Roberts' Weak Welfarism Theorem: A Minor Correction

Peter J. Hammond: `p.j.hammond@warwick.ac.uk`
Dept. of Economics, University of Warwick, Coventry CV4 7AL, UK.

Latest version: 2020 October 9th, typeset from `RobertsThmCRETA.tex`

Abstract:

Roberts' "weak neutrality" or "weak welfarism" theorem concerns Sen social welfare functionals which are defined on an unrestricted domain of utility function profiles and satisfy independence of irrelevant alternatives, the Pareto condition, and a form of weak continuity. Roberts (1980) claimed that the induced welfare ordering on social states has a one-way representation by a continuous, monotonic real-valued welfare function defined on the Euclidean space of interpersonal utility vectors — that is, an increase in this welfare function is sufficient, but may not be necessary, for social strict preference. A counter-example shows that weak continuity is insufficient; a minor strengthening to pairwise continuity is proposed instead and its sufficiency demonstrated.

JEL Classification: D71

Keywords: social welfare functionals, weak welfarism.

1 Introduction: Roberts' Claim

Consider a society with a non-empty finite set N of individuals i.[1] Let X be a domain of at least three social states, and $\mathcal{R}(X)$ the set of all logically possible (complete and transitive) social weak preference orderings R on X. For each $R \in \mathcal{R}(X)$, let P and I denote the corresponding strict preference and indifference relations.

Let $\mathbb{R}^N$ denote the Euclidean space that consists of the Cartesian product of $\#N$ copies of the real line $\mathbb{R}$. A *utility function profile* $\mathbf{u}^N$ is a mapping $X \ni x \mapsto \mathbf{u}^N(x) \in \mathbb{R}^N$. Let $\mathcal{U}^N$ denote the set of all utility function profiles on X. Following Sen (1970, 1977), a *social welfare functional* (or SWFL) f on a domain $\mathcal{D} \subseteq \mathcal{U}^N$ is a mapping $\mathcal{D} \ni \mathbf{u}^N \mapsto f(\mathbf{u}^N) \in \mathcal{R}(X)$ that determines a social preference ordering $R = f(\mathbf{u}^N)$ for each utility function profile in $\mathcal{D}$. With some slight abuse of notation, we let $R(\mathbf{u}^N)$, $P(\mathbf{u}^N)$ and $I(\mathbf{u}^N)$ denote respectively the weak preference, strict preference, and indifference relations associated with $f(\mathbf{u}^N)$.

Given any non-empty subset $A \subset X$:

1. say that two utility function profiles $\mathbf{u}^N, \tilde{\mathbf{u}}^N \in \mathcal{U}^N$ are *equal on* A just in case one has $\mathbf{u}^N(x) = \tilde{\mathbf{u}}^N(x)$ for all $x \in A$;

2. say that two social preference orderings R and $\tilde{R}$ on X are *equal on* A just in case for all $y, z \in A$ one has $y \; R \; z \iff y \; \tilde{R} \; z$.

This paper considers social welfare functionals f satisfying at least the first two of the following three axioms:

Unrestricted domain (U) The domain $\mathcal{D}$ of f is the whole of $\mathcal{U}^N$.

Independence (I) Given any non-empty subset $A \subset X$, if the two utility function profiles $\mathbf{u}^N, \tilde{\mathbf{u}}^N \in \mathcal{U}^N$ are equal on A, then the associated social orderings $f(\mathbf{u}^N), f(\tilde{\mathbf{u}}^N) \in \mathcal{U}^N$ are also equal on A.

Pareto indifference (P^0) In case $y, z \in X$ and $\mathbf{u}^N \in \mathcal{U}^N$ satisfy $\mathbf{u}^N(y) = \mathbf{u}^N(z)$, the associated social indifference relation satisfies $y \; I(\mathbf{u}^N) \; z$.

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 all three conditions (U), (I), and (P^0), then there exists a (complete

[1] Most notation and definitions are based on Roberts (1980). Where appropriate, however, bold letters are used to indicate vectors.

and transitive) *social welfare ordering* R^* on $\mathbb{R}^N$ with the property that $y \; R \; z \iff \mathbf{u}^N(x) \; R^* \; \mathbf{u}^N(y)$. This result plays a prominent role among the results appearing the surveys by Sen (1984), Blackorby, Donaldson and Weymark (1984), d'Aspremont (1985), Mongin and d'Aspremont (1998), and Bossert and Weymark (2004). Both Sen (1977) and d'Aspremont (1985, p. 34) provide complete proofs.[2]

While the Pareto indifference axiom (P^0) is appealing, the impossibility theorem in Arrow (1963) replaces it with the following alternative:

Pareto (P) In case $y, z \in X$ and $\mathbf{u}^N \in \mathcal{U}^N$ satisfy $\mathbf{u}^N(y) \gg \mathbf{u}^N(z)$, the associated strict preference relation $P(\mathbf{u}^N)$ satisfies $y \; P(\mathbf{u}^N) \; z$.[3]

