensurability-measurability assumptions and the restrictiveness of the invariance properties. As Sen [1977b, p. 154] puts it : "... the *less* the information, the *wider* is the range of values over which no discrimination is possible."

To make these ideas precise, we suppose $\mathcal{U}$ has been partitioned into a set $A = \{A_t | t \in T\}$ where T is an indexing set. Each A_t, $t \in T$ is an information set. We require the social-evaluation functional to be invariant within an information set.

<u>Information Invariance</u> : $\forall U, U' \in \mathcal{U}$, if $U, U' \in A_t$, $t \in T$, then $F(U) = F(U')$.

We shall have occasion to consider two different partitions of $\mathcal{U}$. In the first partition, all individual utility functions are numerically significant and fully interpersonally comparable - the finest partition of $\mathcal{U}$ which is possible.

Numerical Comparability : $\forall U,\ U' \in \mathcal{U}$, if $U \in A_t$ and $U' \neq U$, then $U' \notin A_t$, $t \in \mathcal{T}$.

The second partitioning of $\mathcal{U}$ will result in the social-evaluation functional being invariant with respect to common scale changes in the individual utilities and with respect to common changes of origin.

Cardinal Comparability : $\forall U,\ U' \in \mathcal{U}$, $U,\ U' \in A_t$, $t \in \mathcal{T}$, if and only if $\exists\, a, b \in \mathbb{R}$ with $b > 0$ such that $U'(x,i) = a + bU(x,i)$, $\forall (x,i) \in S \times N$.

For each $U \in \mathcal{U}$, R_U is a *social evaluation*. Under certain circumstances, all of these orderings of S can be represented by a single ordering R on $\mathbb{R}^n$, the space of utility n-tuples; that is, there is an isomorphism between the image of the functional F and the ordering R. Specifically, $\forall x, y \in S$, $\forall U \in \mathcal{U}$, xR_Uy if and only if uRu' where $u = (u_1,\ldots,u_n)$, $u' = (u'_1,\ldots,u'_n)$, $u_i = U(x,i)$, and $u'_i = U(y,i)$ for all $i \in N$. When this isomorphism exists, alternatives with the same utilities for the same individuals are indifferent in the social evaluation and non-welfare characteristics of an alternative are ignored. For this reason, this approach to social evaluation is called *welfarism*.

Welfarism places restrictions on the nature of the social-evaluation functional F. Theorem 1 establishes that the following three conditions on F are necessary and sufficient conditions for welfarism.

Unlimited Domain : $\mathcal{U}$ consists of all real-valued extended utility functions on $S \times N$.

Independence of Irrelevant Alternatives : $\forall U, U' \in \mathcal{U}, \forall D \subseteq S$, if $U(x,\cdot) = U'(x,\cdot) \; \forall x \in D$, then $F(U)$ and $F(U')$ coincide on D.

Pareto Indifference : $\forall U \in \mathcal{U}, \forall x, y \in S$, if $U(x,i) = U(y,i) \; \forall i \in N$, then xI_Uy where I_U is the indifference relation corresponding to R_U.

We refer to these three conditions on F as the welfarism axioms.

Theorem 1 : F *satisfies the welfarism axioms if and only if there is an ordering* R *on* $\mathbb{R}^n$ *such that* $\forall x, y \in S, \forall U \in \mathcal{U}, xR_Uy \leftrightarrow uRu'$ *where* $u_i = U(x,i)$ *and* $u_i' = U(y,i), \forall i \in N$.

Proof : Blackorby and Donaldson [1979b, Theorem A.1].

Given a social-evaluation function F, there is only one ordering of $\mathbb{R}^n$, no matter which $U \in \mathcal{U}$ is chosen. However, the ordering of S depends upon the choice of $U \in \mathcal{U}$. Consequently, even if two agents have the same moral preferences (choose the same F), they may still order the elements of S differently if they form their social evaluations on the basis of different extended utility functions.

We conclude this section with two further axioms for F and R.

Continuity : The ordering R is continuous.

This assumption makes it possible to represent R by a continuous *social-welfare function* $W : \mathbb{R}^n \to \mathbb{R}$ so that $uRu' \leftrightarrow W(u) \geqslant W(u')$.

W implicitly defines the *social-evaluation function* $f_U : S \to \mathbb{R}$, by $f_U(x) := W(U(x,1),\ldots,U(x,n)))$ $\forall U \in \mathcal{U}$, $\forall x \in S$.[3] Thus, $xR_Uy \leftrightarrow f_U(x) \geqslant f_U(y)$; that is, f_U represents R_U. If S is a connected subset of $\mathbb{R}^s$ and each $U(\cdot,i)$ is continuous, then the function f_U and the ordering R_U are continuous.

Pareto Preference : $\forall U \in \mathcal{U}$, $\forall x, y \in S$, if $U(x,i) \geqslant U(y,i)$ $\forall i \in N$ and $\exists i \in N$ such that $U(x,i) > U(y,i)$, then xP_Uy where P_U is the strict preference relation corresponding to R_U.

This assumption implies that the social-welfare function W is increasing in each of its arguments. From a social perspective, each person's utility is to count positively. However, nothing in this framework prevents the personal preferences as represented by $U(\cdot,i)$ from displaying malevalence.[4]

3. BENTHAM UTILITIES

Interpersonal comparison of utilities are an important part of Harsanyi's work.[5] Harsanyi assumes that when an individual makes a social (moral) eva-

[3] A social-welfare function has an n-tuple of utilities as arguments in contrast to a social-evaluation function which is defined on social states.

[4] Our modelling of welfarism is based on Maskin [1978]. Variants of Theorem 1 are proved by Guha [1972], Blau [1976], d'Aspremont and Gevers [1977], and Hammond [1979]. For criticisms of welfarism, see Sen [1977a; 1977b, Section 8]. Sen [1977b; 1979] provides a general discussion of the social-evaluation functional approach, emphasizing informational issues. Russell and Wilkinson [1979] provide a textbook account of social-evaluation functionals with welfarism as a maintained hypothesis.

[5] See Harsanyi [1955, Section V; 1977, Section 4.4; 1977, Section 4.10] and Jeffrey [1971; 1974].

luation, he will do so on the basis of interpersonal utility comparisons which are internally consistent. This can be guaranteed in the framework developed in Section 2, by assigning the social-evaluation functional F^k to individual k, $\forall k \in N$, which has as its domain a set of extended utility functions on $S \times N$. Yet to make the theory operational, it is necessary to determine which extended utility functions are appropriate members of $\mathcal{U}$ to use in forming the actual social evaluation of S. This is accomplished by assuming the existence of utility comparisons which possees objective validity. Because of differences in information, the actual comparisons being made may not agree with the objective circumstances and could vary from one observer to the next. We assume that the evaluator is striving for objective validity. Consequently, he would base his evaluation on an extended utility function chosen from the same information set as the objective extended utility function. This assumption is in keeping with the utilitarian tradition of only employing welfare information when making moral decisions. Furthermore, this assumption embodies a principle of consumer sovereignity; subject to his limited information, each evaluator is to use person i's own welfare judgements when comparing (x,i) with (y,i), $\forall x, y \in S$.

