Available online at www.sciencedirect.com
SCIENCE DIRECT®
Economics Letters 88 (2005) 85–90
economics letters
www.elsevier.com/locate/econbase

Harsanyi's utilitarianism via linear programming

Michael Mandler

Department of Economics, Royal Holloway College, University of London, Egham, Surrey TW20 0EX, United Kingdom

Received 23 March 2004; accepted 27 October 2004
Available online 21 April 2005

Abstract

A simple linear programming problem permits a brief and elementary proof of Harsanyi's utilitarianism theorem: a Paretian social welfare function must be a weighted (affine) sum of individual utility functions when individual utilities and social welfare all take the Neumann–Morgenstern form. By adjusting the programming problem slightly, we conclude that the weights on individual utilities are positive or semi-positive when more demanding Pareto principles hold. The reasoning extends easily to cover sets of social choices that equal arbitrary mixture spaces.

Keywords: Utilitarianism; Linear programming; Mixture space; Harsanyi

JEL classification: C6; D6; D7

1. Introduction

Harsanyi (1955) argued that if both social welfare and individual utility functions take the von Neumann–Morgenstern (vNM) form, then any social welfare function that satisfies the Pareto principle must be a weighted sum of individual utility functions. This cannot always be true, since for a society of one individual with the utility $U(p)=p_1+2p_2$ for probabilities over two pure social choices, the social welfare function $2p_1+3p_2$ is Paretian but not a scalar multiple of $U(p)$. Later formulations have therefore rephrased Harsanyi's theorem as stating that a vNM and Paretian social welfare function is a weighted sum of individual utilities plus a constant.

E-mail address: m.mandler@rhul.ac.uk.

0165-1765/$ - see front matter
doi:10.1016/j.econlet.2004.10.017

Domotor (1979) was the first to prove this result without the additional assumptions that Harsanyi implicitly imposed in his original article. Border (1981), reported in Selinger (1986) and Weymark (1994), offers an accessible proof under the mild Pareto principle that if all agents are indifferent between a pair of probability distributions then so is society; see also Hammond (1992). Domotor also showed that stronger Pareto principles lead to positive or semi-positive weights on individual utilities.

This note provides a simple and very short proof that poses Harsanyi's theorem as a linear programming problem; the weights on individual utilities are simply the dual variables or Lagrange multipliers that arise at the solution of the problem, and which exist whether or not a constraint qualification holds (see, e.g., Dantzig (1963, ch. 6, Theorems 3, 4)). One way to show that dual variables exist is to apply the same Farkas lemmas on linear inequalities that Border (1981) utilizes in his proof. But directly invoking the existence of the dual variables shortens the entire argument to two sentences and appeals to the mathematical result most widely known in economic theory. To conclude that the weights on individual utilities are positive or semi-positive requires stronger Pareto principles, and then the necessary Farkas arguments become more intricate, particularly for positive weights (Weymark, 1994; Turunen-Red and Woodland, 1999). The linear programming proofs remain just as short or almost so.

The linear programming proofs and most of the literature on Harsanyi apply to preferences over probability distributions on a finite number of social choices. Harsanyi's theorem also holds when agents and society have preferences over probability measures on a measurable space or, even more generally, over lotteries in a mixture space (Domotor, 1979; Fishburn, 1984; Border, 1985; Hammond, 1992; Coulhon and Mongin, 1989; De Meyer and Mongin, 1995; Zhou, 1997). As we will see, the general mixture space setting reduces readily to the finite case: select a finite number of lotteries and their mixtures and ignore the rest of the domain. To prove the general Harsanyi theorem, one simply notes that if it were false, there would be finitely many lotteries on which social welfare does not equal a linear combination of individual utilities (plus a constant), and that would contradict the finite Harsanyi theorem. Existing proofs of the general Harsanyi theorem, even when relatively short, are not technology free; Border (1985) may be the simplest and even it relies on a separation theorem applied to the set of measurable functions. Since proving Harsanyi's theorem amounts to finding a finite set of weights on individual utilities (even when the utilities themselves are defined on a rich space), it is only fitting to rely on nothing more than finite-dimensional separation, whether expressed from the Farkas or linear programming point of view.

Given the reduction of the mixture case to the finite case, it is straightforward to show again that the weights on individual utilities are positive or semi-positive under appropriately stronger Pareto conditions, results that seem not to have been given elementary proofs.

