Roberts’ weak welfarism theorem: a minor correction

Peter J. Hammond

Received: 24 September 2020 / Accepted: 30 March 2021 © The Author(s) 2021

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 (Rev Econ Stud 47(2):421–439, 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.

1 Introduction: Roberts’ Claim

Consider a society with a non-empty finite set $N$ of individuals. 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 $u^N$ is a mapping $X \ni x \mapsto u^N(x) \in \mathbb{R}^N$. 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 $\mathcal{D} \ni u^N \mapsto f(u^N) \in \mathcal{R}(X)$ that determines a (complete and transitive) social preference ordering $R = f(u^N)$ for each utility function profile in $\mathcal{D}$. With some slight abuse of notation, we let $R(u^N)$, $P(u^N)$ and $I(u^N)$ denote respectively the weak preference, strict preference, and indifference relations associated with $f(u^N)$.

---

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

\[ \text{p.j.hammond@warwick.ac.uk} \]

1 Department of Economics, University of Warwick, Coventry CV4 7AL, UK
Given any non-empty subset $A \subset X$, say that:

1. two utility function profiles $u^N, \tilde{u}^N \in \mathcal{U}^N$ are equal on $A$ just in case one has $u^N(x) = \tilde{u}^N(x)$ for all $x \in A$;
2. 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 $y R z \iff y \tilde{R} z$.

The main part of this paper considers social welfare functionals $f$ which satisfy 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 $u^N, \tilde{u}^N \in \mathcal{D}$ are equal on $A$, then the two associated social orderings $f(u^N), f(\tilde{u}^N) \in R(X)$ are also equal on $A$.

Pareto indifference (P0) : In case $y, z \in X$ and $u^N \in \mathcal{D}$ satisfy $u^N(y) = u^N(z)$, the associated social indifference relation satisfies $y I (u^N) z$.

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 (P0), then there exists a (complete and transitive) social welfare ordering $R^*$ on $\mathbb{R}^N$ with the property that $y R z \iff u^N(y) R^* u^N(z)$. This plays a prominent role among the results appearing in the surveys by Sen (1984), Blackorby et al. (1984), D’Aspremont (1985), Mongin and d’Aspremont (1998), and Bossert and Weymark (2004). Both Sen (1977) and d’Aspremont (D’Aspremont (1985), p. 34) provide complete proofs.2

While the Pareto indifference axiom (P0) is appealing, the impossibility theorem in Arrow (1963) is the most prominent of many results that replace it with the following alternative:

Pareto (P) : In case the two social states $y, z \in X$ and the utility function profile $u^N \in \mathcal{D}$ satisfy $u^N(y) \gg u^N(z)$, the associated strict preference relation $P(u^N)$ on $X$ satisfies $y P(u^N) z$.3

Specifically, under the assumption that individuals’ utility functions 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

---

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 $a, b \in \mathbb{R}^N$ with $a = (a_i)_{i \in N}$ and $b = (b_i)_{i \in N}$, we use the following notation for vector orderings: (i) $a \succeq b$ in case $a_i \geq b_i$ for all $i \in N$; (ii) $a \gg b$ in case $a_i > b_i$ for all $i \in N$; (iii) $a > b$ in case $a \succeq b$ but $a \neq b$. We also let $\mathbb{R}^N_{++}$ denote the set $\{a \in \mathbb{R}^N \mid a \gg 0\}$ of strictly positive vectors in $\mathbb{R}^N$. Springe
case, Roberts (1980, p. 427) specifies an additional condition that can be restated as follows:⁴

**Weak continuity (WC)** : For all utility function profiles $\mathbf{u}^N \in \mathcal{D}$ and all vectors $\epsilon \in \mathbb{R}^N$ with $\epsilon \gg 0$, there exists a utility function profile $\mathbf{u}^N_{\epsilon} \in \mathcal{D}$ satisfying $\epsilon \gg \mathbf{u}^N_{\epsilon}(x) - \mathbf{u}^N(x) \gg 0$ for all social states $x \in X$ with the property that $f(\mathbf{u}^N) = f(\mathbf{u}^N_{\epsilon})$.

