Roberts' Weak Welfarism Theorem: A Minor Correction

Peter J. Hammond

Department of Economics, Stanford University, CA 94305–6072, U.S.A.

e-mail: `hammond@leland.stanford.edu`

Preliminary Version: July 1996; minor revisions March 1997 and September 1999.

1. Introduction

All notation and definitions are taken from Roberts (1980). Let f denote a social welfare functional (SWFL) mapping the domain $\mathcal{U}$ of utility function profiles into social preference orderings on a given set X of at least three social states. An important result in social choice theory with interpersonal comparisons is the "strong neutrality" or "welfarism" result due to d'Aspremont and Gevers (1977) and Sen (1977, p. 1553). This states that, when f satisfies conditions (U) (unrestricted domain), (I) (independence of irrelevant alternatives), and (P^0) (Pareto indifference), then there exists a *social welfare ordering* R^* on the Euclidean space E^N with the property that $x\ R\ y \iff u(x,\cdot)\ R^*\ u(y,\cdot)$. This result plays a prominent role in the surveys by Sen (1984), Blackorby, Donaldson and Weymark (1984), d'Aspremont (1985), Mongin and d'Aspremont (1998), and Bossert and Weymark (1999). Both Sen (1977) and d'Aspremont (1985, p. 34) provide complete proofs.[1]

While Pareto indifference is an appealing condition, the Arrow impossibility theorem replaces it with the alternative ordinary weak Pareto condition (P). To develop a theory general enough to cover this important case, Roberts (1980, p. 428) announced the following result:

Theorem 1. *If f satisfies (U), (I), (P), and (WC) then there exists a continuous real-valued function, W, increasing with an increase in all arguments, with the property that for all $u \in \mathcal{U}$, $x, y \in X$, if $W(u(x,\cdot)) > W(u(y,\cdot))$ then $x\ P\ y$.*

This has come to be known as Roberts' "weak neutrality" or "weak welfarism" theorem. It was cited as an alternative to strong neutrality in many of the surveys mentioned above.

[1] Unfortunately, d'Aspremont's proof, which is otherwise the more elegant of the two, includes a crucial typographical error. The option e should be chosen so that $b \neq e \neq d$.

The unpublished results by Le Breton (1987) and by Bordes and Le Breton (1987) investigating Roberts' theorem for restricted economic domains have recently been amalgamated with related results that appear in Bordes, Hammond and Le Breton (1999).

Of the assumptions included in the above result, Roberts (p. 427) specifies the fourth *weak continuity* condition as follows:

Condition (WC): For all $u \in \mathcal{U}$, $\epsilon \in E^N$, $\epsilon \gg 0$, there exists a $u' \in \mathcal{U}$ satisfying $\epsilon \gg u(x, \cdot) - u'(x, \cdot) \gg 0$ for all $x \in X$ such that $f(u) = f(u')$.

Unfortunately, however, this assumption is insufficient to make Theorem 1 valid. To show this, Section 2 provides a counter example which even satisfies the strict Pareto condition (P*). The same example shows the error in Roberts' attempt to prove his intermediate Lemma 6. Then Section 3 uses a modified form of the alternative "shift invariance" condition due to Roberts (1983, p. 74) himself in order to prove the crucial Lemma 6. This establishes that a slight alteration to Theorem 1 makes it valid.

2. Weak Continuity: A Counter Example

The following is an example of a society with two individuals and a strictly monotone and symmetric quasi-concave function $W : E^2 \to \mathbb{R}$, such that the induced SWFL defined on X by $a \; R \; b \iff W(\mathbf{U}(a)) \geq W(\mathbf{U}(b))$ satisfies conditions (U), (I), (P*) and (WC). The function W has a discontinuity at 0 which gives rise to a thick set $N(0)$.

