Twofold Optimality of the Relative Utilitarian Bargaining Solution

Marcus Pivato

Department of Mathematics, Trent University

9. April 2007

Social Choice and Welfare manuscript No.
(will be inserted by the editor)

Twofold Optimality of the Relative Utilitarian Bargaining Solution

Marcus Pivato

Department of Mathematics, Trent University,
1600 West Bank Drive, Peterborough, Ontario, Canada, K9J 7B8;
e-mail: marcuspivato@trentu.ca

The date of receipt and acceptance will be inserted by the editor

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 (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 which are generated using a certain class of distributions; this is somewhat reminiscent of the results of Harsanyi (1953) and Karni (1998).

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 $\mathcal{A}$ and also over the set of all lotteries between elements in $\mathcal{A}$, and if these lottery preferences satisfy the von Neumann-Morgenstern (vNM) axioms of minimal rationality, then we can define a 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 $\mathcal{B}$ represents an assignment of a vNM utility level to each player, obtainable through some lottery between elements of $\mathcal{A}$. Let $\wp\mathcal{B}$ be the 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 element of $\wp\mathcal{B}$ will be unanimously accepted unless it is Pareto-preferred to $\mathbf{q}$. Thus, the set of admissible bargains is the set $\wp_{\mathbf{q}}\mathcal{B} := \left\{ \mathbf{b} \in \wp\mathcal{B} \,;\, \mathbf{q} \overset{\wp}{\preceq} \mathbf{b} \right\}$, where "$\mathbf{q} \overset{\wp}{\preceq} \mathbf{b}$" means $\mathbf{b}$ is Pareto-preferred to $\mathbf{q}$.

Thus, a von Neumann-Morgenstern *bargaining problem* on $\mathcal{I}$ consists of an ordered pair $(\mathcal{B}, \mathbf{q})$, where $\mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\not\infty}$ 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 will actually assume that $\mathcal{B}$ is *strictly* convex; this involves a slight loss of generality, but it is true for a 'generic' choice of vNM utility functions $\{u_i\}_{i\in\mathcal{I}}$ on $\mathcal{A}$. Let $\mathfrak{B}$ be the set of all strictly convex bargaining problems over $\mathcal{I}$. That is:

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

A *bargaining solution* is a function $\sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\infty}$ such that, for all $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$: (1) $\sigma(\mathcal{B}, \mathbf{q}) \in \mathcal{B}$, and (2) $\sigma(\mathcal{B}, \mathbf{q}) \overset{\wp}{\succeq} \mathbf{q}$. [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 §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.[1]]

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

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

(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

[1] Formally, any bargaining solution can be converted to a social choice function by defining the 'status quo' to be some point of minimal utility for all players (e.g. a Hobbesian 'state of nature'). Conversely, any social choice function can converted into a bargaining solution. Thus, the two concepts are mathematically equivalent; the difference lies in the interpretation. Bargaining problems generally involve relatively small groups of people (e.g. two or three), and require unanimous consent. Social choice problems usually involve large groups (e.g. entire societies), and usually do not require unanimity. These different interpretations may suggest different axioms or modelling assumptions, which may then lead to different solutions.

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, for any $i, j \in \mathcal{I}$, the vNM utility functions u_i and u_j are 'interpersonally comparable'; in other words, if $u_i(a) > u_j(b)$, this somehow means that i is 'happier' under outcome a than j is under outcome b. Nothing in the vNM framework justifies this assertion. Indeed, vNM cardinal utility functions are only well-defined up to affine transformations —that is, if $s \in \mathbb{R}_{\neq}$ and $t \in \mathbb{R}$, then the function $\widetilde{u}_i(a) := s \cdot u_i(a) + t$ 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 change the shape of the bargaining problem $(\mathcal{B}, \mathbf{q})$, and change the outcome of Υ. Thus, the CU solution Υ can be easily manipulated by the players of $\mathcal{I}$, simply by affine-transforming their declared utility functions. Indeed, strictly speaking, Υ is not well-defined within the vNM theory of cardinal utility functions.