Then Roberts (1980, p. 428) claims the following:

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

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

This claim has come to be known as Roberts’ “weak neutrality” or “weak welfarism” theorem.⁵ 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 (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 et al. (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 the two social states $y, z \in X$ and the utility function profile $\mathbf{u}^N \in \mathcal{D}$ satisfy $\mathbf{u}^N(y) \succeq \mathbf{u}^N(z)$, the associated social preferences satisfy $y R(\mathbf{u}^N) z$, with $y P(\mathbf{u}^N) z$ unless $\mathbf{u}^N(y) = \mathbf{u}^N(z)$.

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.

---

⁴ Roberts’ original condition has been restated in a way that seems more appropriate if there is a restricted domain $\mathcal{D} \subset \mathcal{P}^N$ of admissible utility profiles.

⁵ 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).
2 Weak Continuity: A Counter Example

2.1 Definition of a Discontinuous Social Welfare Ordering

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

\[ \mathbb{R}^2 \ni (u_1, u_2) = u \mapsto W(u_1, u_2) = W(u) \to \mathbb{R} \]  

such that the induced SWFL defined on \( X \) by

\[ a R b \iff W(u_1(a), u_2(a)) \geq W(u_1(b), u_2(b)) \]  

satisfies conditions (U), (I), (P*) and (WC). Yet the function \( W \) that we will define has a discontinuity at the origin \( \mathbf{0} = (0, 0) \) which implies a discontinuity in the preference ordering \( R \) on \( \mathbb{R}^2 \) defined by (2). This implies that no continuous function \( W \) can satisfy Claim 1 in this example.

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

\[ w(v) := \min\{v_1 + 2v_2, 2v_1 + v_2\} \]  

Next, partition \( \mathbb{R}^2 \) into the three subdomains \( S_1, S_2, S_3 \) and then define the symmetric function \( S_i \ni v \mapsto W(v) \in \mathbb{R} \) for \( i = 1, 2, 3 \) so that

\[ W(v) := v_1 + v_2 \quad \text{on} \quad S_1 := \{ v \in \mathbb{R}^2 \mid v_1 + v_2 \leq 0 \} \]  

\[ W(v) := 1 + w(v) \quad \text{on} \quad S_2 := \{ v \in \mathbb{R}^2 \mid w(v) > 0 \} \]  

\[ W(v) := \exp\left(\frac{(v_1 + 2v_2)(2v_1 + v_2)}{3(v_1 + v_2)}\right) \quad \text{on} \quad S_3 := \mathbb{R}^2 \setminus (S_1 \cup S_2) \]  

The indifference map corresponding to this function is illustrated in Fig. 1, which shows how:

(i) the closed set \( S_1 \) is separated from \( S_3 \) by the common boundary line \( \{ v \in \mathbb{R}^2 \mid v_1 + v_2 = 0 \} \), which constitutes the closed indifference curve \( W(v) = 0 \);

(ii) the open set \( S_2 \) is separated from \( S_3 \) by the common boundary set \( \{ v \in \mathbb{R}^2 \mid w(v) = 0 \} \);

(iii) \( (0, 0) \) belongs to \( S_1 \), but is a common boundary point of all three subdomains.

Consider first the intermediate subdomain \( S_3 \), which consists of two wedges where \( v_1 + v_2 > 0 \) and \( w(v) \leq 0 \). On \( S_3 \), it follows from (6) that

\[ \ln W(v) = \frac{2(v_1)^2 + 5v_1v_2 + 2(v_2)^2}{3(v_1 + v_2)} = \frac{2(v_1 + v_2)}{3} + \frac{v_1v_2}{3(v_1 + v_2)} \]  

\( \text{Springer} \)
Roberts’ weak welfarism theorem: a minor correction

In $S_3$, because $v_1 + v_2 > 0$, this is a differentiable function of $v$. Its first partial derivative is

$$\frac{\partial}{\partial v_1} \ln W = \frac{2}{3} + \frac{v_2(v_1 + v_2) - v_1v_2}{3(v_1 + v_2)^2} = \frac{2}{3} + \frac{(v_2)^2}{3(v_1 + v_2)^2} > 0$$

By symmetry, its second partial derivative is also positive. This shows that $S_3 \ni v \mapsto W(v) \in \mathbb{R}$ is strictly increasing.

