CRDE

CAHIER 02-2002

IN DEFENSE OF WELFARISM

Charles BLACKORBY,
Walter BOSSERT and
David DONALDSON

**Centre de recherche
et développement en économique**

C.P. 6128, succursale Centre-ville
Montréal QC H3C 3J7

Téléphone : (514) 343-6557
Télécopieur : (514) 343-5831
crde@crde.umontreal.ca
http://www.crde.umontreal.ca/

Université de Montréal

CAHIER 02-2002

IN DEFENSE OF WELFARISM

Charles BLACKORBY[1], Walter BOSSERT[2] and David DONALDSON[3]

[1] University of British Columbia and GREQAM

[2] Centre de recherche et développement en économique (C.R.D.E.) and Département de sciences économiques, Université de Montréal

[3] University of British Columbia

January 2002

This is a revised version of the authors' earlier paper "Welfarism with multiple non-welfare profiles." The authors thank John Weymark for comments and discussions. The paper was presented at the International Symposium on Intergenerational Equity in Tokyo, December 2001. Financial support through a grant from the Social Sciences and Humanities Research Council of Canada is gratefully acknowledged.

RÉSUMÉ

Cet article caractérise l'évaluation sociale welfaristique dans un cadre multiprofil dans lequel, outre les multiples profils d'utilité, on suppose qu'il y a plusieurs profils d'information non welfaristique. Nous prouvons de nouvelles versions des théorèmes de welfarisme dans ce cadre alternatif et nous montrons qu'une propriété très plausible et faible d'anonymat est suffisante pour générer des ordres d'évaluation sociale anonyme.

Mots clés : welfarisme, choix social multiprofil

ABSTRACT

This paper characterizes welfarist social evaluation in a multi-profile setting where, in addition to multiple utility profiles, it is assumed that there are several profiles of non-welfare information. We prove new versions of the welfarism theorems in this alternative framework, and we illustrate that a very plausible and weak anonymity property is sufficient to generate anonymous social-evaluation orderings.

Key words : welfarism, multiple-profile social choice

1. Introduction

Welfarist principles for social evaluation rank social alternatives using information about individual well-being (welfare, utility) alone, ignoring non-welfare information. As a result, those principles regard things such as liberty, freedom of expression or a healthy environment as having instrumental value: they are valuable because of their contribution to well-being.

Welfarism is the normative view that lies behind all of neoclassical welfare economics. Bergson-Samuelson social-welfare functions depend on the utility levels of individuals only and, in addition, Paretian compensation and consumer's-surplus tests focus on preferences as expressed in market behaviour.

Sen [1987] has criticized welfarism on the grounds that preferences or desires may not always be consistent with well-being, noting that individual preferences may be affected by incomplete information and that "the underdog comes to terms with social inequalities by bringing desires in line with feasibilities" (p. 11). Because of this, welfarist principles should be coupled with accounts of well-being, such as those of Griffin [1986] and Sumner [1996], that compensate for information problems and are comprehensive enough to capture all aspects of the good life.[1] In addition, because length of life affects welfare, it is desirable to focus on lifetime well-being rather than well-being in a single period. Without an account of well-being that includes all aspects of the good life, the appeal of our axioms would be significantly diminished.

Conventional social-choice theory employs multiple profiles of welfare (utility) information only: non-welfare information is implicitly fixed. In this setting, welfarism is a consequence of the axioms unlimited (utility) domain, Pareto indifference and binary independence of irrelevant alternatives.[2] Because non-welfare information is fixed, it is impossible to discern the way in which a principle makes use of it. For that, multiple non-welfare profiles are needed.

In this paper, we present a defense of welfarism in a framework in which both social and individual non-welfare information may vary across information profiles. Social non-welfare information may include information about the presence or absence of democratic institutions or freedom of the press. Individual-specific non-welfare information may include length of life, whether the person has a propensity to work hard and whether he or she likes classical music.

Each information profile includes a vector of individual utility functions which represent welfare information, and a vector of functions which describe social and individual

[1] See also Broome [1991] and Mongin and d'Aspremont [1998]. Challenges to welfarism by Sen and others are discussed in Section 6.

[2] See, for example, Blackorby, Bossert and Donaldson [2002], Bossert and Weymark [2002], d'Aspremont and Gevers [1977], Guha [1972], Hammond [1979], Sen [1977, 1979] and Weymark [1998].

non-welfare information.[3] In that setting, the independence axiom is formulated in terms of both welfare and non-welfare information and it, together with unlimited domain and Pareto indifference, is used to make a strong case in favour of welfarism. In addition, a strong and natural argument for anonymity in social evaluation is developed.

A principle for social evaluation is a social-evaluation functional which associates an ordering of the alternatives with each possible information profile. Such a functional is welfarist if and only if there is a single social-evaluation ordering of utility vectors such that, for all information profiles, the ranking of any two alternatives is given by the ranking of the corresponding utility vectors.

Our approach permits a compelling justification of anonymous welfarism. The standard axiom requires the social ordering to be unaffected by a permutation of utility functions across individuals with non-welfare information unchanged. It is possible, however, that some individual may have non-welfare characteristics, such as being hardworking, that are thought to justify special consideration and this lessens the ethical attractiveness of the axiom. By contrast, our anonymity axiom requires the social ordering to be unaffected if *both* utility functions and individual non-utility-information functions are permuted. Together with our other axioms, it implies that the social-evaluation ordering must be anonymous: it ranks all permutations of any utility vector as equally good.

The basic intuition that lies behind welfarist social evaluation is the view that, if one alternative is ranked as better than another, it must be better for at least one person (see Goodin [1991]). Without this requirement, we run the risk of recommending social changes that are empty gestures, benefitting no one and, perhaps, harming some or all. We use this intuition as an axiom which we call minimal individual goodness.

If, in any two alternatives, each person is equally well off, the Pareto-indifference axiom requires the two alternatives to be ranked as equally good. Any social-evaluation functional that satisfies minimal individual goodness necessarily satisfies Pareto indifference (see Blackorby, Bossert and Donaldson [2002]). The latter axiom is employed in our theorems because minimal individual goodness, which is stronger, is not needed.