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 $\{u_i\}_{i\in\mathcal{I}}$. One way to achieve this 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

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

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

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

The *relative utilitarian* bargaining solution $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$ is the point in $\wp_{\mathbf{q}}\mathcal{B}$ which maximizes the value of $U_{\mathcal{B},\mathbf{q}}$.

Relative utilitarianism (RU) is a form 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 obtain $\widetilde{\Upsilon}(\mathcal{B}, \mathbf{q})$, we first apply the rescaling function $F : \mathbb{R}^{\mathcal{I}}_{\neq} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\neq}$ defined

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

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: it does not presuppose some standard, 'objective' way to compare utilities between individuals. 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 §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 §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 randomly generated infinite 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 classic utilitarian bargaining solution Υ to any other bargaining

solution, and she will prefer the relative utilitarian bargaining solution $\widetilde{\Upsilon}$ to any other rescaling-invariant solution.

§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:

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

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)[2] *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 $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$,

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

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)[3] *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 somewhat 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 $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$,

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

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 $(\mathcal{B}, \mathbf{q})$. This leads to the next axiom:

(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 $\rho\sigma + (1 - \rho)\tau \approx \rho\varsigma + (1 - \rho)\tau$.

[2] Segal calls this axiom "Monotonicity".

[3] Segal calls this axiom "Independence".

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 eqn.(1), then

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

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. Our third axiom is a weakened form[4] of Segal's 'Dictatorship Indifference'.

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

The main result of this section is this:

Theorem 1 *Let* $\widetilde{\Upsilon} : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\nearrow}$ *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 that

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

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

(ℓ0) $\Big(\sigma_\rho \approx \sigma_1\Big) \implies \Big(\sigma_\rho \approx \sigma_\mu\Big)$.

[4] Segal also requires indifference amongst 'piecewise mixtures' of dictatorship solutions.

($\boldsymbol{\ell}$1) $\Big(\sigma_\rho \preceq \sigma_1\Big) \implies \Big(\sigma_\rho \preceq \sigma_\mu\Big)$.

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

($\boldsymbol{\ell}$2) $\Big(\sigma_\rho \prec \sigma_1\Big) \implies \Big(\sigma_\rho \prec \sigma_\mu\Big)$.

Finally, note that

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

To see **(L2)**, suppose $\mathbf{0} < \rho < \mu < \mathbf{1}$. If $\sigma_0 \prec \sigma_1$, then Axiom **(SL2)** and eqn.(5) imply that $\sigma_\mu \prec \sigma_1$. By a similar argument, $\sigma_0 \prec \sigma_\rho$. Finally, by a similar argument, $\sigma_\rho \prec \sigma_1$; thus, Fact **($\boldsymbol{\ell}$2)** implies that $\sigma_\rho \prec \sigma_\mu$. 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 $\rho \not\leq \rho'$.*

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

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

Here, $(*)$ is by **(L1)** because $\sigma_0 \preceq \sigma_1$ and $\rho \leq \rho'$. Next, $(\dagger)$ is by Axiom **(SL2)**, because $\sigma_1 \prec \sigma'_1$ and $\mathbf{0} < \rho' < \mathbf{1}$. Finally, (H) is by hypothesis. Thus, we get $\sigma_\rho \prec \sigma_\rho$, which is a contradiction. Thus, it cannot be true that $\rho \leq \rho'$. ◇ Claim 2

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

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

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

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 $U_\sigma(\mathcal{B}, \mathbf{q}) := U_{\mathcal{B},\mathbf{q}}[\sigma(\mathcal{B}, \mathbf{q})]$, for every $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$, where $U_{\mathcal{B},\mathbf{q}}$ is defined as in eqn.(2). Thus, if $\varsigma \in \mathcal{S}$, we write "$U_\sigma \leq U_\varsigma$" 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 $\rho, \rho' : \mathfrak{B} \longrightarrow (0,1)$ such that $\Delta(\sigma, \rho) \cap \Delta(\sigma', \rho') \neq \emptyset$.*

