The Utilitarian Relevance of the Aggregation Theorem

Marc Fleurbaey* and Philippe Mongin†

August 19, 2015

Abstract

Harsanyi invested his Aggregation Theorem and Impartial Observer Theorem with utilitarian sense, but Sen described them as "representation theorems" with little ethical import. This critical view has never been subjected to full analytical scrutinity. The formal argument we provide here supports the utilitarian relevance of the Aggregation Theorem. Following a hint made by Sen himself, we posit an exogeneous utilitarian ordering that evaluates riskless options by the sum of individual utilities and we show that *any* social observer who obeys the conditions of the Aggregation Theorem evaluates social states in terms of a weighted variant of this utilitarian sum. The key step in the argument is to assume that the utilitarian ordering over riskless options can be extended into an ordering over lotteries that also obeys the conditions of the Aggregation Theorem.

Keywords: Utilitarianism, Aggregation Theorem, Impartial Observer Theorem, cardinal utility, VNM utility, Harsanyi, Sen.

JEL Classification: D63, D71, D81.

*Princeton University. Mail: mfleurba@princeton.edu.

†CNRS & HEC Paris. Mail: mongin@greg-hec.com.

1 Introduction

First acclaimed as pathbreaking contributions to social ethics, Harsanyi's Impartial Observer and Aggregation Theorems (1953, 1955) were later criticized by Sen (1977, 1986) for being hardly relevant to the field. Using ethically loaded postulates, such as the so-called Acceptance principle (in the first theorem) or the standard Pareto principle (in the second theorem), along with a von Neumann-Morgenstern (VNM) apparatus of expected utility for both the individuals and the social observer (in either theorem), Harsanyi shows that the observer's VNM utility function equals a weighted sum of individual VNM utility functions, and then claims to have grounded utilitarianism in a new way. Not questioning the formal validity of the theorems, Sen objects to their interpretation. For him, Harsanyi's first theorem is "about utilitarianism in a rather limited sense", and his second theorem, while more informative, remains "primarily a representation theorem" (1986, p. 1123-4). To summarize bluntly, he discards the first theorem and salvages only the mathematical achievement in the second; neither has to do with utilitarianism properly (see also Sen, 1974 and 1977).

Economists have always been divided on the merits of utilitarianism as a redistributive doctrine, and those who would have rejected it anyhow on principled reasons tended to endorse Sen's critique without further examination. At the same time, those who were working on utilitarian lines, whether out of ethical conviction or for mathematical convenience reasons, did not make much effort at responding to it. Sen's critique became better known after Weymark (1991) surveyed the "Harsanyi-Sen debate", but this account of the two positions, although seemingly favouring Sen's, did not bring the issue to a close. More inclined towards Harsanyi's position, Mongin and d'Aspremont (1998), and more explicitly Mongin (2002), proposed buttressing it by a cardinality argument, but this sounded like too direct an approach. All in all, Sen's critique has remained underdiscussed analytically, and the point of the present paper is to confront it with a completely formal response. Limiting ourselves to the Aggregation Theorem, we will enrich its assumptions to the point where its conclusion can be given utilitarian sense, so that Harsanyi's attempt at grounding utilitarianism on VNM theory will appear to be incomplete, but not flawed. It is essential to the interpretation of our result that it be properly located in the context of the "Harsanyi-Sen debate". That is, if it succeeds at all, this is by taking care of Sen's critique, not in grounding utilitarianism *per se*.

Like Harsanyi in the Aggregation Theorem, we suppose that individual preferences satisfy the VNM axioms on risky alternatives, i.e., lotteries. Unlike him, we assume that there already exists a social observer who has formed social preferences on sure alternatives *according to the sum rule of classical utilitarianism*. Positing such an exogeneous utilitarian benchmark seems to us to be the only way of ascertaining whether or not the Aggregation Theorem is relevant to utilitarianism, and the version of the doctrine we need to avoid any circularity can only be that which prevailed before Harsanyi's reform, i.e., classical utilitarianism, as in Bentham, Jevons and Edgeworth, which was unconcerned with risk and uncertainty. Sen precisely points to this direction when he complains thus about the theorem: "There is no independent concept of individual utilities of which social welfare is shown to be the sum, and *as such the result asserts a good deal less than classical utilitarianism does.*" (1986, p. 1123, our emphasis).

