Risky social choice with approximate interpersonal comparisons of well-being

Pivato, Marcus

Department of Mathematics, Trent University

20 September 2010

Risky social choice with approximate interpersonal comparisons of well-being*

Marcus Pivato
Trent University, Canada
marcuspivato@trentu.ca

September 20, 2010

Abstract

We develop a model of social choice over lotteries, where people's psychological characteristics are mutable, their preferences may be incomplete, and approximate interpersonal comparisons of well-being are possible. Formally, we suppose individual preferences are described by a von Neumann-Morgenstern (vNM) preference order on a space of lotteries over psychophysical states; the social planner must construct a vNM preference order on lotteries over social states. First we consider a model when the individual vNM preference order is incomplete (so not all interpersonal comparisons are possible). Then we consider a model where the individual vNM preference order is complete, but unknown to the planner, and thus modeled by a random variable. In both cases, we obtain characterizations of a utilitarian social welfare function.

Most models of social welfare and collective choice take one of two positions. Either there is *no* possibility of interpersonal comparisons of well-being, or there exist *complete* interpersonal comparisons of some specific welfare information. However, starting with Sen (1970b), several authors have considered a compromise between these extremes; while acknowledging that 'precise' interpersonal comparisons of well-being might be impossible, these authors argue that certain 'approximate' interpersonal comparisons are sometimes obvious, and thus, should be incorporated into any reasonable ethical theory.[1] Sen *et al* consider generalizations of the utilitarian social

[*] I am grateful to Özgür Evren, Klaus Nehring, Efe Ok, and Clemens Puppe for their helpful comments on early drafts of this paper. I am especially grateful to Marc Fleurbaey, Franz Dietrich, and two anonymous referees for their many detailed and valuable comments. None of these people are responsible for any errors or deficiencies which remain. The final work on this paper was done while visiting the Université de Montréal Department of Economics; I thank the UdeM and CIREQ for their hospitality. This research was also supported by NSERC grant #262620-2008.

[1] See Sen (1970b, 1972 and Ch.7* of 1970a), Fine (1975), Blackorby (1975), Basu (1980, Ch.6), and Baucells and Shapley (2006, 2008).

welfare function, where each person's preferences could be represented by any member of a family of utility functions, and interpersonal comparisons could be made using any element of a convex cone of weight vectors. The result is a partial ordering over the set of social alternatives which, while incomplete, is still much more complete than the Pareto order. However, these authors still make two other assumptions:

(a) Each person has *complete* preferences (or a complete personal welfare ordering) over the space of social alternatives.

(b) Each person has *fixed* psychological characteristics. Psychological changes are not part of the set of social alternatives. (Thus, we cannot change someone's preferences, or the psychological factors which influence her sense of well-being).

A companion paper (Pivato, 2010a) has developed a model of 'approximate interpersonal welfare comparisons' which relaxes assumptions (a) and (b). In that model, there is a space Φ of 'personal physical states' and a space Ψ of 'personal psychological states', and each individual in society is characterized by an ordered pair $(\psi, \phi) \in \Psi \times \Phi$. Both the physical state ϕ *and* the psychological characteristics ψ are mutable, and hence potential objects of individual or social choice. Pivato (2010a) postulates an incomplete preorder $(\succeq)$ on the space $\Psi \times \Phi$ of all possible psychophysical states. The statement $(\psi_1, \phi_1) \succeq (\psi_2, \phi_2)$ means that psychophysical state (ψ_1, ϕ_1) is objectively better than (or would be universally preferred to) state (ψ_2, ϕ_2).

Pivato (2010b) develops a model of (incomplete) social preferences, based on the interpersonal preorder $(\succeq)$ proposed in Pivato (2010a); the main result is a characterization of the 'approximate maximin' social preorder. Pivato (2010c) instead supposes it is possible to approximately compare the utility *gains* (or losses) which different citizens experience as a result of a policy; this leads to the 'approximate utilitarian' social preorder. However, all of these models apply only to riskless decisions. The present paper introduces risk to the model.

Like Pivato (2010b,c), this paper supposes there is a set $\mathcal{X}$ of 'psychophysical states'. (This could be the space $\Psi \times \Phi$ of Pivato (2010a), but it doesn't have to be.) Each citizen, at each moment, is located at some point in $\mathcal{X}$, so that the social state is described by a vector $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$ (where $\mathcal{I}$ is a set indexing the citizens). Pivato (2010b) posited a preorder $(\succeq)$ on $\mathcal{X}$ (an objective —but possibly incomplete —comparison of the relative welfare or happiness of different psychophysical states), and used this to derive a 'social preorder' $(\trianglerighteq)$ on $\mathcal{X}^{\mathcal{I}}$. To model risky decisions, we must instead consider preferences over lotteries. So, Section 1 of this paper introduces a general model of incomplete interpersonal lottery preferences. Section 2 introduces a model of preference aggregation called a *von Neumann-Morgenstern* (vNM) *social preorder*. This paper's first main result (Theorem 2.1) shows that the 'approximate utilitarian' preorder is a subrelation of every other 'reasonable' vNM social preorder. Section 3 illustrates with an application to bilateral bargaining.

Next, Section 4 introduces an entirely different model of approximate interpersonal preferences over lotteries, in the form of a 'random' vNM utility function on

$\mathcal{X}$. Using this framework, Section 5 contains the paper's other main result: a 'profile-independent' verson of Harsanyi's (1955) Social Aggregation Theorem, with approximate interpersonal comparisons (Theorem 5.2). Finally, Section 6 uses the framework of Section 4 to argue for the instrumental value of personal liberty. All proofs are in an appendix at the end of the paper.

It is first necessary to address a serious philosophical objection to the social aggregation of vNM preferences over lotteries. Harsanyi's (1953) Impartial Observer Theorem, Harsanyi's (1955) Social Aggregation Theorem, and Myerson's (1981) 'no-timing' theorem all interpret the utilitarian social welfare function as a sum of vNM utility functions.[2] However, Sen (1976, 1977) has rejected this approach. Sen distinguishes a person's 'true' utility function (which measures the actual 'welfare' or 'happiness' which she would receive from various alternatives) from her vNM utility function (which conflates her true utility function with her intrinsic risk aversion). Juan and Sue may have the same true utility function, but exhibit different vNM functions, because Juan relishes the thrill of gambling while Sue suffers from anxiety attacks. Thus, the sum of vNM utility functions proposed by Harsanyi (1953, 1955) and Myerson (1981) is the not the 'true' utilitarian SWO which would have been advocated by Bentham or Sidgwick.[3]

There are at least two responses this objection. One response is to argue that, at least for *risky* social decisions, the sum of vNM utility functions *is* the correct objective function for the social planner, because it reflects the aggregate risk preference of the citizens, in addition to the aggregate of their 'true' utilities. If the social planner maximized the expected sum of the citizens' *true* utility functions, instead of the expected sum of their vNM functions, then society might end up taking risks which the citizens themselves would not wish to take (or vice versa); this would be paternalistic.

However, while it seems desirable to respect the 'aggregate risk preference' of the citizens, it isn't clear that the *sum* of vNM functions is the right way to compute this 'aggregate risk preference' —precisely because the vNM functions conflate risk preferences with true utility functions. (If it were somehow possible to isolate the 'pure risk preferences' of each citizen from her vNM function, then perhaps the correct 'aggregator' would be the minimum, or maximum, or product of these risk preferences; this question needs further analysis.) Furthermore, this entire argument falls apart for *riskless* social decisions: why should people's risk preferences influence social choice in a setting without risk?

There is another response to Sen's objection. Instead of summing the vNM utility functions of the citizens themselves, we could sum the vNM utility functions of sympathetic and risk-neutral proxies or delegates acting on their behalf. Imagine that risk-loving Juan appoints a risk-neutral delegate, Janos, while risk-averse Sue appoints

[2] Of course, a vNM preference defines a two-dimensional family of vNM utility functions, which are all affine transformations of one another. So to meaningfully speak of 'summing vNM utility functions', we presume there is some rule which selects a canonical representative from each of these families. (In fact, in the present paper, this issue will never arise.)

[3] See Harsanyi (1975, 1977), Weymark (1991) and Roemer (1996;§4.3) for further discussions of the Sen-Harsanyi debate.

a risk-neutral delegate, Zsuzsanna. Being Juan's delegate, Janos wants exactly the same things Juan wants (at least, in a riskless context), so they have the same true utility function. But since Janos is risk-neutral, his vNM function is equal to his true utility function (and thus, equal to Juan's).[4] Likewise, Zsuzsanna's vNM function is equal to Sue's true utility function. Thus, if we sum the vNM functions of Janos and Zsuzsanna, as proposed by Harsanyi (1953, 1955) and Myerson (1981), then we are actually computing the 'true' utilitarian SWO of Bentham and Sidgwick.

This interposition of imaginary 'delegates' may seem artificial, but in fact it merely extends the reasoning already present in Harsanyi's original thought experiments, which imagine a 'neutral but sympathetic observer' who acts on behalf of the citizens. Thus, throughout this paper, when I speak of the lottery preferences or vNM utility functions of the citizens, I will assume that all citizens are either risk-neutral, or are being represented by risk-neutral 'delegates'. (If the social planner also wishes to respect the 'aggregate risk preference' of the citizens, then this must be done *ex post facto*.)

1 von Neumann-Morgenstern interpersonal preorders

Let $\mathcal{X}$ be a measurable space (i.e. a set equipped with a sigma-algebra), and let $\mathbb{P}(\mathcal{X})$ be the set of all probability measures on $\mathcal{X}$. Let $\mathfrak{P} \subseteq \mathbb{P}(\mathcal{X})$ be any convex subset of $\mathbb{P}(\mathcal{X})$. (For example: $\mathfrak{P}$ might be the set of all probability measures which are absolutely continuous with respect to some fixed reference measure μ. Or, $\mathfrak{P}$ might be the set of all measures satisfying some 'regularity' conditions with respect to a given topology on $\mathcal{X}$.) A *von Neumann-Morgenstern preorder* on $\mathfrak{P}$ is a reflexive, transitive (but possibly incomplete) binary relation ($\succeq$) on $\mathfrak{P}$ which satisfies the following 'Linearity' axiom:

(Lin) For all $\rho, \rho'_1, \rho'_2 \in \mathfrak{P}$ and $s, s' \in (0,1)$ with $s + s' = 1$, $\left(\rho'_1 \preceq \rho'_2\right) \implies \left((s\rho + s'\rho'_1) \preceq (s\rho + s'\rho'_2)\right)$.