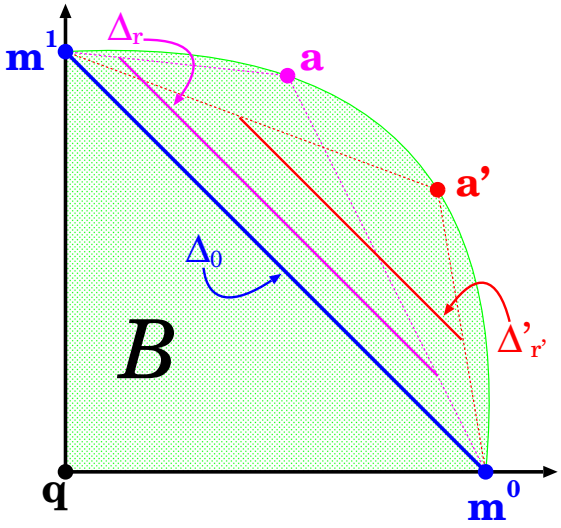

Fig. 1 Claim 5.1.

(b) *$U_\sigma \geq U_{\sigma'}$ 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 eqn.(3)

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

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

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

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

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

Also, let $\Delta_0 := \left\{ \sum_{i\in\mathcal{I}} r_i \mathbf{m}^i \;;\; \mathbf{r} \in [0,1]^{\mathcal{I}} \text{ and } \; \sum_{i\in\mathcal{I}} r_i \;=\; 1 \right\}$. 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 hyperplane segments Δ_r and $\Delta'_{r'}$ both converge to Δ_0; thus, there exist some r and r' such that Δ_r overlaps $\Delta'_{r'}$. $\quad \nabla$ Claim 5.1

Claim 5.2: *$U_{\mathcal{B},\mathbf{q}}[\sigma(\mathcal{B},\mathbf{Q})] \geq U_{\mathcal{B},\mathbf{q}}[\sigma'(\mathcal{B},\mathbf{Q})]$ if and only if $r \leq r'$.*

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

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

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

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 eqn.(6). Finally, $(\spadesuit)$ is because $r' + \sum_{i\in\mathcal{I}} r'_i \;=\; 1$ by definition. Thus, we have

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

Thus,

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

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

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

But clearly, $\rho\sigma \;+\; \sum_{i\in\mathcal{I}} \rho_i\delta_i \in \Delta(\sigma,\rho)$ 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 establishes 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. □

Remark: Our approach is clearly inspired by Segal's (2000) characterization of RU. However, Segal's original paper is not about bargaining solutions, but is instead about a somewhat more abstract class of 'resource allocation schemes'; such a scheme takes any initial bundle of commodities and allocates 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 (partial) 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 multicommodity resource-allocation problem can be reformulated as an abstract bargaining problem (Muthoo, 1999, §2.2). Segal's premise that each individual in society separately derives the optimality of RU is quite similar to our own conclusions in Section 2 (see Theorem 4 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.

2 An *ex ante* Impartial Observer Theorem

In this section we propose a form of Harsanyi's (1953) *Impartial Observer Theorem*[5] in the context of bargaining. Our approach is loosely inspired by Karni (1998); like him, we are troubled by the fact that Harsanyi's definition of '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 economic model of humans as self-regarding rational maximizers.

Instead, we 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. Intuitively, this means that, in the long-term future, she anticipates that she has an equal probability of taking on each of the two or more roles which exist in each bargaining problem —i.e. she has an equal probability of being Vendor or Customer, Landlord or Tenant, Employer or Employee. Under these conditions, she will recognize that the classic utilitarian solution Υ maximizes her *ex ante* μ-expected utility (Proposition 2). If we further require that the bargaining solution be rescaling-invariant, then each person will

[5] See Harsanyi (1953, 1955, 1977), (Weymark, 1991, p.293), (Roemer, 1998, §4.4), Karni and Weymark (1998), or (Karni, 2003, §4).

see that the relative utilitarian solution $\widetilde{\Upsilon}$ maximizes her *ex ante* μ-expected utility (Theorem 4).