Once the classical utilitarian observer is made explicit, it is consistent with Harsanyi's purpose to assume that this observer's representation can be extended to lotteries so as to satisfy the VNM axioms, and that the VNM extension so obtained satisfies the Pareto principle with respect to the individuals' VNM preferences. These conditions are precisely those which the Aggregation Theorem imposes on any social preferences, so that it would be strange if Harsanyi did not regard them as being applicable to utilitarianism. With this assumption, we are able to show that the utility functions that appear in the classical utilitarian sum must also be VNM representations of the individuals' VNM preferences. Once this is proved, it is a short step, using the initial theorem as a lemma, to derive that *any* social observer (i.e., without any prior commitment to utilitarianism) who satisfies the VNM and Pareto conditions must rely on a weighted sum of the utility functions just said, and thus comes close to obeying the utilitarian benchmark. Admittedly, we do not show that the social observer give the individuals equal weights as this benchmark does. However, *weighted utilitarianism*, as d'Aspremont and Gevers (2002) label it, already goes a long way towards utilitarianism proper.[1]

This new result is presented here in several variants. The salient one, Theorem 1, is concerned with sure alternatives and uses the extension to lotteries only as a convenient demonstrative tool. It also specializes the sure alternatives, taking them to be allocations of commodities in a standard economic framework. However, since social choice theory and - under its influence - today's social ethics often envisage abstract domains, we add Theorem 2, in which the sure alternatives are unspecific and lotteries become the primary objects of social evaluation. Harsanyi's (1955, 1977) formal exposition actually privileges this approach. We doubt that classical utilitarians would be at ease with it, but some 20th century utilitarian followers of Harsanyi clearly are (e.g., Hammond, 1982, 1996; more review in Mongin and d'Aspremont, 1998). Both results assume for simplicity that the utilitarian benchmark is defined on a full-dimensional individual utility set. This includes a diversity of preference assumption that lacks in generality, so the appendix

[1] Weighted utilitarianism has been explored in connection with Harsanyi's other theorem by Mongin (2001), and Grant, Kajii, Polak and Safra (2010). Weighted sums have also been used to evaluate social states in generalized forms of utilitarianism, where the weights obey specific normalizations and do not only depend on the names of the individuals; see Karni (1998), Dhillon and Mertens (1999), and Segal (2000).

states the more technical Theorem 3, which generalizes Theorem 1 by requiring a weaker dimensionality assumption.

In sum, we argue that Harsanyi correctly felt that VNM theory could support utilitarianism, but the full argument, which he did not provide, involves *defining utilitarianism exogeneously.* Sen's critique allusively suggests this solution, which we carry out axiomatically.

Section 2 develops Sen's objections in some technical detail. Section 3 sets up the formal framework with the economic assumptions needed for Theorem 1. Section 4 states this result and its lottery variant, Theorem 2. Section 5 briefly discusses how the weighted utilitarian conclusion can be escaped. The appendix explains the mathematical tools and states the more general Theorem 3.

2 Just "representations theorems"?

Sen objects as follows against the use of VNM utility functions for utilitarian purposes:
"The (VNM) values are of obvious importance for protecting individual or social choice under uncertainty, but there is no obligation to talk about (VNM) values only whenever one is talking about individual welfare" (1977, p. 277).
This is but an expression of doubt, but later Sen argues more strongly:
"(Harsanyi's theorem) does not yield utilitarianism as such — only linearity... I feel sad that Harsanyi should continue to believe that his contribution lay in providing an axiomatic justification of utilitarianism with real content." (1977, p. 300).

Here it is again with some detail (this comment was intended for the Impartial Observer Theorem, but if it applies there, it also does the Aggregation Theorem):
"This is a theorem about utilitarianism in a rather limited sense in that the VNM cardinal scaling of utilities covers both (the social and individual utilities) within *one* integrated

system of numbering, and the individual utility numbers do not have independent meaning other than the value associated with each "prize", in predicting choices over lotteries. There is no *independent* concept of individual utilities of which social welfare is shown to be the sum, and as such the results asserts a good deal less than classical utilitarianism does" (1986, p. 1123).

