A Theorem on Utilitarianism

ERIC MASKIN
Harvard University

In a recent paper [1], d'Aspremont and Gevers establish that if a social welfare functional (SWFL) satisfies several reasonable properties, and if interpersonal comparisons of absolute welfare levels are prohibited, although unit interpersonal comparisons of welfare are not (for a discussion of unit and level welfare comparisons, see Sen [9]), the SWFL must be the principle of utilitarianism. In this paper we derive a similar result when full comparability of welfare is permitted, so that both interpersonal levels and units are significant. By characterizing utilitarianism under full comparability, we avoid the somewhat arbitrary informational basis from which the d'Aspremont-Gevers result is derived (although we are still, of course, open to criticism of interpersonal comparisons in general). Our result also gives a new slant to the contrast between utilitarianism and the lexicographic maximin principle (Rawls' generalized difference principle [6]), which Hammond [4] has characterized under full comparability.

In another paper [5], we point out that the theorem is actually an immediate consequence of a proposition from the theory of decision-making under ignorance. Rather than exploiting here the isomorphism between social choice and individual decision theory, however, we shall use results, as well as terminology, from the d'Aspremont-Gevers paper.

Let X be a set of social alternatives containing at least three elements. Let

$$N = \{1, 2, 3, \ldots, n\}$$

be a set of individuals who constitute society. Let $\mathscr{R}$ be the set of all orderings of X and $\mathscr{U}$ the set of all bounded real valued functions on $X \times N$. For $u(\cdot, \cdot) \in \mathscr{U}$, $u(x, i)$ is the utility that the ith individual derives from alternative x. (We implicitly assume here that the underlying preference ordering of X, $\succsim_i$, is, for each individual i, representable by a numerical utility function. This would automatically be the case if X contained at most countably many elements). A social welfare functional (SWFL) f is a mapping $f: \mathscr{U} \to \mathscr{R}$. This definition of a SWFL is equivalent to that given by Sen [8]. Note that it is very similar to Hammond's [4] notion of a generalized social welfare function.

The following are properties which d'Aspremont and Gevers propose that a SWFL f satisfy.

Independence. $\forall u, u' \in \mathscr{U}$, $\forall B \subseteqq X$, if $u(x, \cdot) = u'(x, \cdot)$ for all $x \in B$, then $f(u)$ and $f(u')$ coincide on B.

Strong Pareto Principle. $\forall x, y \in X$, $\forall u \in \mathscr{U}$ (1) if $\forall i \in N$, $u(x, i) = u(y, i)$, then xIy, and (2) if $\forall i \in N$, $u(x, i) \geqq u(y, i)$ and $\exists j \in N$ such that $u(x, j) > u(y, j)$, then xPy, where I and P are, respectively, the indifference and strict preference relations corresponding to $f(u)$.

Anonymity. For any permutation σ of N, if for

$$u, u' \in \mathscr{U}, \forall i \in N, \forall x \in X, u(x, i) = u'(x, \sigma(i)),$$

then $f(u) = f(u')$.

Elimination of Indifferent Individuals. $\forall u^1, u^2 \in \mathscr{U}$ if $\exists M \subseteqq N$ such that

$$\forall i \in M, u^1(\cdot, i) = u^2(\cdot, i) \quad \text{while} \quad \forall j \in N \backslash M, \forall x, y \in X, u^1(x, j) = u^1(y, j)$$

and $u^2(x, j) = u^2(y, j)$, then $f(u^1) = f(u^2)$.

The elimination of indifferent individuals is essentially Debreu's strong separability condition [3]. It is also a weakened form of Strasnick's condition of unanimity [7].

Equity. $\forall u \in \mathcal{U}$, $\forall x, y \in X$, $\forall i, j \in N$ if $\forall g \in N \backslash \{i, j\}$, $u(x, g) = u(y, g)$ and

$$u(y, i) < u(x, i) < u(x, j) < u(y, j)$$

then $xf(u)y$.

The property of equity, while satisfied by Rawl's difference principle, is, in fact, violated by the principle of utilitarianism. We mention it here because it will help to contrast the two principles.