Now imagine some primordial negotiation, where all members of a society must agree upon a single bargaining solution to resolve all their (unknown) future interpersonal conflicts. Assume each person seeks to maximize her expected utility, and reasons in the aforementioned fashion; then the result will be a unanimous consensus to use Υ to solve all future bargaining problems (even if each person uses a different *ex ante* measure in place of μ). If we require that the solution be rescaling-invariant, there will instead be unanimous consensus to use $\widetilde{\Upsilon}$.

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 $\sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\prec}$ is *$\mathcal{A}$-measurable* if $\sigma^{-1}(\mathcal{O}) \in \mathcal{A}$ for every open subset $\mathcal{O} \subset \mathbb{R}^{\mathcal{I}}_{\not\prec}$. 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:

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

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, this yields the following result.

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 $\forall_\mu\ (\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$.*

Proof: Clearly, $\mathbb{E}_\mu(\sigma_\eta) = \frac{1}{I}\mathbb{E}_\mu\left(\sum_{i\in\mathcal{I}} \sigma_i\right)$. Thus, if $\sigma \in \mathcal{S}$ maximizes $\mathbb{E}_\mu[\sigma_\eta]$, then σ must maximize $\mathbb{E}_\mu\left[\sum_{i\in\mathcal{I}} \sigma_i\right]$, which means σ must maximize $\sum_{i\in\mathcal{I}} \sigma_i(\mathcal{B}, \mathbf{q})$, for $\forall_\mu\ (\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$. Thus, $\sigma(\mathcal{B}, \mathbf{q}) = \Upsilon(\mathcal{B}, \mathbf{q})$, for $\forall_\mu\ (\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$. □

Proposition 2 (and classic utilitarianism in general) is objectionable because it requires interpersonal comparison of utility, and such interpersonal comparison is meaningless in the vNM framework. Strictly speaking, any bargaining solution is meaningless in the vNM framework unless it is invariant under any affine rescaling of the player's utility functions (and Υ is not rescaling-invariant).

Formally, let $\mathbf{r} = [r_i]_{i\in\mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$ 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 $\mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\not\supset}$, 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 $\mathbf{r}, \mathbf{q} \in \mathbb{R}^2_{\not\supset}$, 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 $\sigma : \mathfrak{B} \longrightarrow \mathbb{R}^{\mathcal{I}}_{\not\supset}$ is a bargaining solution, then we say that σ is *rescaling invariant* (RI) if, for every $\mathbf{r}, \mathbf{q} \in \mathbb{R}^2_{\not\supset}$ and $(\mathcal{B}, \mathbf{q}_0) \in \mathfrak{B}$, we have $\sigma(\mathbf{r} \times \mathcal{B} + \mathbf{q},\ \mathbf{r} \times \mathbf{q}_0 + \mathbf{q}) \ = \ \mathbf{r} \times \sigma(\mathcal{B}, \mathbf{q}_0) + \mathbf{q}$. For example, $\widetilde{\Upsilon}$ is RI, but Υ is not. Heuristically speaking, RI is a weak form of 'nonmanipulability'; it says that no player can alter the bargaining outcome in her favour by applying an affine transformation to her utility function. For any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$ and $i \in \mathcal{I}$, let $r_i(\mathcal{B}, \mathbf{q}) := \max\{b_i - q_i\ ;\ \mathbf{b} \in \wp_{\mathbf{q}}\mathcal{B}\}$. We define $\widetilde{\mathfrak{B}} \ := \ \{\mathcal{B} \subset \mathbb{R}^{\mathcal{I}}_{\not\supset}\ ;\ \mathcal{B}$ is strictly convex, comprehensive, and compact, and $r_i(\mathcal{B}, \mathbf{0}) = 1$, for all $i \in \mathcal{I}\}$. Let $\widetilde{\mathcal{S}}$ denote the set of all $\mathcal{A}$-measurable, rescaling-invariant bargaining solutions.