In other words, VNM theory provides a cardinalization of utility, both individual and social, which is relevant to preference under uncertainty, but *prima facie* useless for the evaluation of welfare, which is the utilitarian's genuine concern. Of course, classical utilitarianism also presupposes that individual utility functions are cardinal, but there is no reason to conclude that these functions belong to the class of cardinal functions that VNM theory makes available on a completely separate axiomatic basis.

There is another claim in the passage, but it is more subdued, and a comment by Weymark brings it out well:

"No significance should be attached to the linearity or non-linearity of the social welfare function, as the curvature of this function depends solely on whether or not VNM representations are used, and the use of such representations is arbitrary" (1991, p. 315). That is to say, VNM theory deals with preferences taken in an ordinal sense, and it is only for convenience that one usually represents them by means of an expected utility. It is theoretically permissible to replace an individual's VNM utility functions by any non-affine increasing transform, and if one would do so, the social observer's function would not be linear anymore, but *only additively separable*, in terms of individual utility numbers. That is, it would read as $v = \varphi \circ \left(\sum_i \varphi_i^{-1} \circ v_i\right)$, where v, v_i are the chosen increasing transforms of the social and individual VNM utility functions, respectively, and φ, φ_i are the corresponding transformation mappings. This line of criticism offers another interpretation of the claim that Harsanyi proved no more than representation theorems (see Weymark, 1991, p. 305).

Some definitions and notation, which anticipate on the framework of the next section, will help formalize the two objections. If $\succsim$ is a preference relation on a set S and w a real-valued function on S, we say, as usual, that w *represents* $\succsim$ on S, or that w is a *utility function for* $\succsim$ *on* S, iff for all $x, y \in S$,

$$x \succsim y \Longleftrightarrow w(x) \geq w(y).$$

We are in particular concerned with preference relations on a lottery set L. There is an underlying outcome set X, and by the familiar identification of outcomes with sure lotteries, $X \subset L$. If a preference $\succsim$ on L satisfies the VNM axioms, the VNM representation theorem guarantees that there exists a utility function u for $\succsim$ on X with the property that the expectation Eu is a utility function for $\succsim$ on L. Both u and Eu will be called *VNM utility functions*, a standard practice.

The VNM representation theorem also teaches that the set of those u' for which Eu' is a utility function for $\succsim$ on L is exactly the set of *positive affine transforms* (PAT) of the given u, i.e., the set of all $\alpha u + \beta$ with $\alpha > 0$ and $\beta \in \mathbb{R}$. Clearly,

$$\mathcal{U} = \{\varphi \circ u \mid \varphi \text{ positive affine transformation}\} \subsetneq \mathcal{F} = \{\varphi \circ u \mid \varphi \text{ increasing}\}.$$

By the same token, the set of utility functions for $\succsim$ on L that take the form Eu' is

$$\mathcal{U}' = \{\varphi \circ Eu \mid \varphi \text{ positive affine transformation}\} \subsetneq \mathcal{F}' = \{\varphi \circ Eu \mid \varphi \text{ increasing}\}.$$

In all existing versions (see Fishburn's 1982 review), the VNM axioms define an ordinal preference concept, and thus do not by themselves justify selecting a representation in $\mathcal{U}$ or $\mathcal{U}'$ rather than $\mathcal{F}$ or $\mathcal{F}'$. This basic point has led to repeated warnings in decision theory throughout the years. It is perhaps more likely to escape attention in the present collective context, whence the usefulness of *Weymark's reminder*, as we may call it by referring to the above quote.

As Sen's baseline argument involves comparing the Aggregation Theorem with classical utilitarianism, we now introduce this formally. A social observer obeying this doctrine must endow the individuals $i = 1, ..., n$ with welfare indexes u_i^* on X that meaningfully add up, i.e., the imputed u_i^* must be *both cardinally measurable and comparable.* Accordingly, this observer's rule is represented by any element of the set:

$$\mathcal{C} = \left\{ \sum_{i=1}^{n} \varphi_i \circ u_i^* \mid \varphi_i \text{ common positive affine transformations (same } \alpha) \right\}.$$

(For a similar treatment of utilitarianism in social choice theory, see Sen, 1986, and d'Aspremont and Gevers, 2002.)