Indeed, define $w : E^2 \to \mathbb{R}$ by $w(v_1, v_2) := \min\{\, v_1 + 2v_2, 2v_1 + v_2 \,\}$, and then suppose that for all $v = (v_1, v_2) \in E^2$, one has

$$W(v) := \begin{cases} 1 + w(v) & \text{if } w(v) > 0; \\ \exp \dfrac{(v_1 + 2v_2)\,(2v_1 + v_2)}{3(v_1 + v_2)} & \text{if } w(v) \leq 0 \text{ and } v_1 + v_2 > 0; \\ v_1 + v_2 & \text{if } v_1 + v_2 \leq 0. \end{cases}$$

Thus, $W : E^2 \to \mathbb{R}$ is defined for three different regions that are separated in E^2 by the indifference curve $W(v) = 0$ and by the closure of the indifference set $W(v) = 1$, which is made up of two open half-lines emanating from the origin 0. Note that 0 is in the closure of all three regions. The corresponding indifference map is illustrated in Figure 1. The three-dimensional graph of $W(v)$ has a boundary that includes a vertical "cliff" of height 1 at $v = 0$; everywhere else, W is continuous, as is easy to check.

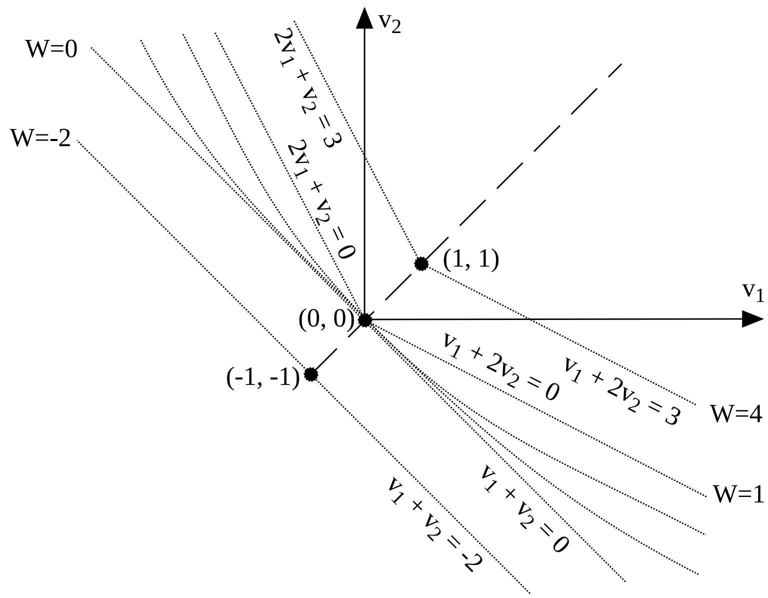

Figure 1

Obviously, the induced SWFL satisfies conditions (U), (I) and (P*). The set $N(0)$ is equal to the middle region where $W(v) \in [0,1]$. Note too that, although $v \in N(0)$ whenever $W(v) \in (0,1]$, one will have $v - \eta \succ -\eta'$ whenever $\eta, \eta' \gg 0$ with η small enough so that $W(v) \geq 0$ because $\eta_1 + \eta_2 \leq v_1 + v_2$. This contradicts Roberts' claim, in the course of trying to prove Lemma 6, that: "as $v + \gamma \in N(v^*)$, (WC) ensures that $v + \gamma - \eta_3 \in N(v^* - \eta_4)$ for some $\ldots \eta_3, \eta_4 \gg 0$."

To verify condition (WC) it is enough to construct, for each $\epsilon = (\epsilon_1, \epsilon_2) \gg 0$, a transformation $v \mapsto (\phi_1^\epsilon(v), \phi_2^\epsilon(v))$ from E^2 into itself satisfying $0 \ll v - \phi^\epsilon(v) \ll \epsilon$, together with the requirement that $W(\phi^\epsilon(v))$ and $W(v)$ are ordinally equivalent because $W(\phi^\epsilon(v)) \equiv \psi^\epsilon(W(v))$ for some strictly increasing transformation $\psi^\epsilon : \mathbb{R} \to \mathbb{R}$. In the following constructions, let $\epsilon_* := \min\{\, \epsilon_1, \epsilon_2 \,\}$ and $e := (1,1)$. Then $\epsilon_* > 0$, of course. The transformation will take the form $\phi^\epsilon(v) := v - \lambda^\epsilon(v)\, e$ for a suitably constructed scalar $\lambda^\epsilon(v)$ in the open interval $(0, \epsilon_*)$.