2. Harsanyi's theorem with a finite number of pure social choices

Each agent j in a finite set J has preferences over lotteries with prizes in a finite set S of social choices. Letting $\Delta = \left\{p \in R_{+}^{\#S} : \sum_{i \in S} p_i = 1\right\}$ denote these lotteries, we assume that each j has a von Neumann–Morgenstern utility $U^j\colon\Delta \rightarrow R$. That is, for each $j \in J$ there is a $u^j \in R^S$ such that, for all $p \in \Delta$, $U^j(p) = p \cdot u^j$. Social preferences over Δ are given by a von Neumann–Morgenstern utility $W\colon\Delta \rightarrow R$: there is a $w \in R^S$ such that $W(p) = p \cdot w$ for all $p \in \Delta$. Let Δ_{++} denote $\Delta \cap R_{++}^{\#S}$.

Definition 1. W and $U=(U^1,\ldots,U^J)$ satisfy

(i) Pareto indifference iff for all $p, p' \in \Delta$, $U^j(p)=U^j(p')$ for each $j \in J \Rightarrow W(p)=W(p')$,
(ii) Semi-strong Pareto iff for all $p, p' \in \Delta$, $U^j(p) \geq U^j(p')$ for each $j \in J \Rightarrow W(p) \geq W(p')$,
(iii) Strong Pareto iff semi-strong Pareto and for all $p, p' \in \Delta$, $U^j(p) \geq U^j(p')$ for each $j \in J$ and $U^i(p) > U^i(p')$ for some $i \in J \Rightarrow W(p) > W(p')$.

Our terminology in Definition 1 follows Weymark (1993). Notice that (iii)⇒(ii)⇒(i).

Theorem 1. *If W and U satisfy Pareto indifference, there are $\alpha \in R^{\#J}$ and $\beta \in R$ such that for all $p \in \Delta$,*

$$W(p) = \sum_{j \in J} \alpha^j U^j(p) + \beta.$$

If W and U satisfy semi-strong Pareto, we may set α to be in $R_+^{\#J}$, and if they satisfy strong Pareto, we may set α to be in $R_{++}^{\#J}$.

The equality in Theorem 1 may also be expressed as $w = \sum_{j \in J} \alpha^j u^j + \beta e$, where e is the vector of $\#S$ 1's. We stress that Theorem 1's second sentence says "may" rather than "must": if, e.g., there are two agents i and j with $U^i = U^j$ then evidently we may always set either α^i or α^j to be negative.

Proof. If semi-strong Pareto holds, then for any given $p' \in \Delta$ the problem

$$\min\ W(p) \text{s.t.} U^j(p) \geq U^j(p'), j \in J, \sum_{s \in S} p_s = 1, p \geq 0,$$

must be solved at p' since otherwise there would be $p'' \in \Delta$ such that $W(p'') < W(p')$ and $U^j(p'') \geq U^j(p')$ for $j \in J$, which would violate semi-strong Pareto. Hence there exist Lagrange multipliers $a=(a^1,\ldots,a^J) \in R_+^{\#J}$ and $b \in R$ that satisfy the problem's first order condition, which, if we select $p' \in \Delta_{++}$, is $w = \sum_{j \in J} a^j u^j + be$.

If Pareto indifference holds, the argument is identical except that for each $j \in J$ we replace each constraint $U^j(p) \geq U^j(p')$ by $U^j(p) = U^j(p')$. This modified problem must be solved at p' since otherwise there would be a $p'' \in \Delta$ such that $W(p'') < W(p')$ and $U^j(p'') = U^j(p')$ for all j. Now the Lagrange multipliers a lie in $R^{\#J}$ rather than $R_+^{\#J}$.

If strong Pareto holds, then given $p' \in \Delta$ and $i \in J$ the problem

$$\max\ U^i(p) \text{s.t.} U^j(p) \geq U^j(p'), j \in J \setminus \{i\}, W(p) \leq W(p'), \sum_{s \in S} p_s = 1, p \geq 0,$$

is solved at p' since otherwise there would be a $p'' \in \Delta$ such that $U^j(p'') \geq U^j(p')$ for all $j \in J$, $U^i(p'') > U^i(p')$ for i, and nevertheless $W(p'') \leq W(p')$. By again setting $p' \in \Delta_{++}$, we conclude that, for each $i \in J$, there are Lagrange multipliers $(\gamma_i^1,\ldots,\gamma_i^{i-1},\gamma_i^{i+1},\ldots\gamma_i^J) \in R_+^{\#J-1}$, $\delta_i \in R$, and $\varepsilon_i \in R_+$ such that

$$u^i + \sum_{j \in J \setminus \{i\}} \gamma_i^j u^j + \delta_i e = \varepsilon_i w.$$