D'Aspremont and Gevers have shown that a SWFL f which satisfies independence and the strong Pareto property induces an ordering $\hat{R}$ of all of $\mathbb{R}^n$ (n-dimensional Euclidean space) in the sense that for any $\bar{x}, \bar{y} \in \mathbb{R}^n$, $\bar{x}\hat{R}\bar{y}$ if and only if $\exists u \in \mathcal{U}$ and $x, y \in X$ such that $(u(x, 1), ..., u(x, n)) = \bar{x}$, $(u(y, 1), ..., u(y, n)) = \bar{y}$, and xRy, where $R = f(u)$. The ordering $\hat{R}$ is said to be continuous if for all $\bar{y} \in \mathbb{R}^n$, the sets $\{\bar{x} \mid \bar{x}\hat{R}\bar{y}\}$ and $\{\bar{x} \mid \bar{y}\hat{R}\bar{x}\}$ are closed. We, therefore, propose the following definition:

Continuity. A SWFL f satisfying independence and the strong Pareto property is continuous if its induced ordering $\hat{R}$ of $\mathbb{R}^n$ is continuous.

Continuity may be considered a reasonable property in that it requires that two alternatives which are nearly as desirable as one another in the eyes of all individuals should bear nearly the same social relationship to all other alternatives. By virtue of Debreu's representation theorem [2], continuity of f implies that f can be represented by a continuous function of individual utilities.

The final condition is not so much a property of f as a description of what kinds of welfare comparisons can be made.

Full Comparability. $\forall u, u' \in \mathcal{U}$ if $\exists b > 0$, $a \in \mathbb{R}$ such that

$$\forall i \in N, \forall x \in X, u(x, i) = bu'(x, i) + a,$$

then $f(u) = f(u')$.

We may now state our result.

Theorem. *If a* SWFL f *satisfies independence, the strong Pareto property, anonymity, elimination of indifferent individuals, continuity, and full comparability, it is the principle of utilitarianism. That is,* $\forall x, y \in X$, $\forall u \in \mathcal{U}$, $xf(u)y$ *if and only if*

$$\Sigma_i u(x, i) \geqq \Sigma_i u(y, i), \quad i = 1, ..., n.$$

One should note that continuity plays a central role in this theorem. With continuity, independence and the strong Pareto property alone, a " generalized " utilitarianism can be inferred. That is, there exists a continuous function $h : \mathbb{R}^n \to \mathbb{R}$ such that $\forall u \in \mathcal{U}$, $\forall x, y \in X$, $xf(u)y$ if and only if $h(u(x, 1), ..., u(x, n)) \geqq h(u(y, 1), ..., u(y, n))$. Elimination of indifferent individuals ensures that h have the form $h(u(x, 1), ..., u(x, n)) = \Sigma_i g_i (u(x, i))$. Anonymity implies that all the g_i's are the same, and full comparability is needed to derive the linearity of the g_i's. It may seem strange to think of full comparability as a restrictive assumption. After all, it is weak enough to permit both unit and level comparisons. One can imagine, however, situations in which invariance of the social ordering need not be invoked whenever individual utilities are shifted by a positive affine transformation. Sen [10], for example, suggests that for some problems there may be a " distinguished point of utilities below which misery dominates ". In such situations, full comparability's invariance, which takes no account of absolute utility levels, would be too strong to make all desired comparisons.

The following table is useful for comparing utilitarianism (U) to the lexicographic maximin (L) when both are considered under the assumption of full comparability.

	L	U
Independence	*	*
Strong Pareto property	*	*
Anonymity	*	*
Elimination of Indifferent Individuals	*	*
Continuity		*
Equity	*	

* denotes satisfaction of the property.

We conclude with a formal proof of the above theorem.

Proof. The proof is essentially an application of Debreu's theorem on additivity separability [3]. By independence and the strong Pareto property, f induces an ordering $\hat{R}$ on $\mathbb{R}^n$. $\hat{R}$ is continuous by hypothesis. By elimination of indifferent individuals, the ordering of $\mathbb{R}^{n-m}$ that $\hat{R}$ induces when m components ($0 \leqq m \leqq n$) are fixed is independent of the level at which the components are fixed. Since $\mathbb{R}^n$ is trivially connected, all the hypotheses of Debreu's theorem are satisfied, and we may conclude that there exist continuous functions $g_1, g_2, \ldots, g_n : \mathbb{R} \to \mathbb{R}$ such that $\forall u \in \mathscr{U}$, $\forall x, y \in X$, $xf(u)y$ if and only if $\Sigma_i g_i(u(x, i)) \geqq \Sigma_i g_i(u(y, i))$. By anonymity, all the g_i's must be the same; i.e. equal to some $g : \mathbb{R} \to \mathbb{R}$. By the strong Pareto property, g is strictly monotone increasing. It remains only to show that g is linear.