The simplest case is when $v_1 + v_2 \leq 0$ and so $W(v) = v_1 + v_2 \leq 0$. Then, define $\lambda^\epsilon(v) := \frac{1}{2}\, \epsilon_*$. In this case it is easy to see that $\phi_1^\epsilon(v) + \phi_2^\epsilon(v) = v_1 + v_2 - \epsilon_* < 0$ and so $W(\phi^\epsilon(v)) = \phi_1^\epsilon(v) + \phi_2^\epsilon(v) = \psi^\epsilon(W(v))$ where $\psi^\epsilon(W) := W - \epsilon_*$ whenever $W \leq 0$.

The second case occurs when $w(v) > 0$ and so $W(v) = 1 + w(v) > 1$. Then, define $\lambda^\epsilon(v) := \frac{1}{6}\min\{\, \epsilon_*, w(v)\,\}$. Clearly, this definition implies that $\lambda^\epsilon(v) \in (0, \epsilon_*)$. Also

$$\phi_1^\epsilon(v) + 2\phi_2^\epsilon(v) = v_1 + 2v_2 - 3\lambda^\epsilon(v) \quad \text{and} \quad 2\phi_1^\epsilon(v) + \phi_2^\epsilon(v) = 2v_1 + v_2 - 3\lambda^\epsilon(v)$$

Because $w(v) := \min\{\, v_1 + 2v_2, 2v_1 + v_2\,\}$ and $\lambda^\epsilon(v) \leq \frac{1}{6}w(v)$, it follows that

$$\min\{\, \phi_1^\epsilon(v) + 2\phi_2^\epsilon(v),\, 2\phi_1^\epsilon(v) + \phi_2^\epsilon(v)\,\} = w(v) - 3\lambda^\epsilon(v) \geq \tfrac{1}{2}\, w(v) > 0$$

Then the definitions of $W(\cdot)$ and of $\lambda^\epsilon(v)$ imply that

$$\begin{aligned} 1 < W(\phi^\epsilon(v)) &= 1 + \min\{\, \phi_1^\epsilon(v) + 2\phi_2^\epsilon(v),\, 2\phi_1^\epsilon(v) + \phi_2^\epsilon(v)\,\} \\ &= 1 + w(v) - 3\lambda^\epsilon(v) = \max\{\, W(v) - \tfrac{1}{2}\,\epsilon_*, \tfrac{1}{2}\,[W(v) + 1]\,\} = \psi^\epsilon(W(v)) \end{aligned}$$

where $\psi^\epsilon(W) := \max\{\, W - \frac{1}{2}\,\epsilon_*, \frac{1}{2}\,(W + 1)\,\}$ whenever $W > 1$.

This leaves the hardest third case, when both $w(v) \leq 0$ and $v_1 + v_2 > 0$. This implies that $0 < W(v) \leq 1$. Now $\lambda^\epsilon(v)$ will be defined implicitly by

$$\ln[W(v - \lambda^\epsilon(v)\, e)] = \ln[W(v)] - \mu\, \epsilon_*$$

for some suitably chosen positive scalar constant μ that is independent of both v and ϵ. Then $\lambda^\epsilon(v)$ will be well defined and positive, with

$$\ln[W(\phi^\epsilon(v))] = \ln[W(v)] - \mu\, \epsilon_* = \ln[\psi^\epsilon(W(v))] < 0$$

where $\psi^\epsilon(W) := W \exp -\mu\, \epsilon_* \in (0, 1)$ whenever $0 < W \leq 1$. It remains only to choose μ so that the corresponding $\lambda^\epsilon(v) < \epsilon_*$. However, $\lambda^\epsilon(v) > 0$ is a value of λ which solves the equation

$$\frac{(v_1 + 2v_2 - 3\lambda)\,(2v_1 + v_2 - 3\lambda)}{3(v_1 + v_2 - 2\lambda)} = \frac{(v_1 + 2v_2)\,(2v_1 + v_2)}{3(v_1 + v_2)} - \mu\, \epsilon_*$$

It follows that $\lambda = \lambda^\epsilon(v) \neq \frac{1}{2}(v_1 + v_2)$ must solve the quadratic equation $q(\lambda) = 0$, where

$$\begin{aligned} q(\lambda) := &(v_1 + v_2)\,(v_1 + 2v_2 - 3\lambda)\,(2v_1 + v_2 - 3\lambda) \\ &- (v_1 + v_2 - 2\lambda)\,(v_1 + 2v_2)\,(2v_1 + v_2) + 3\mu\, \epsilon_*\,(v_1 + v_2)\,(v_1 + v_2 - 2\lambda) \end{aligned}$$