Sum these J equalities and $w=\sum_{j\in J}a^j u^j+be$, where $a\in R_+^{\#J}$ and $b\in R$ are derived as in the semi-strong case:

$$\left(1+\sum_{i\in J}\varepsilon_i\right)w=\sum_{i\in J}\left(1+a^i+\sum_{j\in J\setminus\{i\}}\gamma_j^i\right)u^i+\left(b+\sum_{i\in J}\delta_i\right)e.$$

Dividing by $(1+\sum_{i\in J}\varepsilon_i)$ concludes the proof. □

3. Harsanyi's theorem for an arbitrary mixture space

For $k>1$, let Δ^{k-1} denote $\left\{r\in R_+^k\colon \sum_{i=1}^k r_i=1\right\}$. Let S be a mixture space where for lotteries s_1, $s_2\in S$ and $p\in\Delta^1$, the lottery $p_1s_1+p_2s_2\in S$ denotes their binary mixture.[1] Taking binary mixtures as primitive, one defines finite mixtures by induction: if for any $t_1,\ldots,t_n\in S$ and $q\in\Delta^{n-1}$, $\sum_{i=1}^n q_it_i$ is defined, then for $s_1,\ldots,s_{n+1}\in S$ and $p\in\Delta^n$, set

$$\sum_{i=1}^{n+1}p_is_i=p_{n+1}s_{n+1}+(1-p_{n+1})\sum_{i=1}^{n}\frac{p_i}{1-p_{n+1}}s_i.$$

Each agent j in the finite set J has a utility $V^j\colon S\rightarrow R$ that satisfies the linearity property: for all $p\in\Delta^1$ and s_1, $s_2\in S$, $V^j(p_1s_1+p_2s_2)=p_1V^j(s_1)+p_2V^j(s_2)$. Social preferences are represented by a $W\colon S\rightarrow R$ satisfying the same linearity property. For our purposes, what is important is that if $T\subset S$ consists of the mixtures of a finite number of lotteries in S then $V^j|T$ is an expected utility function. That is, given $s_1,\ldots,s_n\in S$ and $T=\left\{s\in S\colon s=\sum_{i=1}^n q_is_i \text{ for some } q\in\Delta^{n-1}\right\}$, then $V^j|T$ is given by $V^j\left(\sum_{i=1}^n q_is_i\right)=\sum_{i=1}^n q_iV^j(s_i)$.

Our previous definition of the various Pareto principles applies unchanged to W and $V=(V^1,\ldots,V^J)$ if we replace Δ in Definition 1 by S.

Theorem 2. *If W and V satisfy Pareto indifference, there exist $\alpha\in R^{\#J}$ and $\beta\in R$ such that, for all $p\in S$, $W(p)=\sum_{j\in J}\alpha^jV^J(p)+\beta$. If W and V satisfy semi-strong Pareto, we may set α to be $R_+^{\#J}$, and if they satisfy strong Pareto, we may set α to be in $R_{++}^{\#J}$.*

Proof. Viewing V^j and W as elements of the vector space of functions from S to R, let $B=\{Y^1,\ldots,Y^m\}$ be a basis for sp $\{\mathbf{1}, V^1,\ldots,V^J\}$, where sp denotes span and $\mathbf{1}\colon S\rightarrow R$ equals 1 everywhere.

Suppose W and V satisfy Pareto indifference. If the conclusion given in the theorem were false, then $W\notin \text{sp } B$ and so $\{Y^1,\ldots,Y^m,Y^{m+1}\equiv W\}$ would be linearly independent. As we show below, linear independence then implies there are $\hat{s}_1,\ldots,\hat{s}_{m+1}\in S$ such that $\{\hat{y}^1,\ldots,\hat{y}^m,\hat{y}^{m+1}\equiv\hat{w}\}$ is linearly independent, where $\hat{y}^k\in R^{m+1}$ is the vector whose ith coordinate is $Y^k(\hat{s}_i)$. Linear independence implies there is no $a\in R^m$ with $\sum_{j=1}^m a^j\hat{y}^j=\hat{w}$. Since B is a basis for sp $\{\mathbf{1}, V^1,\ldots,V^J\}$, whenever $Y^j(\hat{s})=Y^j(t)$ for $j\in\{1,\ldots,m\}$ then $V^j(s)=V^j(t)$ for all $j\in J$. Hence since W and V satisfy Pareto