Now suppose $\mathcal{X}$ s a set of 'personal psychophysical states'. Each element $x \in \mathcal{X}$ encodes all information about an individual's current psychology (i.e. her personality, mood, knowledge, beliefs, memories, values, desires, etc.) and also all information about her current personal physical state (i.e. her health, wealth, physical location, consumption bundle, sense-data, etc.).[5] Thus, any human being, at any moment in time, resides at some point in $\mathcal{X}$. A preorder on $\mathcal{X}$ thus represents a model of human

[4] Someone might object: even if Janos is risk-neutral, his vNM preferences are simply his 'preferences over lotteries', and not necessarily indicative of his 'true utility function'. I invoke Occam's razor: aside from his true utility function and his 'intrinsic risk preferences', what *else* could influence Janos's preferences over lotteries? Indeed, it seems reasonable to *define* his 'risk preference' to be simply the deviation between his vNM function and his true utility function; then my claim about Janos is true by definition.

[5] Like Pivato (2010b,c), but unlike Pivato (2010a), this model does not assume we can separate someone's 'psychological' state from her 'physical' state. Indeed, if the mind is a function of the brain, then her psychological state is simply one aspect of her physical state.

welfare where individuals may have incomplete preferences (because not all $x, y \in \mathcal{X}$ are comparable), but where 'approximate' interpersonal comparisons of well-being may be possible (because every possible human mind is represented by some subset of $\mathcal{X}$). Pivato (2010a,b) calls this an *interpersonal preorder*.

If $\mathfrak{P} \subseteq \mathbb{P}(\mathcal{X})$, then a vNM preorder $(\succeq)$ on $\mathfrak{P}$ represents a model of approximate interpersonal comparisons and incomplete preferences over lotteries; we will call this a *von Neumann-Morgenstern interpersonal preorder* (vNMIP). For all $x \in \mathcal{X}$, let $\delta_x \in \mathbb{P}(\mathcal{X})$ be the point mass at x (i.e. $\delta_x\{\times\} = 1$), and suppose $\delta_x \in \mathfrak{P}$. Then $(\succeq)$ defines an interpersonal preorder $(\underset{*}{\succeq})$ on $\mathcal{X}$ by $(x \underset{*}{\succeq} y) \iff (\delta_x \succeq \delta_y)$. Thus, a vNMIP is a natural extension of the model of Pivato (2010a,b).

Multiutility representations. For any $\rho \in \mathbb{P}(\mathcal{X})$ and measurable function $v : \mathcal{X} \longrightarrow \mathbb{R}$, we define

$$v^*(\rho) \quad := \quad \int_{\mathcal{X}} v(x) \, \mathrm{d}\rho[x]. \tag{1}$$

Let $\mathcal{V}$ be a collection of measurable functions $v : \mathcal{X} \longrightarrow \mathbb{R}$. We can define a vNMIP $(\succeq)$ on $\mathfrak{P}$ as follows: for any $\rho, \rho' \in \mathfrak{P}$,

$$\Big(\rho \succeq \rho'\Big) \iff \Big(v^*(\rho) \geq v^*(\rho') \text{ for all } v \in \mathcal{V}\Big). \tag{2}$$

We say that $\mathcal{V}$ provides a *multiutility representation* for $(\succeq)$.

Fix a topology on $\mathfrak{P}$. We say that $(\succeq)$ is *continuous* if the following holds: for all $\rho, \rho' \in \mathfrak{P}$, and sequences $\{\rho_n\}_{n=1}^\infty$ and $\{\rho'_n\}_{n=1}^\infty$, if $\rho_n \xrightarrow[n\to\infty]{} \rho$ and $\rho'_n \xrightarrow[n\to\infty]{} \rho'$, and $\rho_n \preceq \rho'_n$ for all $n \in \mathbb{N}$, then $\rho \preceq \rho'$. In several settings, any continuous vNM preorder admits a multiutility representation like (2). For example, suppose $|\mathcal{X}| = N$ is finite, and $\mathfrak{P}$ is the simplex in $\mathbb{R}^N$, which we endow with the obvious Euclidean topology. Then any continuous vNM preorder on $\mathfrak{P}$ has a multiutility representation (Baucells and Shapley, 1998, Theorem 1.8, p.12).

Now let $\mathcal{X}$ be a compact metric space, and let $\mathfrak{P}$ be the space of all Borel probability measures on $\mathcal{X}$, endowed with the weak* topology induced by the set $\mathcal{C}(\mathcal{X})$ of all continuous real-valued functions on $\mathcal{X}$. Then any continuous vNM preorder on $\mathfrak{P}$ has a multiutility representation (Dubra et al., 2004).

Next, let $\mathcal{X}$ be a sigma-compact metric space, and let $\mathfrak{P}_c$ be the space of compactly supported Borel probability measures with the weak* topology induced by $\mathcal{C}(\mathcal{X})$. Then any continuous vNM preorder on $\mathfrak{P}_c$ has an expected multiutility representation (Evren, 2008, Thm.2). Also, if $\mathfrak{P}$ is the space of *all* Borel probability measures on $\mathcal{X}$, and $(\succeq)$ is a continuous vNM preorder on $\mathfrak{P}$ such that the set of point-masses in $\mathfrak{P}$ has a $(\succeq)$-maximal element and a $(\succeq)$-minimal element, then $(\succeq)$ has a multiutility representation (Evren, 2008, Thm.3).

However, let $\mathcal{X} = \mathbb{R}$, and let $\mathfrak{P}$ be the space of Borel probability measures on $\mathbb{R}$, endowed with the weak* topology from the space $\mathcal{C}_b(\mathbb{R})$ of bounded continuous real-valued functions. Then Evren (2008, Prop.1) has shown that some continuous vNM preorders on $\mathfrak{P}$ do *not* admit expected multiutility representations.

2 von Neumann-Morgenstern social preferences

Let $\mathcal{I}$ be a finite set indexing a population. A *social state* is a vector $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$, assigning a psychophysical state x_i to each $i \in \mathcal{I}$. Any policy chosen by the social planner will result in a probability distribution $\boldsymbol{\rho}$ over $\mathcal{X}^{\mathcal{I}}$. To decide the 'best' policy, the social planner must formulate a preference relation $(\trianglerighteq)$ over $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$. For any $\boldsymbol{\rho} \in \mathbb{P}(\mathcal{X}^{\mathcal{I}})$, and any $i \in \mathcal{I}$, let $\rho_i \in \mathbb{P}(\mathcal{X})$ be the lottery on the ith coordinate induced by $\boldsymbol{\rho}$. That is, for any measurable subset $\mathcal{U} \subset \mathcal{X}$,

$$\rho_i[\mathcal{U}] \quad := \quad \boldsymbol{\rho}\big\{\mathbf{x} \in \mathcal{X}^{\mathcal{I}} \ ; \ x_i \in \mathcal{U}\big\}. \tag{3}$$

For any convex subset $\mathfrak{P} \subseteq \mathbb{P}(\mathcal{X})$, let $\mathfrak{P}^{\otimes \mathcal{I}} := \big\{\boldsymbol{\rho} \in \mathbb{P}(\mathcal{X}^{\mathcal{I}}) \ ; \ \rho_i \in \mathfrak{P}, \ \forall \ i \in \mathcal{I}\big\}$; this is a convex subset of $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$.

If $\sigma : \mathcal{I} \longrightarrow \mathcal{I}$ is a permutation, and $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$, then we define $\sigma(\mathbf{x}) := \mathbf{x}'$, where $x'_i := x_{\sigma(i)}$ for all $i \in \mathcal{I}$. For any $\boldsymbol{\rho} \in \mathbb{P}(\mathcal{X}^{\mathcal{I}})$, we define $\sigma(\boldsymbol{\rho}) := \boldsymbol{\rho}'$ as follows:

$$\text{For any measurable subset } \mathcal{U} \subseteq \mathcal{X}^{\mathcal{I}}, \qquad \boldsymbol{\rho}'[\mathcal{U}] \quad := \quad \boldsymbol{\rho}\left[\sigma^{-1}(\mathcal{U})\right]. \tag{4}$$

It is easy to check that $\sigma[\mathfrak{P}^{\otimes \mathcal{I}}] = \mathfrak{P}^{\otimes \mathcal{I}}$. If $(\succeq)$ is a vNMIP on $\mathfrak{P}$, then a $(\succeq)$-*social preference order* (or $(\succeq)$-*vNMSP*) is a preorder $(\trianglerighteq)$ on $\mathfrak{P}^{\otimes \mathcal{I}}$ with the following properties:

(Par) For all $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathfrak{P}^{\otimes \mathcal{I}}$, if $\rho_i \preceq \rho'_i$ for all $i \in \mathcal{I}$, then $\boldsymbol{\rho} \trianglelefteq \boldsymbol{\rho}'$. Also, if $\rho_i \prec \rho'_i$ for all $i \in \mathcal{I}$, then $\boldsymbol{\rho} \triangleleft \boldsymbol{\rho}'$.

(Anon) If $\sigma : \mathcal{I} \longrightarrow \mathcal{I}$ is any permutation, then for all $\boldsymbol{\rho} \in \mathfrak{P}^{\otimes \mathcal{I}}$, $\boldsymbol{\rho} \overset{\triangle}{\equiv} \sigma(\boldsymbol{\rho})$.