Specifically, under the assumption that utilities are ordinally non-comparable, Arrow's impossibility theorem states that (U), (I) and (P) together imply a dictatorship. To develop a theory general enough to cover this important case, Roberts (1980, p. 427) specifies the following additional condition:

Weak continuity (WC) For all $\mathbf{u}^N \in \mathcal{U}^N$ and $\boldsymbol{\epsilon} \in \mathbb{R}^N$ with $\boldsymbol{\epsilon} \gg \mathbf{0}$, there exists a $\mathbf{u}^N_{\boldsymbol{\epsilon}} \in \mathcal{U}^N$ satisfying $\boldsymbol{\epsilon} \gg \mathbf{u}^N(x) - \mathbf{u}^N_{\boldsymbol{\epsilon}} \gg \mathbf{0}$ for all $x \in X$ such that $f(\mathbf{u}^N) = f(\mathbf{u}^N_{\boldsymbol{\epsilon}})$.

Then (p. 428) he claims the following:

Claim 1. *Suppose that f satisfies (U), (I), (P), and (WC). Then there exists a continuous function $\mathbb{R}^N \ni \mathbf{w} \mapsto W(\mathbf{w}) \to \mathbb{R}$, strictly increasing with an increase in all its arguments, with the property that for all $\mathbf{u}^N \in \mathcal{U}^N$ and all $y, z \in X$ one has*

$$W(\mathbf{u}^N(y)) > W(\mathbf{u}^N(z)) \Longrightarrow y \; P(\mathbf{u}^N) \; z$$

This claim has come to be known as Roberts' "weak neutrality" or "weak welfarism" theorem.[4] In many of the surveys mentioned above, it was cited as an alternative to the strong neutrality result of d'Aspremont and Gevers (1977) and Sen (1977, p. 1553). The unpublished results by Le Breton

[2]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 $b \neq e \neq d$.

[3]Given any pair $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$ with $\mathbf{a} = (a_i)_{i \in N}$ and $\mathbf{b} = (b_i)_{i \in N}$, we use the following notation for vector orderings: (i) $\mathbf{a} \geqq \mathbf{b}$ in case $a_i \geq b_i$ for all $i \in N$; (ii) $\mathbf{a} \gg \mathbf{b}$ in case $a_i > b_i$ for all $i \in N$; (iii) $\mathbf{a} > \mathbf{b}$ in case $\mathbf{a} \geqq \mathbf{b}$ but $\mathbf{a} \neq \mathbf{b}$.

[4]The Roberts' theorem which is the topic of this paper concerns social choice in the sense of aggregating preferences. It differs from the Roberts' theorem on revelation of preferences that appeared in Roberts (1979), and has subsequently been discussed by, amongst others, Lavi et al. (2009) and Mishra and Sen (2012).

(1987) and by Bordes and Le Breton (1987) investigating Roberts' theorem for restricted economic domains have since been amalgamated with related results that appear in Bordes, Hammond and Le Breton (2005).

Condition (WC), however, is too weak for the claim to hold. To show this, Section 2 provides a counter example which even satisfies the following familiar condition:

Strict Pareto (P*) In case $y, z \in X$ and $\mathbf{u}^N \in \mathcal{U}^N$ satisfy $\mathbf{u}^N(y) \geqq \mathbf{u}^N(z)$, the associated social preferences satisfy $y\ R(\mathbf{u}^N)\ z$, with $y\ P(\mathbf{u}^N)\ z$ in case $\mathbf{u}^N(y) > \mathbf{u}^N(z)$.

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 in Roberts (1980). This establishes that a slight alteration to Claim 1 makes it valid.

2 Weak Continuity: A Counter Example

The following is an example of a society with two individuals and a strictly increasing and symmetric utilitarian welfare function

$$\mathbb{R}^2 \ni (u_1, u_2) = \mathbf{u} \mapsto W(u_1, u_2) = W(\mathbf{u}) \to \mathbb{R} \tag{1}$$

such that the induced SWFL defined on X by

$$a\ R\ b \Longleftrightarrow W(u_1(a), u_2(a)) \geq W(u_1(b), u_2(b)) \tag{2}$$

satisfies conditions (U), (I), (P*) and (WC). Yet the function W has a discontinuity at the origin $\mathbf{0} = (0, 0)$ which gives rise to a discontinuous induced ordering on $\mathbb{R}^2$. This implies that no continuous function W can satisfy Claim 1 in this example.