Furthermore, consider any sequence $\mathbb{N} \ni n \mapsto v^n = (v^n_1, v^n_2) \in S_3$ which converges to a point $\bar{v} = (\bar{v}_1, \bar{v}_2) \neq (0, 0)$ that lies on the lower boundary of $S_3$ because $\bar{v}_1 + \bar{v}_2 = 0$. Evidently it follows from (7) that, as $n \to \infty$, so $\ln W(v^n_1, v^n_2) \to -\infty$, implying that $W(v^n_1, v^n_2) \to 0$. Also, it follows from (3) and (6) that $W(v) = 1$ when $v$ lies on the upper boundary of $S_3$, where $w(v) = 0$. So the range set $W(S_3)$ must be the whole interval $(0, 1]$.

Fig. 1 Eight level curves of the function $v \mapsto W(v_1, v_2)$
Putting these results for the function \( \mathbb{R}^2 \ni v \mapsto W(v) \) on the subdomain \( S_1 \), together with obvious properties on the subdomains \( S_1 \) and \( S_2 \) where \( W \) is defined by (4) and (5), we see that the respective range sets are the three pairwise disjoint line intervals 

\[
W(S_1) = (−\infty, 0], \quad W(S_2) = (1, \infty), \quad W(S_3) = (0, 1]
\]

From this it is evident that the function \( v \mapsto W(v) \) is strictly increasing throughout \( \mathbb{R}^2 \), and continuous except at \((0, 0)\), where \( W(0, 0) = 0 \).

The three-dimensional graph of \( W(v_1, v_2) \) has a boundary that includes the vertical “cliff” \( \{(0, 0)\} \times [0, 1] \) in \( \mathbb{R}^3 \) of height 1. The base of this cliff is at the origin \((0, 0, 0)\), which is on the graph of \( W \) because \( W(0, 0) = 0 \). Thus, the mapping \( \mathbb{R}^2 \ni v \mapsto W(v) \) is discontinuous at \( v = (0, 0) \). Not only is the function \( W \) discontinuous; so is the preference relation it induces. Indeed, for each fixed \( \bar{v} \in S_3 \) where \( W(\bar{v}) \in (0, 1] \), the upper contour set \( \{(v_1, v_2) \in \mathbb{R}^2 \mid W(v_1, v_2) \geq W(\bar{v})\} \) is not closed because it excludes the point \((0, 0)\) which is in its closure.

### 2.2 Verifying Weak Continuity

Consider now the SWFL \( \mathbb{R}^2 \ni (u_1, u_2) \mapsto f(u_1, u_2) \) defined as in Section 2.1. Obviously, this induced SWFL satisfies conditions (U), (I) and (P*). To verify condition (WC) it is enough to construct, for each fixed \( \epsilon = (\epsilon_1, \epsilon_2) \in \mathbb{R}^2_{++} \), a transformation 

\[
\mathbb{R}^2 \ni v \mapsto \phi^\epsilon(v) = (\phi_1^\epsilon(v), \phi_2^\epsilon(v)) \in \mathbb{R}^2
\]

satisfying 

\[
(0, 0) \ll (v_1, v_2) - \phi^\epsilon(v) \ll \epsilon
\]

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

In the following constructions, given any fixed \( \epsilon = (\epsilon_1, \epsilon_2) \in \mathbb{R}^2_{++} \), let

\[
\epsilon_* := \min\{\epsilon_1, \epsilon_2\} \in \mathbb{R} \quad \text{and} \quad \epsilon := (1, 1) \in \mathbb{R}^2 \tag{8}
\]

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

\[
\mathbb{R}^2 \ni v \mapsto \phi^\epsilon(v) := v - \lambda^\epsilon(v) \epsilon \in \mathbb{R}^2 \tag{9}
\]

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

**Case 1:** The simplest case is when

\[
v_1 + v_2 \leq 0 \quad \text{and so} \quad W(v) = v_1 + v_2 \leq 0 \tag{10}
\]