Choose $u(0<u<1)$ such that $2g(u) = g(1)+g(0)$. u exists and is unique by the monotonicity and continuity of g. Since g is unique only up to positive linear transformations, we may assume without loss of generality that

$$g(0) = 0 \quad \ldots(1)$$

$$g(1) = 1. \quad \ldots(2)$$

By full comparability, $(u, u, 0, 0, \ldots, 0)\hat{I}(1, 0, 0, \ldots, 0)$ implies that

$$((a-b)u+b, (a-b)u+b, b, b, \ldots, b)\hat{I}(a, b, b, \ldots, b) \quad \text{for} \quad a \geqq b$$

where $\hat{I}$ is the indifference relation corresponding to $\hat{R}$.

Therefore,

$$2g((a-b)u+b) = g(a)+g(b) \quad \text{for} \quad a \geqq b. \quad \ldots(3)$$

Using equations (1), (2) and (3), we may successively take

(i) $a = 1$, $b = 0$ to obtain

$$g(u) = \tfrac{1}{2}. \quad \ldots(4)$$

(ii) $a = 1$, $b = $ u to obtain

$$g((2-u)u) = \tfrac{3}{4}. \quad \ldots(5)$$

(iii) $a = 2-u$, $b = 0$ to obtain

$$g(2-u) = \tfrac{3}{2}. \quad \ldots(6)$$

(iv) $a = 2-u$, $b = 1$ to obtain

$$g(u-u^2+1) = \tfrac{5}{4}. \quad \ldots(7)$$

By monotonicity, $u-u^2+1>(2-u)u$. Taking $a = u-u^2+1$ and $b = (2-u)u$ in equation (3), we obtain

$$2g([(u-u^2+1)-(2-u)u]u+(2-u)u) = g((2-u)u)+g(u-u^2+1). \quad \ldots(8)$$

Simplifying using (5) and (7), we conclude that $g(-2u^2+3u) = 1$. By the monotonicity of g and equation (2), $-2u^2+3u = 1$. Thus $u = \frac{1}{2}$ and

$$g(\tfrac{1}{2}(a+b)) = \tfrac{1}{2}(g(a)+g(b)) \quad \text{for all} \quad a, b \in \mathbb{R}. \quad \ldots(9)$$

Using (9) and taking $a = b = 1$, we obtain $g(\frac{1}{2}) = \frac{1}{2}$. Taking $a = 1$, $b = \frac{1}{2}$, we obtain $g(\frac{3}{4}) = \frac{3}{4}$. Taking $a = 0$, $b = \frac{1}{2}$, we obtain $g(\frac{1}{4}) = \frac{1}{4}$. By continuing iteratively, we may show that for any integers n, k, $g(k/2^n) = k/2^n$. The set $\{k/2^n\}$ is dense in $\mathbb{R}$, so, by continuity, $g(x) = x$ for all $x \in \mathbb{R}$. ||

First version received November 1975; *final version accepted March*, 1976 (*Eds*.).

I wish to thank F. H. Hahn and A. K. Sen for valuable comments.

REFERENCES

[1] d'Aspremont, C. and Gevers, L. " Equity and the Informational Basis of Collective Choice ", Faculté des Sciences Économiques et Sociales, Working Paper 7-5, July 1975.

[2] Debreu, G. " Representation of a Preference Ordering by a Numerical Function ", in Thrall, Coombs and Davis (eds.), *Decision Processes* (New York: Wiley, 1954).

[3] Debreu, G. " Topological Methods in Cardinal Utility Theory ", in Arrow, Karlin, Suppes (eds.), *Mathematical Methods in the Social Sciences* (Stanford: Stanford University Press, 1960).

[4] Hammond, P. J. " Equity, Arrow's Conditions and Rawls' Difference Principle ", mimeographed, 1975; forthcoming in *Econometrica.*

[5] Maskin, E. " Decision-making under Ignorance with Implications for Social Choice ", mimeographed, Cambridge University, 1975.

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

[7] Strasnick, S. " Arrow's Paradox and Beyond ", mimeographed, Harvard University, 1975.

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

[9] Sen, A. K. " Informational Bases of Alternative Welfare Approaches: Aggregation and Income Distribution ", *Journal of Public Economics*, 3 (1974), 387-403.

[10] Sen, A. K. " Interpersonal Comparisons of Welfare ", mimeographed, London School of Economics, 1975; forthcoming in a Festchrift for Tibor Scitovsky.