In this paper, we restrict attention to the social evaluation of alternatives and, because of that, we need not be concerned with uncertainty. It is possible, however, to extend welfarist principles so that they are capable of ordering prospects: vectors of alternatives whose components are uncertain. For discussions, see, for example, Blackorby, Bossert and Donaldson [1998, 2002], Blackorby, Donaldson and Weymark [1999, 2002], Broome [1991] and Harsanyi [1953, 1955]. In addition, welfarist principles can be extended so that they can rank alternatives with different populations and population sizes (Blackorby, Bossert and Donaldson [1995, 2002], Broome [1999]).

[3] See also Kelsey [1987] who provides a reformulation of Arrow's [1951, 1963] theorem in a framework where non-welfare information is explicitly modelled.

Because some non-humans are sentient (capable of having experiences), applications of welfarist principles often take account of their well-being. Sidgwick [1907, 1966, p. 414] argues that we should "extend our concern to all the beings capable of pleasure and pain whose feelings are affected by our conduct." We focus on humans in this paper, however; for an extension of welfarist principles to sentient non-humans see, for example, Blackorby and Donaldson [1992].

Section 2 presents our notation and defines social-evaluation functionals. In Section 3, we present our axioms and, in Section 4, we prove that any social-evaluation functional with an unlimited domain that satisfies Pareto indifference and binary independence of irrelevant alternatives must be welfarist, disregarding all non-welfare information. In Section 5, we turn to anonymity and characterize anonymous social-evaluation orderings. Section 6 provides a discussion of challenges to welfarism by Sen and others and Section 7 concludes.

2. Social-evaluation functionals

The set of all positive integers is denoted by $\mathcal{Z}_{++}$ and the set of real numbers by $\mathcal{R}$. For $n \in \mathcal{Z}_{++}$, let $\mathcal{R}^n$ be the n-fold Cartesian product of $\mathcal{R}$. Our notation for vector inequalities is $\geq$, $>$ and $\gg$.

The (fixed) set of individuals is $N = \{1, \ldots, n\}$, $n \in \mathcal{Z}_{++}$. The set of alternatives is X, and we assume that it contains at least three elements.

A utility (welfare) profile is an n-tuple $U = (U_1, \ldots, U_n)$, where $U_i\colon X \to \mathcal{R}$ is the utility function of individual $i \in N$. Utility is an index of individual well-being. The set of all possible utility profiles is $\mathcal{U}$, and we write $U(x) = (U_1(x), \ldots, U_n(x))$ for all $x \in X$ and for all $U \in \mathcal{U}$.

Non-welfare information is described by a profile $K = (K_0, K_1, \ldots, K_n)$, where $K_0\colon X \to \mathcal{S}_0$ is a function that associates social non-welfare information with each alternative in X and, for all $i \in N$, $K_i\colon X \to \mathcal{S}$ associates individual non-welfare information with each alternative in X. The set $\mathcal{S}_0 \neq \emptyset$ is the set of possible values of social non-welfare information, and $\mathcal{S} \neq \emptyset$ is the set of possible values for individual non-welfare information. $\mathcal{S}^n$ is the n-fold Cartesian product of $\mathcal{S}$. The set of all possible profiles of non-welfare information is $\mathcal{K}$ and, for all $x \in X$ and for all $K \in \mathcal{K}$, we define $K(x) = (K_0(x), K_1(x), \ldots, K_n(x))$.

The set of all orderings on X is denoted by $\mathcal{O}$. A social-evaluation functional is a mapping $F\colon \mathcal{D} \to \mathcal{O}$, where $\mathcal{D} \subseteq \mathcal{U} \times \mathcal{K}$ and $\mathcal{D} \neq \emptyset$. We use the notation $\Upsilon = (U, K)$ and, for convenience, we define $R_\Upsilon = F(\Upsilon)$ for all $\Upsilon \in \mathcal{D}$. The asymmetric and symmetric factors of R_Υ are denoted by P_Υ and I_Υ. Furthermore, we write $\Upsilon(x) = (U(x), K(x))$ for all $x \in X$ and all $\Upsilon \in \mathcal{D}$.

3. Axioms

If the social ordering associated with an information profile ranks alternative x as better than alternative y, it is reasonable to ask there to be at least one individual who is better off in x than in y. Minimal individual goodness (Goodin [1991]; see also Blackorby, Bossert and Donaldson [2002]) requires this condition to be met for all profiles and all pairs of alternatives. If this axiom is not satisfied, social changes that are empty gestures, benefitting no one and possibly harming some or all, may be recommended.

Minimal Individual Goodness: For all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$, if $xP_\Upsilon y$, then there exists $j \in N$ such that $U_j(x) > U_j(y)$.

Pareto indifference requires any two alternatives to be ranked as equally good whenever each individual is equally well off in both.

Pareto Indifference: For all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$, if $U(x) = U(y)$, then $xI_\Upsilon y$.

A social-evaluation functional satisfies minimal individual goodness if and only if it satisfies Pareto indifference and Pareto weak preference.

Pareto Weak Preference: For all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$, if $U(x) > U(y)$, then $xR_\Upsilon y$.

Theorem 1 is due to Blackorby, Bossert and Donaldson [2002]; a proof is included for completeness.

Theorem 1: *F satisfies minimal individual goodness if and only if F satisfies Pareto indifference and Pareto weak preference.*

Proof. Suppose F satisfies minimal individual goodness. We first prove by contradiction that Pareto indifference is satisfied. Suppose not. Then there exist $x, y \in X$ and $\Upsilon \in \mathcal{D}$ such that $U(x) = U(y)$ and not $xI_\Upsilon y$. Because R_Υ is complete, either $xP_\Upsilon y$ or $yP_\Upsilon x$. In each case, we obtain a contradiction to minimal individual goodness.

Now suppose F violates Pareto weak preference. Then there exist $x, y \in X$ and $\Upsilon \in \mathcal{D}$ such that $U(x) > U(y)$ and not $xR_\Upsilon y$. By the completeness of R_Υ, we must have $yP_\Upsilon x$, again contradicting minimal individual goodness.