We assume the existence of a *Bentham* [1789] *utility function* $B : S \times N \rightarrow \mathbb{R}$. The Bentham utility function is numerically significant and allows fully interpersonal comparisons of utility.

Utilitarianism specifies an ordering of S for each Bentham utility function. We say that $R_{BU} \in \mathcal{R}$ is a *utilitarian order* for B if and only if

$$xR_{BU}y \leftrightarrow \sum_i B(x,i) \geqslant \sum_i B(y,i), \ \forall x, y \in S. \tag{3.1}$$

The Bentham utility function and the utilitarian ordering (3.1) provide a basis for comparing Harsanyi's three defences of utilitarianism. We consider these defences in the following three sections.

4. SEPARABLE SOCIAL-WELFARE FUNCTIONS

In this section we consider a model of social evaluation based on the separability of social-welfare functions. This problem has been studied by Fleming [1952], Harsanyi [1955, Section II; 1977, Section 4.9], Deschamps and Gevers [1977], and Maskin [1978].

We assume that S is a set of *certain* alternatives. Each agent in society is endowed with a social-evaluation functional $F^k : \mathcal{U} \to \mathcal{R}$, $k \in N$. We adopt the axioms of Section 2 so that F^k is isomorphic to a continuous, increasing social-welfare function $W^k : \mathbb{R}^n \to \mathbb{R}$, $k \in N$.

A social-welfare function which satisfies the strong Pareto condition is said to satisfy *elimination of* (the influence of) *indifferent individuals* if, whenever a group of individuals are indifferent between two alternatives, the social choice between them depends on the utilities of the remaining individuals alone, independent of the utility levels of the first group. Formally, this is the requirement that a social-welfare function satisfies *complete strict separability with respect to* N . To define this concept, we need to first define *strict separability.*

Let $\overline{N} = \{N^c, N^r\}$ be a partition of N , that is, $N^c \cup N^r = N$ and $N^c \cap N^r = \phi$. This partitions the vector of utilities $u = (u_1, \ldots, u_n)$

into $u = (u^c, u^r)$ where $u^c \in \mathbb{R}^{(c)}$ and $u^u \in \mathbb{R}^{(r)}$.[6]

Strict Separability : N^r is strictly separable from its complement in W^k if and only if the set $\{u^r \in \mathbb{R}^{(r)} \,|\, W^k(u^c, u^r) \geq W^k(u^c, \hat{u}^r)\}$ is independent of u^c for all $(u^c, \hat{u}^r) \in \mathbb{R}^n$.

This implies that the conditional ordering on $\mathbb{R}^{(r)}$ generated by W^k when u^c is held fixed is independent of the particular values chosen for the vector u^c. If this independence holds for all partitions of N , then W^k satisfies complete strict separability with respect to N.

Complete Strict Separability with Respect to N : W^k satisfies complete strict separability with respect to N if and only if N^r is strictly separable from its complement in W^k for all $N^r \in 2^N/\phi$.[7]

When combined with continuity, this condition requires additive separability of each W^k.

Theorem 2 : If each W^k, $k \in N$, *is continuous, increasing, and completely strictly separable with respect to* N *, then each social-welfare function can be written as*

$$W^k(u) = \overset{*}{W}{}^k\left(\sum_i \varphi_i^k(u_i) \right), \; k \in N, \tag{4.1}$$

where $\overset{*}{W}{}^k$ *is increasing in its argument.*

[6] See Blackorby, Primont, and Russell [1978, Chapter 3] for a careful discussion of this partitioning as well as an extended discussion of the separability concepts which follow.

[7] $2^N/\phi$ is the set of all non-empty subsets of N. A detailed discussion of this condition in the context of social-welfare functions is provided by Deschamps and Gevers [1977] who consider a number of different interpretations of the set S.

Proof : Gorman [1968, Theorem 1] or Blackorby, Primont, and Russell [1978, Theorem 4.8].

Corollary : If in addition to the assumptions in Theorem 2, each W^k *is symmetric, then each social-welfare function can be written as*

$$W^k = \overset{*}{W}{}^k(\sum_i \varphi^k(u_i)), \quad k \in N. \tag{4.2}$$

There appears to have been a great deal of confusion in the literature concerning functions of the form (4.1) or (4.2). While it is true that $W^k(u)$ and $\sum_i \varphi_i^k(u_i)$ generate the same level surfaces in $\mathbb{R}^n$; their particular numerical representation of these surfaces need only be positive monotone transformations of each other. Consequently, $\overset{*}{W}{}^k$ cannot be arbitrarily set equal to the identity mapping. Furthermore, there is no reason to expect the φ_i^k to be identity mappings either; each person's utility numbers are subjected to a *specific* monotone prior to the summation. In (4.2) these transformations do not depend upon which person's utility we are considering, although they can vary from one evaluator to the next.

We now consider the comparability properties of the utility vectors explicitly. In the case of numerical comparability, W^k is not required to satisfy any invariance restrictions. Consequently, with this informational assumption, it is not possible to strengthen the conclusions of Theorem 2 and its Corollary.

However, if there is imperfect discrimination between the elements of $\mathcal{U}$, the functional form of the social-welfare function will be further restricted. With cardinal comparability we obtain Theorem 3.

<u>*Theorem 3*</u> *: If in addition to the assumptions of* Theorem 2 *, each* F^k *satisfies information invariance with cardinal comparability, then each social-welfare function can be written as*

$$W^k(u) = \overset{**}{W}{}^k\Big(\sum_i \gamma_i^k u_i\Big), \tag{4.3}$$

where $\gamma_i^k > 0, \ \forall\, i, k \in N,$ *and* $\overset{**}{W}$ *is increasing in its argument.*

<u>Proof</u> : Blackorby and Donaldson [1979a, Theorem 11].

<u>*Corollary*</u> *: If in addition to the assumptions in* Theorem 3 *, each* W^k *is symmetric, then each social-welfare function can be written as*

$$W^k(u) = \hat{W}^k\Big(\sum_i u_i\Big) \tag{4.4}$$

where $\hat{W}^k$ *is increasing in its argument.*[8]

Equation (4.3) has the consequence that $uR^k u' \leftrightarrow \sum_i \gamma_i^k u_i \geqslant \sum_i \gamma_i^k u_i'$, or $xR^k_{U^k}y \leftrightarrow \sum_i \gamma_i^i U^k(x,i) \geqslant \sum_i \gamma_i^k U^k(y,i)$, a weighted utilitarian rule *for the function* U^k. Equation (4.4) yields the usual utilitarian rule. We must emphasize that (4.4) determines an ordering of $\mathbb{R}^n$, an ordering which is independent of k. However, the utilitarian order defined in (3.1) is an ordering of S. An ordering of S is obtained from a social-welfare function by employing a particular extended utility function, $U^k \in \mathcal{U}^k$. This choice uniquely determines a social evaluation $R^k_{U^k}$ from the social-

[8] This is essentially the theorem established by Maskin [1978]. To account for the differences in our results and those of Deschamps and Gevers [1977, p. 79] , see Sen [1977b, Section 4].

welfare function. To obtain the utilitarian order (3.1) from the social-welfare function (4.4), it is necessary for each evaluator to pick a U^k which is a positive affine transformation of the Bentham utility function.

