Soc Choice Welfare
DOI 10.1007/s00355-008-0313-0

ORIGINAL PAPER

Twofold optimality of the relative utilitarian bargaining solution

Marcus Pivato

Received: 9 April 2007 / Accepted: 27 March 2008

Abstract Given a bargaining problem, the relative utilitarian (RU) solution maximizes the sum total of the bargainer's utilities, after having first renormalized each utility function to range from zero to one. We show that RU is "optimal" in two very different senses. First, RU is the maximal element (over the set of all bargaining solutions) under any partial ordering which satisfies certain axioms of fairness and consistency; this result is closely analogous to the result of Segal (J Polit Econ 108(3):569–589, 2000). Second, RU offers each person the maximum expected utility amongst all rescaling-invariant solutions, when it is applied to a random sequence of future bargaining problems generated using a certain class of distributions; this is recalls the results of Harsanyi (J Polit Econ 61:434–435, 1953) and Karni (Econometrica 66(6):1405–1415, 1998).

0 Introduction

Let $\mathcal{I}$ be a finite group of individuals, and let $\mathcal{A}$ be a set of social outcomes (e.g. allocations of some finite stock of resources). If each $i \in \mathcal{I}$ has an ordinal preference relation over the set of all lotteries between elements in $\mathcal{A}$, and this preference relation satisfies certain axioms, then von Neumann and Morgenstern showed that there is a ("vNM") cardinal utility function $u_i : \mathcal{A} \longrightarrow \mathbb{R}_{\neq} := [0, \infty)$ such that i's lottery preferences are consistent with maximization of the expected value of u_i. Let $\mathbf{u} := (u_i)_{i \in \mathcal{I}} : \mathcal{A} \longrightarrow \mathbb{R}_{\neq}^{\mathcal{I}}$ be the "joint" utility function, and let $\mathcal{B}$ be the convex, comprehensive closure of the image set $\mathbf{u}(\mathcal{A}) \subset \mathbb{R}_{\neq}^{\mathcal{I}}$; then any element of $\mathcal{B}$ is an assignment of a vNM utility level to each player, obtainable through some lottery between elements of $\mathcal{A}$.

M. Pivato (✉)
Department of Mathematics, Trent University, 1600 West Bank Drive,
Peterborough, ON, Canada K9J 7B8
e-mail: marcuspivato@trentu.ca

For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{\mathcal{I}}_{\not\gamma}$, we write "$\mathbf{x} \stackrel{\wp}{\succeq} \mathbf{y}$" to mean $\mathbf{x}$ is Pareto-preferred to $\mathbf{y}$ (i.e. $\forall i \in \mathcal{I},\ x_i \geq y_i$) and "$\mathbf{x} \stackrel{\wp}{\succ} \mathbf{y}$" to mean $\mathbf{x}$ is strictly Pareto-preferred to $\mathbf{y}$ (i.e. $\forall i \in \mathcal{I},\ x_i > y_i$). Let $\wp\mathcal{B} := \left\{\mathbf{b} \in \mathcal{B}\ ;\ \nexists\, \mathbf{b}' \in \mathcal{B} \text{ with } \mathbf{b}' \stackrel{\wp}{\succ} \mathbf{b}\right\}$ be the (weak) *Pareto frontier* of $\mathcal{B}$. We assume that the members of $\mathcal{I}$ can obtain any social outcome in $\wp\mathcal{B}$, but only through unanimous consent. Let $a_0 \in \mathcal{A}$ represent the "status quo" outcome, which we assume to be Pareto-suboptimal. If $\mathbf{q} := \mathbf{u}(a_0) \in \mathcal{B}$, then no $\mathbf{b} \in \wp\mathcal{B}$ will be unanimously accepted unless $\mathbf{b} \stackrel{\wp}{\succeq} \mathbf{q}$. Thus, the set of admissible bargains is the set $\wp_{\mathbf{q}}\mathcal{B} := \left\{\mathbf{b} \in \wp\mathcal{B}\ ;\ \mathbf{q} \stackrel{\wp}{\preceq} \mathbf{b}\right\}$.

If we forget $\mathcal{A}$ and the utility functions $\{u_i\}_{i\in\mathcal{I}}$, then we are left with an abstract *bargaining problem* on $\mathcal{I}$: an ordered pair $(\mathcal{B}, \mathbf{q})$, where $\mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\not\gamma}$ is convex, compact, and comprehensive, and $\mathbf{q} \in \mathcal{B}$. The problem is to choose some point in $\wp_{\mathbf{q}}\mathcal{B}$ as the social outcome. For simplicity, we assume that $\mathcal{B}$ is *strictly* convex. Let $\mathfrak{B}$ be the set of all strictly convex bargaining problems over $\mathcal{I}$. That is:

$$\mathfrak{B} := \left\{(\mathcal{B}, \mathbf{q})\ ;\ \mathbf{q} \in \mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\not\gamma},\ \text{and } \mathcal{B} \text{ is strictly convex, compact, and comprehensive}\right\}.$$

A *bargaining solution* is a function $\sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\gamma}$ such that, for all $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$: (1) $\sigma(\mathcal{B}, \mathbf{q}) \in \mathcal{B}$, and (2) $\sigma(\mathcal{B}, \mathbf{q}) \stackrel{\wp}{\succeq} \mathbf{q}$. [Condition (1) is normally strengthened to require $\sigma(\mathcal{B}, \mathbf{q}) \in \wp_{\mathbf{q}}\mathcal{B}$; however, we will use the weaker condition so that axiom **(SL)** in Sect. 1 below make sense. Condition (2) reflects the fact that a bargain requires unanimous consent; this distinguishes bargaining solutions from social choice functions, which do not posit a status quo point.]

For example, the *classic utilitarian* (CU) solution $\Upsilon : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\gamma}$ is defined:

$$\Upsilon(\mathcal{B}, \mathbf{q}) := \text{the unique } \mathbf{b} = [b_i]_{i\in\mathcal{I}} \in \wp_{\mathbf{q}}\mathcal{B} \text{ which maximizes } \sum_{i\in\mathcal{I}} b_i.$$

(We have required $\mathcal{B}$ to be strictly convex precisely to guarantee that this maximizer is unique). Myerson (1981) has shown that Υ is the unique bargaining solution which has a useful property of "time independence" when applied to lotteries over unknown future bargaining problems. More broadly construed as a social choice function, classic utilitarianism has several philosophically appealing axiomatic characterizations, due to Harsanyi (1953, 1955, 1977), d'Aspremont and Gevers (1977), Maskin (1978), and Ng (1975, 1985, 2000).

However, CU implicitly assumes that utility functions are "interpersonally comparable", which is unjustified in the vNM framework. Indeed, vNM cardinal utility functions are only well-defined up to affine transformations—that is, if $r_i \in \mathbb{R}_{\not\gamma}$ and $q_i \in \mathbb{R}$, then the function $\widetilde{u}_i(a) := r_i \cdot u_i(a) + q_i$ is "equivalent" to u_i as a description of i's lottery preferences. By applying (distinct) affine-transformations to the utility functions $\{u_i\}_{i\in\mathcal{I}}$, we can reshape the bargaining problem $(\mathcal{B}, \mathbf{q})$, thereby changing the outcome of Υ. In this way, the CU solution Υ can be easily manipulated by the bargainers in $\mathcal{I}$.