(Lin) For all $\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, \boldsymbol{\rho}'_1, \boldsymbol{\rho}'_2 \in \mathfrak{P}^{\otimes \mathcal{I}}$, and $s, s' \in [0,1]$ with $s + s' = 1$, if $\boldsymbol{\rho}_1 \trianglelefteq \boldsymbol{\rho}_2$ and $\boldsymbol{\rho}'_1 \trianglelefteq \boldsymbol{\rho}'_2$, then $(s\boldsymbol{\rho}_1 + s'\boldsymbol{\rho}'_1) \trianglelefteq (s\boldsymbol{\rho}_2 + s'\boldsymbol{\rho}'_2)$.

Axiom (Par) is simply the Pareto axiom, and Axiom (Lin) is just the von Neumann-Morgenstern linearity axiom. Axiom (Anon) makes sense because the elements of $\mathcal{I}$ are merely 'placeholders', with no psychological content —recall that *all* information about the 'psychological identity' of individual i is encoded in x_i. Thus, if $\mathbf{x}, \mathbf{y}$ are two social alternatives, and $x_i \neq y_i$, then it may not make any sense to compare the welfare of x_i with y_i (unless such a comparison is allowed by $(\succeq)$), because x_i and y_i represent *different people* (even though they have the same index). On the other hand, if $x_i = y_j$, then it makes perfect sense to compare x_i with y_j, even if $i \neq j$, because x_i and y_j are in every sense the *same* person (even though this person has different indices in the two social alternatives).[6]

[6] See (Pivato, 2010b, §3) for more discussion of (Anon) and (Par).

Approximate utilitarianism. Fix $\boldsymbol{\rho} \in \mathfrak{P}^{\otimes \mathcal{I}}$. For all $i \in \mathcal{I}$, let $\rho_i \in \mathfrak{P}$ be as in eqn.(3). Define the *per capita average lottery*

$$\overline{\rho} \quad := \quad \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \rho_i \ \in \ \mathfrak{P}. \tag{5}$$

The *approximate utilitarian* $(\succeq)$-vNMSP $(\underset{u}{\trianglerighteq})$ is then defined:

$$\forall \boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathfrak{P}^{\otimes \mathcal{I}}, \qquad \left(\boldsymbol{\rho} \underset{u}{\trianglerighteq} \boldsymbol{\rho}' \right) \iff \left(\overline{\rho} \succeq \overline{\rho}' \right). \tag{6}$$

The first main result of this paper is:

Theorem 2.1 *Let $(\succeq)$ be a vNMIP on $\mathfrak{P}$. Every $(\succeq)$-vNMSP on $\mathfrak{P}^{\otimes \mathcal{I}}$ extends and refines the approximate utilitarian $(\succeq)$-vNMSP $(\underset{u}{\trianglerighteq})$. That is, for any $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathfrak{P}^{\otimes \mathcal{I}}$,*

$$\left(\boldsymbol{\rho} \underset{u}{\trianglerighteq} \boldsymbol{\rho}' \right) \implies \left(\boldsymbol{\rho} \trianglerighteq \boldsymbol{\rho}' \right), \quad \textit{and} \quad \left(\boldsymbol{\rho} \underset{u}{\triangleright} \boldsymbol{\rho}' \right) \implies \left(\boldsymbol{\rho} \triangleright \boldsymbol{\rho}' \right).$$

Theorem 2.1 implies that the set of '$(\trianglerighteq)$-optimal' policies will be closely related to the set of '$(\underset{u}{\trianglerighteq})$-optimal' policies, according to any of the four concepts of 'optimality' which exist for incomplete preorders (Pivato, 2010b, Lemma A.1(g)).

Representation with utility functions. A *utility function* for $(\succeq)$ is a measurable function $u : \mathcal{X} \longrightarrow \mathbb{R}$ such that, for all $\rho_1, \rho_2 \in \mathfrak{P}$,

$$\left(\rho_1 \succeq \rho_2 \right) \quad \implies \quad \left(u^*(\rho_1) \geq u^*(\rho_2) \right), \tag{7}$$

where u^* is defined as in eqn.(1). Let $\mathcal{U}(\succeq)$ be the set of all utility functions for $(\succeq)$. The next result connects $(\underset{u}{\trianglerighteq})$ to the more traditional definition of utilitarianism.

Proposition 2.2 *Let $(\succeq)$ be a vNMIP on $\mathfrak{P}$ and let $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathfrak{P}^{\otimes \mathcal{I}}$.*

(a) *If $\boldsymbol{\rho} \underset{u}{\trianglerighteq} \boldsymbol{\rho}'$, then $\displaystyle\sum_{i \in \mathcal{I}} u^*(\rho_i) \geq \sum_{i \in \mathcal{I}} u^*(\rho_i')$ for all $u \in \mathcal{U}(\succeq)$.*

(b) *Conversely, if $\displaystyle\sum_{i \in \mathcal{I}} u^*(\rho_i) \geq \sum_{i \in \mathcal{I}} u^*(\rho_i')$ for all $u \in \mathcal{U}(\succeq)$, and $(\succeq)$ admits a multiutility representation (2), then $\boldsymbol{\rho} \underset{u}{\trianglerighteq} \boldsymbol{\rho}'$.*

3 Application: Interpersonal utility comparisons and bilateral bargaining

Suppose $\mathcal{X} = \Psi \times \Phi$, where Ψ is a space of 'personal psychological states' and Φ is a space of 'personal physical states'. Thus, at any moment, any person is described by an ordered pair (ψ, ϕ), where ψ encodes her personality, mood, knowledge, beliefs, memories, values, desires, etc., while ϕ encodes her her health, wealth, physical location, consumption bundle, etc. (Pivato, 2010a, §1). For any $\psi \in \Psi$, we suppose that ψ has a complete vNM preference relation $(\underset{\psi}{\succeq})$ defined on the lottery space $\mathbb{P}(\{\psi\} \times \Phi)$, and the vNM interpersonal order $(\succeq)$ agrees with $(\underset{\psi}{\succeq})$ when restricted to $\mathbb{P}(\{\psi\} \times \Phi)$.[7] For any $u : \Psi \times \Phi \longrightarrow \mathbb{R}$, let $u_\psi : \Phi \longrightarrow \mathbb{R}$ be the restriction of u to $\{\psi\} \times \Phi$. For any $\rho \in \mathbb{P}(\Phi)$, define $u(\psi, \rho) := u_\psi^*(\rho)$, as in equation (1).

A social alternative is now an ordered pair $(\boldsymbol{\psi}, \boldsymbol{\phi}) \in \Psi^{\mathcal{I}} \times \mathbb{R}^{\mathcal{I}}$, which assigns a psychophysical state (ψ_i, ϕ_i) to each $i \in \mathcal{I}$. A vNMSP is an (incomplete) preorder $(\trianglerighteq)$ on $\mathbb{P}(\Psi^{\mathcal{I}} \times \Phi^{\mathcal{I}})$. Fix $\boldsymbol{\psi}$ (i.e. allow no psychological changes). Then $(\underset{u}{\trianglerighteq})$ induces a preference order $(\underset{u,\boldsymbol{\psi}}{\trianglerighteq})$ on $\mathbb{P}(\Phi^{\mathcal{I}})$, where, for all $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathbb{P}(\Phi^{\mathcal{I}})$

$$\left(\boldsymbol{\rho} \underset{u,\boldsymbol{\psi}}{\trianglerighteq} \boldsymbol{\rho}'\right) \iff \left((\delta_{\boldsymbol{\psi}} \otimes \boldsymbol{\rho}) \underset{u}{\trianglerighteq} (\delta_{\boldsymbol{\psi}} \otimes \boldsymbol{\rho}')\right). \tag{8}$$

(Here $\delta_{\boldsymbol{\psi}} \in \mathbb{P}(\Psi^{\mathcal{I}})$ is the 'sure thing' probability measure with $\delta_{\boldsymbol{\psi}}\{\boldsymbol{\psi}\} = 1$). If the vNMIP $(\succeq)$ has a multiutility representation (2), then Proposition 2.2 and statement (8) together yield:

$$\left(\boldsymbol{\rho} \underset{u,\boldsymbol{\psi}}{\trianglerighteq} \boldsymbol{\rho}'\right) \iff \left(\sum_{i \in \mathcal{I}} u(\psi_i, \rho_i) \geq \sum_{i \in \mathcal{I}} u(\psi_i, \rho_i'), \text{ for all } u \in \mathcal{U}(\succeq)\right). \tag{9}$$

Fix $u_0 \in \mathcal{U}(\succeq)$. For all $u \in \mathcal{U}(\succeq)$ and $i \in \mathcal{I}$, there exist constants $w_i = w_i(u) \in \mathbb{R}_+$ and $b_i = b_i(u) \in \mathbb{R}$ such that, for all $\phi \in \Phi$, we have $u(\psi_i, \phi) = w_i \cdot u_0(\psi_i, \phi) + b_i$ (because both $u(\psi_i, \bullet)$ and $u_0(\psi_i, \bullet)$ are vNM cardinal utility functions for the vNM preference order $(\underset{\psi_i}{\succeq})$). For any $u \in \mathcal{U}(\succeq)$, define the 'weight vector' $\mathbf{w}(u) := (w_i(u))_{i \in \mathcal{I}} \in \mathbb{R}_+^{\mathcal{I}}$. Next, define $\mathcal{W} := \left\{\mathbf{w}(u)\ ;\ u \in \mathcal{U}(\succeq)\right\} \subseteq \mathbb{R}_+^{\mathcal{I}}$. Then statement (9) becomes:

$$\left(\boldsymbol{\rho} \underset{u,\boldsymbol{\psi}}{\trianglerighteq} \boldsymbol{\rho}'\right) \iff \left(\sum_{i \in \mathcal{I}} w_i \cdot u_0(\psi_i, \rho_i) \geq \sum_{i \in \mathcal{I}} w_i \cdot u_0(\psi_i, \rho_i'), \text{ for all } \mathbf{w} \in \mathcal{W}\right). \tag{10}$$

(The constants $\{b_i(u)\}_{i \in \mathcal{I}}$ are irrelevant because they cancel from both sides of the right-hand inequality in (9), for any fixed $u \in \mathcal{U}(\succeq)$.)