Thus, the particular ordering of S depends not only on the choice of a social-welfare function but also on the choice of a particular $U^k \in \mathcal{U}$. In making a social evaluation we require the choice of U^k to satisfy the *principle of acceptance*.

Principle of Acceptance : Given a partition A of $\mathcal{U}$ and a Bentham utility function B , $R^k_{U^k}$ satisfies the principle of acceptance if and only if U^k and B are in the same information set of A .

For example, suppose that A is the partition of $\mathcal{U}$ for cardinal comparability. Then $R^k_{U^k}$ satisfies the principle of acceptance if and only if $\exists\, a^k, b^k$ such that $U^k(x,i) = a^k B(x,i) + b^k$, $a^k > 0$ for all $(x,i) \in S \times N$.

It is then straightforward to establish the following results.

Theorem 4 : If, in addition to the assumptions of Theorem 2 *, each* F^k *satisfies information invariance with numerical comparability and if social evaluations are required to satisfy the principle of acceptance, then for all* $x, y \in S$, *for all* $B \in \mathcal{U}$,

$$xR^k_B y \leftrightarrow \sum_i \varphi^k_i(B(x,i)) \geqslant \sum_i \varphi^k_i(B(y,i)),\ k \in N, \tag{4.5}$$

where R^k_B *is the* k*th person's social evaluation when* B *is the Bentham utility function.*

Corollary : *If in addition to the assumptions of* Theorem 4 , *each* W^k *is symmetric, then for all* $x, y \in S$, *for all* $B \in \mathcal{U}$

$$xR_B^k y \leftrightarrow \sum_i \varphi^k(B(x,i)) \geqslant \sum_i \varphi^k(B(y,i)), \quad k \in N. \tag{4.6}$$

It is tempting to note that, in (4.5), $\varphi_i^k(B(\cdot,i))$ is ordinally equivalent to $B(\cdot,i)$ for each i . Consequently, if we define $B^k(x,i) := \varphi_i^k(B(x,i))$, (4.5) may be written

$$xR_B^k y \leftrightarrow \sum_i B^k(x,i) \geqslant \sum_i B^k(y,i). \tag{4.7}$$

Harsanyi interprets this as a kind of utilitarianism although, of course, it does *not* agree with the utilitarian ordering of S in (3.1). Additional assumptions are needed for that.

Theorem 5 : *If in addition to the assumptions of* Theorem 3 , *social evaluations are required to satisfy the principle of acceptance, then for all* $x, y \in S$, *for all* $B \in \mathcal{U}$,

$$xR_B^k y \leftrightarrow \sum_i \gamma_i^k B(x,i) \geqslant \sum_i \gamma_i^k B(y,i), \quad k \in N. \tag{4.8}$$

Corollary : *If in addition to the assumptions of* Theorem 5 , *each* W^k *is symmetric, then for all* $x, y \in S$, *for all* $B \in \mathcal{U}$,

$$xR_B^k y \leftrightarrow \sum_i B(x,i) \geqslant \sum_i B(y,i), \quad k \in N. \tag{4.9}$$

Equation (4.9) is the utilitarian social evaluation. It is worth remarking that it is not necessary for there to be complete agreement among agents, in the sense that W^k need not equal $W^{k'}$, for utilitarianism to result. However, even if there is complete agreement in this sense, none of (4.5) - (4.8) are the symmetric utilitarian evaluation (3.1).

In the presence of numerical comparability, Theorem 4 established that social evaluations can be rationalized by a social-welfare function which is additive in the Bentham utilities. To obtain the utilitarian evaluation, an anonymity requirement is coupled with cardinal comparability. It must be stressed that the separability approach, unlike the model presented in Section 6, does not posit the existence of an evaluator who is deprived of certain information on the grounds that it is morally irrelevant. Sen [1977b, p. 1548] has noted that "Axiomatizing moral principles through informational constraints might appear as a confusion of ethical and epistemological considerations." Even if one agrees with Sen that the feasibility of particular comparisons can play a role in determining the acceptance of moral rules, there does not appear to be any compelling reason why cardinal comparability should be chosen to represent the informational constraints. In the model considered in this section, it is the individual agents in society who make the social evaluations. Certainly, an evaluator's personal utility function (on S) should be numerically significant even if his evaluation of some other person's utility function is not numerically significant. Thus, if the informational constraints are supposed to be empirically accurate, it would seem that the partitioning of U into information sets should vary from one observer to the next.

5. THE EXPECTED UTILITY APPROACH

In this section we allow for uncertain events and require social-evaluation functions to satisfy the expected utility hypothesis. In the formation of these social evaluations, attention is restricted to situations where the individual utility functions satisfy the expected utility hypothesis as well. This approach to social evaluation was considered by Harsanyi [1955, Section III; 1977, Section 4.8] and Samuelson [1977, Section 2]. While, in this model, social evaluations are orderings of uncertain alternatives, by considering the restriction of these orderings to the subset of certain events, comparisons with the results of the previous section are possible.

We assume that there are M states of nature so that $X = (x^1, \ldots, x^M) \in S^M$, the M-fold Cartesian product of S,[9] is a complete description of the environment. If $x \in S$ occurs for certain, this is denoted $X^C = (x, \ldots, x)$ where $X^C \in S^C$, the set of certain events. Each person has a social-evaluation functional (for uncertain events) $\tilde{F}^k : \tilde{\mathcal{U}} \to \tilde{\mathcal{R}}$, $k \in N$, where $\tilde{\mathcal{U}}$ is the set of admissible real-valued extended utility functions on $S^M \times N$ with typical element $\tilde{U}$, and $\tilde{\mathcal{R}}$ is the set of orderings of S^M. Throughout this section we maintain the axioms of Section 2 suitably reinterpreted, so that $\forall \tilde{U}^k \in \tilde{\mathcal{U}}$, $\tilde{\mathcal{R}}^k_{\tilde{U}^k} := \tilde{F}^k(\tilde{U}^k)$ can be represented by a continuous social-evaluation function $\tilde{f}^k_{\tilde{U}^k} : S^M \to \mathbb{R}$, $k \in N$ and $\tilde{F}^k$ is isomorphic to a continuous increasing social-welfare function

[9] In this section, we assume that S is a connected subset of $\mathbb{R}^s$.

$\tilde{W}^k : \mathbb{R}^n \to \mathbb{R}$, $k \in N$.

We now assume that $\tilde{f}^k_{\tilde{U}^k}$ satisfies the expected utility hypothesis.[10] Thus, it can be written as

$$\tilde{f}^k_{\tilde{U}^k}(X) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left[\sum_m p^k_m \hat{\tilde{f}}{}^k_{\tilde{U}^k}(x^m)\right] \tag{5.1}$$

where p^k_m is person k's subjective probability that state m will occur, $m = 1,\ldots,M$, $\sum_m p^k_m = 1$, and $\overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}$ is increasing in its argument.[11] The evaluator function $\hat{\tilde{f}}{}^k_{\tilde{U}^k}$ does not depend upon the particular state being considered; consequently

$$\tilde{f}^k_{\tilde{U}^k}(X^C) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left[\hat{\tilde{f}}{}^k_{\tilde{U}^k}(x)\right]. \tag{5.2}$$

Although the welfarism axioms ensure that $\tilde{F}^k$ satisfies unlimited domain, we are only interested in a particular subset of $\tilde{\mathcal{U}}$. We consider only those $\tilde{U}^k \in \tilde{\mathcal{U}}$ which when restricted to person i's utility function (as perceived by k) $\tilde{U}^k(\cdot,\overset{i}{\cancel{k}})$ satisfy the expected utility hypothesis, for all $i, k \in N$.