Finally, suppose F satisfies Pareto indifference and Pareto weak preference but violates minimal individual goodness. Then there exist $x, y \in X$ and $\Upsilon \in \mathcal{D}$ such that $xP_\Upsilon y$ and $U(y) \geq U(x)$. If $U(y) = U(x)$, we obtain a contradiction to Pareto indifference, and if there exists $j \in \{1, \ldots, n\}$ such that $U_j(y) > U_j(x)$, we obtain a contradiction to Pareto weak preference. ■

It is easy to see why Pareto indifference is implied by minimal individual goodness. Suppose that each person is equally well off in x and y. Then it is not the case that at least one person is better off in x and, by minimal individual goodness, it is not the case that x is better than y. In addition, because it is not the case that at least one individual is better off in y, it is not the case that y is better than x. Because the social ordering is, by assumption, complete, x and y are equally good.

In addition to Pareto indifference, an unlimited-domain assumption and binary independence of irrelevant alternatives are usually employed to generate welfarism. Unlimited domain requires that F is capable of producing a social ordering for all possible profiles of welfare and non-welfare information.

Unlimited Domain: $\mathcal{D} = \mathcal{U} \times \mathcal{K}$.

Binary independence of irrelevant alternatives is a condition that ensures consistency across profiles. It requires the social ranking of any two alternatives to depend on the utility information and non-welfare information associated with those two alternatives only. An important property of this axiom is that it does not prevent non-welfare information from being taken into consideration.

Binary Independence of Irrelevant Alternatives: For all $x, y \in X$ and for all $\Upsilon, \bar{\Upsilon} \in \mathcal{D}$, if $\Upsilon(x) = \bar{\Upsilon}(x)$ and $\Upsilon(y) = \bar{\Upsilon}(y)$, then

$$xR_{\Upsilon}y \Leftrightarrow xR_{\bar{\Upsilon}}y.$$

We conclude this section with a formulation of strong neutrality. If the utility vectors for alternatives x and y in one profile are the same as the utility vectors for two (possibly different) alternatives z and w in another, strong neutrality requires the ranking of x and y by the social ordering associated with the first profile to be the same as the ranking of z and w by the social ordering associated with the second.

Strong Neutrality: For all $x, y, z, w \in X$ and for all $\Upsilon, \bar{\Upsilon} \in \mathcal{D}$, if $U(x) = \bar{U}(z)$ and $U(y) = \bar{U}(w)$, then

$$xR_{\Upsilon}y \Leftrightarrow zR_{\bar{\Upsilon}}w.$$

4. Welfarism

Our first step towards proving a welfarism theorem with multiple non-welfare profiles consists of showing that unlimited domain, Pareto indifference and binary independence of irrelevant alternatives together imply that the social ordering cannot depend on non-welfare information. It is easy to see why this is the case if there are four or more alternatives.

	Welfare Information				Non-Welfare Information			
	x	y	z	w	x	y	z	w
Profile Υ	u	v			k	ℓ		
Profile Υ^1	u	v	u	v	k	ℓ	$\bar{k}$	$\bar{\ell}$
Profile Υ^2	u	v	u	v	$\bar{k}$	$\bar{\ell}$	$\bar{k}$	$\bar{\ell}$
Profile $\bar{\Upsilon}$	u	v			$\bar{k}$	$\bar{\ell}$		

Table 1

In Table 1, x, y, z and w are distinct alternatives, entries under the welfare-information heading are utility vectors and entries under the non-welfare-information heading are non-welfare-information vectors. In profile Υ, utility vectors for x and y are $u \in \mathcal{R}^n$ and $v \in \mathcal{R}^n$ and non-welfare information vectors for x and y are $k \in \mathcal{S}_0 \times \mathcal{S}^n$ and $\ell \in \mathcal{S}_0 \times \mathcal{S}^n$. In profile $\bar{\Upsilon}$, utility vectors for x and y are the same, but the non-welfare-information vectors may be different and are denoted by $\bar{k}$ and $\bar{\ell}$. Information for all other alternatives is unspecified and can be anything in the domain.

We show that the ranking of x and y by R_Υ, the ordering corresponding to profile Υ, is the same as the ranking of x and y by $R_{\bar{\Upsilon}}$, the ordering corresponding to profile $\bar{\Upsilon}$. To do so, we construct two other profiles which are feasible by the unlimited-domain axiom. Profile Υ^1 coincides with profile Υ on x and y but is specified for z and w. By binary independence of irrelevant alternatives, the rankings of x and y by R_Υ and R_{Υ^1} are the same. Because the pairs (x, z) and (y, w) have the same utility vectors, Pareto indifference requires R_{Υ^1} to declare x and z to be equally good and y and w to be equally good. Consequently, the two pairs are ranked in the same way by R_{Υ^1}. Profiles Υ^1 and Υ^2 coincide on z and w and, by binary independence, the rankings of z and w by R_{Υ^1} and R_{Υ^2} are identical. In addition, Pareto indifference requires R_{Υ^2} to rank the pairs (x, y) and (z, w) in the same way. Because profiles Υ^2 and $\bar{\Upsilon}$ coincide on x and y, binary independence requires the rankings of x and y by R_{Υ^2} and $R_{\bar{\Upsilon}}$ to be the same. Together, these observations prove the result.

The above discussion provides only a partial demonstration. Most of the complexity in the following proof is a consequence of the possibility that $|X| = 3$.

Theorem 2: *If F satisfies unlimited domain, Pareto indifference and binary independence of irrelevant alternatives, then, for all $x, y \in X$ and for all $\Upsilon, \bar{\Upsilon} \in \mathcal{D}$ such that $U(x) = \bar{U}(x)$ and $U(y) = \bar{U}(y)$,*

$$x R_\Upsilon y \Leftrightarrow x R_{\bar{\Upsilon}} y. \tag{1}$$

Proof. Let $x, y \in X$ and $\Upsilon, \bar{\Upsilon} \in \mathcal{D}$ be such that $U(x) = \bar{U}(x)$ and $U(y) = \bar{U}(y)$. Let $u = U(x) = \bar{U}(x)$, $v = U(y) = \bar{U}(y)$, $k = K(x)$, $\ell = K(y)$, $\bar{k} = \bar{K}(x)$ and $\bar{\ell} = \bar{K}(y)$. Because X contains at least three alternatives, there exists $z \in X \setminus \{x, y\}$. By unlimited domain, we can define the profiles Υ^1, Υ^2, Υ^3 and Υ^4 as follows. Let $\Upsilon^1(x) = (u, k)$, $\Upsilon^1(y) = (v, \ell)$, $\Upsilon^1(z) = (v, \bar{\ell})$, $\Upsilon^2(x) = (u, k)$, $\Upsilon^2(y) = (v, \bar{\ell})$, $\Upsilon^2(z) = (v, \bar{\ell})$, $\Upsilon^3(x) = (u, k)$, $\Upsilon^3(y) = (v, \bar{\ell})$, $\Upsilon^3(z) = (u, \bar{k})$, $\Upsilon^4(x) = (u, \bar{k})$ $\Upsilon^4(y) = (v, \bar{\ell})$ and $\Upsilon^4(z) = (u, \bar{k})$.