Thus, Nash (1950), Kalai and Smorodinsky (1975), and others have insisted that any meaningful bargaining solution must be *rescaling invariant*—that is, invariant under any affine transformations of the utility functions $\{u_i\}_{i\in\mathcal{I}}$. Formally, let $\mathbf{r} = [r_i]_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}_{\neq}$ and $\mathbf{q} = [q_i]_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$. If $\mathbf{b} = [b_i]_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$, then we define $\mathbf{r} \times \mathbf{b} := [r_i \cdot b_i]_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$, and $\mathbf{b} + \mathbf{q} := [b_i + q_i]_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$. If $\mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\neq}$, then define $\mathbf{r} \times \mathcal{B} := \{\mathbf{r} \times \mathbf{b}\ ;\ \mathbf{b} \in \mathcal{B}\}$ and $\mathcal{B} + \mathbf{q} := \{\mathbf{b} + \mathbf{q}\ ;\ \mathbf{b} \in \mathcal{B}\}$. If $(\mathcal{B}, \mathbf{q}_0) \in \mathfrak{B}$, and $\mathbf{r}, \mathbf{q} \in \mathbb{R}^2_{\neq}$, then $(\mathbf{r} \times \mathcal{B} + \mathbf{q},\ \mathbf{r} \times \mathbf{q}_0 + \mathbf{q})$ represents the "same" bargaining problem as $(\mathcal{B}, \mathbf{q}_0)$, encoded using a different (but equivalent) vNM utility function for each $i \in \mathcal{I}$. If $\sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\neq}$ is a bargaining solution, then we say that σ is *rescaling invariant* (RI) if, for every $\mathbf{r}, \mathbf{q} \in \mathbb{R}^2_{\neq}$ and $(\mathcal{B}, \mathbf{q}_0) \in \mathfrak{B}$, we have $\sigma(\mathbf{r} \times \mathcal{B} + \mathbf{q},\ \mathbf{r} \times \mathbf{q}_0 + \mathbf{q}) = \mathbf{r} \times \sigma(\mathcal{B}, \mathbf{q}_0) + \mathbf{q}$. Thus, no one can manipulate the outcome of σ by applying an affine transformation to her utility function.

One way to achieve RI is to "renormalize" the functions $\{u_i\}_{i\in\mathcal{I}}$ to each range from zero to one, and then apply the classic utilitarian solution to this renormalized problem; this yields the *relative utilitarian* bargaining solution. Formally, let $(\mathcal{B}, \mathbf{q})$ be a bargaining problem on $\mathcal{I}$. For every $i \in \mathcal{I}$, let

$$M_i := \max\left\{b_i\ ;\ \mathbf{b} \in \wp_{\mathbf{q}}\mathcal{B}\right\}. \tag{1}$$

be i's *dictatorship* utility level. Define the "renormalized" joint utility function $U_{\mathcal{B},\mathbf{q}} : \mathbb{R}^{\mathcal{I}}_{\neq} \longrightarrow \mathbb{R}$ by:

$$U_{\mathcal{B},\mathbf{q}}(\mathbf{b}) := \sum_{i\in\mathcal{I}} \frac{b_i - q_i}{M_i - q_i} \tag{2}$$

The *relative utilitarian* (RU) bargaining solution $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$ is the point in $\wp_{\mathbf{q}}\mathcal{B}$ which maximizes the value of $U_{\mathcal{B},\mathbf{q}}$. Clearly, $\widetilde{\Upsilon}$ is RI; thus, $\widetilde{\Upsilon}$ is a variant of utilitarianism which obviates the problem of interpersonal utility comparison by effectively legislating that each bargainer's status quo utility is "morally equivalent" to every other bargainer's status quo utility; likewise, each bargainer's dictatorship utility is "morally equivalent" to every other bargainer's dictatorship utility. In other words, to get $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$, we first apply the rescaling function $F : \mathbb{R}^{\mathcal{I}}_{\neq} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\neq}$ defined

$$F(\mathbf{x})_i := \frac{x_i - q_i}{M_i - q_i}, \qquad \forall\, i \in \mathcal{I}.$$

Thus, $F(\mathbf{q}) = \mathbf{0}$, and if $\widetilde{\mathcal{B}} := F(\mathcal{B})$, then $\widetilde{M}_i = 1$ for all $i \in \mathcal{I}$. We then apply the classic utilitarian solution Υ to the rescaled problem $(\widetilde{\mathcal{B}}, \mathbf{0})$. We then have $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{q}) = F^{-1}\left[\Upsilon(\widetilde{\mathcal{B}}, \mathbf{0})\right]$.

Like Υ—and unlike the *egalitarian* solution of Kalai (1977) and the *relative egalitarian* solution of Kalai and Smorodinsky (1975)—$\widetilde{\Upsilon}$ is willing to make cost/benefit tradeoffs which decrease one person's surplus so as to increase someone else's surplus, as long as the benefits (to the recipient's utility) exceed the costs (to the donor's utility). However, like the Nash (1950) and Kalai–Smorodinsky solutions (and unlike Υ or egalitarianism), $\widetilde{\Upsilon}$ is rescaling-invariant. As a social choice function, RU admits several appealing axiomatic characterizations, due to Cao (1982), Dhillon (1998),

and Dhillon and Mertens (1999). Also, Karni (1998) has characterized RU using a modified version of Harsanyi's (1953) impartial observer theorem, while Segal (2000) has shown that RU is optimal in a certain sense, when used as a "resource allocation policy".

We will show that the RU bargaining solution is "optimal" in two distinct ways. In Sect. 1, we develop a variant of Segal's (2000) argument. Theorem 1 states that, if "$\preceq$" is a partial ordering over the set of all bargaining solutions, and "$\preceq$" satisfies certain reasonable axioms of "fairness" and "consistency", then $\widetilde{\Upsilon}$ is a maximal element under "$\preceq$"; furthermore, $\widetilde{\Upsilon}$ is the *only* solution which is maximal for every such ordering. Finally, if "$\preceq$" is a *total* ordering, then $\widetilde{\Upsilon}$ dominates every other bargaining solution. Thus, any arbitrator with "reasonable" preferences over the set of bargaining solutions would, upon reflection, decide that $\widetilde{\Upsilon}$ was the best solution. Although our conclusion is philosophically very similar to Segal's, it is not logically equivalent (because our framework and axioms are not logically equivalent to his). We believe that our framework is technically simpler than Segal's, while our conclusion is slightly stronger.

In Sect. 2, we develop a variant of Harsanyi's (1953) impartial observer theorem. We imagine that a society must select a single bargaining solution to apply to a random sequence of future bargaining problems, and that each player foresees equal probability that she will take on each "role" in each of these bargaining problems. Under the standard vNM assumption that a person wishes to maximize her long-term expected utility, we will show that she will prefer the relative utilitarian bargaining solution $\widetilde{\Upsilon}$ to any other rescaling-invariant solution.

Sections 1 and 2 are logically independent, and can be read in either order.

1 Dictatorship indifference

Recall that $\mathcal{I}$ is a finite population of individuals and $\mathfrak{B}$ is the set of all strictly convex bargaining problems over $\mathcal{I}$. Let $\mathcal{S}$ be the set of all bargaining solutions defined on $\mathfrak{B}$. That is:

$$\mathcal{S} := \left\{ \sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\nearrow} \;;\; \forall\, (\mathcal{B}, \mathbf{q}) \in \mathfrak{B},\; \sigma(\mathcal{B}, \mathbf{q}) \in \mathcal{B} \text{ and } \sigma(\mathcal{B}, \mathbf{q}) \overset{\wp}{\geq} \mathbf{q} \right\}.$$

Imagine an arbitrator who is trying to decide which bargaining solution to employ. This arbitrator has moral intuitions, which cause her to prefer some bargaining solutions to others. Formally, we can express this by saying that her moral intuitions induce a *preference ordering* "$\preceq$" over $\mathcal{S}$. We will show that, if "$\preceq$" satisfies certain "reasonable" axioms, then the relative utilitarian bargaining solution will be the *maximal* element in $\mathcal{S}$ according to the ordering "$\preceq$".