[10] See Arrow [1965] and Blackorby, Davidson and Donaldson [1977] for a discussion of the expected utility hypothesis where separability argument like those employed in Section 4 are used. The discussion in Drèze [1974] is also relevant.

[11] As noted in the discussion of (4.1) and (4.2), it is not legitimate to set $\overset{*}{\tilde{f}}{}^k_{U^k}$ equal to the identity mapping. The term in square brackets in (5.1) is ordinally equivalent to $\tilde{f}^k_{\tilde{U}^k}$, but not necessarily identical to $\tilde{f}^k_{\tilde{U}^k}$.

The subjective probabilities can vary in two quite different ways. The firsc way is obvious from the notation; the subjective probabilities may differ from one individual to another. The second way is not so clear; although each individual has but one social-evaluation functional $\tilde{F}^k$, $k \in N$, the probabilities may vary depending upon which $\tilde{U}^k \in \tilde{\mathcal{U}}$ is chosen. In order to be able to compare our results with those of Harsanyi, who assumes the existence of objective prcbabilities, we assume that all *subjective probabilities coincide across agents* and that in (5.1) $p_m^k = p_m$, $\forall k \in N$, $m = 1,\ldots,M$. Hence, for those $\tilde{U}^k(\cdot,i)$ which satisfy the expected utility hypothesis (and to which we are restricting our attention) we may write

$$\tilde{U}^k(X,i) = \overset{*}{\tilde{U}}{}^k_i \left[\Sigma p_m \hat{\tilde{U}}^k(x^m,i) \right] \tag{5.3}$$

where $\overset{*}{\tilde{U}}{}^k_i$ is increasing in its argument.

When attention is restricted to certain alternatives, then we may write

$$\tilde{U}^k(X^C,i) = \overset{*}{\tilde{U}}{}^k_i \left[\hat{\tilde{U}}^k(x,i) \right]. \tag{5.4}$$

This provides sufficient information for us to prove the main result of this section.

<u>Theorem 6</u> : *If the social-evaluation function* $\tilde{f}^k_{\tilde{U}^k}$ *and each* $\tilde{U}^k(\cdot,i)$ *satisfy the expected utility hypothesis, the subjective probabilities coincide and are positive, and the range of each* $\tilde{U}^k(\cdot,i)$ *is an interval, then the social-evaluation function may be written as*

$$\tilde{f}^k_{\tilde{U}^k}(X) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left(\sum_i a_i^k \sum_m p_m \hat{\tilde{U}}^k(x^m,i) + b^k\right), \tag{5.5}$$

where each a_i^k *is positive.*

<u>Proof</u> : From Theorem 1, we may write

$$\begin{aligned}\tilde{f}^k_{\tilde{U}^k}(X^C) &= \tilde{W}^k(\tilde{U}^k(X^C,1),\ldots,\tilde{U}^k(X^C,n)) \\ &= \tilde{W}^k(\overset{*}{\tilde{U}}{}^k_1[\hat{\tilde{U}}^k(x,1)],\ldots,\overset{*}{\tilde{U}}{}^k_n[\hat{\tilde{U}}^k(x,n)]) \quad \text{[using (5.4)]} \qquad (5.6) \\ &= \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left[\hat{\tilde{f}}{}^k_{\tilde{U}^k}(x)\right] \quad \text{[using (5.2)].}\end{aligned}$$

Using (5.1) and the coincidence of subjective probabilities,

$$\tilde{f}^k_{\tilde{U}^k}(X) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left[\sum_m p_m \hat{\tilde{f}}{}^k_{\tilde{U}^k}(x^m)\right]. \tag{5.7}$$

This, in conjunction with the last equality in (5.6), yields

$$\tilde{f}^k_{\tilde{U}^k}(X) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left[\sum_m p_m \overset{*}{\tilde{f}}{}^{k^{-1}}_{\tilde{U}^k}(\tilde{W}^k\,[\overset{*}{\tilde{U}}{}^k_1(\hat{\tilde{U}}^k(x^m,1)),\ldots,\overset{*}{\tilde{U}}{}^k_n(\hat{\tilde{U}}^k(x^m,n))])\right]. \tag{5.8}$$

From Theorem 1 and (5.3)

$$\begin{aligned}\tilde{f}^k_{\tilde{U}^k}(X) &= \tilde{W}^k(\tilde{U}^k(X,1),\ldots,\tilde{U}^k(X,n)) \\ &= \tilde{W}^k\left(\overset{*}{\tilde{U}}{}^k_1\left[\sum_m p_m \hat{\tilde{U}}^k(x^m,1)\right],\ldots,\overset{*}{\tilde{U}}{}^k_n\left[\sum_m p_m \hat{\tilde{U}}^k(x^m,n)\right]\right). \qquad (5.9)\end{aligned}$$

By equating (5.8) and (5.9) we have a functional equation in the social-welfare function $\tilde{W}^k$. To see this more clearly we introduce some transformations. Define $y_m^k = (y_{1m}^k,\ldots,y_{nm}^k)$, $m = 1,\ldots,M$ by

$$y^k_{im} := p_m \, \hat{\tilde{U}}^k(x^m, i), \tag{5.10}$$

$\check{W}^k_m$ by

$$\check{W}^k_m(y^k_m) := p_m \, \overset{*}{\tilde{f}}{}^{k\,-1}_{\tilde{U}^k} \left(\tilde{W}^k \left[\overset{*}{\tilde{U}}{}^k_1(y^k_{1m}/p_m), \ldots, \overset{*}{\tilde{U}}{}^k_n(y^k_{nm}/p_m)\right]\right), \tag{5.11}$$

and $\check{\overline{W}}^k$ by

$$\check{\overline{W}}^k\left(\sum_m y^k_m\right) := \overset{*}{\tilde{f}}{}^{k\,-1}_{\tilde{U}^k}\left[\tilde{W}^k\left(\overset{*}{\tilde{U}}{}^k_1\left(\sum_m y^k_{1m}\right), \ldots, \overset{*}{\tilde{U}}{}^k_n\left(\sum_m y^k_{nm}\right)\right)\right]. \tag{5.12}$$

Equating (5.8) and (5.9), inverting $\overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}$, and substituting (5.10) - (5.12) into the result yields

$$\sum_m \check{W}^k_m(y^k_m) = \check{\overline{W}}^k\left(\sum_m y^k_m\right). \tag{5.13}$$

Equation (5.13) is a Pexider equation whose solution for $\check{\overline{W}}^k$ is [12]

$$\check{\overline{W}}^k(y^k_m) = \sum_i a^k_i \, y^k_{im} + b^k. \tag{5.14}$$

Using (5.9), (5.10), and (5.12), this yields

$$\tilde{f}^k_{U^k}(X) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left(\sum_i a^k_i \sum_m p_m \, \hat{\tilde{U}}^k(x^m, i) + b^k\right), \tag{5.5}$$

where each a^k_i is positive as $\tilde{W}$ is increasing.