By binary independence of irrelevant alternatives, we have

$$xR_{\Upsilon}y \Leftrightarrow xR_{\Upsilon^1}y.$$

By Pareto indifference, $yI_{\Upsilon^1}z$ and it follows that

$$xR_{\Upsilon^1}y \Leftrightarrow xR_{\Upsilon^1}z.$$

Using binary independence again, we obtain

$$xR_{\Upsilon^1}z \Leftrightarrow xR_{\Upsilon^2}z.$$

By Pareto indifference, $zI_{\Upsilon^2}y$ and, therefore,

$$xR_{\Upsilon^2}z \Leftrightarrow xR_{\Upsilon^2}y.$$

Now binary independence implies

$$xR_{\Upsilon^2}y \Leftrightarrow xR_{\Upsilon^3}y.$$

By Pareto indifference, $xI_{\Upsilon^3}z$ and it follows that

$$xR_{\Upsilon^3}y \Leftrightarrow zR_{\Upsilon^3}y.$$

Using binary independence again, we obtain

$$zR_{\Upsilon^3}y \Leftrightarrow zR_{\Upsilon^4}y.$$

By Pareto indifference, $zI_{\Upsilon^4}x$ and it follows that

$$zR_{\Upsilon^4}y \Leftrightarrow xR_{\Upsilon^4}y.$$

Using binary independence once more, we obtain

$$xR_{\Upsilon^4}y \Leftrightarrow xR_{\bar{\Upsilon}}y.$$

Combining the above equivalences, (1) results. ■

If two profiles have the same welfare profiles, Theorem 2 demonstrates that the corresponding social orderings must be identical. The following theorem is an immediate consequence.

Theorem 3: *If F satisfies unlimited domain, Pareto indifference and binary independence of irrelevant alternatives, then there exists a functional $f : \mathcal{U} \to \mathcal{O}$ such that, for all $\Upsilon = (U, K) \in \mathcal{D}$, $F(\Upsilon) = f(U)$.*

Theorem 3 shows that a social-evaluation functional is equivalent to a functional on a single-non-welfare-profile domain. Because our axioms reduce to the standard ones on the smaller domain, it is straightforward to show that Pareto indifference and binary independence of irrelevant alternatives together are equivalent to strong neutrality if F satisfies unlimited domain (see, for example, Blau [1976], Bossert and Weymark [2002], d'Aspremont and Gevers [1979], Guha [1972] and Sen [1977]).

Theorem 4: *Suppose F satisfies unlimited domain. F satisfies Pareto indifference and binary independence of irrelevant alternatives if and only if F satisfies strong neutrality.*

Proof. First, suppose that F satisfies strong neutrality. That binary independence of irrelevant alternatives is satisfied follows from setting $x = z$ and $y = w$ in the definition of strong neutrality. To show that Pareto indifference is implied, let $U = \bar{U}$ and $y = z = w$. Strong neutrality implies that $xR_\Upsilon y$ if and only if $yR_\Upsilon y$ whenever $U(x) = U(y)$. Because R_Υ is reflexive, this implies $xI_\Upsilon y$.

Now suppose that F satisfies unlimited domain, Pareto indifference and binary independence of irrelevant alternatives. By Theorem 2, we know that non-welfare information is irrelevant. Consider two profiles $\Upsilon, \bar{\Upsilon} \in \mathcal{D}$ and four (not necessarily distinct) alternatives $x, y, z, w \in X$ such that $U(x) = \bar{U}(z) = u$ and $U(y) = \bar{U}(w) = v$.

By unlimited domain, there exist profiles $\Upsilon^1, \Upsilon^2, \Upsilon^3, \Upsilon^4 \in \mathcal{D}$ such that $U^1(x) = u$, $U^1(y) = v$, $U^1(w) = v$, $U^2(x) = u$, $U^2(w) = v$, $U^3(x) = u$, $U^3(y) = v$, $U^3(z) = u$, $U^4(y) = v$ and $U^4(z) = u$.

By binary independence of irrelevant alternatives,

$$xR_\Upsilon y \Leftrightarrow xR_{\Upsilon^1} y.$$

By Pareto indifference, $yI_{\Upsilon^1} w$ and, therefore,

$$xR_{\Upsilon^1} y \Leftrightarrow xR_{\Upsilon^1} w.$$

Using binary independence of irrelevant alternatives again, we obtain

$$xR_{\Upsilon^1} w \Leftrightarrow xR_{\Upsilon^2} w.$$

By binary independence of irrelevant alternatives,

$$xR_{\Upsilon^2} w \Leftrightarrow xR_{\Upsilon^3} y.$$

By Pareto indifference, $xI_{\Upsilon^3}z$ and, therefore,

$$xR_{\Upsilon^3}y \Leftrightarrow zR_{\Upsilon^3}y.$$

By binary independence of irrelevant alternatives,

$$zR_{\Upsilon^3}y \Leftrightarrow zR_{\Upsilon^4}y$$

and, using binary independence of irrelevant alternatives once more, we obtain

$$zR_{\Upsilon^4}y \Leftrightarrow zR_{\bar{\Upsilon}}w.$$

Combining the above equivalences, we obtain

$$xR_{\Upsilon}y \Leftrightarrow zR_{\bar{\Upsilon}}w,$$

and strong neutrality is satisfied. ■

Given unlimited domain and our assumption that X contains at least three elements, strong neutrality is equivalent to the existence of a social-evaluation ordering R on $\mathcal{R}^n$ which can be used to rank the alternatives in X for any profile $\Upsilon \in \mathcal{D}$.[4] The asymmetric and symmetric factors of R are P and I. Combined with Theorem 4, this observation yields the following welfarism theorem.[5]

Theorem 5: *Suppose F satisfies unlimited domain. F satisfies Pareto indifference and binary independence of irrelevant alternatives if and only if there exists a social-evaluation ordering R on $\mathcal{R}^n$ such that, for all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$,*

$$xR_{\Upsilon}y \Leftrightarrow U(x)RU(y). \tag{2}$$

Proof. Clearly, if there exists a social-evaluation ordering R such that (2) is satisfied for all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$, then F satisfies Pareto indifference and binary independence of irrelevant alternatives.