Recall that a *partial ordering* on $\mathcal{S}$ is a relation "$\preceq$" which is *transitive* (i.e. for all $\sigma, \varsigma, \tau \in \mathcal{S}$, if $\sigma \preceq \varsigma \preceq \tau$ then $\sigma \preceq \tau$) and *reflexive* (i.e. for all $\sigma \in \mathcal{S}$, we have $\sigma \preceq \sigma$). If $\sigma \preceq \varsigma$ and $\varsigma \preceq \sigma$, then we write "$\sigma \approx \varsigma$". If $\sigma \preceq \varsigma$ and $\varsigma \not\preceq \sigma$, then we write "$\sigma \prec \varsigma$". We say that "$\preceq$" is a *total* ordering if, for any $\sigma, \varsigma \in \mathcal{S}$, either $\sigma \preceq \varsigma$ or $\varsigma \preceq \sigma$. We do *not* assume that "$\preceq$" is a total ordering. In other words, for

any arbitrary $\sigma, \varsigma \in \mathcal{S}$, it may be the case that neither $\sigma \preceq \varsigma$ nor $\varsigma \preceq \sigma$ (i.e. σ and ς are *incomparable*).

If $\sigma \in \mathcal{S}$, then σ is *maximal* if there exists no other $\varsigma \in \mathcal{S}$ such that $\sigma \prec \varsigma$. We say σ *dominates* $\mathcal{S}$ if, for all $\varsigma \in \mathcal{S}$, we have $\varsigma \preceq \sigma$. Clearly, any dominant element is maximal. However, in general, $(\mathcal{S}, \preceq)$ may not have any maxima; even if it has one, the maximum might not be unique; and even if $(\mathcal{S}, \preceq)$ has a unique maximum, this maximum might not be dominant. Conversely, even a dominant maximum might not be unique. However, if "$\preceq$" is a *total* ordering on $\mathcal{S}$, then any maximum is dominant.

We will assume that "$\preceq$" satisfies three axioms: Global Pareto, Strong Linearity, and Dictatorship Indifference. The first of these axioms is quite plausible; it says that a reasonable arbitrator would prefer a bargaining solution ς to another bargaining solution σ, if ς was systematically Pareto-superior to σ:

(GP) (Global Pareto) *Let $\sigma, \varsigma \in \mathcal{S}$. Suppose that, for all $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, we have $\sigma(\mathcal{B}, \mathbf{q}) \overset{\wp}{\preceq} \varsigma(\mathcal{B}, \mathbf{q})$. Then $\sigma \preceq \varsigma$. Furthermore, if there exists some $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$ such that $\sigma(\mathcal{B}, \mathbf{q}) \prec \varsigma(\mathcal{B}, \mathbf{q})$, then $\sigma \prec \varsigma$.*

To formulate the second axiom, suppose that $\sigma_0, \sigma_1 \in \mathcal{S}$ are two bargaining solutions. For any $r \in [0, 1]$, we define the bargaining solution $\sigma_r := r\sigma_1 + (1-r)\sigma_0$ as follows: for any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$,

$$\sigma_r(\mathcal{B}, \mathbf{q}) := r\sigma_1(\mathcal{B}, \mathbf{q}) + (1 - r)\sigma_0(\mathcal{B}, \mathbf{q}).$$

Heuristically, σ_r represents a "randomized" bargaining solution: with probability r we will apply solution σ_1, while with probability $(1 - r)$ we will apply solution σ_0. This perhaps provides a "compromise" solution which combines the (dis)advantages of σ_0 and σ_1. The von Neumann–Morgenstern theory of cardinal utility says that preferences should be "linear" with respect to such convex combinations. This suggests the following axiom:

(WL) (Weak Linearity) *Let $\sigma, \varsigma, \tau \in \mathcal{S}$. Let $r \in (0, 1)$.*

- *If $\sigma \prec \varsigma$, then $r\sigma + (1 - r)\tau \prec r\varsigma + (1 - r)\tau$.*
- *If $\sigma \approx \varsigma$, then $r\sigma + (1 - r)\tau \approx r\varsigma + (1 - r)\tau$.*

However, we will actually require a stronger form of linearity. Let $\rho : \mathfrak{B} \longrightarrow [0, 1]$ be some "weight function". We define the bargaining solution $\sigma_\rho := \rho\sigma_1 + (1 - \rho)\sigma_0$ as follows: for any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$,

$$\sigma_\rho(\mathcal{B}, \mathbf{q}) := \rho(\mathcal{B}, \mathbf{q}) \cdot \sigma_1(\mathcal{B}, \mathbf{q}) + [1 - \rho(\mathcal{B}, \mathbf{q})] \cdot \sigma_0(\mathcal{B}, \mathbf{q}).$$

Thus σ_ρ is a "randomized" bargaining solution, where with probability ρ we apply solution σ_1, while with probability $(1 - \rho)$ we apply solution σ_0. However, the value of ρ might depend on the bargaining problem $(\mathcal{B}, \mathbf{q})$. We require:

(SL) (Strong Linearity) *Let $\sigma, \varsigma, \tau \in \mathcal{S}$ and let $\rho : \mathfrak{B} \longrightarrow [0, 1]$.*
(SL1) *If $\sigma \preceq \varsigma$, then $\rho\sigma + (1 - \rho)\tau \preceq \rho\varsigma + (1 - \rho)\tau$.*
Furthermore, suppose that $\rho : \mathfrak{B} \longrightarrow (0, 1)$. Then
(SL2) *If $\sigma \prec \varsigma$, then $\rho\sigma + (1 - \rho)\tau \prec \rho\varsigma + (1 - \rho)\tau$.*

Note that **(SL1)** immediately implies:

(SL0) *If* $\rho : \mathfrak{B} \longrightarrow [0,1]$, *and* $\sigma \approx \varsigma$, *then* $\rho\sigma + (1-\rho)\tau \;\approx\; \rho\varsigma + (1-\rho)\tau$.

Also, note that **(SL)** implies **(WL)**; just set $\rho \equiv r$.

To state the last axiom, we define the *dictatorship* bargaining solutions δ_j for each $j \in \mathcal{I}$ as follows: for any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, if M_j is as in (1), then

$$\delta_j(\mathcal{B}, \mathbf{q}) := \mathbf{m}^j = [m_i^j]_{i \in \mathcal{I}}, \quad \text{where } m_j^j := M_j, \quad \text{and } m_i^j := q_i \text{ for all } i \neq j. \quad (3)$$

In other words, δ_j is the solution which always gives all surplus utility to player j, and leaves all other bargainers with their status quo. We require:

(DI) (Dictatorship Indifference) *For all* $i, j \in \mathcal{I}$, $\delta_i \approx \delta_j$.

The main result of this section is this:

Theorem 1 *Let* $\widetilde{\Upsilon} : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\neq}$ *be the relative utilitarian bargaining solution.*

(a) *If "*$\preceq$*" is any partial ordering on* $\mathcal{S}$ *which satisfies axioms* **(GP)**, **(SL)** *and* **(DI)**, *then* $\widetilde{\Upsilon}$ *is a maximal element of* $\mathcal{S}$ *with respect to "*$\preceq$*".*

(b) $\widetilde{\Upsilon}$ *is the only element of* $\mathcal{S}$ *which is maximal for every ordering satisfying* **(GP)**, **(SL)**, *and* **(DI)**.

(c) *If "*$\preceq$*" is a total ordering on* $\mathcal{S}$ *which satisfies* **(GP)**, **(SL)** *and* **(DI)**, *then* $\widetilde{\Upsilon}$ *is a dominant, maximal element of* $\mathcal{S}$.