Because $v_1 + v_2 > 0$, note that $q(0) = 3\mu\, \epsilon_*\,(v_1 + v_2)^2 > 0$ and also $q(\lambda) \to \infty$ as $\lambda \to \infty$. In addition

$$q(\tfrac{1}{2}(v_1 + v_2)) = -\tfrac{1}{4}(v_1 + v_2)\,(v_1 - v_2)^2$$

and

$$
\begin{aligned}
q(\epsilon_*) &= -9(v_1+v_2)^2\,\epsilon_* + 9(v_1+v_2)\,\epsilon_*^2 \\
&\qquad + 2(v_1+2v_2)\,(2v_1+v_2)\,\epsilon_* + 3\mu\,\epsilon_*\,(v_1+v_2)^2 - 6\mu\,(v_1+v_2)\,\epsilon_*^2 \\
&= (9-6\mu)\,(v_1+v_2)\,\epsilon_*^2 + (3\mu-9)\,(v_1+v_2)^2\,\epsilon_* + 2(2v_1^2+5v_1v_2+2v_2^2)\,\epsilon_* \\
&= (v_1+v_2)\,\epsilon_*\,[(9-6\mu)\,\epsilon_* + (3\mu-5)\,(v_1+v_2)] + 2v_1v_2\,\epsilon_*
\end{aligned}
$$

Because $v_1+v_2>0$ but $w(v)\le 0$, it follows that v_1 and v_2 have opposite signs. Hence, $q(\frac{1}{2}(v_1+v_2))<0$ and also $q(\epsilon_*)<0$ whenever $9<6\mu<10$. So choosing any fixed $\mu\in\left(\frac{3}{2},\frac{5}{3}\right)$ guarantees that $q(\lambda)=0$ has one real root $\lambda^\epsilon(v)$ in the open interval between 0 and $\min\{\,\epsilon_*,\frac{1}{2}(v_1+v_2)\,\}$; there is another irrelevant real root with $\lambda>\frac{1}{2}(v_1+v_2)$. In particular, $\lambda^\epsilon(v)\in(0,\epsilon_*)$, as required.

Finally, therefore, $W(\phi^\epsilon(v))\equiv\psi^\epsilon(W(v))$ where

$$
\psi^\epsilon(W) := \begin{cases} W - \frac{1}{2}\,(\epsilon_1+\epsilon_2) & \text{if } W\le 0; \\ W\,\exp(-\mu\,\epsilon_*) & \text{if } 0<W\le 1; \\ \max\{\,W-\frac{1}{2}\,\epsilon_*,\frac{1}{2}\,(W+1)\,\} & \text{if } W>1. \end{cases}
$$

In particular, ψ^ϵ is strictly increasing for each $\epsilon\gg 0$. ▮

3. Pairwise Continuity: A New Sufficient Condition

Roberts (1983, p. 74) later introduced the following *shift invariance* assumption:

Condition (SI): For all $u\in\mathcal{U}$, $\epsilon\in E^N$, $\epsilon\gg 0$, there exists an $\epsilon'\in E^N$ with $\epsilon'\gg 0$ and a $u'\in\mathcal{U}$ such that $\epsilon\gg u(x,\cdot)-u'(x,\cdot)\gg\epsilon'$ for all $x\in X$, and $f(u)=f(u')$.

As he states in a footnote: "Shift invariance is slightly stronger than … (WC). … The strengthening allows one to deal with problems that are akin to the existence of poles in a consumer's indifference map." Indeed, it is this footnote that suggested to me how the above counter example might be constructed. However, when proving Lemma A.5, it seems that Roberts (1983, p. 90) in the end reverses the order of some quantifiers and actually uses the following *uniform shift invariance* assumption:

Condition (USI): For all $\epsilon\in E^N$, $\epsilon\gg 0$, there exists an $\epsilon'\in E^N$ with $\epsilon'\gg 0$ for which, whenever $u\in\mathcal{U}$, there exists a $u'\in\mathcal{U}$ such that $\epsilon\gg u(x,\cdot)-u'(x,\cdot)\gg\epsilon'$ for all $x\in X$, and $f(u)=f(u')$.