Now suppose F satisfies unlimited domain, Pareto indifference and binary independence of irrelevant alternatives. By Theorem 4, F satisfies strong neutrality. We complete the proof by constructing the social-evaluation ordering R. For all $u, v \in \mathcal{R}^n$, let uRv if and only if there exist a profile $\Upsilon \in \mathcal{D}$ and two alternatives $x, y \in X$ such that $U(x) = u$, $U(y) = v$ and $xR_{\Upsilon}y$. Strong neutrality implies that non-welfare information is irrelevant and that the relative ranking of any two utility vectors u and v does not depend on the

[4] Gevers [1979] uses the term social-welfare ordering for R.

[5] See d'Aspremont and Gevers [1977] and Hammond [1979] for a version with a single non-welfare profile. Bordes, Hammond and Le Breton [1997] and Weymark [1998] prove variants of this theorem with specific domain restrictions, again in the single-non-welfare-profile case.

profile Υ or on the alternatives x and y used to generate u and v. Therefore, R is well-defined. That R is reflexive and complete follows immediately because R_Υ is reflexive and complete for all $\Upsilon \in \mathcal{D}$. It remains to show that R is transitive. Suppose $u, v, q \in \mathcal{R}^n$ are such that uRv and vRq. By unlimited domain and the assumption that X contains at least three alternatives, there exist a profile $\Upsilon \in \mathcal{D}$ and three alternatives $x, y, z \in X$ such that $U(x) = u$, $U(y) = v$ and $U(z) = q$. Because $U(x)RU(y)$ and $U(y)RU(z)$, it follows that $xR_\Upsilon y$ and $yR_\Upsilon z$ by definition of R. Because R_Υ is transitive, we have $xR_\Upsilon z$. Hence, $U(x)RU(z)$ or, equivalently, uRq. ■

Theorem 5 implies that, for all $x, y \in X$ and all $\Upsilon \in \mathcal{D}$, $xP_\Upsilon y$ if and only if $U(x)PU(y)$ and $xI_\Upsilon y$ if and only if $U(x)IU(y)$.

5. Anonymity

A principle for social evaluation may be welfarist and, at the same time, fail to be impartial. That would be the case, for example, if a weighted sum of utilities of those who are alive were used to rank alternatives with a weight of 2 for the utility of person 1 and a weight of 1 for all other utilities. If there is a single non-welfare profile, such a principle might be justified by the fact that person 1 is hardworking in every alternative.

In such an environment, the anonymity axiom that is commonly used requires the social ordering to be unchanged if utility functions are permuted across individuals (see Sen [1970a], Blackorby, Bossert and Donaldson [2002]). Although this produces the desired result, the permutation of utility functions does not change non-welfare information and, as a consequence, the case for anonymous welfarism is not compelling.

We employ a more compelling anonymity axiom. It requires the social ordering to be unchanged if *both* utility functions and individual non-welfare information functions are permuted across individuals.

Anonymity: For all $\Upsilon, \bar{\Upsilon} \in \mathcal{D}$, if $K_0 = \bar{K}_0$ and there exists a bijection $\rho\colon N \to N$ such that $U_i = \bar{U}_{\rho(i)}$ and $K_i = \bar{K}_{\rho(i)}$ for all $i \in N$, then $R_\Upsilon = R_{\bar{\Upsilon}}$.

Anonymity is easily defended because it allows non-welfare information to matter. All that is ruled out is the claim that an individual's identity justifies special treatment, no matter what non-welfare information obtains.

An ordering R on $\mathcal{R}^n$ is anonymous if and only if, for all $u \in \mathcal{R}^n$ and for all bijections $\rho\colon N \to N$,

$$uI(u_{\rho(1)}, \ldots, u_{\rho(n)}).$$

Together with unlimited domain, Pareto indifference and binary independence of irrelevant alternatives, anonymity is sufficient to ensure that the social-evaluation functional is welfarist and anonymous.

Theorem 6: *Suppose F satisfies unlimited domain. F satisfies Pareto indifference, binary independence of irrelevant alternatives and anonymity if and only if there exists an anonymous social-evaluation ordering R on $\mathcal{R}^n$ such that, for all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$,*

$$xR_\Upsilon y \Leftrightarrow U(x)RU(y). \tag{3}$$

Proof. Clearly, the existence of an anonymous social-evaluation ordering R such that (3) is satisfied for all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$ implies that F satisfies the required axioms.

Conversely, suppose F satisfies unlimited domain, Pareto indifference, binary independence of irrelevant alternatives and anonymity. By Theorem 5, there exists a social-evaluation ordering R on $\mathcal{R}^n$ such that (3) is satisfied for all $x, y \in X$ and for all $\Upsilon \in \mathcal{D}$. It remains to show that R must be anonymous.

For $j, k \in N$ with $j \neq k$, define the transposition bijection $\bar{\rho}_{jk}: N \to N$ by $\bar{\rho}_{jk}(j) = k$, $\bar{\rho}_{jk}(k) = j$ and $\bar{\rho}_{jk}(i) = i$ for all $i \in N \setminus \{j, k\}$. For $u \in \mathcal{R}^n$ and $j, k \in N$ with $j \neq k$, let $\bar{u}^{jk} = (u_{\bar{\rho}_{jk}(1)}, \ldots, u_{\bar{\rho}_{jk}(n)})$. By unlimited domain, there exist $\Upsilon \in \mathcal{D}$ and $x, y \in X$ such that $U(x) = u$ and $U(y) = \bar{u}^{jk}$. Let $\bar{\Upsilon}^{jk} = ((U_{\bar{\rho}_{jk}(1)}, \ldots, U_{\bar{\rho}_{jk}(n)}), (K_0, K_{\bar{\rho}_{jk}(1)}, \ldots, K_{\bar{\rho}_{jk}(n)}))$. By anonymity, $R_\Upsilon = R_{\bar{\Upsilon}^{jk}}$.