Lemma 3 **(a)** *There is a natural bijection $\Phi : \widetilde{\mathfrak{B}} \times \mathbb{R}^{\mathcal{I}}_{\not\supset} \times \mathbb{R}^{\mathcal{I}}_{\not\supset} \longrightarrow \mathfrak{B}$ defined by* $\Phi(\widetilde{\mathcal{B}}, \mathbf{r}, \mathbf{q}) \ := \ (\mathbf{r} \times \widetilde{\mathcal{B}}, \mathbf{q})$.

(b) *If $\sigma \in \widetilde{\mathcal{S}}$, then σ is determined entirely by its values on $\widetilde{\mathfrak{B}}$.*

(c) *$\widetilde{\Upsilon}$ is the unique element of $\widetilde{\mathcal{S}}$ which maximizes the value of $\sum_{i\in\mathcal{I}} \sigma_i(\mathcal{B}, \mathbf{0})$ for every $\mathcal{B} \in \widetilde{\mathfrak{B}}$.* □

Proof: **(b)** and **(c)** follow from **(a)**. To prove **(a)**, it suffices to show that, for any $(\mathcal{B}, \mathbf{q}) \in \mathfrak{B}$ there is a unique $\widetilde{\mathcal{B}} \in \widetilde{\mathfrak{B}}$ and a unique $\mathbf{r} \in \mathbb{R}^{\mathcal{I}}_{\not\supset}$ such that $(\mathcal{B}, \mathbf{q}) = \mathbf{r} \times (\widetilde{\mathcal{B}}, \mathbf{0}) + \mathbf{q}$. To see this, let $\mathbf{r} := [r_i]_{i\in\mathcal{I}}$, where $r_i = r_i(\mathcal{B}, \mathbf{q})$. Let $\mathbf{r}^{-1} := [r_i^{-1}]_{i\in\mathcal{I}}$, and define $\widetilde{\mathcal{B}} := \mathbf{r}^{-1} \times (\mathcal{B} - \mathbf{q})$. Then $\mathcal{B} = (\mathbf{r} \times \widetilde{\mathcal{B}}) + \mathbf{q}$. Thus $(\mathcal{B}, \mathbf{q}) = \mathbf{r} \times (\widetilde{\mathcal{B}}, \mathbf{0}) + \mathbf{q}$. Uniqueness is clear. □

Let $\widetilde{\mathcal{A}}$ be a sigma-algebra on $\widetilde{\mathfrak{B}}$, let Φ be as in Lemma 3(a), and assume that Φ is measurable with respect to $\mathcal{A}$, $\widetilde{\mathcal{A}}$, and the Borel sigma-algebra on $\mathbb{R}^{\mathcal{I}}_{\not\supset} \times \mathbb{R}^{\mathcal{I}}_{\not\supset}$. Let $\widetilde{\mu}$ be a probability measure on $\widetilde{\mathfrak{B}}$, let $\overline{\mu}$ be a probability measure on $\mathbb{R}^{\mathcal{I}}_{\not\supset} \times \mathbb{R}^{\mathcal{I}}_{\not\supset}$, and let $\mu := \Phi(\widetilde{\mu} \times \overline{\mu})$. Thus, a μ-random bargaining problem in $\mathfrak{B}$ is obtained by first generating a $\widetilde{\mu}$-random bargaining problem in $\widetilde{\mathfrak{B}}$, and then applying an independent, $\overline{\mu}$-random rescaling to this problem. For all $i \in \mathcal{I}$, let

$\overline{r}_i := \int_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} r_i \; d\overline{\mu}[\mathbf{r}, \mathbf{q}]$. We say that $\overline{\mu}$ is *anonymous* if there is some constant $\overline{r}$ such that $\overline{r}_i = \overline{r}$ for all $i \in \mathcal{I}$. This means every coordinate receives the same average rescaling (in particular, this will be true if $\overline{\mu}$ is any measure on $\mathbb{R}^{\mathcal{I}}_{\not\sim}$ which is invariant under any transitive group of permutations of the $\mathcal{I}$-indexed coordinate axes).