In particular, suppose $\mathcal{I} = \{1, 2\}$; then $\mathcal{W} \subseteq \mathbb{R}_+^2$. Let $\underline{A} := \inf\{w_1/w_2\ ;\ \mathbf{w} \in \mathcal{W}\}$ and $\overline{A} := \sup\{w_1/w_2\ ;\ \mathbf{w} \in \mathcal{W}\}$. As shown in Figure 1, define a preorder $(\underset{u,\boldsymbol{\psi}}{\blacktriangleright})$ on $\mathbb{R}^2$

[7] This is a 'nonpaternalism' condition, analogous to axiom (IP1) in Pivato (2010a).

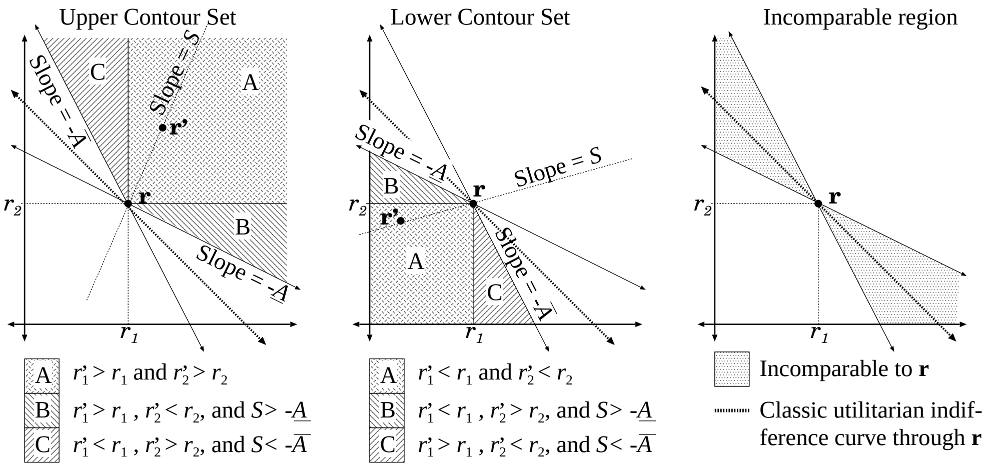
Upper Contour Set
Lower Contour Set
Incomparable region
Slope = -$\underline{A}$
Slope = S
Slope = -$\overline{A}$
A
B
C
r
r'
r_1
r_2
A $r'_1 > r_1$ and $r'_2 > r_2$
B $r'_1 > r_1$, $r'_2 < r_2$, and $S > -\underline{A}$
C $r'_1 < r_1$, $r'_2 > r_2$, and $S < -\overline{A}$
A $r'_1 < r_1$ and $r'_2 < r_2$
B $r'_1 < r_1$, $r'_2 > r_2$, and $S > -\underline{A}$
C $r'_1 > r_1$, $r'_2 < r_2$, and $S < -\overline{A}$
Incomparable to **r**
Classic utilitarian indifference curve through **r**

Figure 1: Contour sets for the relation $(\underset{u,\psi}{\blacktriangleright})$. *Left:* the upper contour set of $\mathbf{r}$. *Middle:* the lower contour set of $\mathbf{r}$. Each contour set contains three disjoint regions, corresponding to the three possible conditions implying the relation $\mathbf{r}' \underset{u,\psi}{\blacktriangleright} \mathbf{r}$ (or vice versa). *Right:* The incomparable regions $\left\{\mathbf{r}' \in \mathbb{R}^2 \; ; \; \mathbf{r}' \not\bowtie \mathbf{r}\right\}$. For reference, we also show the indifference curve of the classical utilitarian SWO.

as follows: for all $\mathbf{r}, \mathbf{r}' \in \mathbb{R}^2$,

$$\left(\mathbf{r}' \underset{u,\psi}{\blacktriangleright} \mathbf{r}\right) \iff \left(\begin{matrix}\text{either (A)} & r'_1 \geq r_1 \text{ and } r'_2 \geq r_2; \\ \text{or} \quad \text{(B)} & r'_1 \geq r_1,\; r'_2 \leq r_2 \text{ and } S \geq -\underline{A}; \\ \text{or} \quad \text{(C)} & r'_1 \leq r_1,\; r'_2 \geq r_2, \text{ and } S \leq -\overline{A}\end{matrix}\right), \tag{11}$$

where $S \;:=\; \dfrac{r'_2 - r_2}{r'_1 - r_1}$ is the slope of the line through $\mathbf{r}$ and $\mathbf{r}'$.

Proposition 3.1 *Let $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathbb{P}(\Phi^{\mathcal{I}})$, and for $i \in \{1, 2\}$, let $r_i := u_0(\psi_i, \rho_i)$ and $r'_i := u_0(\psi_i, \rho'_i)$, to obtain vectors $\mathbf{r}$ and $\mathbf{r}'$ in $\mathbb{R}^2$. Then $\left(\boldsymbol{\rho}' \underset{u,\psi}{\trianglerighteq} \boldsymbol{\rho}\right) \iff \left(\mathbf{r}' \underset{u,\psi}{\blacktriangleright} \mathbf{r}\right)$.*

Bilateral bargaining. Let $\mathcal{B} \subset \mathbb{R}^2$ be some compact, convex set —for example, the set of feasible utility profiles in a bilateral bargaining problem. Classic bargaining solutions prescribe a small (usually singleton) subset of $\mathcal{B}$ —typically by maximizing some social welfare order defined on $\mathbb{R}^2$. However, an incomplete preorder like $(\underset{u,\psi}{\blacktriangleright})$ may not have any 'maximal' points in $\mathcal{B}$. Instead, the appropriate bargaining solution in this context is the *weakly undominated set*:

$$\mathsf{wkUnd}\left(\mathcal{B}, \underset{u,\psi}{\blacktriangleright}\right) \quad := \quad \left\{\mathbf{b}^* \in \mathcal{B} \; ; \; \mathbf{b}^* \underset{u,\psi}{\not\blacktriangleleft} \mathbf{b}, \; \forall \; \mathbf{b} \in \mathcal{B}\right\}.$$

If $\mathbf{b} \in \mathcal{B}$, then $\mathbf{b}$ is weakly $(\underset{u,\psi}{\blacktriangleright})$-undominated iff the wedge $\left\{\mathbf{r}' \in \mathbb{R}^2 \; ; \; \mathbf{r}' \underset{u,\psi}{\blacktriangleright} \mathbf{r}\right\}$ shown in Figure 1(A) intersects $\mathcal{B}$ only at $\mathbf{b}$. Thus, if $\mathcal{P}$ is the Pareto frontier of $\mathcal{B}$, then $\mathsf{wkUnd}\left(\mathcal{B}, \underset{u,\psi}{\blacktriangleright}\right) \subseteq \mathcal{P}$. Furthermore, if $\mathbf{b} \in \mathcal{P}$, and T is the slope of the tangent line

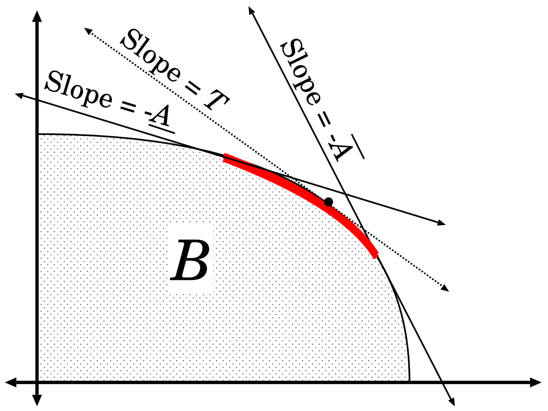

Figure 2: The approximate utilitarian bargaining solution $\mathsf{wkUnd}\left(\mathcal{B}, \underset{u,\psi}{\underline{\blacktriangleright}}\right)$.

to $\mathcal{P}$ at $\mathbf{b}$, then $\mathbf{b} \in \mathsf{wkUnd}\left(\mathcal{B}, \underset{u,\psi}{\underline{\blacktriangleright}}\right)$ if and only if $-\overline{A} \leq T \leq -\underline{A}$;[8] see Figure 2. If $(\trianglerighteq)$ is any other vNMSP, and we define a preorder $\underset{\psi}{\underline{\blacktriangleright}}$ on $\mathbb{R}^2$ by analogy to Proposition 3.1, then Theorem 2.1 implies that $(\underset{\psi}{\underline{\blacktriangleright}})$ must extend and refine $(\underset{u,\psi}{\underline{\blacktriangleright}})$. Thus, $\mathsf{wkUnd}\left(\mathcal{B}, \underset{\psi}{\underline{\blacktriangleright}}\right) \subseteq \mathsf{wkUnd}\left(\mathcal{B}, \underset{u,\psi}{\underline{\blacktriangleright}}\right)$ (Pivato, 2010b, Lemma A.1(g)). Thus, the bargaining solution proposed by any vNMSP must be contained in the approximate utilitarian bargaining solution.

4 Stochastic utility functions

We will now consider a completely different model of approximate interpersonal utility comparisons. Suppose there exists a complete vNM preference order $(\succeq)$ on $\mathbb{P}(\mathcal{X})$, described by a vNM utility function $u : \mathcal{X} \longrightarrow \mathbb{R}$, which, in principle, would allow us to make precise interpersonal comparisons of well-being. However, the exact structure of $(\succeq)$ is unknown to us. We can model this by representing u as a random variable. Formally: let Ω be a probability space, and treat u as a measurable function $u : \mathcal{X} \times \Omega \longrightarrow \mathbb{R}$. Call u a *stochastic utility function*.