Because $U(x) = \bar{U}^{jk}(y) = u$ and $U(y) = \bar{U}^{jk}(x) = \bar{u}^{jk}$, we have

$$uR\bar{u}^{jk} \Leftrightarrow xR_\Upsilon y \Leftrightarrow yR_{\bar{\Upsilon}^{jk}}x \tag{4}$$

and

$$\bar{u}^{jk}Ru \Leftrightarrow yR_\Upsilon x \Leftrightarrow xR_{\bar{\Upsilon}^{jk}}y. \tag{5}$$

Because $R_\Upsilon = R_{\bar{\Upsilon}^{jk}}$, (4) and (5) together imply

$$uR\bar{u}^{jk} \Leftrightarrow \bar{u}^{jk}Ru$$

and, because R is complete, both $uR\bar{u}^{jk}$ and $\bar{u}^{jk}Ru$ are true, so $uI\bar{u}^{jk}$.

Now let $v = (u_{\rho(1)}, \ldots, u_{\rho(n)})$ for any bijection $\rho: N \to N$. Then there exist a finite number of transposition bijections such that ρ is the composition of those bijections. By repeated application of the above argument, uIv. ■

The anonymity axiom used in this section is not the only possible one. An alternate axiom applies to each profile separately. If the associated utility and individual non-welfare-information vectors for any one alternative are the same permutation of the corresponding vectors for a second, the axiom requires the two alternatives to be ranked as equally good. Neither it nor anonymity requires non-welfare information to be ignored and, in the presence of our other axioms, both imply anonymous welfarism.

6. Challenges to welfarist social evaluation

In this section, we discuss some of the most important criticisms of welfarist social evaluation.

Welfarist principles for social evaluation are sometimes criticized as taking a narrow view of being a person, seeing them as "locations of their respective utilities" only (Sen and Williams [1982, p. 4]). The use of comprehensive accounts of lifetime well-being, such as those of Griffin [1986] and Sumner [1996], which attempt to take account of everything in which individual people have an interest, is, in our view at least, sufficient to respond to this criticism.

Griffin's 'list' view of well-being focuses on an enumeration of basic elements of the good life such as enjoyment, freedom from anxiety, good health, pleasure and the absence of pain, limbs and senses that work well, length of life (when it is worth living), autonomy, liberty, understanding, accomplishment, satisfying work and good human relationships. A ceteris paribus increase in any element on the list increases well-being. But individual people may differ in the way that the items on the list contribute to their welfare. Sumner's account focuses on happiness and it is equated with life satisfaction "which has both an affective component (experiencing the conditions of your life as fulfilling and rewarding) and a cognitive component (judging that your life is going well for you)" (p. 172). Like Griffin, Sumner allows for many determinants of well-being.

Sen [1987, p. 11] criticizes preference and desire accounts of well-being on the grounds that "the battered slave, the broken unemployed, the hopeless destitute, the tamed housewife, may have the courage to desire little." This observation points to the need for full-information qualifications in those accounts. All of the best ones, such as those of Broome [1991], Griffin [1986], Mongin and d'Aspremont [1998], and Sumner [1996], do this. Sumner also includes an autonomy qualification. Well-being is identified with what the individual would prefer if he or she were fully informed and, possibly, autonomous. It follows that actual preferences may not always be consistent with individual well-being.

A theory of well-being which focuses on 'functionings and capabilities' is presented by Sen [1985]. Functionings are the 'doings and beings' a person achieves. Refining the list of possible functions to the list actually used is seen as a valuational exercise and aggregation of the items on the resulting list is influenced by individual differences. Sen's view is similar to Griffin's but he adds another dimension. Capabilities are opportunities to achieve various functionings and they are seen as valuable in themselves.[6] The presence of capabilities on Sen's list gives him a way to value individual liberty.

Sen's theory can be employed in a welfarist context, nevertheless. What is needed is an individual goodness relation which ranks all the possible combinations of functionings

[6] Nussbaum [2000a,b] focuses almost exclusively on capabilities. For a discussion of Sen's approach, see Sumner [1996, pp. 60–68].

and capabilities. Although the resulting view of well-being would be more objective than the ones we favour, there would be no difficulty in using it with welfarist principles.[7]

Fixed-population utilitarianism exhibits indifference to inequality of well-being, and this has prompted the complaint that "persons do not count as individuals in this any more than individual petrol tanks do in the analysis of the national consumption of petroleum" (Sen and Williams [1982, p. 4]). Although utilitarianism exhibits no aversion to utility inequality, it *is* averse to inequality of consumption. We can be reasonably sure that, for people in good health, the value of an additional dollar's worth of consumption declines as consumption increases. For that reason, a transfer of consumption from rich to poor—without indirect effects—increases total utility and is seen as good by utilitarianism. In addition, it is easy to see that the utilitarian principle is concerned with the needs of the sick and disabled. Moreover, the criticism of Sen and Williams does not apply to welfarist principles that are averse to utility inequality: they rank more equal distributions of utility as better than less equal ones.

An interesting challenge to welfarism that makes use of widely held moral intuitions has been offered by Sen [1979] and it is summarized in Table 2.[8] In x, person 1 is poor and hungry and person 2 is rich and has plenty of food. Alternative y results from a transfer of food from person 2 to person 1. In it, both total utility and minimum utility rise and, as long as other people are unaffected, any weakly inequality-averse welfarist principle declares y to be better than x. In alternative z, person 1, who is a sadist, receives no transfer of food, but is allowed to torture person 2. The utility levels are not based on poorly informed or non-autonomous preferences but are supposed to represent well-being accurately.

	x	y	z
Person 1	40	70	70
Person 2	100	80	80

Table 2

Because utility levels are the same in y and z, Pareto indifference requires them to be ranked as equally good. Consequently, z and x must be ranked in the same way that

[7] Kolm [1972, 1996] presents a theory of individual well-being in which preferences, which represent happiness, are completely determined by individual characteristics. The resulting goodness relation is called a fundamental preference relation for the individual and it provides an account of well-being that can be used in welfarist social evaluation.