Instead of (WC) or (SI), I shall use the following *pairwise continuity* assumption which weakens (USI):

Condition (PC): For all $\epsilon \in E^N$, $\epsilon \gg 0$, there exists an $\epsilon' \in E^N$ with $\epsilon' \gg 0$ for which, whenever $u \in \mathcal{U}$ and $x, y \in X$ satisfy $x \; P \; y$, there exists $u' \in \mathcal{U}$ such that $u'(x, \cdot) \ll u(x, \cdot) - \epsilon'$, $u'(y, \cdot) \gg u(y, \cdot) - \epsilon$, and $x \; P' \; y$.

Like shift invariance, this condition strengthens weak continuity because the same strictly positive vector ϵ' must work *simultaneously* for all $x, y \in X$. Like uniform shift invariance, it also strengthens shift invariance because the same strictly positive vector ϵ' must also work for all $u \in \mathcal{U}$. On the other hand, pairwise continuity weakens even weak continuity to the extent that u' can depend on the pair $x, y \in X$, and also only one-way strict inequalities need be satisfied.

Of course, just as with Roberts' (WC) and (SI) conditions, (USI) and so (PC) is certainly satisfied if f is invariant under the set of all shift transformations taking the form $u'(x, i) \equiv \alpha + u(x, i)$ (all $i \in N$, $x \in X$) with $\alpha \in \mathbb{R}$ independent of i (e.g., cardinal full comparability with invariant units). However, none of the four conditions (WC), (SI), (USI) and (PC) need be satisfied if each utility function can have both positive and negative values and if f is invariant only under the set of all transformations taking the form $u'(x, i) \equiv \beta_i \, u(x, i)$ with $\beta_i > 0$ (all $i \in N$, $x \in X$). This explains why Blackorby and Donaldson (1982) and also Tsui and Weymark (1997) imposed other continuity conditions in considering ratio-scale invariant social welfare functionals.[2]

With condition (PC) replacing (WC), Roberts' Lemma 6 will be proved via the following two separate lemmas:

LEMMA 6A. *If f satisfies (U), (I) and (P), then for all $v, v', \eta, \eta' \in E^N$ with $\eta, \eta' \gg 0$ one has $v \in N(v') \to v + \eta \in M(v' - \eta')$.*[3]

PROOF: Suppose that x, y, z are three distinct elements of X. By condition (U), there exists $u \in \mathcal{U}$ such that

$$v + \eta \gg u(x, \cdot) \gg u(y, \cdot) \gg v \quad \text{and} \quad v' \gg u(z, \cdot) \gg v' - \eta'$$

[2] I owe this to John Weymark, as well as the observation that the remark following Roberts' Lemma 8 is also incorrect. Note, however, that if each $u \in \mathcal{U}$ has strictly positive (resp. negative) values, one can work instead with $\log u(x, i)$ (resp. $-\log[-u(x, i)]$) as a transformed utility function.

[3] This is the correct "preliminary result" in Roberts' discussion of Lemma 6. However, the proof provided seems incomplete.

Now $z\ P\ y$ would imply that $v' \succ v$. So $v \in N(v') \to y\ R\ z$. Then condition (P) implies that $x\ P\ y$, and so $v \in N(v') \to x\ P\ z$ because R is transitive. From this it follows that $v \in N(v') \to v + \eta \succ v' - \eta'$. ∎

LEMMA 6B. *If f satisfies (U), (I), (P) and (PC), then:*

(a) if $\epsilon \gg 0$ and v, v' satisfy $v + \eta \in M(v' + \epsilon)$ for all $\eta \gg 0$, then $v \in M(v')$;

(b) $\nexists v, v^*, \gamma \in E^N$ with $\gamma \gg 0$ such that $v, v + \gamma \in N(v^*)$.