Proof **(a)** If $\rho, \mu : \mathfrak{B} \longrightarrow [0,1]$ are two weight functions, then we write "$\rho \leq \mu$" if, for all $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, we have $\rho(\mathcal{B}, \mathbf{q}) \leq \mu(\mathcal{B}, \mathbf{q})$. Thus, "$\rho \not\leq \mu$" means there is some $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$ with $\rho(\mathcal{B}, \mathbf{q}) > \mu(\mathcal{B}, \mathbf{q})$. Finally, we write "$\rho < \mu$" if, for all $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, we have $\rho(\mathcal{B}, \mathbf{q}) < \mu(\mathcal{B}, \mathbf{q})$. Let $\mathbf{0}, \mathbf{1} : \mathfrak{B} \longrightarrow \{0,1\}$ be the constant zero and constant one functions. Thus, $\rho : \mathfrak{B} \longrightarrow (0,1)$ iff $\mathbf{0} < \rho < \mathbf{1}$. If $\sigma_0, \sigma_1 \in \mathcal{S}$, and $\rho : \mathfrak{B} \longrightarrow [0,1]$, recall that we define $\sigma_\rho := \rho\sigma_1 + (1-\rho)\sigma_0$.

Claim 1 Let $\sigma_0, \sigma_1 \in \mathcal{S}$. Let $\rho, \mu : \mathfrak{B} \longrightarrow [0,1]$, with $\rho \leq \mu$.

(L0) If $\sigma_0 \approx \sigma_1$ then $\sigma_0 \approx \sigma_\rho \approx \sigma_\mu \approx \sigma_1$.
(L1) If $\sigma_0 \preceq \sigma_1$ then $\sigma_0 \preceq \sigma_\rho \preceq \sigma_\mu \preceq \sigma_1$.
(L2) Suppose $\mathbf{0} < \rho < \mu < \mathbf{1}$. If $\sigma_0 \prec \sigma_1$ then $\sigma_0 \prec \sigma_\rho \prec \sigma_\mu \prec \sigma_1$.

Proof Define $\nu : \mathfrak{B} \longrightarrow [0,1]$ by $\nu(\mathcal{B}, \mathbf{q}) := \dfrac{\mu(\mathcal{B}, \mathbf{q}) - \rho(\mathcal{B}, \mathbf{q})}{1 - \rho(\mathcal{B}, \mathbf{q})}$. It is easy to check:

$$\sigma_\mu = \nu\sigma_1 + (1-\nu)\sigma_\rho \quad \text{and} \quad \sigma_\rho = \nu\sigma_\rho + (1-\nu)\sigma_\rho. \quad (4)$$

Thus, Axioms **(SL0)** and **(SL1)** and (4) imply:

(ℓ0) $\big(\sigma_\rho \approx \sigma_1\big) \implies \big(\sigma_\rho \approx \sigma_\mu\big)$.
(ℓ1) $\big(\sigma_\rho \preceq \sigma_1\big) \implies \big(\sigma_\rho \preceq \sigma_\mu\big)$.

Furthermore, if $\mathbf{0} < \rho < \mu < \mathbf{1}$, then $\mathbf{0} < \nu < \mathbf{1}$, in which case **(SL2)** implies:

(ℓ2) $\big(\sigma_\rho \prec \sigma_1\big) \implies \big(\sigma_\rho \prec \sigma_\mu\big)$.

Finally, note that

$$\sigma_\mu := \mu\sigma_1 + (1-\mu)\sigma_0 \quad \text{and} \quad \sigma_1 = \mu\sigma_1 + (1-\mu)\sigma_1. \tag{5}$$

To see **(L2)**, suppose $\mathbf{0} \prec \rho \prec \mu \prec \mathbf{1}$. If $\sigma_0 \prec \sigma_1$, then Axiom **(SL2)** and (5) imply that $\sigma_\mu \prec \sigma_1$. Similarly, $\sigma_0 \prec \sigma_\rho$. Finally, by a similar argument, $\sigma_\rho \prec \sigma_1$; thus, Fact **($\boldsymbol{\ell}$2)** implies $\sigma_\rho \prec \sigma_\mu$. This establishes **(L2)**. To get **(L1)**, replace all "$\prec$" with "$\preceq$" and use Axiom **(SL1)** and Fact **($\boldsymbol{\ell}$1)**. To get **(L0)**, replace all "$\preceq$" with "$\approx$" and use Axiom **(SL0)** and Fact **($\boldsymbol{\ell}$0)**. □ (Claim 1)

Claim 2 Let $\sigma_0, \sigma_1, \sigma_1' \in \mathcal{S}$, with $\sigma_0 \preceq \sigma_1 \prec \sigma_1'$. Let $\rho, \rho' : \mathfrak{B} \longrightarrow (0,1)$, and let $\sigma_\rho := \rho\sigma_1 + (1-\rho)\sigma_0$ and $\sigma'_{\rho'} := \rho'\sigma_1' + (1-\rho')\sigma_0$. If $\sigma'_{\rho'} \approx \sigma_\rho$, then $\rho \not\preceq \rho'$.

Proof (by contradiction) Suppose $\rho \preceq \rho'$. Let $\sigma_{\rho'} := \rho'\sigma_1 + (1-\rho')\sigma_0$. Then

$$\sigma_\rho \overset{\preceq}{\underset{(*)}{}} \sigma_{\rho'} \overset{\prec}{\underset{(\dagger)}{}} \sigma'_{\rho'} \underset{(H)}{\approx} \sigma_\rho.$$

Here, $(*)$ is by **(L1)** because $\sigma_0 \preceq \sigma_1$ and $\rho \preceq \rho'$. Next, $(\dagger)$ is by Axiom **(SL2)**, because $\sigma_1 \prec \sigma_1'$ and $\mathbf{0} \prec \rho' \prec \mathbf{1}$. Finally, (H) is by hypothesis. Thus, we get $\sigma_\rho \prec \sigma_\rho$, which is a contradiction. Thus, it is false that $\rho \preceq \rho'$. □ (Claim 2)

$$\text{Let } \Delta := \left\{ \sum\nolimits_{i\in\mathcal{I}} \rho_i \delta_i \ ; \ \forall\, i \in \mathcal{I},\ \rho_i : \mathfrak{B} \longrightarrow [0,1], \text{ and } \sum\nolimits_{i\in\mathcal{I}} \rho_i \equiv \mathbf{1} \right\}.$$

Claim 3 All elements of Δ are "$\preceq$"-indifferent.

Proof Use **(L0)** and Axiom **(DI)**. □ (Claim 3)

For any $\sigma \in \mathcal{S}$ and $\rho : \mathfrak{B} \longrightarrow [0,1]$, let

$$\Delta(\sigma, \rho) := \left\{ \rho\sigma + \sum_{i\in\mathcal{I}} \rho_i \delta_i \ ; \ \forall\, i \in \mathcal{I},\ \rho_i : \mathfrak{B} \longrightarrow [0,1], \text{ and } \rho + \sum_{i\in\mathcal{I}} \rho_i \equiv \mathbf{1} \right\}.$$

Claim 4 For any fixed σ and ρ, all elements of $\Delta(\sigma, \rho)$ are "$\preceq$"-indifferent.

Proof Use Axiom **(SL0)** and Claim 3. □ (Claim 4)

For any $\sigma \in \mathcal{S}$, we define $U_\sigma : \mathfrak{B} \longrightarrow \mathbb{R}_{\neq}$ by $U_\sigma(\mathcal{B}, \mathbf{q}) := U_{\mathcal{B},\mathbf{q}}\left[\sigma(\mathcal{B}, \mathbf{q})\right]$, for every $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, where $U_{\mathcal{B},\mathbf{q}}$ is defined as in (2). Thus, if $\varsigma \in \mathcal{S}$, we write "$U_\sigma \preceq U_\varsigma$" if $U_{\mathcal{B},\mathbf{q}}[\sigma(\mathcal{B}, \mathbf{Q})] \leq U_{\mathcal{B},\mathbf{q}}[\varsigma(\mathcal{B}, \mathbf{Q})]$, for all $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$.

Claim 5 Let $\sigma, \sigma' \in \mathcal{S}$.