With the definitions just given, it is impossible to conclude that $u_i^* \in \mathcal{U}_i$, the set of i's VNM representations on X, or that $\sum_i u_i \in \mathcal{C}$ when $(u_1, ..., u_n)$ is a vector of such representations. Since VNM utility values do not have to measure utilitarian welfare, if Harsanyi proves that the social utility function is a sum of individual VNM utility values, this says nothing for utilitarianism. The gap remains even if one makes the reasonable assumption that *u_i^* is a utility function for $\succsim_i$ on X*, for this is not sufficient to deliver cardinal equivalence with u_i; that is, one only gets $u_i^* \in \mathcal{F}_i$, the set of i's general representations on X; and similarly, the assumption does not make $\sum_i u_i$ (or any weighted variant of this sum) a member of the set $\mathcal{C}$. This formal point is the most important one we read in Sen's quotes, so we will refer to it as to *Sen's point.*

In sum, two related, but distinct problems stand in the way of Harsanyi's utilitarian interpretation of his results. It has been noted in the literature that these problems would vanish if one could ground the utility functions u_i and u_i^* on a common basis of cardinal preference, assuming that cardinal preference can formalized properly. Indeed, if the utilitarian cardinalization rests on a genuine preference basis and the VNM cardinalization can be reduced to that basis, this cardinalization escapes irrelevance (*pace* Weymark) and it coincides with the welfare interpretation needed for utilitarianism (*pace* Sen).

Technically, this approach relies on defining cardinal preference to be a relation on *pairs* of sure alternatives, i.e., $(x, y) \succsim_i^* (z, w)$, and assuming that such comparisons of preference differences are made in coherence with u_i^*. Then a connecting axiom is introduced to ensure that the VNM cardinalization u_i is also coherent with $\succsim_i^*$. This axiom will imply that for all $x, y, w, z \in X$,

$$(1/2)u_i(x) + (1/2)u_i(y) \geq (1/2)u_i(z) + (1/2)u_i(w) \text{ iff}$$
$$u_i^*(x) - u_i^*(z) \geq u_i^*(w) - u_i^*(y). \quad (*)$$

Mongin (2002) develops this strategy, which was already suggested, but without axiomatic detail, by Weymark (1991, p. 308) and Mongin and d'Aspremont (1998, p. 435).[2]

While this approach makes Harsanyi's position logically consistent, it is question-begging as an argument for his position, because the connecting axiom is too direct a way of getting the problematic equivalence $(*)$. Moreover, for most economists, preference is an ordinal concept by definition, as it is tightly connected in their eyes with choice, which they view as being itself an exclusively ordinal concept. There is no evidence in Harsanyi that he meant to depart from this tenet.

In this paper, we dispense with $(*)$ or its axiomatic counterpart, and we use an indirect argument instead. As will be seen, it also takes care of both Sen's point and Weymark's reminder.

3 The framework

We consider a set $X \subseteq \mathbb{R}^{mn}$, the elements of which are potentially feasible allocations of m commodities to the $n \geq 2$ individuals. Departing from basic consumer theory, which takes

[2]Harvey (1999) goes to the extreme of the present reinterpretation in dispensing with VNM theory altogether and basing a revised form of the Aggregation Theorem on cardinal preferences exclusively.

$X = \mathbb{R}_+^{mn}$, we do not require X to be a Cartesian product.[3] We require connectedness, which is much less restrictive than convexity.

Assumption 1: X is a path-connected subset of $\mathbb{R}^{mn}$.

The VNM apparatus can now be introduced formally. Concerning the social observer, all axiomatizations of VNM theory in expectational form work; take any one in Fishburn (1982). However, concerning the individuals, we need *continuous* VNM utility functions, a property which these systems do not normally provide, so we turn to Grandmont's (1972), which was set up for that purpose.

Define $\mathcal{B}(X)$ to be the set of Borelian sets of X, i.e., the σ-algebra generated by the open subsets of X, and take the set $L = \Delta(X)$ of all probability measures on the measurable space $(X, \mathcal{B}(X))$. By a standard assumption, this set is endowed with the topology of weak convergence, which makes it a metric space. Now, a *continuous VNM preference relation* $\succsim$ on L is by definition an ordering that satisfies two conditions (as usual, we write $p \sim q$ and $p \succ q$ for the symmetric and asymmetric parts of $\succsim$).