PROOF: (a) Given $\epsilon \gg 0$, let $\epsilon' \gg 0$ be specified as in the statement of condition (PC). Choose $\eta \gg 0$ so that $\eta \ll \epsilon'$. Because $v + \eta \in M(v' + \epsilon)$, there exist $u \in \mathcal{U}$ and $x, y \in X$ such that $x\ P\ y$ while

$$v + \eta \gg u(x, \cdot) \quad \text{and} \quad u(y, \cdot) \gg v' + \epsilon$$

By condition (PC), there exists $u' \in \mathcal{U}$ such that $x\ P'\ y$ while

$$u'(x, \cdot) \ll u(x, \cdot) - \epsilon' \quad \text{and} \quad u'(y, \cdot) \gg u(y, \cdot) - \epsilon$$

But then

$$u'(x, \cdot) \ll v + \eta - \epsilon' \ll v \quad \text{and} \quad u'(y, \cdot) \gg v'$$

Hence $v \succ v'$.

(b) Suppose that $v + \gamma \in N(v^*)$. By definition of $N(\cdot)$, it follows that $v^* \in N(v + \gamma)$. Choose any $\gamma' \gg 0$ satisfying $\gamma' \ll \gamma$. Now Lemma 6A implies that $v^* + \eta \in M(v + \gamma')$ for all $\eta \gg 0$. So part (a) implies that $v^* \in M(v)$. In particular, $v \notin N(v^*)$. ∎

4. Conclusion

Roberts' (1980) weak neutrality or welfarism theorem is indeed "both important and useful" (p. 428). The minor errors in its statement and in the proof of Lemma 6 are simple to correct by replacing condition (WC) with the new condition (PC) stated in Section 3.

An open question is whether Roberts' (1983) Theorem 1 holds under shift invariance (SI) instead of uniform shift invariance (USI), which is stronger than (PC). However, even (USI) is weak enough that having to impose it instead of (WC) or (SI) would do little to detract from the significance or wide applicability of Roberts' theorem. Only in the case of ratio-scale measurability of utilities that can change sign is the theorem inapplicable.

References

Blackorby, C. and D. Donaldson (1982) "Ratio-Scale and Translation-Scale Full Interpersonal Comparability without Domain Restrictions: Admissible Social-Evaluation Functions", *International Economic Review*, **23**, pp. 249–268.

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

Bordes, G., P.J. Hammond and M. Le Breton (1999), "Social Welfare Functionals on Restricted Domains and in Economic Environments" working paper, Department of Economics, Stanford University; forthcoming in the *Journal of Public Economic Theory.*

Bordes, G. and M. Le Breton (1987), "Roberts's Theorem in Economic Environments: (2) The Private Goods Case" preprint, Faculté des Sciences Economiques et de Gestion, Université de Bordeaux I.

Bossert, W and J.A. Weymark (1999) "Utility in Social Choice," in S. Barberà, P.J. Hammond, and C. Seidl (eds.) *Handbook of Utility Theory, Vol. II: Applications and Extensions* (Dordrecht: Kluwer Academic Publishers), ch. 20 (forthcoming).

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

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

Le Breton, M. (1987) "Roberts' Theorem in Economic Environments: 1. — The Public Goods Case," preprint, Faculté des Sciences Economiques et de Gestion, Université de Rennes.

Mongin, P. and C. d'Aspremont (1998) "Utility Theory and Ethics," in S. Barberà, P.J. Hammond, and C. Seidl (eds.) *Handbook of Utility Theory, Vol. 1: Principles* (Dordrecht: Kluwer Academic Publishers), ch. 10, pp. 371–481.

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

ROBERTS, K.W.S. (1983) "Social Choice Rules and Real Valued Representations", *Journal of Economic Theory*, **29**, pp. 72–94.

SEN, A.K. (1977) "On Weights and Measures: Informational Constraints in Social Welfare Analysis", *Econometrica*, **45**, pp. 1539–72; reprinted in Sen (1982).

SEN, A.K. (1982) *Choice, Welfare and Measurement* (Oxford: Basil Blackwell, and Cambridge, Mass.: MIT Press).

SEN, A.K. (1984) "Social Choice Theory", in K.J. Arrow and M.D. Intriligator (ed.) *Handbook of Mathematical Economics, Vol. 3* (Amsterdam: North-Holland) ch. 22, pp. 1073–1181.

TSUI, K.-Y. AND J.A. WEYMARK (1997) "Social Welfare Orderings for Ratio-Scale Measurable Utilities", *Economic Theory*, **9**, pp. 241–256.