In this case, define \( \lambda^\epsilon(v) := \frac{1}{2} \epsilon_* \) for \( \epsilon_* \) given by (8). Then (9) implies that

\[
\phi^\epsilon(v) = (v_1 - \frac{1}{2} \epsilon_*, v_2 - \frac{1}{2} \epsilon_*)
\]
So (9) and (10) imply that $\phi_1^e(v) + \phi_2^e(v) = v_1 + v_2 - \epsilon_+ \leq -\epsilon_* < 0$. Now, whenever $v_1 + v_2 \leq 0$, it follows that

$$W(\phi^e(v)) = \phi_1^e(v) + \phi_2^e(v) = \psi^e(W(v))$$

provided we define

$$\psi^e(W) := W - \epsilon_* \quad \text{for all } W \leq 0 \tag{11}$$

**Case 2:** This case occurs when

$$w(v) > 0 \quad \text{and so } W(v) = 1 + w(v) > 1 \tag{12}$$

In this case, define

$$\lambda^e(v) := \frac{1}{6} \min\{\epsilon_-, w(v)\} \tag{13}$$

Clearly, this definition implies that $\lambda^e(v) \in (0, \epsilon_+).$ Also

$$\phi_1^e(v) + 2\phi_2^e(v) = v_1 + 2v_2 - 3\lambda^e(v) \tag{14}$$

and

$$2\phi_1^e(v) + \phi_2^e(v) = 2v_1 + v_2 - 3\lambda^e(v) \tag{15}$$

But $w(v) := \min\{v_1 + 2v_2, 2v_1 + v_2\}$ and (13) implies that $\lambda^e(v) \leq \frac{1}{6} w(v)$. So it follows from (14) and (15) that

$$\min\{\phi_1^e(v) + 2\phi_2^e(v), 2\phi_1^e(v) + \phi_2^e(v)\} = W(v) - 3\lambda^e(v) \geq \frac{1}{2} w(v) > 0 \tag{16}$$

So the definition of $\mathbb{R}^2 \ni v \mapsto W(v) \in \mathbb{R}$ in (5) and of $\mathbb{R}^2 \ni v \mapsto \lambda^e(v) \in \mathbb{R}$ in (13) imply, together with (16), that

$$1 < W(\phi^e(v)) = 1 + \min\{\phi_1^e(v) + 2\phi_2^e(v), 2\phi_1^e(v) + \phi_2^e(v)\}$$

$$= 1 + w(v) - 3\lambda^e(v)$$

$$= 1 + w(v) - \frac{1}{2} \min\{\epsilon_-, w(v)\}$$

$$= \max\{1 + w(v) - \frac{1}{2} \epsilon_-, 1 + \frac{1}{2} w(v)\}$$

$$= \max\{W(v) - \frac{1}{2} \epsilon_-, \frac{1}{2} [W(v) + 1]\}$$

It follows that $W(\phi^e(v)) = \psi^e(W(v))$ provided that we define

$$\psi^e(W) := \max\{W - \frac{1}{2} \epsilon_-, \frac{1}{2} (W + 1)\} \quad \text{for all } W > 1 \tag{17}$$

**Case 3:** This leaves the hardest third case, when

$$w(v) \leq 0 \quad \text{and also } v_1 + v_2 > 0 \tag{18}$$

In this case, the definition in (6) implies that $0 < W(v) \leq 1$.

Fix any $v = (v_1, v_2) \in \mathbb{R}^2$ satisfying (18). Then, given any $\epsilon \in \mathbb{R}^2$ satisfying $\epsilon \gg 0$, consider the non-empty open interval of $\mathbb{R}$ defined by
Now consider the function $g$ defined on the open interval in (19) by

$$I^c(v) := (0, \min\{\epsilon_*, \frac{1}{2}(v_1 + v_2)\}) = (0, \epsilon_*) \cap (0, \frac{1}{2}(v_1 + v_2)) \quad (19)$$

Now let $\mu$ denote a suitably chosen positive scalar constant which is independent of both $v$ and $\epsilon_*$, and whose possible range will be specified later. For each $v \in \mathbb{R}^2$ that satisfies the inequalities (18), because $g$ is strictly decreasing and positive, we can define $\lambda^c(v)$ implicitly as the unique value of $\lambda$ that solves the equation

$$g(\lambda) = W(v - \lambda e) = W(v) \exp(-\mu \epsilon_*) \quad (24)$$

Then $\lambda^c(v)$ will be well defined and positive, with

$$W(\phi^c(v)) = W(v) \exp(-\mu \epsilon_*) = \psi^c(W(v)) < 1$$

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

It remains only to choose $\mu > 0$ so that the corresponding solution to equation (24) exists in the open interval $I^c(v)$ defined by (19) and so satisfies $\lambda^c(v) < \epsilon_*$. Because we are assuming that the inequalities (18) hold, definition (6) implies that any $\lambda^c(v)$ satisfying (24) and (19) must be a value of $\lambda$ 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 $\lambda$, so we can clear fractions to obtain the quadratic equation $q(\lambda) = 0$, where