(*Continuity*) For all $p \in L$, the sets

$$\{p' \in L : p' \succsim p\} \text{ and } \{p' \in L : p \succsim p'\}$$

are closed in L.

(*Independence*) For all $p, q, r \in L$ and all $\lambda \in (0, 1]$, $p \sim q$ iff $\lambda p + (1-\lambda)r \sim \lambda q + (1-\lambda)r$.

Grandmont's Theorems 2 and 3 (1972, p. 48-49) apply to X and L as special cases. They ensure that there is a continuous and bounded utility function $u(x)$ for $\succsim$ on X such that the expectation $v(p) = Eu(p)$ is a utility function for $\succsim$ on L. It is also the case that v is continuous, and that the set of u' such that $\succsim$ is represented by Eu' is exactly the set

[3] A Cartesian product is ill-suited when the list of commodities includes public goods or services exchanged between individuals, so that individual consumptions exhibit technical dependencies. Even in the case of private goods, it may be inappropriate if X takes the availability of resources into account.

of PAT of u.

Assumption 2: Each $i = 1, ..., n$ is endowed with a continuous VNM preference relation $\succsim_i$ on L.

Now to the exogenous utilitarian social preferences. We fix a vector of functions on X, $U^* = (u_1^*, ..., u_n^*)$, to represent the cardinally measurable and comparable utility functions that a utilitarian social observer would associate with the individuals, and accordingly, we formally define the *classical utilitarian social preference ordering* $\succsim^*$ on X by

$$x \succsim^* y \text{ iff } \sum_{i=1}^{n} u_i^*(x) \geq \sum_{i=1}^{n} u_i^*(y).$$

We need two technical conditions on U^*.

Assumption 3: For each $i = 1, ..., n$, u_i^* is continuous on X.

Assumption 4: The image set $U^*(X)$ has a nonempty connected interior $U^*(X)^\circ$ in $\mathbb{R}^n$ such that $U^*(X) \subseteq \overline{U^*(X)^\circ}$, i.e., this image set is included in the closure of its interior.

It would be equivalent to impose these assumptions on any collection of PAT $\varphi_i \circ u_i^*$ having the same α for all i, so that they make utilitarian sense. Assumption 3 is mild and standard, but Assumption 4 less so. In one respect, it simply complements Assumptions 1 and 3, which entail that $U^*(X)$ is connected, by a common regularity assumption. In another respect, it requires $U^*(X)$ to have full affine dimension n.[4] This is not demanding under standard microeconomic conditions. If there are private consumption goods, each individual is concerned only with how much he consumes, and free disposal is allowed, then throwing away someone's allocation will change his utility without affecting the others'. However, if there are only pure public goods, Assumption 4 requires sufficient diversity of individual preferences, and for instance, no two individuals can be alike in the

[4] The affine dimension of the set $U^*(X)$ is the linear dimension of the translated set $U^*(X) - x_0$ for any choice of $x_0 \in U^*(X)$. A full affine dimension excludes that any of the u_i is constant.

utilitarian observer's eyes.

Finally, we relate the utilitarian and VNM halves of the construction to each other.

Assumption 5: For each $i = 1, ..., n$, u_i^* is a utility function for $\succsim_i$ on X.

Crucially, this imposes no more than *ordinal* equivalence on u_i^* and u_i, whereas *cardinal* equivalence may not hold between them; if we assumed the latter right away, we would in essence fall back on the equivalence $(*)$ of last section.

That u_i^* and u_i are ordinally equivalent means that $u_i = f_i \circ u_i^*$ for some increasing function f_i on $u_i^*(X)$. Actually, in view of the previous assumptions and the following lemma, each f_i must be continuous.

Lemma 1 *Suppose that g and h are continuous real-valued functions defined on some path-connected set X and f is an increasing real-valued function such that $h = f \circ g$; then, f is also continuous.*

(We skip the proof of this fact. A stronger form, which does not require f to be increasing, can actually be established; see the working paper version of this article.)

We complete the groundwork for the next section by stating a functional equation theorem due to Rado and Baker (1987, Theorem 1; see also their Corollary 1). These authors actually formulate it for $n = 2$ but it extends to the general case, as the appendix shows.[5] Let T be an open connected subset of $\mathbb{R}^n$, $n \geq 2$. Define $T_+ = \left\{ \sum_{i=1}^{n} z_i \mid (z_1, ..., z_n) \in T \right\}$ and $T_i = \{z_i \mid (z_1, ..., z_n) \in T\}$.