Example 4.1 (a) Let Ψ be a space of personal psychological states, and Φ is a space of personal physical states, as in §3. Suppose that each $\psi \in \Psi$ has a (complete) vNM preference order $(\underset{\psi}{\succeq})$ on $\mathbb{P}(\Phi)$, which is known to us, and which can be described by a (known) vNM utility function $v_\psi : \Phi \longrightarrow \mathbb{R}$. However, the different utility functions in $\{v_\psi\}_{\psi \in \Psi}$ are expressed on different 'scales', and the correct interpersonal calibration is unknown to us. Formally: for all $\psi \in \Psi$, there are (unknown) constants $a_\psi > 0$ and $b_\psi \in \mathbb{R}$, such that, for all $\phi \in \Phi$, the 'true' well-being of (ψ, ϕ) is given by $a_\psi \, v_\psi(\phi) + b_\psi$. We don't know the vectors $\mathbf{a} := (a_\psi)_{\psi \in \Psi} \in \mathbb{R}_+^\Psi$ and $\mathbf{b} := (b_\psi) \in \mathbb{R}^\Psi$ so we model them as random variables. Thus, in this model, $\Omega := \mathbb{R}_+^\Psi \times \mathbb{R}^\Psi$ (with some probability

[8] If $\mathbf{b}$ is a corner point of $\mathcal{P}$, then this inequality must hold for *all* tangent lines at $\mathbf{b}$.

measure), $\mathcal{X} := \Psi \times \Phi$, and the stochastic utility function $u : \mathcal{X} \times \Omega \longrightarrow \mathbb{R}$ is defined by $u(\psi, \phi, \mathbf{a}, \mathbf{b}) := a_\psi \, v_\psi(\phi) + b_\psi$, for all $(\psi, \phi, \mathbf{a}, \mathbf{b}) \in \Psi \times \Phi \times \mathbb{R}_+^\Psi \times \mathbb{R}^\Psi$.

(b) Suppose we *know* how to calibrate the utility functions $\{v_\psi\}_{\psi \in \Psi}$ relative to one another; thus, we can assemble a global utility function $v : \Psi \times \Phi \longrightarrow \mathbb{R}$, so in principle we could make precise interpersonal comparisons. However, we have incomplete knowledge of the true psychological type of each person (as in a Bayesian game). There is a space Ξ of 'publicly visible' personality types, in addition to the space of 'true' psychological types Ψ. (For example, the fact that someone appears outwardly cheerful or morose is encoded in Ξ. The fact that she is truly happy or unhappy is encoded in Ψ). If a person's visible personality is $\xi \in \Xi$, then her true psychological type $\psi(\xi) \in \Psi$ is unknown to us, and thus modelled as a random variable. Formally: let Ω be a probability space, and let $\psi : \Xi \times \Omega \longrightarrow \Psi$ be a measurable function. Then define $u : \Xi \times \Phi \times \Omega \longrightarrow \mathbb{R}$ by $u(\xi, \phi, \omega) := v[\psi(\xi, \omega), \ \phi]$. Thus, if $\mathcal{X} := \Xi \times \Phi$, we obtain a stochastic utility function $u : \mathcal{X} \times \Omega \longrightarrow \mathbb{R}$. ◇

5 A stochastic social aggregation theorem

Let $\mathcal{A}$ be a set of social alternatives, and let $\mathbb{P}(\mathcal{A})$ be the set of lotteries over $\mathcal{A}$. Let $\mathcal{I}$ be a set of individuals. Harsanyi (1955, 1976) presented the following argument for utilitarianism.

Social Aggregation Theorem. *For each $i \in \mathcal{I}$, let $(\underset{i}{\succeq})$ be a vNM preference relation on $\mathbb{P}(\mathcal{A})$, represented by vNM utility function $u_i : \mathcal{A} \longrightarrow \mathbb{R}$. Let $(\trianglerighteq)$ be the social planner's vNM preference relation over $\mathbb{P}(\mathcal{A})$, and suppose $(\trianglerighteq)$ satisfies:*

(Par) *For any $\rho, \rho' \in \mathbb{P}(\mathcal{A})$, if $\rho \underset{i}{\succeq} \rho'$ for all $i \in \mathcal{I}$, then $\rho \trianglerighteq \rho'$.*

Then there are nonnegative constants $\{c_i\}_{i \in \mathcal{I}} \subset \mathbb{R}_+$ such that $(\trianglerighteq)$ is represented by the vNM utility function $U : \mathcal{A} \longrightarrow \mathbb{R}$ defined by $U(a) := \sum_{i \in \mathcal{I}} c_i u_i(a)$ for all $a \in \mathcal{A}$. □

Unfortunately, because of its 'single-profile' framework, the SAT is *not* an argument for utilitarianism. It does *not* prescribe a particular weighted utilitarian social welfare function which the social planner must employ, independent of the profile of individual vNM preferences. Instead, the SAT says that, *given* a profile $\{\underset{i}{\succeq}\}_{i \in \mathcal{I}}$ of individual vNM preferences, and given a collective vNM preference $(\trianglerighteq)$ (generated through whatever means), if $(\trianglerighteq)$ satisfies (Par) for the profile $\{\underset{i}{\succeq}\}_{i \in \mathcal{I}}$, then $(\trianglerighteq)$ can always be 'rationalized' as weighted utilitarianism *ex post facto*, by a suitable choice of weights $\{c_i\}_{i \in \mathcal{I}}$. These weights might depend on the particular profile $\{\underset{i}{\succeq}\}_{i \in \mathcal{I}}$. A proper characterization of utilitarianism must specify some weights independent of the particular profile $\{\underset{i}{\succeq}\}_{i \in \mathcal{I}}$, and use them for all conceivable profiles.[9]

[9] See Weymark (1991) or Mongin (1994) for further discussion.

The 'stochastic utility function' model of §4 yields a profile-independent version of the SAT with approximate interpersonal comparisons. For any $\boldsymbol{\rho} \in \mathbb{P}(\mathcal{X}^{\mathcal{I}})$ and $i \in \mathcal{I}$, let $\rho_i \in \mathbb{P}(\mathcal{X})$ be the projection of $\boldsymbol{\rho}$ onto the ith coordinate, defined in eqn.(3) of §2. Fix $\omega \in \Omega$, and let $u^*(\rho_i, \omega)$ be the ρ_i-expected value of u, given ω. That is:

$$u^*(\rho_i, \omega) \quad := \quad \int_{\mathcal{X}} u(x_i, \omega) \ \mathrm{d}\rho_i[x_i].$$

Given ω, assume that individual i has a preference relation $(\underset{\omega,i}{\succeq})$ over $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$ defined by $(\boldsymbol{\rho} \underset{\omega,i}{\preceq} \boldsymbol{\rho}') \iff (u^*(\rho_i, \omega) \leq u^*(\rho'_i, \omega))$. This is a vNM preference relation on $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$, with vNM utility function $u_i^\omega : \mathcal{X}^{\mathcal{I}} \longrightarrow \mathbb{R}$ defined by $u_i^\omega(\mathbf{x}) := u(x_i, \omega)$.

Suppose the social planner wishes to formulate a complete preorder $(\trianglerighteq)$ over $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$. If $(\trianglerighteq)$ satisfied the vNM axioms, then it could be represented by a vNM utility function $U : \mathcal{X}^{\mathcal{I}} \longrightarrow \mathbb{R}$. The problem is that the correct choice of U may depend on the true value of ω, which is unknown to the planner. For any measurable subset $\mathcal{S} \subseteq \Omega$, if the planner 'observes' $\mathcal{S}$ (i.e. if she acquires enough information to know that $\omega \in \mathcal{S}$), then we suppose she formulates a vNM preference relation $(\underset{\mathcal{S}}{\trianglerighteq})$ on $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$, described by a vNM utility function $U_{\mathcal{S}} : \mathcal{X}^{\mathcal{I}} \longrightarrow \mathbb{R}$. Let $\mathfrak{S}$ be the sigma-algebra on Ω and let $\pi : \mathfrak{S} \longrightarrow [0, 1]$ be the probability measure governing the random variable ω. We suppose that the family $\{U_{\mathcal{S}}\}_{\mathcal{S} \in \mathfrak{S}}$ of utility functions on $\mathcal{X}^{\mathcal{I}}$ satisfies the following 'Bayesian consistency' condition:

(Bayes) For any $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$ and any countable collection $\{\mathcal{S}_n\}_{n=1}^{\infty} \subset \mathfrak{S}$ of disjoint measurable sets, if $\mathcal{S} = \bigsqcup_{n=1}^{\infty} \mathcal{S}_n$ and $\pi(\mathcal{S}) > 0$, then $U_{\mathcal{S}}(\mathbf{x}) \ = \ \dfrac{1}{\pi(\mathcal{S})} \sum_{n=1}^{\infty} \pi(\mathcal{S}_n) \, U_{\mathcal{S}_n}(\mathbf{x})$.

Intuitively, this says that the family $\{U_{\mathcal{S}}\}_{\mathcal{S} \in \mathfrak{S}}$ behaves as if $U_{\mathcal{S}}(\mathbf{x})$ is the expected value of the unknown 'true' social utility of $\mathbf{x}$, conditioned on the observation $\mathcal{S}$. Indeed, we have the following:

Lemma 5.1 *Suppose the family* $\{U_{\mathcal{S}}\}_{\mathcal{S} \in \mathfrak{S}}$ *satisfies* (Bayes). *Then there exists a measurable function* $U : \mathcal{X}^{\mathcal{I}} \times \Omega \longrightarrow \mathbb{R}$ *such that, for any* $\mathcal{S} \in \mathfrak{S}$ *and* $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$,

$$U_{\mathcal{S}}(\mathbf{x}) \quad = \quad \frac{1}{\pi(\mathcal{S})} \int_{\mathcal{S}} U_\omega(\mathbf{x}) \ \mathrm{d}\pi[\omega]. \tag{12}$$

Intuitively, $U_\omega : \mathcal{X} \longrightarrow \mathbb{R}$ is the vNM utility function which the planner would employ if she knew that the true value was ω. Let $(\underset{\omega}{\trianglerighteq})$ be the vNM preference relation on $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$ represented by U_ω. For all $\omega \in \Omega$, we assume $(\underset{\omega}{\trianglerighteq})$ satisfies the following axioms:

(Par) For all $\boldsymbol{\rho}, \boldsymbol{\rho}' \in \mathbb{P}(\mathcal{X}^{\mathcal{I}})$, if $\boldsymbol{\rho} \underset{\omega,i}{\preceq} \boldsymbol{\rho}'$ for all $i \in \mathcal{I}$, then $\boldsymbol{\rho} \underset{\omega}{\trianglelefteq} \boldsymbol{\rho}'$.