$$q(\lambda) := (v_1 + v_2)(v_1 + 2v_2 - 3\lambda)(2v_1 + v_2 - 3\lambda) - (v_1 + v_2)(2v_1 + v_2)(v_1 + v_2 - 2\lambda) + 3\mu \epsilon_*(v_1 + v_2)(v_1 + v_2 - 2\lambda) \quad (25)$$

Now, note that when $\lambda = 0$ the first two terms on the right-hand side of (25) cancel. Because $v_1 + v_2 > 0$, it follows that
\[ q(0) = 3 \mu \varepsilon_* (v_1 + v_2)^2 > 0 \] (26)

In addition, simple calculation shows that
\[ q\left(\frac{1}{2}(v_1 + v_2)\right) = -\frac{1}{4}(v_1 + v_2)(v_1 - v_2)^2 \] (27)

Finally, some much more tedious but still routine algebraic manipulation shows that
\[ q(\varepsilon_*) = (v_1 + v_2)\varepsilon_* [(9 - 6\mu)\varepsilon_* + (3\mu - 5)(v_1 + v_2)] + 2v_1 v_2 \varepsilon_* \] (28)

Because \( v_1 + v_2 > 0 \) but \( w(v) \leq 0 \), 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 (27) implies that \( q(\frac{1}{2}(v_1 + v_2)) < 0 \) whereas (28) implies that for any \( \mu \) satisfying \( 9 < 6\mu < 10 \) one has \( q(\varepsilon_*) < 0 \). 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 a unique root \( \lambda^e(v) \) in the open interval \( I^e(v) = (0, \min\{\varepsilon_*, \frac{1}{2}(v_1 + v_2)\}) \) defined by (19).\(^6\) In particular, for each \( v \) satisfying (18), the root \( \lambda^e(v) \) of \( q(\lambda) = 0 \) that we have found lies in \( (0, \varepsilon_*) \), as required.

Finally, putting all the three different cases together gives \( W(\phi^e(v)) \equiv \psi^e(W(v)) \), where \( \phi^e(v) = v - \lambda^e(v)e \), and then
\[
\psi^e(W) := \begin{cases} 
W - \frac{1}{2}(\varepsilon_1 + \varepsilon_2) < 0 & \text{if } W \leq 0; \\
W \exp(-\mu \varepsilon_*) \in (0, 1] & \text{if } 0 < W \leq 1; \\
\max\{W - \frac{1}{2} \varepsilon_*, \frac{1}{2} (W + 1)\} > 1 & \text{if } W > 1.
\end{cases}
\]

In particular, \( \psi^e \) is strictly increasing in \( W \) for each fixed \( e \gg 0 \). \(\square\)

### 2.3 Diagnosis

Next, to see where his proof erred, we introduce a definition that incorporates some more notation from Roberts (1980, pp. 425–426).

**Definition 1** Given the SWFL \( f \), the strict preference relation \( > \) on utility vectors in \( \mathbb{R}^N \) is defined so that \( a > b \) just in case there exist a utility function profile \( u^N \in \mathcal{D} \) and two social states \( x, y \in X \) with \( xf(u^N) \leq y \) such that \( a \gg u(x) \) and \( u(y) \gg b \). Then, for each \( v^* \in \mathbb{R}^N \), the three sets \( L(v^*) \), \( M(v^*) \) and \( N(v^*) \) are defined respectively by
\[
L(v^*) := \{ v \in \mathbb{R}^N \mid v^* > v \} \quad (29)
\]
\[
M(v^*) := \{ v \in \mathbb{R}^N \mid v > v^* \} \quad (30)
\]
\[
N(v^*) := \mathbb{R}^N \backslash [L(v^*) \cup M(v^*)] \quad (31)
\]

\(^6\) Because \( q(\lambda) \to +\infty \) as \( \lambda \to +\infty \), the quadratic equation \( q(\lambda) = 0 \) has a second real root that satisfies \( \lambda > \max\{\varepsilon_*, \frac{1}{2}(v_1 + v_2)\} \). But this root is irrelevant.
Now, in the above example 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 \( \eta, \eta' \succ 0 \) with \( \eta \) 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 \( v + \gamma \in N(v^*) \), [condition] (WC) ensures that \( v + \gamma - \eta_3 \in N(v^*-\eta_4) \) for some \( \epsilon \gg \eta_3, \eta_4 \gg 0 \) where \( \epsilon \) is subject to choice.”