(a) There exist weight functions $\rho, \rho' : \mathfrak{B} \longrightarrow (0,1)$ with $\Delta(\sigma, \rho) \cap \Delta(\sigma', \rho') \neq \emptyset$.
(b) Let ρ and ρ' be as in part (a). Then $U_\sigma \geq U_{\sigma'}$ if and only if $\rho \leq \rho'$.

Proof Fix $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$. For all $i \in \mathcal{I}$, let $\mathbf{m}^i$ be as in (3)

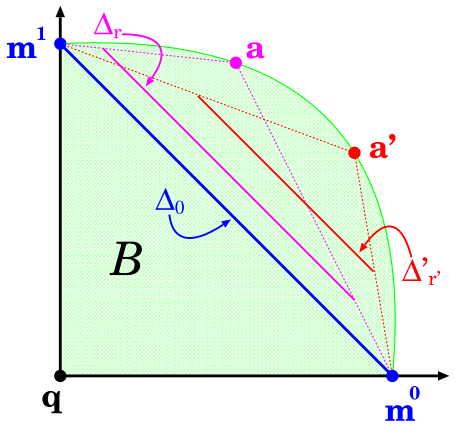

Fig. 1 Claim 5.1

Claim 5.1 There exist $\mathbf{r}, \mathbf{r}' \in [0,1]^{\mathcal{I}}$ and $r, r' \in (0,1)$ with $r + \sum_{i\in\mathcal{I}} r_i = 1 = r' + \sum_{i\in\mathcal{I}} r'_i = 1$ such that

$$r\sigma(\mathcal{B},\mathbf{q}) + \sum_{i\in\mathcal{I}} r_i \mathbf{m}^i = r'\sigma'(\mathcal{B},\mathbf{q}) + \sum_{i\in\mathcal{I}} r'_i \mathbf{m}^i. \tag{6}$$

Proof Let $\mathbf{a} := \sigma(\mathcal{B},\mathbf{q})$ and $\mathbf{a}' := \sigma'(\mathcal{B},\mathbf{q})$. As shown in Fig. 1, for any fixed $r, r' \in [0,1]$, let

$$\Delta_r := \left\{ r\mathbf{a} + \sum_{i\in\mathcal{I}} r_i \mathbf{m}^i \;;\; \mathbf{r} \in [0,1]^{\mathcal{I}} \text{ and } r + \sum_{i\in\mathcal{I}} r_i = 1 \right\};$$

$$\Delta'_{r'} := \left\{ r'\mathbf{a}' + \sum_{i\in\mathcal{I}} r'_i \mathbf{m}^i \;;\; \mathbf{r}' \in [0,1]^{\mathcal{I}} \text{ and } r' + \sum_{i\in\mathcal{I}} r'_i = 1 \right\};$$

$$\text{and} \quad \Delta_0 := \left\{ \sum_{i\in\mathcal{I}} r_i \mathbf{m}^i \;;\; \mathbf{r} \in [0,1]^{\mathcal{I}} \text{ and } \sum_{i\in\mathcal{I}} r_i = 1 \right\}.$$

Then Δ_r and $\Delta'_{r'}$ are hyperplane segments parallel to Δ_0 (and thus, to each other). Furthermore, as $r, r' \to 0$, the segments Δ_r and $\Delta'_{r'}$ both converge to Δ_0; thus, there exist r and r' such that Δ_r overlaps $\Delta'_{r'}$. □ (Claim 5.1)

Claim 5.2 $U_{\mathcal{B},\mathbf{q}}[\sigma(\mathcal{B},\mathbf{Q})] \geq U_{\mathcal{B},\mathbf{q}}\left[\sigma'(\mathcal{B},\mathbf{Q})\right] \iff r \leq r'$ in Claim 5.1.

Proof If $\mathbf{r}, \mathbf{r}' \in [0,1]^{\mathcal{I}}$ and $r, r' \in (0,1)$ are as in Claim 5.1, then

$$1 + r \cdot \left[U_{\mathcal{B},\mathbf{q}}(\mathbf{a}) - 1\right] = (1-r) + rU_{\mathcal{B},\mathbf{q}}(\mathbf{a}) \underset{(\diamond)}{=} rU_{\mathcal{B},\mathbf{q}}(\mathbf{a}) + \sum_{i\in\mathcal{I}} r_i$$

$$
\begin{aligned}
&\underset{(*)}{=} r U_{\mathcal{B},\mathbf{q}}(\mathbf{a}) + \sum_{i\in\mathcal{I}} r_i U_{\mathcal{B},\mathbf{q}}(\mathbf{m}^i) \quad \underset{(\mathrm{L})}{=} \quad U_{\mathcal{B},\mathbf{q}}\left(r\mathbf{a} + \sum_{i\in\mathcal{I}} r_i \mathbf{m}^i\right)\\
&\underset{(\dagger)}{=} U_{\mathcal{B},\mathbf{q}}\left(r'\mathbf{a}' + \sum_{i\in\mathcal{I}} r'_i \mathbf{m}^i\right) \quad \underset{(\mathrm{L})}{=} \quad r' U_{\mathcal{B},\mathbf{q}}(\mathbf{a}') + \sum_{i\in\mathcal{I}} r'_i U_{\mathcal{B},\mathbf{q}}(\mathbf{m}^i)\\
&\underset{(*)}{=} r' U_{\mathcal{B},\mathbf{q}}(\mathbf{a}') + \sum_{i\in\mathcal{I}} r'_i \quad \underset{(\spadesuit)}{=} \quad (1-r') + r' U_{\mathcal{B},\mathbf{q}}(\mathbf{a}')\\
&= 1 + r' \cdot \left[U_{\mathcal{B},\mathbf{q}}(\mathbf{a}') - 1\right].
\end{aligned}
$$

Here, $(\diamond)$ is because $r + \sum_{i\in\mathcal{I}} r_i = 1$ by definition. $(*)$ is because $U_{\mathcal{B},\mathbf{q}}(\mathbf{m}^i) = 1$ for all $i \in \mathcal{I}$ by definition. (L) is because $U_{\mathcal{B},\mathbf{q}}$ is linear, and $(\dagger)$ is by (6). Finally, $(\spadesuit)$ is because $r' + \sum_{i\in\mathcal{I}} r'_i = 1$ by definition. Thus, we have

$$
r \cdot \left[U_{\mathcal{B},\mathbf{q}}(\mathbf{a}) - 1\right] = r' \cdot \left[U_{\mathcal{B},\mathbf{q}}(\mathbf{a}') - 1\right].
$$

Thus,

$$
\left(U_{\mathcal{B},\mathbf{q}}(\mathbf{a}) \geq U_{\mathcal{B},\mathbf{q}}(\mathbf{a}')\right) \iff \left(U_{\mathcal{B},\mathbf{q}}(\mathbf{a}) - 1 \geq U_{\mathcal{B},\mathbf{q}}(\mathbf{a}') - 1\right) \iff \left(r \leq r'\right),
$$

as desired. □ (Claim 5.2)

So, for each $(\mathcal{B},\mathbf{q}) \in \mathfrak{B}$, set $\rho(\mathcal{B},\mathbf{q}) := r$ and $\rho'(\mathcal{B},\mathbf{q}) := r'$, and define $\rho_i(\mathcal{B},\mathbf{q}) := r_i$ and $\rho'_i(\mathcal{B},\mathbf{q}) := r'_i$ for all $i \in \mathcal{I}$, where these values are as in Claim 5.1. Then

$$
\rho\sigma + \sum_{i\in\mathcal{I}} \rho_i \delta_i = \rho'\sigma' + \sum_{i\in\mathcal{I}} \rho'_i \delta_i.
$$

But $\rho\sigma + \sum_{i\in\mathcal{I}} \rho_i\delta_i \in \Delta(\sigma,\rho)$ and $\rho'\sigma' + \sum_{i\in\mathcal{I}} \rho'_i\delta_i \in \Delta(\sigma',\rho')$. Thus, $\Delta(\sigma,\rho) \cap \Delta(\sigma',\rho') \neq \emptyset$. This proves part **(a)**. Part **(b)** follows from Claim 5.2. □ (Claim 5)

Claim 6 Let $\sigma, \sigma \in \mathcal{S}$. If $\sigma \prec \sigma'$, then $U_\sigma \not\geq U_{\sigma'}$.