(Anon) If $\sigma : \mathcal{I} \longrightarrow \mathcal{I}$ is any permutation, then for all $\boldsymbol{\rho} \in \mathbb{P}(\mathcal{X}^{\mathcal{I}})$, $\boldsymbol{\rho} \overset{\triangle}{\underset{\omega}{\equiv}} \sigma(\boldsymbol{\rho})$ [where $\sigma(\boldsymbol{\rho})$ is defined by eqn.(4) in §2].

(Nonindiff) The ordering $(\underset{\omega}{\trianglerighteq})$ is not totally indifferent over $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$.

(Welf) There exists a function $F : \mathbb{R}^{\mathcal{I}} \longrightarrow \mathbb{R}$ such that, for any $\omega \in \Omega$ and $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$, if $r_i := u(x_i, \omega)$ for all $i \in \mathcal{I}$, then $U_\omega(\mathbf{x}) = F(\mathbf{r})$.

The meanings of axioms (Par), (Anon), and (Nonindiff) are clear. Axiom (Welf) says that the function U is 'formally welfarist' —that is, $U_\omega(\mathbf{x})$ is entirely determined by the values of $u(x_i, \omega)$ (for all $i \in \mathcal{I}$), independent of ω.[10] Loosely speaking, this ensures that U cannot assign more 'weight' to some values of ω than others.

For any $\mathcal{S} \in \mathfrak{S}$, define $\overline{u}_{\mathcal{S}} : \mathcal{X} \longrightarrow \mathbb{R}$ by

$$\overline{u}_{\mathcal{S}}(x) \quad := \quad \frac{1}{\pi(\mathcal{S})} \int_{\mathcal{S}} u(x, \omega) \; \mathrm{d}\pi[\omega], \qquad \text{for all } x \in \mathcal{X}. \tag{13}$$

In words: $\overline{u}_{\mathcal{S}}(x)$ is the *expected value* of the random variable $u(x)$, conditional on observing the event $\mathcal{S}$. The second main result of this paper is as follows:

Theorem 5.2 *Let* $\{(\underset{\mathcal{S}}{\trianglerighteq})\}_{\mathcal{S} \in \mathfrak{S}}$ *be a* $\mathfrak{S}$*-indexed collection of vNM preference relations on* $\mathbb{P}(\mathcal{X}^{\mathcal{I}})$ *satisfying axioms* (Bayes)*,* (Par)*,* (Anon)*,* (Nonindiff)*, and* (Welf)*. Then for any* $\mathcal{S} \in \mathfrak{S}$*, the vNM relation* $(\underset{\mathcal{S}}{\trianglerighteq})$ *is represented by the vNM utility function* $\overline{U}_{\mathcal{S}} : \mathcal{X}^{\mathcal{I}} \longrightarrow \mathbb{R}$ *defined by*

$$\overline{U}_{\mathcal{S}}(\mathbf{x}) \quad := \quad \sum_{i \in \mathcal{I}} \overline{u}_{\mathcal{S}}(x_i), \qquad \textit{for all } \mathbf{x} \in \mathcal{X}^{\mathcal{I}}. \tag{14}$$

This model obviates the 'single-profile' criticism of Harsanyi's original SAT. By hypothesis, $\mathcal{X} \times \Omega$ encodes the space of all possible human psychologies *which could ever exist*; hence u encodes all possible vNM preference relations which could ever manifest in *any* profile. Thus, Theorem 5.2 does not presuppose any particular profile; it prescribes $\overline{U}_{\mathcal{S}}$ as the social welfare function which the social planner must employ when she observes event $\mathcal{S}$, independent of the 'profile' u which actually obtains.

For practical purposes, this model does not require the social planner to have precise information about people's true preferences. The hidden variable ω could contain a lot of information; in terms of Example 4.1(a), the model is even applicable when Ψ is trivial, so that *all* information about people's true preferences is hidden from the social planner. However, the model *does* require the social planner to have a correct model of the probability distribution of preferences, even if she doesn't know which preferences actually obtain (i.e. the planner must know the vNM function $u : \mathcal{X} \times \Omega \longrightarrow \mathbb{R}$, even if she doesn't know the true value of ω).

6 The value of liberty

Welfarist social choice theory has been criticized for not recognizing the value of personal liberty.[11] Let $\mathcal{A}$ be a set of alternatives, and suppose individual i has utility

[10] d'Aspremont and Gevers (2002; §3.3.1, pp.489-494) discuss 'formal welfarism'.

[11] Dowding and van Hees (2009) provide a good summary of this debate.

function $u_i : \mathcal{A} \longrightarrow \mathbb{R}$. Let a^* be the u_i-maximal element of $\mathcal{A}$. Intuitively, we feel that a social policy which allows i to choose a^* herself is more desirable than a social policy which forces a^* upon her —even though both policies yield the same utility for i. Formally, we can imagine a policy which allows i to choose any element from some subset $\mathcal{F} \subseteq \mathcal{A}$; the larger $\mathcal{F}$ is, the more 'freedom' it offers i, and hence, the more desirable the policy.

However, this account is puzzling, because by definition, elements of the set $\mathcal{A}$ are supposed to encode *all* information relevant to i's happiness or well-being, as measured by u_i. Furthermore, any 'freedom' offered by $\mathcal{F}$ is clearly a function of the 'quality' of the elements of $\mathcal{F}$ as well as their quantity. For example, if $\mathcal{F}'$ is obtained by adding an extremely undesirable option (e.g. 'execution at dawn') to $\mathcal{F}$, then we would not feel that $\mathcal{F}'$ offers i 'more freedom' than $\mathcal{F}$. This is because when i 'freely chooses' an element from $\mathcal{F}$, we suppose that what she really does is solve an optimization problem; adding options which are obviously grossly suboptimal does not enhance her optimization opportunities. However, if this 'optimization' view of free choice is correct, then once $\mathcal{F}$ contains the global optimum a^*, it seems superfluous to add any other options, because *any* other element of $\mathcal{A}$ is suboptimal, relative to a^*. Hence any measure of 'freedom' which accounts for the 'quality' of elements in $\mathcal{F}$ leads us back to welfarism. In short: 'optimality vitiates liberty'.

The 'stochastic utility' model of §4 furnishes at least two rebuttals of this 'vitiation'. The first is an argument originating with Mill (1859). Liberty is salutary because by making choices, i cultivates her ability to process and evaluate complex information and forecast the long-term consequences of her actions; she also develops her self-confidence and her sense of personal responsibility. Furthermore, there is a certain kind of satisfaction which she can achieve only by exercising personal autonomy. Formally, suppose $\mathcal{X} = \Psi \times \Phi$ as in Example 4.1(a), and individual i currently has psychophysical state (ψ_i, ϕ_i). The 'vitiation' argument implicitly assumed that $\mathcal{A} \subseteq \{\psi\} \times \Phi$. But in fact the structure of $\mathcal{A}$ depends on who does the choosing. If the planner chooses ϕ^*, then i changes from state (ψ_i, ϕ_i) to (ψ_i, ϕ^*). But by allowing i herself to identify and freely choose ϕ^*, the planner triggers a change from her current psychological state ψ_i to a new and better state $\widehat{\psi_i}$ —a psychological change which could not be caused by any other means. Since $(\widehat{\psi_i}, \phi^*) \succ (\psi_i, \phi^*)$, it is socially better to allow i to choose ϕ^* freely.

However, the 'salutary' defense of liberty has its limits. First, there are diminishing returns. Consider a person whose rational choice skills are already highly developed; perhaps someone who has made many complex decisions while occupying a position of great responsibility. Is she really going to experience much personal growth by deciding what to eat for breakfast?

Second, as observed by Sen (1997, §3), people sometimes don't *want* the responsibility of choice. For example, at a dinner party, a guest may wish she could sit in the most comfortable chair, or could take the choicest cake from a communal plate. However, she would prefer that these decisions were 'forced' upon her (e.g. that the host insists, or that the other guests take all the other chairs or cakes first), because to voluntarily take the best chair or cake would appear gauche, and violate her internal norms of politesse and/or altruism. Of course, one might argue that freedom and responsibility are still salutary, whether people want them or not. But people are sometimes faced with terrible dilemmas (e.g. ‘Sophie’s Choice’ scenarios) where, instead of gratifying and edifying, the exercise of personal choice is agonizing and psychologically destructive.

Third, this argument suggests that there is no moral difference between true liberty and a convincing illusion of liberty —as long as the citizens *believe* they are free, we have fully captured the psychological gains provided by freedom. For example, stage magicians and con artists often appear to offer their subjects a ‘free choice’, when in fact the outcome is completely determined in advance. This does not seem like freedom worth having.

But there is another, entirely different rebuttal of the ‘optimality vitiates liberty’ argument: it assumes that we *know* the u_i-optimal element of $\mathcal{A}$, because we know u_i. In reality, our knowledge of u_i is imperfect. Even in a purely welfarist framework, liberty then acquires instrumental value: by offering i a larger feasible set $\mathcal{F}$ to freely choose from, we increase the probability that $\mathcal{F}$ contains her true optimum a^* (which is unknown to us); more generally, we increase the *expected value* of $\max_{a\in\mathcal{F}} u_i(a)$.[12]

As in §4, let Ω be a probability space, and $u : \mathcal{X} \times \Omega \longrightarrow \mathbb{R}$ be a stochastic utility function. Suppose that social policy does not determine a single point $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$; instead, a social policy determines, for each $i \in \mathcal{I}$, some subset $\mathcal{F}_i \subseteq \mathcal{X}$, leaving i the freedom to choose any element of $\mathcal{F}_i$. Presumably i chooses $\arg\max_{x_i\in\mathcal{F}_i} u(x_i, \omega)$. Let us refer to the collection $(\mathcal{F}_i)_{i\in\mathcal{I}}$ as a ***freedom allocation***.[13] Given a choice between two freedom allocations $\mathbf{F} := (\mathcal{F}_i)_{i\in\mathcal{I}}$ and $\mathbf{F}' := (\mathcal{F}'_i)_{i\in\mathcal{I}}$, a utilitarian social planner will choose $\mathbf{F}$ over $\mathbf{F}'$ if it offers a higher expected utility sum, conditional on individual optimization; that is, if

$$\int_\Omega \sum_{i\in\mathcal{I}} \max_{x_i\in\mathcal{F}_i} u(x_i,\omega)\ \mathrm{d}\omega \quad > \quad \int_\Omega \sum_{i\in\mathcal{I}} \max_{x'_i\in\mathcal{F}'_i} u(x'_i,\omega)\ \mathrm{d}\omega.$$

Thus, even ignoring the potentially salutary effects of personal autonomy, a stochastic utilitarian may deem it socially optimal to maximize the liberty of citizens.