### 3 A new sufficient condition

#### 3.1 Statement of pairwise continuity

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

**Condition (SI)**: Given any \( \epsilon \in \mathbb{R}^N_+ \), for all profiles \( u^N \in \mathcal{D} \), there exists an \( \epsilon' \in \mathbb{R}^N_+ \) and a profile \( \tilde{u}^N \in \mathcal{D} \) such that \( f(u^N) = f(\tilde{u}^N) \) and, for all \( x \in X \), one has \( \epsilon \gg u^N(x) - \tilde{u}^N(x) \gg \epsilon' \).

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 ....” However, when proving his Lemma A.5, it seems that Roberts (1983, p. 90) in the end reverses the order of the quantified statements “for all profiles \( u^N \in \mathcal{U}^N \) and “there exists an \( \epsilon' \in \mathbb{R}^N_+ \)” and actually uses the following uniform shift invariance assumption:

**Condition (USI)**: Given any \( \epsilon \in \mathbb{R}^N_+ \), there exists an \( \epsilon' \in \mathbb{R}_+^N \) for which, for all profiles \( u^N \in \mathcal{D} \), there exists a profile \( \tilde{u}^N \in \mathcal{D} \) such that \( f(u^N) = f(\tilde{u}^N) \) and, for all \( x \in X \), one has \( \epsilon \gg u^N(x) - \tilde{u}^N(x) \gg \epsilon' \).

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

**Condition (PC)**: Given any \( \epsilon \in \mathbb{R}_+^N \), there exists an \( \epsilon' \in \mathbb{R}_+^N \) for which, for all profiles \( u^N \in \mathcal{D} \), given any pair \( x, y \in X \) satisfying \( x P(u^N) y \), there exists a profile \( \tilde{u}^N \in \mathcal{D} \) with \( \tilde{u}^N(x) \ll u^N(x) - \epsilon' \) and \( \tilde{u}^N(y) \gg u^N(y) - \epsilon \) such that \( x P(\tilde{u}^N) y \).

Like shift invariance, condition (PC) strengthens weak continuity because the same strictly positive vector \( \epsilon' \) must work simultaneously for all \( x, y \in X \). Like uniform shift invariance, it also strengthens shift invariance because the same strictly positive vector \( \epsilon' \) must also work for all profiles \( u^N \in \mathcal{D} \). On the other

---

7 Indeed, it is this footnote that suggested to me how the above counter example might be constructed.
hand, condition (PC) weakens even condition (WC), as well as condition (USI), to
the extent that the profile $\tilde{u}^N$ can depend on the pair $x, y \in X$, and also because it
requires only one-way strict inequalities which, moreover, are different for $x$ and $y$.

3.2 Sufficiency of pairwise continuity

With condition (PC) replacing (WC), Lemma 6 of Roberts (1980) will be proved via
the following two separate lemmas that involve definitions (30) and (31):

Lemma 1 If $f$ satisfies (U), (I) and (P), then for all $v, v', \eta, \eta' \in \mathbb{R}^N$ with $\eta, \eta' \gg 0$, one has $v \in N(v') \implies v + \eta \in M(v' - \eta')$.\footnote{This is the correct “preliminary result” in Roberts’ discussion of Lemma 6. However, it seemed that the proof provided there could benefit from more detail.}

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

$$v + \eta \gg u^N(x) \gg u^N(y) \gg v \text{ and } v' \gg u^N(z) \gg v' - \eta'$$ (32)

By Definition 1, if $z P(u^N) y$ were true, it would imply that $v' > v$. So
$v \in N(v') \implies y R(u^N) z$. Then the Pareto condition (P) implies that $x P(u^N) y$, and
so $v \in N(v') \implies x P(u^N) z$ because $R(u^N)$ is transitive. From (32) it follows that
$v \in N(v') \implies v + \eta > v' - \eta'$.