Indeed, first define the function $\mathbb{R}^2 \ni (v_1, v_2) \mapsto w(v_1, v_2) \in \mathbb{R}$ by

$$w(v_1, v_2) := \min\{v_1 + 2v_2, 2v_1 + v_2\} \tag{3}$$

Then define $\mathbb{R}^2 \ni (v_1, v_2) = \mathbf{v} \mapsto W(v_1, v_2) = W(\mathbf{v}) \in \mathbb{R}$ by

$$W(\mathbf{v}) := \begin{cases} 1 + w(\mathbf{v}) & \text{if } w(\mathbf{v}) > 0; \\ \exp \dfrac{(v_1 + 2v_2)\,(2v_1 + v_2)}{3(v_1 + v_2)} & \text{if } w(\mathbf{v}) \leq 0 \text{ and } v_1 + v_2 > 0; \\ v_1 + v_2 & \text{if } v_1 + v_2 \leq 0. \end{cases} \tag{4}$$

Thus, the composite function W is defined for three different regions of $\mathbb{R}^2$ separated by: (i) the indifference curve $W(v_1, v_2) = 0$; and (ii) the closure of the indifference set $W(v_1, v_2) = 1$, which is made up of two open half-lines emanating from the origin $(0, 0)$. Note that $(0, 0)$ is in the closure of all three regions. The corresponding indifference map is illustrated in Figure 1. The three-dimensional graph of $W(v_1, v_2)$ has a boundary that includes a vertical "cliff" of height 1 at $(v_1, v_2) = (0, 0)$ where $W(0, 0) = 0$, yet every neighbourhood of $(0, 0)$ has points $\mathbf{v}$ where $w(\mathbf{v}) > 0$, implying that $W(\mathbf{v}) > 1$. Everywhere apart from the origin, however, the mapping $\mathbf{v} \mapsto W(\mathbf{v})$ is continuous, as is easy to check.

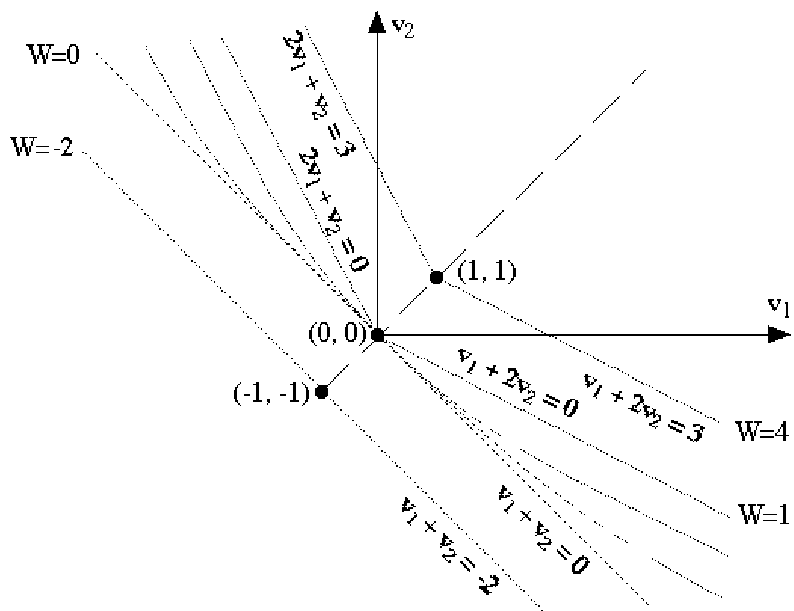

Figure 1: Level curves of the social welfare function $(v_1, v_2) \mapsto W(v_1, v_2)$

Nor can this discontinuity be removed by a strictly increasing transformation. To see this, consider any point $\bar{\mathbf{v}} = (\bar{v}_1, \bar{v}_2) \in \mathbb{R}^2$ satisfying $w(\bar{\mathbf{v}}) < 0$ and $\bar{v}_1 + \bar{v}_2 > 0$, implying that $0 < W(\bar{\mathbf{v}}) < 1$. Now, every neighbourhood of $(0, 0)$ has points $\mathbf{v}$ where $w(\mathbf{v}) > 0$ and so $W(\mathbf{v}) > 1$. So after applying any strictly increasing transformation $\mathbb{R} \ni W \mapsto \tilde{W} = \psi(W) \in \mathbb{R}$ to the function $\mathbf{v} \mapsto W(\mathbf{v})$, one has

$$\tilde{W}(0) = \psi(0) < \tilde{W}(\bar{\mathbf{v}}) = \psi(W(\bar{\mathbf{v}})) < \psi(1) < \psi(W(\mathbf{v})) = \tilde{W}(\mathbf{v})$$

This shows that the discontinuity at $(0, 0)$ cannot be removed.

Consider now the SWFL $\mathcal{U}^2 \ni (u_1, u_2) \mapsto f(u_1, u_2)$ defined by (2), (3) and (4). Obviously, this induced SWFL satisfies conditions (U), (I) and (P*). To verify condition (WC) it is enough to construct, for each $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2) \gg (0, 0)$, a transformation

$$\mathbb{R}^2 \ni \mathbf{v} \mapsto \boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v}) = (\phi_1^{\boldsymbol{\epsilon}}(\mathbf{v}), \phi_2^{\boldsymbol{\epsilon}}(\mathbf{v})) \in \mathbb{R}^2$$

satisfying

$$(0, 0) \ll (v_1, v_2) - \boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v}) \ll \boldsymbol{\epsilon}$$

together with the requirement that $\mathbf{v} \mapsto W(\boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v}))$ and $\mathbf{v} \mapsto W(\mathbf{v})$ are ordinally equivalent welfare functions in the sense that there exists a strictly increasing transformation $\mathbb{R} \ni \mathbf{w} \mapsto \psi^{\boldsymbol{\epsilon}}(\mathbf{w}) \in \mathbb{R}$ for which $W(\boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v})) \equiv \psi^{\boldsymbol{\epsilon}}(W(\mathbf{v}))$.