Proof Let $\rho, \rho' : \mathfrak{B} \longrightarrow (0,1)$ be as in Claim 5(a). Fix some $\delta_* \in \Delta_0$. Let $\delta := \rho\sigma + (1-\rho)\delta_*$ and $\delta' := \rho'\sigma' + (1-\rho')\delta_*$.

Claim 6.1 $\delta \approx \delta'$.

Proof Find $\delta_\sharp \in \Delta(\sigma,\rho) \cap \Delta(\sigma',\rho')$; this exists by Claim 5(a). Then we have $\delta \approx \delta_\sharp \approx \delta'$, where both "$\approx$" are by Claim 4, because $\delta \in \Delta(\sigma,\rho)$ and $\delta' \in \Delta(\sigma',\rho')$. Thus, $\delta \approx \delta'$, because "$\approx$" is transitive. □ (Claim 6.1)

But $\sigma \prec \sigma'$, so Claims 2 and 6.1 imply that $\rho \not\leq \rho'$. But then Claim 5(b) implies that $U_\sigma \not\geq U_{\sigma'}$. □ (Claim 6)

Claim 7 $\widetilde{\Upsilon}$ is a maximal element of "$\preceq$".

Proof (by contradiction) Suppose $\widetilde{\Upsilon}$ is not maximal; then there is some $\sigma \in \mathcal{S}$ with $\widetilde{\Upsilon} \prec \sigma$. But then Claim 6 says that $U_{\widetilde{\Upsilon}} \not\geq U_{\sigma}$, which means there is some $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$ such that $U_{\mathcal{B},\mathbf{q}}\left[\widetilde{\Upsilon}(\mathcal{B}, \mathbf{Q})\right] < U_{\mathcal{B},\mathbf{q}}\left[\sigma(\mathcal{B}, \mathbf{Q})\right]$. But this contradicts the fact that $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{Q})$ always maximizes $U_{\mathcal{B},\mathbf{q}}$ by definition of $\widetilde{\Upsilon}$. □ (Claim 7)

(b) Suppose $\sigma \in \mathcal{S}$ is maximal for *every* ordering satisfying **(GP)**, **(SL)**, and **(DI)**. We must show that $\sigma = \widetilde{\Upsilon}$.

Fix $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, and consider the ordering "$_{\mathcal{B}}\preceq_{\mathbf{q}}$" defined by:

$$\left(\sigma \; {}_{\mathcal{B}}\preceq_{\mathbf{q}} \; \sigma'\right) \iff \left(U_{\mathcal{B},\mathbf{q}}\left[\sigma(\mathcal{B}, \mathbf{q})\right] \leq U_{\mathcal{B},\mathbf{q}}\left[\sigma'(\mathcal{B}, \mathbf{q})\right]\right).$$

It is easy to check that "$_{\mathcal{B}}\preceq_{\mathbf{q}}$" satisfies **(GP)**, **(SL)**, and **(DI)**. If σ is maximal for "$_{\mathcal{B}}\preceq_{\mathbf{q}}$", then we must have $\sigma(\mathcal{B}, \mathbf{q}) = \widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$, because $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$ is the unique point which maximizes the value of $U_{\mathcal{B},\mathbf{q}}$ in $\wp_{\mathbf{q}}\mathcal{B}$.
Since we can do this for any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, we conclude that $\sigma = \widetilde{\Upsilon}$.
(c) follows from **(a)**, because maxima are always dominant in total orderings. To see that **(c)** is nonvacuous, however, we must show that there exists a total ordering which satisfies **(GP)**, **(SL)**, and **(DI)**. However, for any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, the ordering "$_{\mathcal{B}}\preceq_{\mathbf{q}}$" in the proof of **(b)** is such a total ordering. □

Our approach is clearly inspired by Segal's (2000) characterization of RU, but differs in both its interpretation and its formal content.

Interpretive differences. Segal's original paper is not about bargaining solutions at all, but is instead about resource-allocation *policies*: rules which take any initial bundle of commodities and allocate it amongst two or more competing claimants, whose preferences are encoded by cardinal utility functions over commodity bundles. Also, instead of positing an arbitrator, Segal imagines that each member of society separately develops a preference ordering satisfying certain axioms, based on her personal moral intuitions (formally, this just involves replacing the symbol "$\preceq$" with"$\preceq_i$" for some $i \in \mathcal{I}$). He concludes that all members of society, after due consideration, would separately but unanimously endorse relative utilitarianism.

Segal's resource-allocation framework introduces considerable technical complexity, but it does not provide any greater generality, because any resource-allocation problem can be reformulated as an abstract bargaining problem (Muthoo, 1999, Sect. 2.2). Segal's premise that each individual in society separately deduces the optimality of RU is quite similar to our own conclusions in Sect. 2 (see Theorem 3 below). However, this premise is unrealistic in the present context, because the key axiom needed for Segal's result (and for our Theorem 1) is Dictatorship Indifference. Axiom **(DI)** requires that each person recognize that her *own* dictatorship is just as morally objectionable as anyone else's. This places a rather heavy burden on the "fairmindedness" and "objectivity" of each bargainer. Indeed, history suggests that even great champions of egalitarianism and democracy often seem to feel that, while any dictatorship is evil, their *own* dictatorship is "not quite as evil" as someone else's. We feel that **(DI)** is not a realistic axiom for the *bargainers*, but it is a reasonable axiom for a neutral *arbitrator*; that is why we have formulated our model in this way.

Formal differences. If $\mathcal{C}$ is the set of commodities, then $\mathbb{R}^{\mathcal{C}}_{\gneq}$ is the space of collective commodity bundles, and $\mathbb{R}^{\mathcal{C}\times\mathcal{I}}_{\gneq}$ is the space of commodity allocations to the members of $\mathcal{I}$. Segal defines a ***policy*** as a function from $\mathbb{R}^{\mathcal{C}}_{\gneq}$ into the set of lotteries over $\mathbb{R}^{\mathcal{C}\times\mathcal{I}}_{\gneq}$. Segal requires these policy functions to be Borel measurable, whereas we impose no regularity conditions on our set $\mathcal{S}$ of bargaining solutions. If $\mathcal{F}$ is the space of measurable policies, then Segal also requires $\preceq$ to be a *total* ordering on $\mathcal{F}$, whereas we allow $\preceq$ to be any partial ordering on $\mathcal{S}$.

Each of our three axioms corresponds roughly to an axiom of Segal, but only roughly, because the underlying formalisms differ. Our axiom **(GP)** corresponds to Segal's *Monotonicity* **(M)**. Segal's *Independence* **(I)** corresponds to our axiom **(WL)**, but we required the stronger axiom **(SL)**. On the other hand, Segal's Dictatorship Indifference **(D)** is stronger than our **(DI)**, because he also requires indifference amongst "piecewise mixtures" of dictatorship solutions. Finally, Segal requires a fourth axiom, **(C)**: the ordering $\preceq$ must to be *continuous* relative to a certain topology on $\mathcal{F}$. He first approximates each policy with one which is "piecewise constant" in a certain sense. He then obtains relationships between such approximations using **(M)**, **(I)** and his stronger **(D)**. Finally, he needs **(C)** to show that these relationships also hold in the limit as the approximations converge to the original policies. Because our **(SL)** is stronger than Segal's **(I)**, we can get away with a weaker form of **(DI)**, and we can sidestep the "approximation" strategy altogether, so that our approach requires no topology.