Part (b) of the following Lemma is a minor restatement of the conclusion of
Lemma 6 in Roberts (1980):

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

(a) if $\epsilon, v, v'$ in $\mathbb{R}^N$ satisfy $\epsilon \gg 0$ as well as $v + \eta \in M(v' + \epsilon)$ for all $\eta \gg 0$, then $v \in M(v')$;
(b) for all $v, v^* \in \mathbb{R}^N$ that satisfy $v \in N(v^*)$, there is no $\gamma \in \mathbb{R}^N_{++}$ such that
$v + \gamma \in N(v^*)$.

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 $v + \eta \in M(v' + \epsilon)$, condition (U) and Defini-
tion 1 imply that there exist $u^N \in \varnothing$ and $x, y \in X$ such that $x P(u^N) y$ while

$$v + \eta \gg u^N(x) \text{ and } u^N(y) \gg v' + \epsilon$$ (33)

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

$$\tilde{u}^N(x) \ll u^N(x) - \epsilon' \text{ and } \tilde{u}^N(y) \gg u^N(y) - \epsilon$$ (34)
But then (33) and (34) together imply that $\tilde{u}^N(x) \ll v + \eta - \epsilon' \ll v$, because of the choice of $\eta$; they also imply that $\tilde{u}^N(y) \gg v'$. Because $x P(\tilde{u}^N)y$, it follows from Definition 1 that $v > v'$.

(b) Suppose that $v + \gamma \in N(v^*)$. Definition (31) implies that $v^* \in N(v + \gamma)$. Choose any $\gamma' \gg 0$ satisfying $\gamma' \gg \gamma$. Now Lemma 1 implies that $v^* + \eta \in M(v + \gamma')$ for all $\eta \gg 0$. So part (a) implies that $v^* \in M(v)$. In particular, $v \not\in N(v^*)$.

3.3 Invariance conditions and pairwise continuity

Just as with Roberts’ (WC) and (SI) conditions, condition (USI) and so (PC) is certainly satisfied if $f$ is invariant under a set of transformations of individual utility functions large enough to include all shifts of 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 $i$. Of the six invariance classes $\Phi$ presented on p. 423 of Roberts (1980), this is true of the first five, namely (ONC), (CNC), (OLC), (CUC), and (CFC).

It remains to discuss the sixth cardinal ratio-scale or (CRS) class, as well as the broader non-comparable ratio-scale or (NRS) invariance class that includes all transformations of the form $\tilde{u}_i(x) \equiv \beta_i u_i(x)$ (for all $i \in N$ and $x \in X$) with $\beta_i > 0$ for each individual $i \in N$. Indeed, since the unrestricted domain condition (U) allows each utility function to have both positive and negative values, there are difficulties with the key pairing continuity condition (PC) in this case.10

When Roberts (1980, p. 423) discusses the (CRS) class, he states that “For simplicity, it will be assumed that welfares are always strictly positive.” This, of course, violates condition (U), but almost all the relevant results are easily restated for any restricted domain $\mathcal{D} \subset \mathbb{R}^N$ large enough to include all profiles of utility functions whose values are strictly positive. Nevertheless, the key pairwise continuity condition (PC) is violated because, no matter how small $\epsilon' \in \mathbb{R}^N_{++}$ may be, for any $x \in X$ there will always be a strictly positive-valued profile $u^N \in \mathcal{D}$ such that $u^N(x) \ll \epsilon'$; this makes it impossible to find any profile $\tilde{u}^N$ of strictly positive-valued utility functions that satisfies $\tilde{u}^N(x) \ll u^N(x) - \epsilon'$.

When utilities are restricted to have strictly positive values throughout $X$, however, one can work instead with $\ln u_i(x)$ as a transformed utility function. Then ratio-scale invariance for utilities is equivalent to invariance under additive shifts of their logarithms. Moreover, condition (PC) can be restated for these transformed utility functions. A similar trick works when utility functions are restricted to have strictly negative values; one works instead with $-\ln[-u_i(x)]$ as a transformed utility function. The (CRS) and (NRS) invariance classes present problems only when utilities are allowed to change signs, or to become zero.