□

[12] See the excellent discussion in Eichhorn [1978, pp. 49-52]. Eichhorn's proof assumes (5.13) holds for all of $\mathbb{R}^n$, which is not guaranteed by our assumptions since the range of $\tilde{U}^k(\cdot, i)$ could be a strict subset of $\mathbb{R}$ for some i. However, our continuity and range assumptions ensure that the same solution obtains over our restricted domain.

Corollary 1 : *If, in addition to the assumptions of* Theorem 6, $\tilde{W}^k$ *is symmetric in its arguments then the social-evaluation function may be written as*

$$\tilde{f}^k_{\tilde{U}^k}(X) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left(a^k \sum_i \sum_m p_m \hat{\tilde{U}}^k(x^m,i) + b^k\right), \tag{5.15}$$

where $a^k > 0$.

When considering certain alternatives (5.5) and (5.15) have a slightly simpler structure. The former becomes

$$\tilde{f}^k_{\tilde{U}^k}(X^C) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left(\sum_i a^k_i \hat{\tilde{U}}^k(x,i) + b^k\right), \tag{5.16}$$

and the latter becomes

$$\tilde{f}^k_{\tilde{U}^k}(X^C) = \overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}\left(a^k \sum_i \hat{\tilde{U}}^k(x,i) + b^k\right). \tag{5.17}$$

Some insight into the nature of this result can be gleaned from examining the equality of (5.8) and (5.9) in the special case where $\overset{*}{\tilde{f}}{}^k_{\tilde{U}^k}$ and each $\overset{*}{\tilde{U}}{}^k_i$, $i \in N$ are identity maps. This yields

$$\sum_m p_m \tilde{W}^k[\hat{\tilde{U}}^k(x^m,1),\ldots,\hat{\tilde{U}}^k(x^m,n)] = \tilde{W}^k\left(\sum_m p_m \hat{\tilde{U}}^k(x^m,1),\ldots,\sum_m p_m \hat{\tilde{U}}^k(x^m,n)\right). \tag{5.18}$$

The left side of (5.18) is the expected value of social welfare to be derived from W while the right side is the level of social welfare associated with assigning each person his expected utility from X. In this special case the order in which the expectation operator and the welfare function operator are employed is of no consequence. The generalization of this accounts for the linear structure found in (5.5) and (5.15) - (5.17).

Before proceeding, one more remark about the proof seems justified. The probabilities, $p_1,\ldots,p_m$, are completely determined when $\tilde{f}^k_{\tilde{U}^k}$ satisfies the expected utility hypothesis. Thus, throughout the proof of Theorem 6, these probabilities are constant. Both Harsanyi [1977, Section 4.8] and Samuelson [1977, Section 2] employ proofs which require the probabilities to take on specific values independent of $\tilde{f}^k_{\tilde{U}^k}$. While this procedure is formally correct, it does seem inconsistent with the spirit of their work, particularly if the probabilities are interpreted to be objective, as appears to be the case with Harsanyi.

We turn now to the utilitarian implications of Theorem 6.

From (5.16), we have, for all X^C, $Y^C \in S^C$,

$$X^C \; \tilde{R}^k_{\tilde{U}^k} \; Y^C \leftrightarrow \sum_i a^k_i \, \hat{\tilde{U}}^k(x,i) \geqslant \sum_i a^k_i \, \hat{\tilde{U}}^k(y,i), \tag{5.19}$$

where $\tilde{R}^k_{\tilde{U}^k}$ is person k's social evaluation for the extended utility function $\tilde{U}^k$. This is a weighted utilitarian evaluation *for the functions* $\overset{*}{\tilde{U}}{}^{k\,-1}_{i}(\tilde{U}^k)$. [See (5.4).]

To determine the choice of the extended utility function to use in forming the social evaluation, we now apply the (slightly modified) principle of acceptance to this rule. To do this, we must define Bentham utilities over S^C. This is done with the Bentham utility function on S, and we have

$$\tilde{B}(X^C,i) := B(x,i) \tag{5.20}$$

$\forall X^C \in S^C$ and all $i \in N$. The principle of acceptance in Section 4 is simply modified to the new domain. We can now prove :

Theorem 7 *: If, in addition to the assumptions of* Theorem 6, $\tilde{F}^k$ *satisfies information invariance with numerical comparability and if social evaluations satisfy the principle of acceptance over* S^C *, then*

$$X^C \tilde{R}^k_{\tilde{B}} Y^C \leftrightarrow \sum_i a^k_i \overset{*}{\tilde{U}}{}^{k\;-1}_i [B(x,i)] \geqslant \sum_i a^k_i \overset{*}{\tilde{U}}{}^{k\;-1}_i [B(y,i)], \tag{5.21}$$

where $a^k_i > 0 \; \forall i, k \in N$ *and* $\tilde{R}^k_{\tilde{B}}$ *is the* k*th person's social ordering of* S^C.

Proof : Given numerical comparability, the principle of acceptance requires (using (5.4)) that

$$\tilde{U}^k(X^C,i) = \overset{*}{\tilde{U}}{}^k_i[\tilde{\tilde{U}}{}^k(x,i)] = \tilde{B}(X^C,i) = B(x,i), \tag{5.22}$$

for all $X^C \in S^C$ and $i, k \in N$. Solving for $\tilde{\tilde{U}}{}^k$ and using (5.16) yields (5.21).

□

Symmetry of the ordering $\tilde{R}^k_{\tilde{B}}$ in the *Bentham* utilities yields :

Theorem 8 *: If, in addition to assumptions of* Theorem 7, $\tilde{F}^k$ *is symmetric in the Bentham utilities (any permutation of Bentham utilities among people is a matter of social indifference), then*

$$X^C \tilde{R}^k_{\tilde{B}} Y^C \leftrightarrow \sum_i g^k[B(x,i)] \geqslant \sum_i g^k[B(y,i)]. \tag{5.23}$$

Proof : Symmetry of (5.21) in the Bentham utilities yields

$$a_i^k \overset{*}{\tilde{U}}{}_i^{k^{-1}}(t) = a_j^k \overset{*}{\tilde{U}}{}_j^{k^{-1}}(t) =: g^k(t), \ \forall i, \ j \in N. \tag{5.24}$$

□

In Theorems 7 and 8, we assumed numerical comparability so that $\tilde{U}^k$ and $\tilde{B}$ coincide on S^C. This is the strongest comparability assumption (between $\tilde{B}$ and $\tilde{U}^k$) that we can use. Yet in the presence of this complete welfare information, it is clear that we have not been able to argue for utilitarianism. Rather, the additively separable (5.19) and (5.21) are obtained, with distributional preferences over Bentham utilities dictated by $\{a_i^k\}$, $\{\overset{*}{\tilde{U}}{}_i^{k^{-1}}\}$, and $\{g^k\}$.