In the following constructions, let

$$\epsilon_* := \min\{\epsilon_1, \epsilon_2\} \in \mathbb{R} \text{ and } \mathbf{e} := (1, 1) \in \mathbb{R}^2 \tag{5}$$

Then $\epsilon_* > 0$, of course. The transformation will take the form

$$\mathbb{R}^2 \ni \mathbf{v} \mapsto \boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v}) := \mathbf{v} - \lambda^{\boldsymbol{\epsilon}}(\mathbf{v})\, \mathbf{e} \in \mathbb{R}^2 \tag{6}$$

for a suitably constructed scalar function $\mathbb{R}^2 \ni \mathbf{v} \mapsto \lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) \in \mathbb{R}$ taking values in the open interval $(0, \epsilon_*)$.

Case 1: The simplest case is when

$$v_1 + v_2 \leq 0 \text{ and so } W(\mathbf{v}) = v_1 + v_2 \leq 0 \tag{7}$$

In this case, define $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) := \frac{1}{2}\, \epsilon_*$ for ϵ_* given by (5). Then it is easy to see from (5), (6), and (7) that $\phi_1^{\epsilon}(\mathbf{v}) + \phi_2^{\epsilon}(\mathbf{v}) = v_1 + v_2 - \epsilon_* \leq -\epsilon_* < 0$. Now, whenever $v_1 + v_2 \leq 0$, it follows that

$$W(\boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v})) = \phi_1^{\boldsymbol{\epsilon}}(\mathbf{v}) + \phi_2^{\boldsymbol{\epsilon}}(\mathbf{v}) = \psi^{\boldsymbol{\epsilon}}(W(\mathbf{v}))$$

provided we define

$$\psi^{\boldsymbol{\epsilon}}(W) := W - \epsilon_* \text{ for all } W \leq 0 \tag{8}$$

Case 2: This case occurs when

$$w(\mathbf{v}) > 0 \text{ and so } W(\mathbf{v}) = 1 + w(\mathbf{v}) > 1 \tag{9}$$

In this case, define

$$\lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) := \tfrac{1}{6} \min\{\epsilon_*, w(\mathbf{v})\} \tag{10}$$

Clearly, this definition implies that $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) \in (0, \epsilon_*)$. Also

$$\begin{aligned} \phi_1^{\boldsymbol{\epsilon}}(\mathbf{v}) + 2\phi_2^{\boldsymbol{\epsilon}}(\mathbf{v}) &= v_1 + 2v_2 - 3\lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) \\ \text{and } 2\phi_1^{\boldsymbol{\epsilon}}(\mathbf{v}) + \phi_2^{\boldsymbol{\epsilon}}(\mathbf{v}) &= 2v_1 + v_2 - 3\lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) \end{aligned}$$

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

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

Then the definitions of $\mathbb{R}^2 \ni \mathbf{v} \mapsto W(\mathbf{v}) \in \mathbb{R}$ and of $\mathbb{R}^2 \ni \mathbf{v} \mapsto \lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) \in \mathbb{R}$ in (4) and (10) imply that

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

It follows that $W(\psi^{\boldsymbol{\epsilon}}(\mathbf{v})) = \psi^{\boldsymbol{\epsilon}}(W(\mathbf{v}))$ provided that we define

$$\psi^{\boldsymbol{\epsilon}}(W) := \max\{W - \tfrac{1}{2}\,\epsilon_*, \tfrac{1}{2}\,(W + 1)\} \text{ for all } W > 1 \tag{11}$$

Case 3: This leaves the hardest third case, when

$$w(\mathbf{v}) \leq 0 \text{ and also } v_1 + v_2 > 0 \tag{12}$$

In this case, the definition in (4) implies that $0 < W(\mathbf{v}) \leq 1$.

Fix any $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ satisfying (12). Then, given any $\boldsymbol{\epsilon} \in \mathbb{R}^2$ satisfying $\boldsymbol{\epsilon} \gg 0$, consider the non-empty open interval of $\mathbb{R}$ defined by

$$I^{\boldsymbol{\epsilon}}(\mathbf{v}) := (0, \min\{\epsilon_*, \tfrac{1}{2}(v_1 + v_2)\}) = (0, \epsilon_*) \cap (0, \tfrac{1}{2}(v_1 + v_2)) \tag{13}$$

Now consider the function g defined on the open interval in (13) by

$$I^{\boldsymbol{\epsilon}}(\mathbf{v}) \ni \lambda \mapsto g(\lambda) := W(\mathbf{v} - \lambda\, \mathbf{e}) \in \mathbb{R} \tag{14}$$