[1] Formally, a mixture space is defined via a function from $S\times S\times\Delta^1$ to S whose image at (s_1, s_2, p) is denoted $p_1s_1+p_2s_2$ and where, for all s_1, $s_2\in S$ and p, $q\in\Delta^1$, (i) $p_1s_1+p_2s_2=p_2s_2+p_1s_1$, (ii) $1s_1+0s_2=s_1$, and (iii) $p_1(q_1s_1+q_2s_2)+p_2s_2=p_1q_1s_1+(1-p_1q_1)s_2$.

indifference, Pareto indifference holds for W and Y^j, $j\in\{1,\ldots, m\}$ and also therefore for W and the Y^j when restricted to $\left\{s\in S\colon s=\sum_{i=1}^{m+1} q_i\hat{s}_i \text{ for some } q\in\Delta^m\right\}$. Since $W\left(\sum_{i=1}^{m+1} q_i\hat{s}_i\right)=q\cdot\hat{w}$ and $Y^j\left(\sum_{i=1}^{m+1} q_i\hat{s}_i\right)=q\cdot\hat{y}^j$ for all $q\in\Delta^m$, Theorem 1 implies there are $a'\in R^m$ and $b'\in R$ such that $\sum_{j=1}^{m} a^{j\prime}\hat{y}^j+b'e=\hat{w}$, where e is the vector of $m+1$ 1's. Since $b'e\in\text{sp}\ \{\hat{y}^1,\ldots,\hat{y}^m\}$ there exists $a\in R^m$ such that $\sum_{j=1}^{m} a^j\hat{y}^j=\hat{w}$. This contradiction establishes the first sentence of the Theorem.

We have invoked the fact that if $X^1,\ldots, X^n$ are linearly independent real-valued functions on S, then there are $t_1,\ldots, t_n\in S$ such that $\{\hat{x}^1,\ldots,\hat{x}^n\}$ is linearly independent, where $\hat{x}^k\in R^n$ has the ith coordinate $X^k(t_i)$. This practically goes without proof but we give the details. Since $\{X^1,\ldots, X^n\}$ is linearly independent, X^1 cannot everywhere equal 0 and so there is $t_1\in S$ such that $\{\hat{x}_1^1\equiv X^1(t_1)\}$ is linearly independent. Given $\{t_1,\ldots, t_n\}$, X^k, and $h\leq n$, let $\hat{x}_h^k$, $k=1,\ldots, n$, denote the vector in R^h with ith coordinate $X^k(t_i)$, $i=1,\ldots, h$. Proceeding by induction, suppose for any $h\in\{1,\ldots, n-1\}$ that there are $t_1,\ldots, t_h$ such that $\{\hat{x}_h^1,\ldots,\hat{x}_h^h\}$ is linearly independent. Then there must be a $\mu\in R^h$ such that $\sum_{j=1}^{h}\mu^j\hat{x}_h^j=\hat{x}_h^{h+1}$. Since $X^1,\ldots, X^h, X^{h+1}-\sum_{j=1}^{h}\mu^jX^j$ are linearly independent, there is a t_{h+1} such that $X^{h+1}(t_{h+1})-\sum_{j=1}^{h}\mu^jX^j(t_{h+1})\neq 0$. So $\{\hat{x}_{h+1}^1,\ldots,\hat{x}_{h+1}^h,\hat{x}_{h+1}^{h+1}-\sum_{j=1}^{h}\mu^j\hat{x}_{h+1}^j\}$ and hence $\left\{\hat{x}_{h+1}^1,\ldots,\hat{x}_{h+1}^{h+1}\right\}$ are linearly independent.

We turn to the semi-strong and strong cases. Since B is linearly independent, we may pick r_1, ..., $r_m\in S$ such that $\bar{B}=\{\bar{y}^1,\ldots,\bar{y}^m\}$ is linearly independent, where each $\bar{y}^k\in R^m$ has ith coordinate equal to $Y^k(r_i)$, $i=1,\ldots, m$. Define $\bar{v}^j,\bar{w}\in R^m$ by setting their ith coordinates equal to $V^j(r_i)$ and $W(r_i)$, respectively.