Theorem 4 *Let $\overline{\mu}$ be an anonymous probability measure on $\mathbb{R}^{\mathcal{I}}_{\not\sim}$, 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 $\overline{q}_i := \int_{\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}} q_i \; d\overline{\mu}[\mathbf{r}, \mathbf{q}]$. Then

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

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 $\widetilde{\sigma}$ must maximize $\mathbb{E}_{\widetilde{\mu}} \left[\sum_{i \in \mathcal{I}} \widetilde{\sigma}_i \right]$, which means $\widetilde{\sigma}$ must maximize the value of $\sum_{i \in \mathcal{I}} \widetilde{\sigma}_i(\mathcal{B})$ 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 $\mu = \Phi(\widetilde{\mu} \times \overline{\mu})$. □

Remark: (a) The key assumptions of Theorem 4 —that $\mu = \Phi(\widetilde{\mu} \times \overline{\mu})$, and $\overline{\mu}$ is an anonymous measure on $\mathbb{R}^{\mathcal{I}}_{\not\sim} \times \mathbb{R}^{\mathcal{I}}_{\not\sim}$ —are fairly restrictive. It is possible to prove a similar theorem for an arbitrary measure μ on $\mathfrak{B}$; however it yields a kind of 'generalized' relative utilitarianism, where for each $\mathcal{B} \in \widetilde{\mathfrak{B}}$, we maximize a

weighted utilitarian sum $\sum_{i\in\mathcal{I}} r_i(\mathcal{B})\sigma_i(\mathcal{B})$, where the weights $r_i(\mathcal{B})$ depend on $\mathcal{B}$, and are determined by the measure μ. To obtain RU, we must have $r_i(\mathcal{B}) = r_j(\mathcal{B})$ for all $i, j \in \mathcal{I}$; the assumptions of Theorem 4 are the most natural hypotheses yielding this condition.

(b) Theorem 4 says that $\widetilde{\Upsilon}$ is the unique bargaining solution in $\widetilde{\mathcal{S}}$ which is *ex ante* optimal for each person. However, clearly, $\widetilde{\Upsilon}$ is not *ex post* optimal: once a person learns the specific bargaining problem which confronts her, she can probably find some other solution in $\widetilde{\mathcal{S}}$ which will give her higher utility for *this* problem. Thus, any implementation of RU based on Theorem 4 must include a mechanism to extract irrevocable commitments to RU from all players at the *ex ante* stage, and make defection from RU highly costly at the *ex post* stage. (Note that each player will find it *ex ante* optimal to make such a commitment, as long as she is assured that *every other* player must also make such a commitment.)

(c) Our analysis of RU bargaining assumes that it is possible to obtain true information about the utility functions of the bargainers. Of course this is false. Sobel (2001) has studied the Nash equilibria of the game which results when players are allowed to strategically misrepresent their utility functions in RU bargaining.

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. Review of Economic Studies 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. Journal of Political Economy 61 (434-435).

Harsanyi, J., 1955. Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. Journal of Political Economy 63, 309–321.

Harsanyi, J., 1977. Rational behaviour and bargaining equilibrium in games and social situations. Cambridge UP, Cambridge, UK.

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, J. A., 1998. An informationally parsimonious impartial observer theorem. Soc. Choice Welf. 15 (3), 321–332.

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

Muthoo, A., 1999. Bargaining theory with applications. Cambridge UP, Cambridge, UK.

Myerson, R. B., 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., October 1975. Bentham or Bergson? Finite sensibility, utility functions, and social welfare functions. Review of Economic Studies 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 and Decision 49 (4), 299–312.

Roemer, J. E., 1998. Theories of Distributive Justice. Harvard UP, Cambridge, MA.

Segal, U., 2000. Let's agree that all dictatorships are equally bad. Journal of Political Economy 108 (3), 569–589.

Sobel, J., 2001. Manipulation of preferences and relative utilitarianism. Games Econom. Behav. 37 (1), 196–215.

Weymark, J. A., 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In: Elster, J., Roemer, J. (Eds.), Interpersonal comparisons of well-being. Cambridge UP, Cambridge, UK, pp. 255–320.