Because definition (3) implies that W is strictly increasing as a function of two variables, it follows that g is strictly decreasing. Also, when $\lambda > 0$, it is evident that

$$w(\mathbf{v} - \lambda\,\mathbf{e}) < w(\mathbf{v}) \leq 0 \tag{15}$$

On the other hand, when $\lambda < \frac{1}{2}(v_1 + v_2)$, because $e_1 = e_2 = 1$, one has

$$(v_1 - \lambda e_1) + (v_2 - \lambda e_2) = v_1 + v_2 - 2\lambda > 0 \tag{16}$$

So for all $\lambda \in I^{\epsilon}(\mathbf{v})$ the inequalities (15) and (16) imply that the 2-vector $\mathbf{v} - \lambda\,\mathbf{e}$ satisfies (12). It follows that $W(\mathbf{v} - \lambda\,\mathbf{e})$ is also defined as in Case 3, so (4) implies that

$$g(\lambda) = W(\mathbf{v} - \lambda\,\mathbf{e}) = \exp \frac{(v_1 + 2v_2 - 3\lambda)\,(2v_1 + v_2 - 3\lambda)}{3(v_1 + v_2 - 2\lambda)} \tag{17}$$

Now let μ denote a suitably chosen positive scalar constant which is independent of both v and ϵ, and whose possible range will be specified later. For each $\mathbf{v} \in \mathbb{R}^2$ that satisfies the inequalities (12), because g is strictly decreasing and positive, we can define $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v})$ implicitly as the unique value of λ that solves the equation

$$g(\lambda) = W(\mathbf{v} - \lambda\,\mathbf{e}) = W(\mathbf{v}) \exp(-\mu\,\epsilon_*) \tag{18}$$

Then $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v})$ will be well defined and positive, with

$$W(\boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v})) = W(\mathbf{v}) \exp(-\mu\,\epsilon_*) = \psi^{\boldsymbol{\epsilon}}(W(\mathbf{v})) < 1$$

where $\psi^{\boldsymbol{\epsilon}}(W) := W \exp(-\mu\,\epsilon_*) \in (0, 1)$ whenever $0 < W \leq 1$.

It remains only to choose $\mu > 0$ so that the corresponding solution to equation (18) satisfies $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v}) < \epsilon_*$. In fact, we find a solution in the open interval $I^{\boldsymbol{\epsilon}}(\mathbf{v})$ defined by (13). Because we are assuming that the inequalities (12) hold, definition (4) implies that any $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v})$ satisfying (18) and (13) must be a value of λ which solves the equation

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

But $v_1 + v_2 > 2\lambda > 0$ in the relevant interval of values of λ, so we can clear fractions to obtain the quadratic equation $q(\lambda) = 0$, where

$$\begin{aligned} q(\lambda) := (v_1 + v_2)\,(v_1 + 2v_2 - 3\lambda)\,(2v_1 + v_2 - 3\lambda) \\ - (v_1 + v_2 - 2\lambda)\,(v_1 + 2v_2)\,(2v_1 + v_2) \\ + 3\mu\,\epsilon_*\,(v_1 + v_2)\,(v_1 + v_2 - 2\lambda) \end{aligned} \tag{19}$$

Now, note that when $\lambda = 0$ the first two terms on the right-hand side of (19) cancel. Because $v_1 + v_2 > 0$, it follows that

$$q(0) = 3\mu\,\epsilon_*\,(v_1 + v_2)^2 > 0 \tag{20}$$

In addition, simple calculation shows that

$$q\left(\tfrac{1}{2}(v_1 + v_2)\right) = -\tfrac{1}{4}(v_1 + v_2)\,(v_1 - v_2)^2 \tag{21}$$

Finally, some much more tedious but still routine algebraic manipulation shows that

$$q(\epsilon_*) = (v_1 + v_2)\,\epsilon_*\,[(9 - 6\mu)\,\epsilon_* + (3\mu - 5)\,(v_1 + v_2)] + 2v_1 v_2\,\epsilon_* \tag{22}$$

Because $v_1 + v_2 > 0$ but $w(\mathbf{v}) \leq 0$, it follows from (3) that v_1 and v_2 have opposite signs. In particular $v_1 \neq v_2$ and $v_1 v_2 < 0$. Then (21) implies that $q(\frac{1}{2}(v_1 + v_2)) < 0$ whereas (22) implies that for any μ satisfying $9 < 6\mu < 10$ one has $q(\epsilon_*) < 0$. So choosing any fixed $\mu \in \left(\frac{3}{2}, \frac{5}{3}\right)$ guarantees that, by the intermediate value theorem, the quadratic equation $q(\lambda) = 0$ has one real root $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v})$ in the open interval $I^{\boldsymbol{\epsilon}}(\mathbf{v}) = (0, \min\{\,\epsilon_*, \frac{1}{2}(v_1 + v_2)\,\}$ defined by (13).[5] In particular, for each $\mathbf{v}$ satisfying (12), the root $\lambda^{\boldsymbol{\epsilon}}(\mathbf{v})$ of $q(\lambda) = 0$ that we have found lies in $(0, \epsilon_*)$, as required.