[8] See also Roemer [1996, p. 30].

y and x are. It follows that, if food redistribution is good, torture is also good. Sen's conclusion is that we should abandon the Pareto indifference axiom along with welfarist social evaluation.

Sen's example appeals to a moral intuition that sees torture as bad. That intuition rests on beliefs about the welfare consequences of torture, however. People who are tortured suffer terribly at the time of their ordeal and for many years afterward if they survive. The intuition is linked, in the example, to two implausible claims of fact: that torture can be a substitute for decent nutrition and that a loss of food to the rich person is just as bad as being tortured. If we really could make starving people better off in this way, the world's problems would be easier to solve. In addition, we are rightly suspicious about the implicit assumption that there are no indirect effects of the torture. No reasonable government would pass a law that allows one particular (named) person to torture another. Rather, a change in law that applies more widely would be required. Such a legal change is likely to raise the level of bad behaviour substantially, with disastrous consequences for present and future generations.

Sen's example provides an illustration of the connection between welfarism and minimal individual goodness. If y (food redistribution) is ranked above z (torture), no one is better off in y and minimal individual goodness is not satisfied. Now suppose that, in z, utility levels were 71 and 81 instead of 70 and 80. If y is ranked as better than z, the weak Pareto principle is also violated.

An interesting challenge to welfarism has appeared in recent years. It replaces concern for well-being with concern for *opportunities* for well-being on the grounds that individual people are responsible for their choices (in certain circumstances they may be thought to be responsible for their preferences as well).[9] In practice, welfarists often agree that the provision of opportunities is socially warranted, but their concern is with actual well-being. If autonomy is a significant aspect of well-being, people must be free to make important choices for themselves, and this provides a constraint which restricts the feasible set of social possibilities. By way of analogy, parents typically try to provide opportunities for their children, but that does not mean that opportunities are what they care about.

Although welfarist principles are consistent with rights (see Mill [1861, 1969, Chapter 5]), they are not consistent with unconditional rights. An example of this is Sen's [1970a, Chapter 6, 1970b] Paretian liberal paradox.[10] We present a variant of one of Sen's examples in Table 3.

[9] See, for example, Arneson [1989, 2000a,b], Roemer [1996] and, for an unsympathetic critique, Anderson [1999]. See also Fleurbaey and Maniquet [2002].

[10] See also Gaertner, Pattanaik and Suzumura [1992].

	Welfare Information		
	x	y	z
Who Attends	Person 1	Person 2	No One
Profile Υ	$(1,1)$	$(3,3)$	$(3,1)$
Profile $\bar{\Upsilon}$	$(2,3)$	$(1,2)$	$(3,1)$

Table 3

Person 2 wants to see a violent movie but her friend person 1 does not. There are many alternatives but we focus on three which, apart from movie attendance, are identical. In x, person 1 sees the movie but person 2 does not, in y, person 2 sees the movie but person 1 does not and, in z, neither sees the movie. The utility pairs in the table (person 1's utility first) represent the individual goodness relations for both people. In profile Υ, for example, person 1's utility level is 3 and person 2's utility level is 1 in alternative z.

Because movie attendance is thought to be a private matter, rights are assigned: person 1's utilities decide the ranking of x and z and person 2's utilities decide the ranking of y and z. In profile Υ, each person's utility is unaffected by the other's behaviour: person 1 is better off not seeing the movie and person 2 is better off seeing it. Because person 2's utility decides the ranking of y and z, y is better than z and, because person 1's utility decides the ranking of x and z, z is better than x. Transitivity requires y to be better than x.

But should such rights be unconditional, applying to all possible profiles? In profile $\bar{\Upsilon}$, each person's utility depends, in part, on the behaviour of the other. Person 1's utility is highest in z but x is better than y for him because he believes that the movie may have an adverse effect on person 2's character. Person 2 is worst off in z but ranks x above y: she thinks that attending the movie would help person 1 face the realities of life.

Because person 2's utility decides the ranking of y and z, y is better than z. And, because person 1's utility decides the ranking of x and z, z is better than x. Transitivity requires y to be better than x, but x is ranked as better than y by weak Pareto. In this case, rights assignments are incompatible with the weak Pareto axiom. Because all commonly used welfarist principles satisfy weak Pareto, it follows that they cannot accommodate unconditional rights over all profiles.

Rights of the type Sen discusses are typically justified, in part, on the grounds that the choices under consideration affect the individual with the right much more than they affect others, a condition that is met in some, but not all, profiles. In the example of

Table 3, the assigned rights can be justified if individual preferences correspond roughly to those of profile Υ. This suggests that rights assignments should be, to some extent, profile-dependent. If this view is accepted, Sen's paradox disappears.

7. Conclusion

Variants of the welfarism theorem can be proven on several different domains. If the domain consists of a single profile, the theorem requires Pareto indifference only: unlimited domain and binary independence of irrelevant alternatives are not needed because there is only one profile (Blackorby, Donaldson and Weymark [1990]). And, as is well known, the theorem is true with multiple welfare profiles and a single non-welfare-information profile.

In both these cases, it would be wrong to conclude that non-welfare information is irrelevant. In the single-profile case, if all of the utility vectors are distinct, Pareto indifference imposes no restriction and it might be true that the principle uses only non-welfare information in ranking alternatives. This is consistent with the formal definition of welfarism; a single ordering of utility vectors exists and it can be used to order the elements of X. A similar observation can be made in the single-non-welfare-profile case.

When the domain of the social-evaluation functional consists of multiple profiles of welfare and non-welfare information, no such ambiguity exists. As Theorem 2 indicates, any principle with an unlimited domain must ignore non-welfare information. Our version of the welfarism theorem is, therefore, more powerful in this sense.

On a multi-profile domain, the welfarism theorem implies that any principle for social evaluation with an unlimited domain that uses non-welfare information must fail to satisfy Pareto indifference or binary independence of irrelevant alternatives. If it does not satisfy independence, it must be inconsistent across profiles. Because independence applies only to pairs of profiles for which welfare *and* non-welfare information coincide on a pair of alternatives, such inconsistency is not easily defended. On the other hand, if it does not satisfy Pareto indifference, it must also fail to satisfy minimal individual goodness. Such principles can have little ethical appeal as long as the account of well-being that is employed is a comprehensive one.