---

9 This subsection in particular owes much to John Weymark and the anonymous referee.

10 There are also difficulties with Lemma 8 of Roberts (1980). We note that Blackorby and Donaldson (1982) as well as Tsui and Weymark (1997), in their work on ratio-scale invariant social welfare functionals, imposed a more direct continuity condition on a social preference ordering which is defined on the space of utility vectors.
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 3.1.

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. With zero as the utility of the infinite set of potential people who, in each possible social state, never come into existence, this happens to be exactly the setting that we consider in Chichilnisky et al. (2020). But in that paper we extend the kind of original position due to Vickrey (1945) and Har- sanyi (1953; 1955). Then we use an “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, during John Weymark’s term as 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. Finally, I am grateful for the most helpful comments of a diligent referee.

Declarations

Conflict of interest The author declares that there is no conflict of interest.

Funding Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.
References

Arrow KJ (1963) Social choice and individual values, 2nd edn. Yale University Press, New Haven
Blackorby C, Donaldson D (1982) Ratio-scale and translation-scale full interpersonal comparability without domain restrictions: admissible social-evaluation functions. Int Econ Rev 23(2):249–268
Blackorby C, Donaldson D, Weymark JA (1984) Social choice with interpersonal utility comparisons: a diagrammatic introduction. Int Econ Rev 25(2):327–356
Bordes G, Hammond PJ, Le Breton M (2005) Social welfare functionals on restricted domains and in economic environments. J Public Econ Theory 7(1):1–25
Bordes G, Le Breton M (1987) Roberts’s theorem in economic environments: (2) The private goods case. Faculté des Sciences Economiques et de Gestion, Université de Bordeaux I, Preprint
Bossert, W, Weymark JA (2004) Utility in social choice. In: Barberà, S, Hammond, PJ, Seidl, C (eds) Handbook of utility theory, vol. II: Applications and extensions. Kluwer, Dordrecht, ch 20, pp 1099–1177
Chichilnisky G, Hammond PJ, Stern N (2020) Fundamental utilitarianism and intergenerational equity with extinction discounting. Soc Choice Welf 54(2–3):397–427
D’Aspremont C (1985) Axioms for social welfare orderings. In: Hurwicz, L, Schmeidler, D, Sonnenschein, H (eds) Social goals and social organization: Essays in memory of Elisha Pazner. Cambridge University Press, Cambridge, ch 1, pp 19–76
D’Aspremont C, Gevers L (1977) Equity and the informational basis of collective choice. Rev Econ Stud 44(2):199–209
Harsanyi JC (1953) Cardinal utility in welfare economics and in the theory of risk-taking. J Polit Econ 61(5):434–435
Harsanyi JC (1955) Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy 63(4):309–321
Lavi R, Mu’alem A, Nisan N (2009) Two simplified proofs for Roberts’ theorem. Soc Choice Welf 32(3):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, Sen A (2012) Roberts’ theorem with neutrality: a social welfare ordering approach. Games Econ Behav 75(1):283–298
Mongin P, d’Aspremont C (1998) Utility theory and ethics. In: Barberà S, Hammond PJ, Seidl C (eds) Handbook of utility theory, vol. I: Principles. Kluwer, Dordrecht, ch 10, pp 371–481
Roberts KWS (1979) The characterization of implementable choice rules. In: Laffont J-J (ed) Aggregation and revelation of preferences. North-Holland, Amsterdam, pp 321–349
Roberts KWS (1980) Interpersonal comparability and social choice theory. Rev Econ Stud 47(2):421–439
Roberts KWS (1983) Social choice rules and real valued representations. J Econ Theory 29(1):72–94
Sen AK (1970) Collective choice and social welfare. Holden-Day, San Francisco
Sen AK (1977) On weights and measures: informational constraints in social welfare analysis. Econometrica 45(7):1539–72
Sen AK (1984) Social choice theory. In: Arrow KJ, Intriligator MD (eds) Handbook of mathematical economics, Vol. 3. North-Holland, Amsterdam, ch 22, pp 1073–1181
Tsui K-Y, Weymark JA (1997) Social welfare orderings for ratio-scale measurable utilities. Econ Theor 10(2):241–256
Vickrey WS (1945) Measuring marginal utility by reactions to risk. Econometrica 13(4):319–333

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.