If we restrict the information available to agents so that $\tilde{F}^k$ satisfies information invariance with cardinal comparability, $\tilde{U}^k$ need only be a positive affine transformation of $\tilde{B}$. It is straightforward to establish that with this comparability assumption, the functions g^k in (5.23) can be set equal to the identity mapping. For brevity, we do not state this formally. Thus, with cardinal comparability, it is possible to obtain the utilitarian order in the expected utility model, as was done in the previous section. Again this is not a particularly satisfactory result; the use of cardinal comparability means that morally relevant information is being ignored through the use of an arbitrary information restriction.

Returning to the case of full numerical comparability, so far we have asked that $\tilde{U}^k$ and $\tilde{B}$ coincide only over S^C. It remains to be seen whether the principle of acceptance can strengthen Theorems 7 and 8 when the principle's domain is extended to S^M. To do this, we ask $\widetilde{B}$ to be defined on $S^M \times N$ and require each $\widetilde{B}(\cdot,i)$ to satisfy the expected utility hypothesis. Then

$$\widetilde{B}(X,i) = b^i \left[\sum_m p_m \overset{*}{\widetilde{B}}(x^m,i) \right] \qquad (5.25)$$

where each b^i is increasing in its argument. If the principle of acceptance is required to hold over S^M and we have numerical comparability, then $\widetilde{B} = \widetilde{U}^k$. In this case, we have

$$\begin{aligned} \widetilde{U}^k(X,i) \geqslant \widetilde{U}^k(Y,i) &\leftrightarrow \widetilde{B}(X,i) \geqslant \widetilde{B}(Y,i) \\ &\leftrightarrow \sum_m p_m \widetilde{\widetilde{U}}^k(x^m,i) \geqslant \sum_m p_m \widetilde{\widetilde{U}}^k(y^m,i) \quad \text{[using (5.3)]} \qquad (5.26) \\ &\leftrightarrow \sum_m p_m \overset{*}{\widetilde{B}}(x^m,i) \geqslant \sum_m p_m \overset{*}{\widetilde{B}}(y^m,i) \quad \text{[using (5.25)]} \end{aligned}$$

for all $X, Y \in S^M$ and $i \in N$.

A standard result in expected utility theory tells us that for each i, there exist $\gamma^k_i > 0$, δ^k_i such that, for each $x \in S$,

$$\widetilde{\widetilde{U}}^k(x,i) = \gamma^k_i \overset{*}{\widetilde{B}}(x,i) + \delta^k_i. \qquad (5.27)$$

We now prove :

<u>*Theorem 9*</u> *: If, in addition to the assumptions of* Theorem 7 *, social evaluations satisfy the principle of acceptance over* S^M *, then*

$$X^C \, \widetilde{R}^k_{\widetilde{B}} \, Y^C \leftrightarrow \sum_i \bar{a}^k_i \, b^{i^{-1}} [B(x,i)] \geqslant \sum_i \bar{a}^k_i \, b^{i^{-1}} [B(y,i)] \qquad (5.28)$$

where $\bar{a}^k_i > 0$ *for all* $k, i \in N$.

Proof : Reproducing (5.19)

$$X^C \tilde{R}^k_{\tilde{U}^k} Y^C \leftrightarrow \sum_i a_i^k \hat{\tilde{U}}^k(x,i) \geqslant \sum_i a_i^k \hat{\tilde{U}}^k(y,i), \qquad (5.19)$$

and from (5.27),

$$X^C \tilde{R}^k_{\tilde{B}} Y^C \leftrightarrow \sum_i \bar{a}_i^k \overset{*}{\tilde{B}}(x,i) \geqslant \sum_i \bar{a}_i^k \overset{*}{\tilde{B}}(y,i) \qquad (5.29)$$

where $\bar{a}_i^k = a_i^k \gamma_i^k$ for all $i, k \in N$.

Since $B(x,i) = \tilde{B}(X^C,i) = b^i[\overset{*}{\tilde{B}}(x,i)]$ for all x and i from (5.25), the theorem is proved.

□

In (5.21), the functions $\{\tilde{U}_i^k\}$ may vary from one evaluator to another. However, the analogous functions $\{b^i\}$ in (5.28) are common across evaluators. Individuals attitudes toward risk (in Bentham utilities) dictate the curvature properties of the social evaluation.

Theorem 10 : If, in addition to the assumptions of Theorem 8, *social evaluations satisfy the principle of acceptance over* S^M *then*

$$X^C \tilde{R}^k_{\tilde{B}} Y^C \leftrightarrow \sum_i g[B(x,i)] \geqslant \sum_i g[B(y,i)], \qquad (5.30)$$

where g *is increasing in its argument.*

Proof : The assumptions of Theorem 8 and the principle of acceptance over S^M imply symmetry in (5.28); hence

$$\bar{a}_i^k \, {b^i}^{-1}(t) = \bar{a}_j^k \, {b^j}^{-1}(t) \qquad (5.31)$$

for all $i, j, k \in N$, and for all t in the range of B. Rewrite (5.31) as

$$\frac{b^{i^{-1}}(t)}{b^{j^{-1}}(t)} = \frac{a_j^{-k}}{a_i^{-k}} \; . \tag{5.32}$$

Since the left side of (5.32) is independent of k, so is the right side and

$$\frac{a_j^{-k}}{a_i^{-k}} = \frac{\alpha_j}{\alpha_i} \,, \quad \text{say.} \tag{5.33}$$

Substituting into (5.31) yields

$$\alpha_i b^{i^{-1}}(t) = \alpha_j b^{j^{-1}}(t) =: g(t), \quad \text{say,} \tag{5.34}$$

which must be independent of i and j.

Inserting (5.34) into (5.38) yields (5.30).

□

Symmetry (Theorem 10) therefore, requires a common attitude toward risk in Bentham utilities by all individuals.[13]

Again, the orderings we have obtained are not utilitarian since there is no reason to assume that $(a_i^{-k} b^{i^{-1}})$ or g is an affine transform.

[13] It is apparent that the requirement that individual attitudes toward risk should dictate the evaluator's distributional preferences is incompatible with symmetry (anonymity) unless individuals share a common attitude toward risk in Bentham utilities. The ethical desirability of this sort of "consumers' sovereignty" is open to question.

Further, in (5.21) and (5.23), there is no necessary agreement with the transforms used in the additively separable representations found in the previous section. Consequently, with numerically comparability, the expected utility approach and the separable social-welfare function approach will typically result in *different* orderings of S even if the same Bentham utilities for certain events are used in both cases.

6. EQUAL PROBABILITIES BEHIND A VEIL OF IGNORANCE

In this section, we assume uncertainty about the state of nature that occurs (as we did before) and add an additional uncertainty, the person that the evaluator will turn out to be. In this model, the latter information is viewed to be morally irrelevant. The evaluator (person k) is thought of as being behind a veil of ignorance.[14] The evaluator pursues his own interest in this situation, but the knowledge that he is to be each individual with probability $1/n$ guarantees preferences over social states that resemble moral preferences. Models of this sort have been considered by Vickery [1945, Section III; 1960] and Harsanyi [1953; 1955, Section III; 1977, Section 4.1].

[14] For simplicity we use Rawls' (1971) terminology to describe the choice situation. This use of the word "ignorance" departs from its formal definition in the theory of choice under ignorance where, as with Rawls, agents are deprived of any information which would allow probabilities to be attached to states of the world. Both Rawls and Harsanyi assume that agents are self-interested but differ in their interpretation of what constitutes morally irrelevant knowledge. For Rawls, morally irrelevant knowledge includes attitudes towards risk and uncertainty.