Appendix: Proofs

Proof of Theorem 2.1. Without loss of generality, suppose $\mathcal{I} = [1 \dots I]$, and define the permutation $\sigma : \mathcal{I} \longrightarrow \mathcal{I}$ by $\sigma(i) := (i + 1) \bmod I$. Define $\widehat{\boldsymbol{\rho}} := \frac{1}{I}\sum_{n=0}^{I-1} \sigma^n(\boldsymbol{\rho})$ and

[12] This is a social choice analog of the concept of ‘preference for flexibility’ developed by Kreps (1979) and others in the setting of individual choice.

[13] If $i \neq j$, then generally, $\mathcal{F}_i \neq \mathcal{F}_j$, because people do not have complete freedom to modify their own psychology. There will also be other constaints on the sorts of freedom allocations the planner can offer (e.g. resource constraints). Finally, this model unrealistically assumes that each $i \in \mathcal{I}$ can choose a point in $\mathcal{F}_i$ independent of the choices made by other $j \in \mathcal{I}$. In reality, the agents might interact (e.g. trade) and their choices will be interdependent, resulting in an I-player game.

$\widehat{\boldsymbol{\rho}}' := \dfrac{1}{I}\displaystyle\sum_{n=0}^{I-1} \sigma^n(\boldsymbol{\rho}')$. (Here, $\sigma^2 = \sigma \circ \sigma$, $\sigma^3 = \sigma \circ \sigma \circ \sigma$, etc.) Then

$$\boldsymbol{\rho} \quad = \quad \frac{1}{I}\sum_{n=0}^{I-1} \boldsymbol{\rho} \quad \overset{\triangle}{\equiv} \quad \frac{1}{I}\sum_{n=0}^{I-1} \sigma^n(\boldsymbol{\rho}) \quad = \quad \widehat{\boldsymbol{\rho}}. \tag{15}$$

Here "$\overset{\triangle}{\equiv}$" is by I-fold application of axiom (Lin), because $\boldsymbol{\rho} \overset{\triangle}{\equiv} \sigma^n(\boldsymbol{\rho})$ for all $n \in \mathbb{N}$, by axiom (Anon). By a similar argument, $\boldsymbol{\rho}' \overset{\triangle}{\equiv} \widehat{\boldsymbol{\rho}}'$. Meanwhile, for all $i \in \mathcal{I}$, we have

$$\hat{\rho}_i \quad = \quad \overline{\rho} \qquad \text{and} \qquad \hat{\rho}'_i \quad = \quad \overline{\rho}', \tag{16}$$

where $\overline{\rho}$ and $\overline{\rho}'$ are the per capita average lotteries of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$, as defined in eqn.(5). Thus,

$$\begin{aligned}
&\left(\boldsymbol{\rho} \underset{u}{\trianglerighteq} \boldsymbol{\rho}'\right) \underset{(*)}{\Longleftrightarrow} \left(\overline{\rho} \succeq \overline{\rho}'\right) \underset{(\dagger)}{\Longleftrightarrow} \left(\hat{\rho}_i \succeq \hat{\rho}'_i \text{ for all } i \in \mathcal{I}\right) \\
&\qquad \underset{(\ddagger)}{\Longrightarrow} \left(\widehat{\boldsymbol{\rho}} \trianglerighteq \widehat{\boldsymbol{\rho}}'\right) \underset{(\diamond)}{\Longleftrightarrow} \left(\boldsymbol{\rho} \trianglerighteq \boldsymbol{\rho}'\right). \\
\text{Likewise, } &\left(\boldsymbol{\rho} \underset{u}{\triangleright} \boldsymbol{\rho}'\right) \underset{(*)}{\Longleftrightarrow} \left(\overline{\rho} \succ \overline{\rho}'\right) \underset{(\dagger)}{\Longleftrightarrow} \left(\hat{\rho}_i \succ \hat{\rho}'_i \text{ for all } i \in \mathcal{I}\right) \\
&\qquad \underset{(\ddagger)}{\Longrightarrow} \left(\widehat{\boldsymbol{\rho}} \triangleright \widehat{\boldsymbol{\rho}}'\right) \underset{(\diamond)}{\Longleftrightarrow} \left(\boldsymbol{\rho} \triangleright \boldsymbol{\rho}'\right).
\end{aligned}$$

Here, $(*)$ is by defining formula (6), $(\dagger)$ is by eqn.(16), $(\ddagger)$ is by axiom (Par), and $(\diamond)$ is by eqn.(15) and the transitivity of $(\trianglerighteq)$. □

Proof of Proposition 2.2. Let $\overline{\rho}$ and $\overline{\rho}'$ be the per capita average lotteries of $\boldsymbol{\rho}$ and $\boldsymbol{\rho}'$, as defined in eqn.(5). For any measurable $u : \mathcal{X} \longrightarrow \mathbb{R}$, we have

$$u^*(\overline{\rho}) \quad = \quad \frac{1}{|\mathcal{I}|}\sum_{i \in \mathcal{I}} u^*(\rho_i) \qquad \text{and} \qquad u^*(\overline{\rho}') \quad = \quad \frac{1}{|\mathcal{I}|}\sum_{i \in \mathcal{I}} u^*(\rho'_i), \tag{17}$$

because the function $u^* : \mathfrak{P} \longrightarrow \mathbb{R}$ is linear. Thus,

$$\begin{aligned}
&\left(\boldsymbol{\rho} \underset{u}{\trianglerighteq} \boldsymbol{\rho}'\right) \underset{(*)}{\Longleftrightarrow} \left(\overline{\rho} \succeq \overline{\rho}'\right) \underset{(\dagger)}{\Longrightarrow} \left(u^*(\overline{\rho}) \geq u^*(\overline{\rho}'), \text{ for all } u \in \mathcal{U}(\succeq)\right) \\
&\qquad \underset{(\diamond)}{\Longleftrightarrow} \left(\sum_{i \in \mathcal{I}} u^*(\rho_i) \geq \sum_{i \in \mathcal{I}} u^*(\rho'_i), \text{ for all } u \in \mathcal{U}(\succeq)\right),
\end{aligned}$$

as desired. Here, $(*)$ is by defining formula (6); $(\dagger)$ is by statement (7), and $(\diamond)$ is by eqn.(17). This proves (a)

If $(\succeq)$ admits a multiutility representation (2), then clearly $\mathcal{V} \subseteq \mathcal{U}(\succeq)$; furthermore, (2) remains true if we substitute $\mathcal{V} := \mathcal{U}(\succeq)$. Thus, the implication "$\underset{(\dagger)}{\Longrightarrow}$" changes to "$\Longleftrightarrow$". This proves (b). □

Proof of Proposition 3.1. For any $\mathbf{w} \in \mathcal{W}$, we have

$$\begin{aligned}\sum_{i\in\mathcal{I}} w_i \cdot u_0(\psi_i, \rho_i') - \sum_{i\in\mathcal{I}} w_i \cdot u_0(\psi_i, \rho_i) \quad &= \quad \Big(w_1 r_1' + w_2 r_2'\Big) - \Big(w_1 r_1 + w_2 r_2\Big) \\ = \quad w_1 \cdot (r_1' - r_1) + w_2 \cdot (r_2' - r_2) \quad &= \quad w_2 \cdot \left(\left(\frac{w_1}{w_2}\right) \cdot (r_1' - r_1) + (r_2' - r_2) \right).\end{aligned}$$

Thus, statement (10) becomes

$$\begin{aligned}\Big(\boldsymbol{\rho}' \underset{u,\boldsymbol{\psi}}{\trianglerighteq} \boldsymbol{\rho}\Big) \quad &\Longleftrightarrow \quad \Big(\Big(\tfrac{w_1}{w_2}\Big) \cdot (r_1' - r_1) + (r_2' - r_2) \geq 0, \text{ for all } \mathbf{w} \in \mathcal{W}\Big) \\ &\Longleftrightarrow \quad \begin{pmatrix} \text{either (A)} \ \ (r_1' - r_1) \geq 0 \text{ and } (r_2' - r_2) \geq 0; \text{ or} \\ \text{(B)} \ \ (r_1' - r_1) \geq 0 \geq (r_2' - r_2) \text{ and } \underline{A} \cdot (r_1' - r_1) \geq (r_2 - r_2'); \text{ or} \\ \text{(C)} \ \ (r_2' - r_2) \geq 0 \geq (r_1' - r_1) \text{ and } (r_2' - r_2) \geq \overline{A} \cdot (r_1 - r_1'). \end{pmatrix} \quad (18)\end{aligned}$$

If $S := \frac{r_2' - r_2}{r_1' - r_1}$, then condition (B) in statement (18) is equivalent to $r_1' \geq r_1$, $r_2' \leq r_2$ and $S \geq -\underline{A}$. Meanwhile, condition (C) is equivalent to $r_1' \leq r_1$, $r_2' \geq r_2$, and $S \leq -\overline{A}$.