Finally, putting all three cases together gives $W(\boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v})) \equiv \psi^{\boldsymbol{\epsilon}}(W(\mathbf{v}))$, where $\boldsymbol{\phi}^{\boldsymbol{\epsilon}}(\mathbf{v}) = \mathbf{v} - \lambda^{\boldsymbol{\epsilon}}(\mathbf{v})\,\mathbf{e}$, and then

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

In particular, $\psi^{\boldsymbol{\epsilon}}$ is strictly increasing in W for each fixed $\boldsymbol{\epsilon} \gg 0$. □

Next, to see where his proof erred, we adapt some more notation from Roberts (1980, pp. 425–6). The relation $\succ$ on utility vectors in $\mathbb{R}^N$ is defined so that $\mathbf{a} \succ \mathbf{b}$ iff there exist a utility function profile $\mathbf{u}^N \in \mathcal{U}$ and two social states $x, y \in X$ with $x\ f(\mathbf{u})\ y$ such that $\mathbf{a} \gg \mathbf{u}(x)$ and $\mathbf{u}(y) \gg \mathbf{b}$. Then, for each $\mathbf{v}^* \in \mathbb{R}^N$, the three sets $L(\mathbf{v}^*)$, $M(\mathbf{v}^*)$ and $N(\mathbf{v}^*)$ are defined respectively by

$$L(\mathbf{v}^*) := \{\mathbf{v} \in \mathbb{R}^N \mid \mathbf{v}^* \succ \mathbf{v}\} \tag{23}$$

$$M(\mathbf{v}^*) := \{\mathbf{v} \in \mathbb{R}^N \mid \mathbf{v} \succ \mathbf{v}^*\} \tag{24}$$

$$N(\mathbf{v}^*) := \mathbb{R}^N \setminus [L(\mathbf{v}^*) \cup M(\mathbf{v}^*)] \tag{25}$$

[5] Because $q(\lambda) \to +\infty$ as $\lambda \to +\infty$, the quadratic equation $q(\lambda) = 0$ has a second irrelevant real root that satisfies $\lambda > \max\{\,\epsilon_*, \frac{1}{2}(v_1 + v_2)\,\}$.

Now, in the above example the set $N(\mathbf{0})$ is equal to the middle region where $W(\mathbf{v}) \in [0,1]$. Note too that, although $\mathbf{v} \in N(\mathbf{0})$ whenever $W(\mathbf{v}) \in (0,1]$, one will have $\mathbf{v} - \boldsymbol{\eta} \succ -\boldsymbol{\eta}'$ whenever $\boldsymbol{\eta}, \boldsymbol{\eta}' \gg \mathbf{0}$ with $\boldsymbol{\eta}$ small enough so that $W(\mathbf{v}) \geq 0$ 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 $\mathbf{v} + \boldsymbol{\gamma} \in N(\mathbf{v}^*)$, (WC) ensures that $\mathbf{v} + \boldsymbol{\gamma} - \boldsymbol{\eta}_3 \in N(\mathbf{v}^* - \boldsymbol{\eta}_4)$ for some $\boldsymbol{\epsilon} \gg \boldsymbol{\eta}_3, \boldsymbol{\eta}_4 \gg \mathbf{0}$ where $\boldsymbol{\epsilon}$ is subject to choice."

3 A New Sufficient Condition

Roberts (1983, p. 74) later introduced a *shift invariance* condition which can be slightly restated as follows:

Condition (SI) For all profiles $\mathbf{u}^N \in \mathcal{U}^N$ and all $\boldsymbol{\epsilon} \in \mathbb{R}^N$ with $\boldsymbol{\epsilon} \gg \mathbf{0}$, there exists an $\boldsymbol{\epsilon}' \in \mathbb{R}^N$ with $\boldsymbol{\epsilon}' \gg \mathbf{0}$ and a profile $\tilde{\mathbf{u}}^N \in \mathcal{U}^N$ such that $f(\mathbf{u}^N) = f(\tilde{\mathbf{u}}^N)$ and, for all $x \in X$, one has $\boldsymbol{\epsilon} \gg \mathbf{u}^N(x) - \tilde{\mathbf{u}}^N(x) \gg \boldsymbol{\epsilon}'$.

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 …."[6] However, when proving his 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 $\boldsymbol{\epsilon} \in \mathbb{R}^N$ with $\boldsymbol{\epsilon} \gg \mathbf{0}$, there exists an $\boldsymbol{\epsilon}' \in \mathbb{R}^N$ with $\boldsymbol{\epsilon}' \gg 0$ for which, given any profile $\mathbf{u}^N \in \mathcal{U}^N$, there exists a profile $\tilde{\mathbf{u}}^N \in \mathcal{U}^N$ such that $f(\mathbf{u}^N) = f(\tilde{\mathbf{u}}^N)$ and, for all $x \in X$, one has $\boldsymbol{\epsilon} \gg \mathbf{u}^N(x) - \tilde{\mathbf{u}}^N(x) \gg \boldsymbol{\epsilon}'$.