As in Section 2, we let S be the set of certain outcomes. $S \times N$ is the set of social "stations". A typical member of $S \times N$ is (x,i), "being person i in state x". There are M states of nature, and an uncertain outcome is $X \in S^M$. For uncertainty about one's name, there are n (the cardinality of N) states of identity, and an uncertain outcome is $I = (i^1, i^2, \ldots, i^n) \in N^n$ with each i^ℓ experienced with probability $1/n$. If I consists of any permutation of the elements of N, then the probability of being a given person is $1/n$. If each $i^\ell = i$, then there is no uncertainty over who one is to be. In this case, we write $I^C \in N^C$. An overall uncertain outcome is an event in $S^M \times N^n$ with outcomes (x^m, i^ℓ) occurring with (subjective) probability p_m/n. This uncertain outcome is written $(X,I) = [(x^1,i^1),(x^2,i^1),\ldots,(x^M,i^1);\ldots;(x^1,i^n),(x^2,i^n),\ldots,(x^M,i^n)]$.

Person k $(k \in N)$ is assumed to have an ordering $(\vec{R}^k)$ over $S^M \times N^n$ represented by the function $\vec{V}^k$, so that

$$\begin{aligned}(X,I)\,\vec{R}^k\,(Y,J) \leftrightarrow\ & \vec{V}^k[(x^1,i^1),\ldots,(x^m,i^\ell),\ldots,(x^M,i^n)] \\ & \geqslant \vec{V}^k[(y^1,j^1),\ldots,(y^m,j^\ell),\ldots,(y^M,j^n)].\end{aligned} \tag{6.1}$$

We assume that k's preferences satisfy the expected utility hypothesis, so that we can write

$$\vec{V}^k[(X,I)] = \overset{*}{V}\left[\sum_\ell \sum_m \frac{p_m}{n}\, \hat{\vec{V}}^k(x^m,i^\ell)\right]. \tag{6.2}$$

Now suppose that $X = X^C$, or that x occurs for certain, and that $I = I^N = [1,2,\ldots,n]$, or that k will be person i with probability $1/n$. This induces an ordering on S, and we write

$$x\,\hat{R}^k\,y \leftrightarrow (X^C, I^N)\,\bar{R}^k\,(Y^C, I^N). \tag{6.3}$$

Theorem 11 : If k*'s preferences over* $S^M \times N^n$ *satisfy the expected utility hypothesis and if the ordering* $\hat{R}^k$ *is defined by* (6.3) *, then*

$$x\,\hat{R}^k\,y \leftrightarrow \sum_i \hat{\bar{V}}^k(x,i) \geqslant \sum_i \hat{\bar{V}}^k(y,i). \tag{6.4}$$

The proof is immediate, given (6.2) and the definition of X^C and I^N.

We now relate the functions $\bar{V}^k$, $\hat{\bar{V}}^k$, $\forall k \in N$ to the (extended) Bentham utility functions $\tilde{B}$ and B . As in Section 5, $\tilde{B}(X,j)$ is the (Bentham) utility j gets from X and $B(x,j) = \tilde{B}(X^C,j)$. We assume that, for each j, $\tilde{B}(\cdot,j)$ satisfies the expected utility hypothesis. Therefore, (reproducing (5.25)),

$$\tilde{B}(X,j) = b^j\left[\sum_m p_m \overset{*}{B}(x^m,j)\right]. \tag{5.25}$$

Further, we would like to require that each person's ordering $\bar{R}^j$ over $S^M \times N^n$ agree with (5.25) over all events where he is himself for certain. That is,

$$(X,J^C)\,\bar{R}^j\,(Y,J^C) \leftrightarrow \tilde{B}(X,j) \geqslant \tilde{B}(Y,j) \tag{6.5}$$

for all $j \in N$ and all $X, Y \in S^M$. As well, we would like person k's ordering over all events where he is j for certain to coincide with agent j's ordering of the same events. Thus,

$$(X,J^C)\,\bar{R}^k\,(Y,J^C) \leftrightarrow (X,J^C)\,\bar{R}^j\,(Y,J^C) \tag{6.6}$$

for all $j, k \in N$ and for all $X, Y \in S^M$. We define the *principle of acceptance for the veil of ignorance* as satisfaction of the requirement that $\vec{V}^k[(\cdot,J^C)]$ and $\tilde{B}(\cdot,j)$ belong to the same information set . Thus, for numerical comparability, $\tilde{B}(X,J) = \vec{V}^k[(X,J^C)] \quad \forall X \in S^M$. This principle implies (6.5) and (6.6) for any comparability rule (including ordinal non-comparability).

From (5.25), (6.2), and (6.5) we notice that

$$\sum_m p_m \hat{\vec{V}}^k(x^m,j) \geqslant \sum_m p_m \hat{\vec{V}}^k(y^m,j) \leftrightarrow \sum_m p_m \overset{*}{\tilde{B}}(x^m,j) \geqslant \sum_m p_m \overset{*}{\tilde{B}}(y^m,j) \tag{6.7}$$

for all $j \in N$ and for all $X, Y \in S^M$. It follows that there exist $\gamma_j^k > 0$ and δ_j^k such that

$$\hat{\vec{V}}^k(x,j) = \gamma_j^k \overset{*}{\tilde{B}}(x,j) + \delta_j^k \tag{6.8}$$

for all $j \in N$ and for all $x \in S$. We also know, given numerical comparability and the principle of acceptance that

$$\vec{V}^k[(X,J^C)] = \tilde{B}(X,j) \tag{6.9}$$

for all $X \in S^M$ and all $k, j \in N$. This enables us to prove :

<u>*Theorem 12*</u> : *If, in addition to the assumptions of Theorem* 11, k*'s preferences satisfy the principle of acceptance with numerical comparability, then*

$$x \hat{R}^k y \leftrightarrow \sum_i \gamma_i^k \overset{*}{\tilde{B}}(x,i) \geqslant \sum_i \gamma_i^k \overset{*}{\tilde{B}}(y,i) \tag{6.10}$$

and

$$x \hat{R}^k y \leftrightarrow \sum_i \gamma_i^k \, b^{i^{-1}} [B(x,i)] \geqslant \sum_i \gamma_i^k \, b^{\;-1} [B(y,i)] \qquad (6.11)$$

for all x, y ∈ S *and for all* k ∈ N.

Proof : (6.10) is immediate from (6.2), (6.4), and (6.8). From the fact that $\tilde{B}(X^c,i) = B(x,i)$ and (5.25),

$$B(x,i) = b^{\overset{*}{i}}[\tilde{B}(x,i)] \qquad (6.12)$$

and (6.11) is established from (6.10).

□

This result is identical to the result of Section 5 when the principle of acceptance is required to hold on S^M. Again the ordering is not utilitarian since there is no reason to assume that $\tilde{b}^i$ is affine.

Symmetry in the Bentham utilities will require

$$x \hat{R}^k y \leftrightarrow \sum_i g[B(x,i)] \geqslant \sum_i g[B(y,i)], \qquad (6.13)$$

the *same* result as that proved in Theorem 10. Again, we do not get the utilitarian ordering.