REFERENCES

Anderson, E., 1999, What is the point of equality?, *Ethics* **109**, 287–337.
Arneson, R., 1989, Equality of opportunity for welfare, *Philosophical Studies* **56**, 77–93.
Arneson, R., 2000a, Luck egalitarianism and prioritarianism, *Ethics* **110**, 339–349.
Arneson, R., 2000b, Perfectionism and politics, *Ethics* **111**, 37–63.
Arrow, K., 1951 (second ed. 1963), *Social Choice and Individual Values,* Wiley, New York.

Blackorby, C., W. Bossert and D. Donaldson, 1995, Intertemporal population ethics: critical-level utilitarian principles, *Econometrica* **63**, 1303–1320.

Blackorby, C., W. Bossert and D. Donaldson, 1998, Uncertainty and critical-level population principles, *Journal of Population Economics* **11**, 1–20.

Blackorby, C., W. Bossert and D. Donaldson, 2002, Utilitarianism and the theory of justice, in *Handbook of Social Choice and Welfare,* K. Arrow, A. Sen and K. Suzumura, eds., Elsevier, Amsterdam, forthcoming.

Blackorby, C. and D. Donaldson, 1992, Pigs and guinea pigs: a note on the ethics of animal exploitation, *Economic Journal* **102**, 1345–1369.

Blackorby, C., D. Donaldson and J. Weymark, 1990, A welfarist proof of Arrow's theorem, *Recherches Economique de Louvain* **56**, 259–286.

Blackorby, C., D. Donaldson and J. Weymark, 1999, Harsanyi's social aggregation theorem for state-contingent alternatives, *Journal of Mathematical Economics* **32**, 365–387.

Blackorby, C., D. Donaldson and J. Weymark, 2002, Social aggregation and the expected utility hypothesis, in *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls,* M. Salles and J. Weymark, eds., Cambridge University Press, Cambridge, forthcoming.

Blau, J., 1976, Neutrality, monotonicity, and the right of veto: a comment, *Econometrica* **44**, 603.

Bordes, G., P. Hammond and M. Le Breton, 1997, Social welfare functionals on restricted domains and in economic environments, unpublished manuscript, Stanford University, Department of Economics.

Bossert, W. and J. Weymark, 2002, Utility in social choice, in *Handbook of Utility Theory, vol. 2: Extensions,* S. Barberà, P. Hammond and C. Seidl, eds., Kluwer, Dordrecht, forthcoming.

Broome, J., 1991, *Weighing Goods,* Basil Blackwell, Oxford.

Broome, J., 1999, *Weighing Lives,* unpublished manuscript.

d'Aspremont, C. and L. Gevers, 1977, Equity and the informational basis of collective choice, *Review of Economic Studies* **44**, 199–209.

Fleurbaey, M. and F. Maniquet, 2002, Compensation and responsibility, in *Handbook of Social Choice and Welfare,* K. Arrow, A. Sen and K. Suzumura, eds., Elsevier, Amsterdam, forthcoming.

Gaertner, W., P. Pattanaik and K. Suzumura, 1992, Individual rights revisited, *Economica* **59**, 161–177.

Gevers, L., 1979, On interpersonal comparability and social welfare orderings, *Econometrica* **47**, 75–89.

Goodin, R., 1991, Utility and the good, in *A Companion to Ethics,* P. Singer, ed., Blackwell, Oxford, 241–248.

Griffin, J., 1986, *Well-Being: Its Meaning, Measurement, and Moral Importance,* Clarendon, Oxford.

Guha, A., 1972, Neutrality, monotonicity, and the right of veto, *Econometrica* **40**, 821–826.

Hammond, P., 1979, Equity in two person situations: some consequences, *Econometrica* **47**, 1127–1135.

Harsanyi, J., 1953, Cardinal utility in welfare economics and in the theory of risk-taking, *Journal of Political Economy* **61**, 434–435.

Harsanyi, J., 1955, Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility, *Journal of Political Economy* **63**, 309–321.

Kelsey, D., 1987, The role of information in social welfare judgements, *Oxford Economic Papers* **39**, 301–317.

Kolm, S.-C., 1972, *Justice et Equité,* Editions C.N.R.S., Paris.

Kolm, S.-C., 1996, *Modern Theories of Justice,* MIT Press, Cambridge.

Mill, J., 1969 (originally published in 1861), Utilitarianism, in *Collected Works of John Stuart Mill, vol. X,* University of Toronto Press, Toronto, 203–259.

Mongin, P. and C. d'Aspremont, 1998, Utility theory in ethics, in *Handbook of Utility Theory, vol. 1: Principles,* S. Barberà, P. Hammond and C. Seidl, eds., Kluwer, Dordrecht, 371–481.

Nussbaum, M., 2000a, *Women and Human Development: The Capabilities Approach,* Cambridge University Press, Cambridge.

Nussbaum, M., 2000b, The central human capabilities, *Ethics* **111**, 6–7.

Roemer, J., 1996, *Theories of Distributive Justice,* Harvard University Press, Cambridge.

Sen, A., 1970a, *Collective Choice and Social Welfare,* Holden-Day, San Francisco.

Sen, A., 1970b, The impossibility of a Paretian liberal, *Journal of Political Economy* **78**, 152–157.

Sen, A., 1977, On weights and measures: informational constraints in social welfare analysis, *Econometrica* **45**, 1539–1572.

Sen, A., 1979, Personal utilities and public judgements: or what's wrong with welfare economics?, *Economic Journal* **89**, 537–558.

Sen, A., 1985, *Commodities and Capabilities,* Elsevier, Amsterdam.

Sen, A., 1987, *The Standard of Living,* Cambridge University Press, Cambridge.

Sen, A. and B. Williams, 1982, Introduction, in *Utilitarianism and Beyond,* A. Sen and B. Williams, eds., Cambridge University Press, Cambridge, 1–21.

Sidgwick, H., 1966 (originally published in 1907), *The Methods of Ethics,* Dover, New York.

Sumner, L., 1996, *Welfare, Happiness, and Ethics,* Clarendon, Oxford.

Weymark, J., 1998, Welfarism on economic domains, *Mathematical Social Sciences* **36**, 251–268.