Lemma 2 *Suppose that $f : T_+ \to \mathbb{R}$ and $f_i : T_i \to \mathbb{R}$, $i = 1, ..., n$ satisfy the equation*

$$f(\sum_{i=1}^{n} z_i) = \sum_{i=1}^{n} f_i(z_i)$$

[5] Rado and Baker's results are also reported in Aczel (1987, p. 80). When restating Harsanyi's Aggregation Theorem for state-contingent alternatives, Blackorby, Donaldson and Weymark (1999) use them for the same reason as we do here, i.e., they do not impose a Cartesian product domain on f and the f_i.

for all $(z_1, ..., z_n) \in T$*. Suppose that one of the* f*,* f_i *is bounded from above or from below in an interval of its domain. Then, there exist scalars* $a, b_1,...,b_n$ *such that*

$$f(z) = az + \sum_{i=1}^{n} b_i,$$

$$f_i(z) = az + b_i, \qquad i = 1, ..., n.$$

Notice that if one of the f, f_i is constant, this sets $a = 0$, and the remaining functions are also constant. Clearly, this case must be excluded if one is to make informative use of Lemma 2.

4 Weighted utilitarianism

The Aggregation Theorem was first stated by Harsanyi (1955, 1977) and rigorously proved and developed by later authors. The lottery set L and the VNM axioms in its statement can be taken in all the ways covered by Fishburn (1982). The theorem relies on a Pareto condition that can also be formulated variously. Given individual preference relations $\succsim_i$, $i = 1, ..., n$, and a social preference relation $\succsim$, all being defined on L, let us say that *Pareto indifference* holds if, for all $p, q \in L$,

$$p \sim_i q, i = 1, ..., n \Rightarrow p \sim q,$$

and that *Strong Pareto* holds if, in addition to Pareto indifference, for all $p, q \in L$,

$$p \succsim_i q, i = 1, ..., n \;\&\; \exists i : p \succ_i q \Rightarrow p \succ q.$$

The Aggregation Theorem is often stated in terms of Pareto indifference alone, but here we adopt a more assertive form based on Strong Pareto.[6]

[6] Along with further Paretian variants, it is proved in Weymark (1993) and De Meyer and Mongin (1995).

Lemma 3 *(The Aggregation Theorem) Suppose that there are individual preference relations* $\succsim_1, ..., \succsim_n$ *and a social preference relation* $\succsim$ *satisfying the VNM axioms on a lottery set L, and suppose also that Pareto indifference holds. Then, for every choice of VNM utility functions* $v, v_1, ..., v_n$ *for* $\succsim, \succsim_1, ..., \succsim_n$ *on L, there are real numbers* $a_1, ..., a_n$ *and b such that*

$$v = \sum_{i=1}^{n} a_i v_i + b.$$

If Strong Pareto holds, there exist $a_i > 0$, $i = 1, ..., n$. *The* a_i *and b are unique if and only if the* $v_1, ..., v_n$ *are affinely independent.*

The following assumption is the cornerstone of our conceptual and mathematical argument. It takes the VNM and Pareto conditions that Harsanyi imposes on any social observer to be valid for a lottery extension of the given utilitarian preferences on sure outcomes. As stressed in the introduction, it is consistent with Harsanyi's position to offer the benefit of the two conditions to the utilitarian observer. Moreover, the Sen critique, as we have delineated it, does not question these conditions by themselves, so that we do not contradict it either by applying them here.

Assumption H: The utilitarian social preference $\succsim^*$ on X can be extended to a preference $\succsim^{*ext}$ on L that satisfies the VNM axioms as well as Pareto indifference with respect to the $\succsim_i$.

Our main result imposes the two conditions to an arbitrary preference relation $\succsim$ on L as in the original theorem, but we get more from them owing to Assumption H: $\succsim$ can be represented on X by a weighted formula that makes utilitarian sense.