To obtain the utilitarian ordering, it is necessary, as in the previous sections, to restrict the allowable information to that consistent with cardinal comparability. While it is granted that a central feature of the present model is the restriction of the information available to the evaluator, it is only morally irrelevant information which is to be so restricted. On these grounds one can justify the ignorance about one's place in society; we do not believe this objective provides a basis for ignoring the welfare information neglected by cardinal comparability.

7. CONCLUDING REMARKS

A key feature of our analysis is the distinction between a social-welfare function which orders utility n-tuples and a social-evaluation functional which orders social states. Even if all agents have the same social-welfare function, by utilizing different extended utility functions, different agents can obtain different social evaluations. Our principle of acceptance requires each evaluator to use an extended utility function which is informationally-equivalent to the objectively-given Bentham utility function.

Claude d'Aspremont has pointed out to us that it is not essential for our results that we have a Bentham utility function as an objective standard. It is sufficient that all agents use the same information set in forming their social evaluations; this can be accomplished without the use of objective utilities. As before, $U^k(x,i)$ is the value person k attributes to being person i in state x. In place of the principle of acceptance, we can use a *principle of consensus*; there is agreement among evaluators on the choice of the relevant information set to use in social decision-making.

We have presented three models of extended preferences and social evaluation. In each case a series of assumptions has been employed and its implications explored. For each set of assumptions, the result has been a social-evaluation function which is additively separable in the individual Bentham utilities. In each case, given complete welfare information, the social ranking is not the utilitarian one. To obtain the utilitarian order, in each model it is necessary to restrict the usable welfare information to that allowed by cardinal comparability. We have not found any persuasive justification for this information restriction in the literature.

We therefore conclude that Harsanyi's arguments are effective (and persuasive in the presence of the welfarism axioms) arguments for additive separability rather than for utilitarianism. Unfortunately, this requirement leaves a good deal of freedom for the social evaluator. He may adopt the utilitarian rule, being indifferent to all distributions of a given amount of total utility, or adopt preferences such as

$$x R y \leftrightarrow \sum_i \frac{(B(x,i))^r}{r} \geqslant \sum_i \frac{(B(y,i))^r}{r} \qquad (7.1)$$

(as long as the range of B is restricted to $\mathbb{R}_{++}$) which, as $r \rightarrow -\infty$ approaches the maximin rule. Thus, Harsanyi's arguments are compatible with a very wide range of distributional judgements. However, in the models of Sections 5 and 6, these judgements are not the creation of the evaluator alone; they arise from the attitude toward risk of the individuals in a society. If symmetry in the Bentham utilities is imposed, we have

$$x R y \leftrightarrow \sum_i g(B(x,i)) \geqslant \sum_i g(B(y,i)) \qquad (7.2)$$

where g represents a common social attitude to risk in the Bentham utilities. In this case, it is not all obvious whether the choice of g should be based on the preferences of individuals or whether it should be the subject of a separate moral decision.

REFERENCES

Arrow, K.J. [1951], *Social Choice and Individual Values*, New York : Wiley.

Arrow, K.J. [1965], *Aspects of the Theory of Risk-Bearing*, Helsinki : Yrjö Jahnsson Foundation.

Bentham, J. [1789], *An Introduction to the Principles of Morals and Legislation*, London : Payne.

Blackorby, C., R. Davidson and D. Donaldson [1977], "A Homiletic Exposition of the Expected Utility Hypothesis," *Economica* 44, 351-358.

Blackorby, C. and D. Donaldson [1979a], "Interpersonal Comparability of Origin- or Scale-Independent Utilities : Admissable Social Evaluation Functionals," Discussion Paper No. 79-04, Department of Economics, University of British Columbia.

Blackorby, C. and D. Donaldson [1979b], "Moral Criteria for Evaluating Population Change," Discussion Paper No. 79-08, Department of Economics, University of British Columbia.

Blackorby, C., D. Primont and R.R. Russell [1978], *Duality, Separability and Functional Structure : Theory and Economic Applications*, New York : North-Holland.

Blau, J.H. [1976], "Neutrality, Monotonicity, and the Right of Veto : A Comment," *Econometrica* 44, 603.

d'Aspremont, C. and L. Gevers [1977], "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies* 44, 199-209.

Deschamps, R. and L. Gevers [1977], "Separability, Risk-Bearing, and Social Welfare Judgements," *European Economic Review* 10, 77-94.

Drèze, J.H. [1974], "Axiomatic Theories of Choice, Cardinal Utility and Subjective Probability : A Review," in J.H. Drèze, ed., *Allocation Under Uncertainty : Equilibrium and Optimality*, London : Macmillan, 3-23.

Eichhorn, W. [1978], *Functional Equations in Economics*, Reading, Mass : Addison-Wesley.

Fleming, M. [1952], "A Cardinal Concept of Welfare," *Quarterly Journal of Economics* 66, 366-384.

Gorman, W.M. [1968], "The Structure of Utility Functions," *Review of Economic Studies* 32, 369-390.

Hammond, P.J. [1979], "Equity in Two-Person Situations : Some Consequences," *Econometrica* 47, 1127-1136.

Harsanyi, J.C. [1953], "Cardinal Utility in Welfare Economics and in the Theory of Risk-Bearing," *Journal of Political Economy* 61, 434-435. Reprinted in Harsanyi [1976].

Harsanyi, J.C. [1955], "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy* 63, 309-321. Reprinted in Harsanyi [1976].

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

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

Jeffrey, R.C. [1971], "On Interpersonal Utility Theory," *Journal of Philosophy* 68, 647-656.

Jeffrey, R.C. [1974], "Remarks on Interpersonal Utility Theory," in S. Stenlund, ed., *Logical Theory and Semantic Analysis*, Dordrecht : D. Reidel, 35-44.

Maskin, E. [1978], "A Theorem on Utilitarianism," *Review of Economic Studies* 45, 93-96.

Guha, A.S. [1972], "Neutrality, Monotonicity, and the Right of Veto," *Econometrica* 40, 821-826.

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

Russell, R.R. and M. Wilkinson [1979], *Microeconomics : A Synthesis of Modern and Neoclassical Theory*, New York : Wiley.

Samuelson, P.A. [1977], "When is it Ethically Optimal to Allocate Money Income in Stipulated Fractional Shares," in A.S. Blinder and P. Friedman, eds., *Natural Resources, Uncertainty, and General Equilibrium Systems : Essays in Memory of Rafael Lusky*, New York : Academic Press, 175-195.

Sen, A.K. [1977a], "The Poverty of Welfarism," *Intermountain Economic Review* 8, 1-13.

Sen, A.K. [1977b], "On Weights and Measures : Informational Constraints in Social Welfare Analysis," *Econometrica* 45, 1539-1572.

Sen, A.K. [1979], "Social Choice Theory," forthcoming in K.J. Arrow and M. Intriligator, eds., *Handbook of Mathematical Economics*, vol. III, Amsterdam : North-Holland.

Vickery, W. [1945], "Measuring Marginal Utility by Reactions to Risk," *Econometrica* 13, 319-333.

Vickery, W. [1960], "Utility, Strategy and Social Decision Rules," *Quarterly Journal of Economics* 74, 507-535.