2 A rescaling-invariant impartial observer theorem

We now develop a variant of Harsanyi's (1953) *impartial observer theorem*[1] in the context of bargaining. Our approach is loosely inspired by Karni (1998); like him, we are troubled that Harsanyi's "impartiality" implicitly requires interpersonal comparability of utility functions. We are also troubled by Harsanyi's premise that fairminded individuals can and will temporarily pretend ignorance of their own circumstances so as to obtain social consensus; this is inconsistent with the standard assumption that people are self-regarding maximizers.

Instead, imagine a person who anticipates that, in the long-term future, she will be involved in multiple bargaining interactions involving $\mathcal{I}$ individuals (including herself). At present, she cannot predict the specific shape of these future bargaining problems; or which other people will be involved in each one. Instead, she posits an ex ante probability distribution μ over the set $\mathfrak{B}$ of all possible bargaining problems, and she imagines that she will encounter an infinite sequence of independent random bargaining problems generated according to μ. She further assumes that her "roles" in these bargaining problems (that is, which axis represents her utility) are independent, uniformly distributed, $\mathcal{I}$-valued random variables. Thus, in the long-term future, she anticipates that she has an equal probability of playing each role in each bargaining problem—i.e. she has an equal probability of being vendor or customer,

[1] See Harsanyi (1953, 1955, 1977), Roemer (1998, Sect. 4.4), Karni and Weymark (1998), or Karni (2003, Sect. 4).

landlord or tenant, employer or employee. Under these conditions, she will see that the relative utilitarian solution $\widetilde{\Upsilon}$ maximizes her ex ante μ-expected utility over all rescaling-invariant (RI) solutions (Theorem 3); hence she will prefer it to any other RI solution. If all members of society reason in a similar fashion (each perhaps using a different ex ante measure μ), then the result will be a unanimous consensus to use $\widetilde{\Upsilon}$ to solve all future bargaining problems.

Formally, let $\mathcal{I}$ be a finite set of indices, representing "bargaining roles" (for example, in a labour contract negotiation, we might have $\mathcal{I} = \{0, 1\}$ where 0 represents the worker and 1 represents the employer). Let $\mathfrak{B}$ be the set of all convex bargaining problems over $\mathcal{I}$. If $\mathcal{A}$ is a sigma-algebra of subsets of $\mathfrak{B}$, then a *probability measure* on $(\mathfrak{B}, \mathcal{A})$ is a countably additive function $\mu : \mathcal{A} \longrightarrow [0, 1]$ such that $\mu[\mathfrak{B}] = 1$. If $P(\mathfrak{b})$ is some statement which could be either true or false for each $\mathfrak{b} \in \mathfrak{B}$, then we write, "$P(\mathfrak{b})$, for $\forall_\mu\ \mathfrak{b} \in \mathfrak{B}$" to mean that the set $\mathfrak{F} := \{\mathfrak{b} \in \mathfrak{B}\ ;\ P(\mathfrak{b}) \text{ is false}\}$ is in $\mathcal{A}$, and $\mu[\mathfrak{F}] = 0$. A bargaining solution $\sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\sim}$ is *$\mathcal{A}$-measurable* if $\sigma^{-1}(\mathcal{O}) \in \mathcal{A}$ for every open subset $\mathcal{O} \subset \mathbb{R}^{\mathcal{I}}_{\not\sim}$. If we write $\sigma := (\sigma_i)_{i \in \mathcal{I}}$, then, for all $i \in \mathcal{I}$, we can compute the *μ-expected value* of i's utility under solution σ:

$$\mathbb{E}_\mu(\sigma_i) := \int_{\mathfrak{B}} \sigma_i(\mathcal{B}, \mathbf{q})\ \mathrm{d}\mu[\mathcal{B}, \mathbf{q}].$$

In contemplating a sequence of unknown future bargaining problems, you might expect that sometimes you will play one role and sometimes the other (for example, in future labour negotiations, sometimes you will be a worker, and sometimes an employer). If η is some probability distribution on $\mathcal{I}$, then let $\sigma_\eta := \sum_{i \in \mathcal{I}} \eta\{i\}\sigma_i$ be the η-expected value of σ, assuming you receive payoff σ_i with probability $\eta\{i\}$. If $\mathcal{S}$ denotes the set of all $\mathcal{A}$-measurable bargaining solutions, it is easy to prove a version of Harsanyi's theorem:

Proposition 2 *Let η be the uniform probability distribution on $\mathcal{I}$, and let μ be any probability distribution on $\mathfrak{B}$. If $\sigma \in \mathcal{S}$ maximizes the value of $\mathbb{E}_\mu(\sigma_\eta)$ over $\mathcal{S}$, then $\sigma(\mathcal{B}, \mathbf{q}) = \Upsilon(\mathcal{B}, \mathbf{q})$, for $\forall_\mu\ (\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$.*

Proposition 2 (and classic utilitarianism in general) is objectionable because it implicitly assumes interpersonal comparability of utility, which is meaningless in the vNM framework. As argued in the introduction, a bargaining solution should be RI (and Υ is not). We seek an RI version of Proposition 2. For any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$ and $i \in \mathcal{I}$, let $r_i(\mathcal{B}, \mathbf{q}) := \max\left\{b_i - q_i\ ;\ \mathbf{b} \in \wp_{\mathbf{q}}\mathcal{B}\right\}$. Let $\widetilde{\mathfrak{B}} := \{\mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\not\sim}\ ;\ \mathcal{B} \text{ is strictly convex, comprehensive, and compact, and } r_i(\mathcal{B}, \mathbf{0}) = 1, \text{ for all } i \in \mathcal{I}\}$. Let $\widetilde{\mathcal{S}}$ denote the set of all $\mathcal{A}$-measurable, rescaling-invariant bargaining solutions. There is a natural bijection $\Phi : \widetilde{\mathfrak{B}} \times \mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim} \longrightarrow \mathfrak{B}$ defined by

$$\Phi(\widetilde{\mathcal{B}}, \mathbf{r}, \mathbf{q}) := (\mathbf{r} \times \widetilde{\mathcal{B}}, \mathbf{q}).$$

Thus, if $\sigma \in \widetilde{\mathcal{S}}$, then σ is determined entirely by its values on $\widetilde{\mathfrak{B}}$. In particular, $\widetilde{\Upsilon}$ is the unique element of $\widetilde{\mathcal{S}}$ which maximizes $\sum_{i \in \mathcal{I}} \sigma_i(\mathcal{B}, \mathbf{0})$ for every $\mathcal{B} \in \widetilde{\mathfrak{B}}$.

Let $\widetilde{\mathcal{A}}$ be a sigma-algebra on $\widetilde{\mathfrak{B}}$ such that Φ is measurable with respect to $\mathcal{A}$, $\widetilde{\mathcal{A}}$, and the Borel sigma-algebra on $\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}$. Let $\widetilde{\mu}$ be a probability measure on $\widetilde{\mathfrak{B}}$, let $\overline{\mu}$ be a probability measure on $\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}$, and let $\mu := \Phi(\widetilde{\mu} \times \overline{\mu})$. Thus, a μ-random bargaining problem in $\mathfrak{B}$ is obtained by taking a $\widetilde{\mu}$-random problem $\mathcal{B} \in \widetilde{\mathfrak{B}}$ and applying an independent, $\overline{\mu}$-random rescaling to $\mathcal{B}$. For all $i \in \mathcal{I}$, let $\overline{r}_i := \int_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} r_i \, \mathrm{d}\overline{\mu}[\mathbf{r}, \mathbf{q}]$. We say $\overline{\mu}$ is *anonymous* if there is some constant $\overline{r}$ such that $\overline{r}_i = \overline{r}$ for all $i \in \mathcal{I}$. Thus every coordinate receives the same average rescaling (in particular, this is true if $\overline{\mu}$ is any measure on $\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}$ which is invariant under a transitive group of permutations of the first $\mathcal{I}$ coordinate axes).