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

Condition (PC) For all $\boldsymbol{\epsilon} \in \mathbb{R}^N$ with $\boldsymbol{\epsilon} \gg \mathbf{0}$, there exists an $\boldsymbol{\epsilon}' \in \mathbb{R}^N$ with $\boldsymbol{\epsilon}' \gg 0$ for which, given any profile $\mathbf{u}^N \in \mathcal{U}^N$ and any pair $x, y \in X$ for which $x \; P(\mathbf{u}^N) \; y$, there exists a profile $\tilde{\mathbf{u}}^N \in \mathcal{U}^N$ with $\tilde{\mathbf{u}}^N(x) \ll \mathbf{u}^N(x) - \boldsymbol{\epsilon}'$ and $\tilde{\mathbf{u}}^N(y) \gg \mathbf{u}^N(y) - \boldsymbol{\epsilon}$ such that $x \; P(\tilde{\mathbf{u}}^N) \; y$.

[6]Indeed, it is this footnote that suggested to me how the above counter example might be constructed.

Like shift invariance, this condition strengthens weak continuity because the same strictly positive vector $\boldsymbol{\epsilon}'$ must work *simultaneously* for all $x, y \in X$. Like uniform shift invariance, it also strengthens shift invariance because the same strictly positive vector $\boldsymbol{\epsilon}'$ must also work for all profiles $\mathbf{u}^N \in \mathcal{U}^N$. On the other hand, condition (PC) weakens even condition (WC) to the extent that the profile $\tilde{\mathbf{u}}^N$ 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, condition (USI) and so (PC) is certainly satisfied if f is invariant under the set of all shift transformations of individual utility functions that take the form $\tilde{u}_i(x) \equiv \alpha + u_i(x)$ (for all $i \in N$ and $x \in X$) with $\alpha \in \mathbb{R}$ independent of i. This is true, for example, if individual utilities satisfy 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 that take the form $\tilde{u}_i(x) \equiv \beta_i \, u_i(x)$ (for all $i \in N$ and $x \in X$) with each individual's $\beta_i > 0$. 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.[7] With condition (PC) replacing (WC), Roberts' Lemma 6 will be proved via the following two separate lemmas that involve definitions (24) and (25):

Lemma 1. *If f satisfies (U), (I) and (P), then for all $\mathbf{v}, \mathbf{v}', \boldsymbol{\eta}, \boldsymbol{\eta}' \in \mathbb{R}^N$ with $\boldsymbol{\eta}, \boldsymbol{\eta}' \gg \mathbf{0}$, one has $\mathbf{v} \in N(\mathbf{v}') \Longrightarrow \mathbf{v} + \boldsymbol{\eta} \in M(\mathbf{v}' - \boldsymbol{\eta}')$.*[8]

Proof. Suppose that x, y, z are three distinct elements of X. By condition (U), there exists a profile $\mathbf{u}^N \in \mathcal{U}^N$ such that

$$\mathbf{v} + \boldsymbol{\eta} \gg \mathbf{u}^N(x) \gg \mathbf{u}^N(y) \gg \mathbf{v} \text{ and } \mathbf{v}' \gg \mathbf{u}^N(z) \gg \mathbf{v}' - \boldsymbol{\eta}' \qquad (26)$$

Now $z \; P(\mathbf{u}^N) \; y$ would imply that $\mathbf{v}' \succ \mathbf{v}$. So $\mathbf{v} \in N(\mathbf{v}') \Longrightarrow y \; R(\mathbf{u}^N) \; z$. Then the Pareto condition (P) implies that $x \, P(\mathbf{u}^N) \, y$, and so $\mathbf{v} \in N(\mathbf{v}') \Longrightarrow x \; P(\mathbf{u}^N) \; z$ because $R(\mathbf{u}^N)$ is transitive. From (26) it follows that $\mathbf{v} \in N(\mathbf{v}') \Longrightarrow \mathbf{v} + \boldsymbol{\eta} \succ \mathbf{v}' - \boldsymbol{\eta}'$. □

[7] 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 we limit the domain of utility profiles $\mathbf{u}^N$ to those that have strictly positive (resp. negative) values throughout X, then one can work instead with $\ln u_i(x)$ (resp. $-\ln[-u_i(x)]$) as a transformed utility function.

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

Part (b) of the following Lemma is a minor restatement of the conclusion of Lemma 6 in Roberts (1983):

Lemma 2. *If f satisfies (U), (I), (P) and (PC), then:*

(a) if $\boldsymbol{\epsilon}, \mathbf{v}, \mathbf{v}'$ in $\mathbb{R}^N$ satisfy $\boldsymbol{\epsilon} \gg \mathbf{0}$ as well as $\mathbf{v} + \boldsymbol{\eta} \in M(\mathbf{v}' + \boldsymbol{\epsilon})$ for all $\boldsymbol{\eta} \gg \mathbf{0}$, then $\mathbf{v} \in M(\mathbf{v}')$;

(b) for all $\mathbf{v}, \mathbf{v}'$ in $\mathbb{R}^N$ that satisfy $\mathbf{v} \in N(\mathbf{v}^*)$, there is no $\boldsymbol{\gamma} \in \mathbb{R}^N$ with $\boldsymbol{\gamma} \gg \mathbf{0}$ such that $\mathbf{v} + \boldsymbol{\gamma} \in N(\mathbf{v}^*)$.