Lemma. *If Pareto indifference holds and* $(\text{a,b})\in R^{\#J+1}$ *satisfies* $\sum_{j\in J} a^j\bar{v}^j+be=\bar{w}$, *then* $\sum_{j\in J} a^jV^j+b\mathbf{1}=W$.

Proof of lemma. We have shown already that Pareto indifference implies $W\in\text{sp}\ \{\mathbf{1}, V^1,\ldots, V^J\}$. Hence, since B is a basis for sp $\{\mathbf{1}, V^1,\ldots, V^J\}$, there exists $a'\in R^m$ such that $\sum_{j=1}^{m} a^{j\prime}Y^j=W$. Hence $\sum_{j=1}^{m} a^{j\prime}\bar{y}^j=\bar{w}$, and since $\bar{B}$ is linearly independent, a' is the only solution to this equality. Similarly there is a $\gamma\in R^m$ such that $\sum_{j=1}^{m}\gamma^jY^j=\sum_{j\in J} a^jV^j+b\mathbf{1}$ and therefore $\sum_{j=1}^{m}\gamma^j\bar{y}^j=\bar{w}$. So $\gamma=a'$ and $\sum_{j\in J} a^jV^j+b\mathbf{1}=W$, proving the lemma.

If semi-strong Pareto holds, apply Theorem 1 to the utilities $\bar{V}^j$ and social welfare $\bar{W}$, each defined on Δ^{m-1}, and given respectively by $\bar{V}^j(q)=q\cdot\bar{v}^j, j\in J$, and $\bar{W}(q)=q\cdot\bar{w}$ to conclude there are $a\in R_+^{\#J}$ and $b\in R$ such that $\sum_{j\in J} a^j\bar{v}^j+be=\bar{w}$. The lemma then implies $\sum_{j\in J} a^jV^j+b\mathbf{1}=W$. If strong Pareto holds, apply Theorem 1 to conclude that $a\in R_{++}^{\#J}$, and again apply the lemma. □

Acknowledgement

My thanks to John Weymark for several helpful suggestions.

References

Border, K., 1981. Notes on von Neumann–Morgenstern Social Welfare Functions, Mimeo. California Institute of Technology.
Border, K., 1985. More on Harsanyi's utilitarian cardinal welfare theorem. Social Choice and Welfare 1, 279–281.
Coulhon, T., Mongin, P., 1989. Social choice theory in the case of von Neumann–Morgenstern utilities. Social Choice and Welfare 6, 175–187.
Dantzig, G., 1963. Linear Programming and Extensions. Princeton University Press, Princeton.

De Meyer, B., Mongin, P., 1995. A note on affine aggregation. Economics Letters 47, 177–183.
Domotor, Z., 1979. Ordered sum and tensor product of linear utility structures. Theory and Decision 11, 375–399.
Fishburn, P., 1984. On Harsanyi's utilitarian cardinal welfare theorem. Theory and Decision 17, 21–28.
Hammond, P., 1992. Harsanyi's utilitarian theorem: a simpler proof and some ethical considerations. In: Selten, R. (Ed.), Rational Interaction: Essays in Honor of John C. Harsanyi. Springer, Berlin, pp. 187–196.
Harsanyi, J., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy 63, 309–321.
Selinger, S., 1986. Harsanyi's aggregation theorem without selfish preferences. Theory and Decision 20, 53–62.
Turunen-Red, A., Woodland, A., 1999. On economic applications of the Kuhn–Fourier theorem. In: Wooders, M. (Ed.), Topics in Mathematical Economics and Game Theory: Essays in Honor of Robert J. Aumann. American Mathematical Society, Providence, RI, pp. 257–276.
Weymark, J., 1993. Harsanyi's social aggregation theorem and the weak Pareto principle. Social Choice and Welfare 10, 209–221.
Weymark, J., 1994. Harsanyi's social aggregation theorem with alternative Pareto principles. In: Eichhorn, W. (Ed.), Models and Measurement of Welfare and Inequality. Springer, Berlin, pp. 869–887.
Zhou, L., 1997. Harsanyi's utilitarianism theorems: general societies. Journal of Economic Theory 72, 198–207.