Thus, the right side of statement (18) is equivalent to the right side of statement (11). □

Proof of Lemma 5.1. Fix $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$. Define the function $\mu_{\mathbf{x}} : \mathfrak{S} \longrightarrow \mathbb{R}$ by $\mu_{\mathbf{x}}[\mathcal{S}] := U_{\mathcal{S}}(\mathbf{x}) \cdot \pi[\mathcal{S}]$, for all $\mathcal{S} \in \mathfrak{S}$. Axiom (Bayes) says that $\mu_{\mathbf{x}}$ is countably additive (i.e. $\mu_{\mathbf{x}}[\bigsqcup_{n=1}^{\infty} \mathcal{S}_n] = \sum_{n=1}^{\infty} \mu_{\mathbf{x}}[\mathcal{S}_n]$); hence it is a sigma-finite signed measure (because π is a probability measure and $|U_{\mathcal{S}}(\mathbf{x})| < \infty$ for all $\mathcal{S} \in \mathfrak{S}$). Clearly, $\mu_{\mathbf{x}}$ is absolutely continuous relative to π [i.e. $(\pi[\mathcal{S}] = 0) \Longrightarrow (\mu_{\mathbf{x}}[\mathcal{S}] = 0)$]. Thus, the Radon-Nikodym Theorem (Conway, 1990, Thm.C.7, p.380) says there is a $\mathfrak{S}$-measurable function $f_{\mathbf{x}} : \Omega \longrightarrow \mathbb{R}$ such that $\mu_{\mathbf{x}}[\mathcal{S}] = \int_{\mathcal{S}} f_{\mathbf{x}}(\omega) \, \mathrm{d}\pi[\omega]$ for all $\mathcal{S} \in \mathfrak{S}$. Now define $U : \mathcal{X}^{\mathcal{I}} \times \Omega \longrightarrow \mathbb{R}$ by $U_\omega(\mathbf{x}) := f_{\mathbf{x}}(\omega)$, for all $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$ and $\omega \in \Omega$. Then for any $\mathbf{x} \in \mathcal{X}^{\mathcal{I}}$ and $\mathcal{S} \in \mathfrak{S}$, we have

$$U_{\mathcal{S}}(\mathbf{x}) \quad = \quad \frac{\mu_{\mathbf{x}}[\mathcal{S}]}{\pi[\mathcal{S}]} \quad = \quad \frac{1}{\pi[\mathcal{S}]} \int_{\mathcal{S}} f_{\mathbf{x}}(\omega) \, \mathrm{d}\pi[\omega] \quad = \quad \frac{1}{\pi(\mathcal{S})} \int_{\mathcal{S}} U_\omega(\mathbf{x}) \, \mathrm{d}\pi[\omega],$$

which yields eqn.(12). □

Proof of Theorem 5.2. For any $\omega \in \Omega$, if the vNM preference relation $(\underset{\omega}{\trianglerighteq})$ satisfies (Par), then Harsanyi's SAT implies that $(\underset{\omega}{\trianglerighteq})$ can be represented as maximizing the expected value of a vNM utility function $\widetilde{U}_\omega : \mathcal{X}^{\mathcal{I}} \longrightarrow \mathbb{R}$ of the form:

$$\widetilde{U}_\omega(\mathbf{x}) \quad := \quad \sum_{i\in\mathcal{I}} c_i^\omega \cdot u_i^\omega(\mathbf{x}) \quad = \quad \sum_{i\in\mathcal{I}} c_i^\omega \cdot u(x_i, \omega), \quad \text{for all } \mathbf{x} \in \mathcal{X}^{\mathcal{I}},$$

for some nonnegative constants $\{c_i^\omega\}_{i\in\mathcal{I}} \subset \mathbb{R}_+$. Axiom (Nonindiff) says at least one these constants is nonzero, while (Anon) implies that they must all be equal; hence we can assume without loss of generality that $c_i^\omega = 1$ for all $i \in \mathcal{I}$, so that $\widetilde{U}_\omega(\mathbf{x}) = \sum_{i\in\mathcal{I}} u(x_i,\omega)$ for all $\mathbf{x} \in \mathcal{X}^\mathcal{I}$ and $\omega \in \Omega$.

Now, U_ω and $\widetilde{U}_\omega$ represent the same vNM preference relation ($\underset{\omega}{\trianglerighteq}$), so there exist constants $a(\omega) > 0$ and $b(\omega) \in \mathbb{R}$ such that $U_\omega = a(\omega)\,\widetilde{U}_\omega + b(\omega)$. That is:

$$U_\omega(\mathbf{x}) \;=\; b(\omega) + a(\omega)\sum_{i\in\mathcal{I}} u(x_i,\omega), \quad \text{for all } \mathbf{x} \in \mathcal{X}^\mathcal{I} \text{ and } \omega \in \Omega.$$

Axiom (Welf) then implies that $a(\omega_1) = a(\omega_2)$ and $b(\omega_1) = b(\omega_2)$ for all $\omega_1, \omega_2 \in \Omega$. Thus, there are constants $a > 0$ and $b \in \mathbb{R}$ such that

$$U_\omega(\mathbf{x}) \;=\; b + a\sum_{i\in\mathcal{I}} u(x_i,\omega), \quad \text{for all } \mathbf{x} \in \mathcal{X}^\mathcal{I} \text{ and } \omega \in \Omega. \tag{19}$$

Substituting (19) into (12), we get:

$$\begin{aligned} U_\mathcal{S}(\mathbf{x}) \quad &= \quad \frac{1}{\pi(\mathcal{S})}\int_\mathcal{S} \left(b + a\sum_{i\in\mathcal{I}} u(x_i,\omega)\right) \mathrm{d}\pi[\omega] \\ &= \; b + a\sum_{i\in\mathcal{I}} \frac{1}{\pi(\mathcal{S})}\int_\mathcal{S} u(x_i,\omega)\,\mathrm{d}\pi[\omega] \quad = \quad b \;+\; a\sum_{i\in\mathcal{I}} \overline{u}_\mathcal{S}(x_i), \end{aligned} \tag{20}$$

and where $\overline{u}_\mathcal{S}$ is defined as in eqn.(13). But clearly the vNM utility function $U_\mathcal{S}$ in eqn.(20) is equivalent to the vNM utility function $\overline{U}_\mathcal{S}$ in eqn.(14). □

References

Anand, P., Pattanaik, P., Puppe, C. (Eds.), 2009. Rational and Social Choice: An overview of new foundations and applications. Oxford UP, Oxford, UK.

Basu, K., 1980. Revealed preference of government. Cambridge UP, Cambridge, UK.

Baucells, M., Shapley, L. S., 1998. Multiperson utility. Tech. Rep. Working paper No. 779, Department of Economics, UCLA, available at `http://webprofesores.iese.edu/mbaucells`.

Baucells, M., Shapley, L. S., 2006. Multiperson utility: the linearly independent case. (preprint)Available at `http://webprofesores.iese.edu/mbaucells`.

Baucells, M., Shapley, L. S., 2008. Multiperson utility. Games Econom. Behav. 62 (2), 329–347.

Blackorby, C., 1975. Degrees of cardinality and aggregate partial orderings. Econometrica 43 (5–6), 845–852.

Butts, R., Hintikka, J. (Eds.), 1977. Foundational problems in the social sciences. D. Reidel, Dordrecht.

Conway, J. B., 1990. A course in functional analysis, 2nd Edition. Vol. 96 of Graduate Texts in Mathematics. Springer-Verlag, New York.

Dowding, K., van Hees, M., 2009. Freedom of choice. In: Anand et al. (2009), Ch. 15, pp. 374–392.

Dubra, J., Maccheroni, F., Ok, E. A., 2004. Expected utility theory without the completeness axiom. J. Econom. Theory 115 (1), 118–133.

Elster, J., Roemer, J. E. (Eds.), 1991. Interpersonal comparisons of well-being. Cambridge UP, Cambridge, UK.

Evren, Ö., 2008. On the existence of expected multi-utility representations. Econom. Theory 35 (3), 575–592.

Fine, B., 1975. A note on "Interpersonal aggregation and partial comparability". Econometrica 43, 169–172.

Harsanyi, J., 1953. Cardinal utility in welfare economics and in the theory of risk-taking. J. Political Economy 61 (434-435).

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

Harsanyi, J. C., 1975. Nonlinear social welfare functions: Do welfare economists have a special exemption from Bayesian rationality? Theory and Decision 6, 311–332.

Harsanyi, J. C., 1976. Essays on Ethics, Social Behaviour, and Scientific Explanation. D. Reidel, Dordrecht.

Harsanyi, J. C., 1977. Nonlinear social welfare functions: A rejoinder to Professor Sen. In: Butts and Hintikka (1977).

Kreps, D. M., 1979. A representation theorem for "preference for flexibility". Econometrica 47 (3), 565–577.

Mill, J. S., 1859. On Liberty. Penguin, London, (Reprinted 1985).

Mongin, P., 1994. Harsanyi's aggregation theorem: multi-profile version and unsettled questions. Soc. Choice Welf. 11 (4), 331–354.

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

Pivato, M., 2010a. Approximate interpersonal comparisons of well-being. (preprint).

Pivato, M., 2010b. Aggregation of incomplete ordinal preferences with approximate interpersonal comparisons. (preprint).

Pivato, M., 2010c. Quasi-utilitarian social evaluation with approximate interpersonal comparison of welfare gains. (preprint).

Roemer, J. E., 1996. Theories of distributive justice. Harvard UP, Cambridge, MA.

Sen, A. K., 1970a. Collective choice and social welfare. Holden Day, San Francisco.

Sen, A. K., 1970b. Interpersonal aggregation and partial comparability. Econometrica 38, 393–409.

Sen, A. K., 1972. Interpersonal comparison and partial comparability: A correction. Econometrica 40 (5), 959.

Sen, A. K., 1976. Welfare inequalities and Rawlsian axiomatics. Theory and Decision 7, 243–262.

Sen, A. K., 1977. Non-linear social welfare functions: A reply to Professor Harsanyi. In: Butts and Hintikka (1977).

Weymark, J. A., 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In: Elster and Roemer (1991), pp. 255–320.