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

$$\mathbf{v} + \boldsymbol{\eta} \gg \mathbf{u}^N(x) \text{ and } \mathbf{u}^N(y) \gg \mathbf{v}' + \boldsymbol{\epsilon}$$

By condition (PC), there exists $\tilde{\mathbf{u}}^N \in \mathcal{U}^N$ such that $x \; P(\tilde{\mathbf{u}}^N) \; y$ while

$$\tilde{\mathbf{u}}^N(x) \ll \mathbf{u}^N(x) - \boldsymbol{\epsilon}' \text{ and } \tilde{\mathbf{u}}^N(y) \gg \mathbf{u}^N(y) - \boldsymbol{\epsilon}$$

But then

$$\tilde{\mathbf{u}}^N(x) \ll \mathbf{v} + \boldsymbol{\eta} - \boldsymbol{\epsilon}' \ll \mathbf{v} \text{ and } \tilde{\mathbf{u}}^N(y) \gg \mathbf{v}'$$

It follows that $\mathbf{v} \succ \mathbf{v}'$.

(b) Suppose that $\mathbf{v} + \boldsymbol{\gamma} \in N(\mathbf{v}^*)$. By definition (25), it follows that $\mathbf{v}^* \in N(\mathbf{v} + \boldsymbol{\gamma})$. Choose any $\boldsymbol{\gamma}' \gg \mathbf{0}$ satisfying $\boldsymbol{\gamma}' \ll \boldsymbol{\gamma}$. Now Lemma 1 implies that $\mathbf{v}^* + \boldsymbol{\eta} \in M(\mathbf{v} + \boldsymbol{\gamma}')$ for all $\boldsymbol{\eta} \gg \mathbf{0}$. So part (a) implies that $\mathbf{v}^* \in M(\mathbf{v})$. In particular, $\mathbf{v} \notin N(\mathbf{v}^*)$. □

4 Conclusion

The weak neutrality or welfarism theorem due to Roberts (1980) is indeed "both important and useful" (p. 428). The minor errors in its statement and in the proof of the key Lemma 6 are not very difficult to correct by replacing the weak continuity condition (WC) with the new pairwise continuity condition (PC) stated here in Section 3.

An open question is whether the closely related Theorem 1 of Roberts (1983) 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 does Roberts' theorem seem inapplicable. This happens to be exactly the setting that we considered in Chichilnisky et al. (2020). But there we use an original position or "impartial benefactor" argument to derive a utilitarian SWFL more directly.

Acknowledgements

An earlier version appeared as Stanford University Department of Economics Working Paper No. 99-021, which is available at `https://web.stanford.edu/~hammond/roberts.pdf`. Later, when John Weymark was an editor of *Social Choice and Welfare*, he very kindly encouraged me to submit a revised version. This seems an excellent occasion on which to follow his suggestion at last.

References

Arrow, K.J. (1963) *Social Choice and Individual Values, 2nd edn.* (New Haven: Yale University Press).

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: 249–268.

Blackorby, C. D. Donaldson and J.A. Weymark (1984) "Social choice with interpersonal utility comparisons: A diagrammatic introduction" *International Economic Review* 25: 327–356.

Bordes, G., P.J. Hammond and M. Le Breton (2005), "Social welfare functionals on restricted domains and in economic environments" *Journal of Public Economic Theory* 7: 1–25.

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 (2004) "Utility in social choice," in *Handbook of Utility Theory, Vol. II: Applications and Extensions* by S. Barberà, P.J. Hammond, and C. Seidl, eds. (Kluwer: Dordrecht).

Chichilnisky, G., P.J. Hammond and N. Stern (2020) "Fundamental utilitarianism and intergenerational equity with extinction discounting" *Social Choice and Welfare* 54: 397–427.

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

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

Lavi, R., A. Mu'alem and N. Nisan (2009) "Two simplified proofs for Roberts' theorem" *Social Choice and Welfare* 32: 407–423.

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.

Mishra, D. and A. Sen (2012) "Roberts' theorem with neutrality: A social welfare ordering approach" *Games and Economic Behavior* 75: 283–298.

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

Roberts, K.W.S. (1979) "The characterization of implementable choice rules" in *Aggregation and Revelation of Preferences* J.-J. Laffont (ed.) (Amsterdam: North-Holland) 321–349.

Roberts, K.W.S. (1980) "Interpersonal comparability and social choice theory" *Review of Economic Studies* 47: 421–439.

Roberts, K.W.S. (1983) "Social choice rules and real valued representations" *Journal of Economic Theory* 29: 72–94.

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

Sen, A.K. (1977) "On weights and measures: Informational constraints in social welfare analysis" *Econometrica* 45: 1539–72.

Sen, A.K. (1984) "Social choice theory" in *Handbook of Mathematical Economics, Vol. 3* K.J. Arrow and M.D. Intriligator (eds.) (North-Holland: Amsterdam) 22: 1073–1181.

Tsui, K.-Y. and J.A. Weymark (1997) "Social welfare orderings for ratio-scale measurable utilities" *Economic Theory* 9: 241–256.