Theorem 3 *Let $\overline{\mu}$ be an anonymous probability measure on $\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}$, let $\widetilde{\mu}$ be a probability measure on $\widetilde{\mathfrak{B}}$, and let $\mu := \Phi(\widetilde{\mu} \times \overline{\mu})$. Let η be the uniform probability distribution on $\mathcal{I}$. If $\sigma \in \widetilde{\mathcal{S}}$ maximizes the value of $\mathbb{E}_\mu(\sigma_\eta)$ over $\widetilde{\mathcal{S}}$, then $\sigma(\mathcal{B}, \mathbf{q}) = \widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$, for $\forall_\mu \, (\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$.*

Proof Define $\widetilde{\sigma} : \widetilde{\mathfrak{B}} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\sim}$ by $\widetilde{\sigma}(\mathcal{B}) := \sigma(\mathcal{B}, \mathbf{0})$ for all $\mathcal{B} \in \widetilde{\mathfrak{B}}$. Fix $i \in \mathcal{I}$, and let $\overline{q}_i := \int\limits_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} q_i \, \mathrm{d}\overline{\mu}[\mathbf{r}, \mathbf{q}]$. Then

$$
\begin{aligned}
\mathbb{E}_\mu(\sigma_i) &= \int\limits_{\mathfrak{B}} \sigma_i(\mathcal{B}, \mathbf{q}) \;\, \mathrm{d}\mu[\mathcal{B}, \mathbf{q}] \;\underset{(\diamond)}{=}\; \int\limits_{\widetilde{\mathfrak{B}}} \int\limits_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} \sigma_i(\mathbf{r} \times \widetilde{\mathcal{B}}, \mathbf{q}) \, \mathrm{d}\overline{\mu}[\mathbf{r}, \mathbf{q}] \, \mathrm{d}\widetilde{\mu}[\widetilde{\mathcal{B}}] \\
&\underset{(*)}{=} \int\limits_{\widetilde{\mathfrak{B}}} \int\limits_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} \left(r_i \sigma_i(\widetilde{\mathcal{B}}, \mathbf{0}) + q_i \right) \mathrm{d}\overline{\mu}[\mathbf{r}, \mathbf{q}] \, \mathrm{d}\widetilde{\mu}[\widetilde{\mathcal{B}}] \\
&\underset{(\dagger)}{=} \overline{q}_i + \int\limits_{\widetilde{\mathfrak{B}}} \sigma_i(\widetilde{\mathcal{B}}, \mathbf{0}) \left(\int\limits_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} r_i \, \mathrm{d}\overline{\mu}[\mathbf{r}, \mathbf{q}] \right) \mathrm{d}\widetilde{\mu}[\widetilde{\mathcal{B}}] \\
&\underset{(\ddagger)}{=} \overline{q}_i + \int\limits_{\widetilde{\mathfrak{B}}} \overline{r} \, \widetilde{\sigma}_i(\widetilde{\mathcal{B}}) \, d\widetilde{\mu}[\widetilde{\mathcal{B}}] = \overline{q}_i + \overline{r} \, \mathbb{E}_{\widetilde{\mu}}(\widetilde{\sigma}_i). \qquad (7)
\end{aligned}
$$

Here, $(\diamond)$ is because $\mu = \Phi(\widetilde{\mu} \times \overline{\mu})$, $(*)$ is because σ is RI, $(\dagger)$ is by definition of $\overline{q}_i$, and $(\ddagger)$ is because $\overline{\mu}$ is anonymous. Thus,

$$
\begin{aligned}
\mathbb{E}_\mu(\sigma_\eta) &= \frac{1}{I} \sum_{i \in \mathcal{I}} \mathbb{E}_\mu(\sigma_i) \;\underset{(7)}{=}\; \frac{1}{I} \sum_{i \in \mathcal{I}} \overline{q}_i + \frac{1}{I} \sum_{i \in \mathcal{I}} \overline{r} \, \mathbb{E}_{\widetilde{\mu}}(\widetilde{\sigma}_i) \\
&= \frac{1}{I} \sum_{i \in \mathcal{I}} \overline{q}_i + \frac{\overline{r}}{I} \, \mathbb{E}_{\widetilde{\mu}} \left(\sum_{i \in \mathcal{I}} \widetilde{\sigma}_i \right).
\end{aligned}
$$

Thus, if $\sigma \in \widetilde{\mathcal{S}}$ maximizes $\mathbb{E}_{\mu}[\sigma_j]$, then $\widetilde{\sigma}$ maximizes $\mathbb{E}_{\widetilde{\mu}}\left[\sum_{i\in\mathcal{I}} \widetilde{\sigma}_i\right]$, which means $\widetilde{\sigma}$ maximizes $\sum_{i\in\mathcal{I}} \widetilde{\sigma}_i(\mathcal{B})$ for $\forall_{\widetilde{\mu}}\ \mathcal{B} \in \widetilde{\mathfrak{B}}$. Thus, $\sigma(\mathcal{B}, \mathbf{0}) = \widetilde{\Upsilon}(\mathcal{B}, \mathbf{0})$, for $\forall_{\widetilde{\mu}}\ \mathcal{B} \in \widetilde{\mathfrak{B}}$. Thus, $\sigma(\mathcal{B}, \mathbf{q}) = \widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$, for $\forall_{\mu}\ (\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, because $\mu = \Phi(\widetilde{\mu} \times \overline{\mu})$. □

References

Cao X (1982) Preference functions and bargaining solutions. In: Proceeedings of the 21st IEEE conference on decision and control, vol 1, pp 164–171

d'Aspremont C, Gevers L (1977) Equity and the informational basis of collective choice. Rev Econ Stud 44:199–209

Dhillon A (1998) Extended Pareto rules and relative utilitarianism. Soc Choice Welf 15(4):521–542

Dhillon A, Mertens J-F (1999) Relative utilitarianism. Econometrica 67(3):471–498

Harsanyi J (1953) Cardinal utility in welfare economics and in the theory of risk-taking. J Polit Econ 61:434–435

Harsanyi J (1955) Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. J Polit Econ 63:309–321

Harsanyi J (1977) Rational behaviour and bargaining equilibrium in games and social situations. Cambridge University Press, Cambridge

Kalai E (1977) Proportional solutions to bargaining situations: interpersonal utility comparisons. Econometrica 45(7):1623–1630

Kalai E, Smorodinsky M (1975) Other solutions to Nash's bargaining problem. Econometrica 43:513–518

Karni E (1998) Impartiality: definition and representation. Econometrica 66(6):1405–1415

Karni E (2003) Impartiality and interpersonal comparisons of variations in well-being. Soc Choice Welf 21(1):95–111

Karni E, Weymark JA (1998) An informationally parsimonious impartial observer theorem. Soc Choice Welf 15(3):321–332

Maskin E (1978) A theorem on utilitarianism. Rev Econ Stud 45(1):93–96

Muthoo A (1999) Bargaining theory with applications. Cambridge university Press, Cambridge

Myerson RB (1981) Utilitarianism, egalitarianism, and the timing effect in social choice problems. Econometrica 49(4):883–897

Nash J (1950) The bargaining problem. Econometrica 18:155–162

Ng Y-K (1975) Bentham or Bergson? Finite sensibility, utility functions, and social welfare functions. Rev Econ Stud 42:545–569

Ng Y-K (1985) The utilitarian criterion, finite sensibility, and the weak majority preference principle. A response Soc Choice Welf 2(1):37–38

Ng Y-K (2000) From separability to unweighted sum: a case for utilitarianism. Theory Decis 49(4):299–312

Roemer JE (1998) Theories of Distributive Justice. Harvard University Press, Cambridge

Segal U (2000) Let's agree that all dictatorships are equally bad. J Polit Econ 108(3):569–589