Theorem 1 *Let Assumptions 1–5 and H hold. Then, for any preference relation* $\succsim$ *on L satisfying the VNM axioms, if Pareto indifference holds between* $\succsim$ *and the* $\succsim_i$, *there are unique constants* a_i, $i = 1, ..., n$, *such that the VNM utility functions for* $\succsim$ *on X are* $\sum_i a_i u_i^*$ *and PATs. If* $\succsim$ *satisfies the Strong Pareto condition, the* a_i *are positive.*

Proof. Let u and u_i be VNM utility functions for $\succsim^{*ext}$ and $\succsim_i$ on X, respectively. From section 3, u_i can be taken to be continuous, and there is no loss of generality in also supposing that for some $\overline{x} \in X$, $u(\overline{x}) = 0 = u_i(\overline{x})$, $i = 1, ..., n$. By Lemma 3 applied to the corresponding VNM functions on L, there are constants b_i, $i = 1, ..., n$ s.t. $Eu = \sum_{i=1}^{n} b_i Eu_i$, hence by restricting this equation to X,

$$u = \sum_{i=1}^{n} b_i u_i.$$

There are increasing functions f_i, f on the utility sets $u_i^*(X)$, $\sum u_i^*(X)$ s.t. $u_i = f_i \circ u_i^*$ and $u = f \circ \sum_{i=1}^{n} u_i^*$, so that the equation becomes:

$$f \circ \sum_{i=1}^{n} u_i^* = \sum_{i=1}^{n} b_i f_i \circ u_i^*.$$

Since the left-hand side is increasing in every u_i^*, and each of them is non-constant (see fn 4), necessarily $b_i > 0$ for all i.

As X is path-connected, the f_i are continuous by Lemma 1, and it then follows from the last form of the equation that f is also continuous. Defining $f_i' = b_i f_i$, we rewrite the equation as

$$f \circ \left(\sum_{i=1}^{n} u_i^* \right) = \sum_{i=1}^{n} f_i' \circ u_i^*,$$

or

$$f \left(\sum_{i=1}^{n} z_i \right) = \sum_{i=1}^{n} f_i'(z_i),$$

for all $(z_1, ..., z_n) \in U^*(X) \subseteq \mathbb{R}^n$.

Consider the subset $T = U^*(X)^\circ$. It is a nonempty, open connected subset of $\mathbb{R}^n$, and f is continuous, so we can apply Lemma 2 to the functional equation by restricting it to T. It follows that there exist constants a and c_1,...,c_n s.t.

$$(1)\ \forall z \in T_+,\ f(z) = az + \sum_{i=1}^{n} c_i,$$
$$(2)\ \forall z \in T_i,\ \ f_i'(z) = az + c_i,\ i = 1, ..., n,$$

where T_+, T_i are defined as in Lemma 2. Since none of the f, f_i is constant, we have that $a > 0$.

A stronger result actually holds:

$$(1')\ \forall z \ \in \ [U^*(X)]_+ ,\ f(z) = az + \sum_{i=1}^{n} c_i,$$
$$(2')\ \forall z \ \in \ [U^*(X)]_i ,\ f_i'(z) = az + c_i,\ i = 1, ..., n.$$

To prove $(1')$ from (1), take $z \in [U^*(X)]_+$. There is $(z_1, ..., z_n) \in U^*(X)$ s.t. $z = \sum_{i=1}^n z_i$. As $(z_1, ..., z_n) \in \overline{T}$ by assumption, there is in T a sequence $(z_1^l, ..., z_n^l)$, $l \in \mathbb{N}$, s.t. $(z_1, ..., z_n) = \lim_{l\to\infty}(z_1^l, ..., z_n^l)$ and $z = \lim_{l\to\infty} \sum_{i=1}^n z_i^l$. Now, since f is continuous on $[U^*(X)]_+$,

$$f(z) = \lim_{l\to\infty} f(\sum_{i=1}^{n} z_i^l) = \lim_{l\to\infty} a \sum_{i=1}^{n} z_i^l + \sum_{i=1}^{n} c_i = az + \sum_{i=1}^{n} c_i,$$

which establishes $(1')$. The proof of $(2')$ from (2) is similar.

Equation $(1')$ and the definition of f entail that, for all $x \in X$,

$$(1'')\ u(x) = a \sum_{i=1}^{n} u_i^*(x) + \sum_{i=1}^{n} c_i,$$

i.e., u is a PAT of $\sum_i u_i^*$. Similarly, for $i = 1, ..., n$, equations $(2')$ and the definitions of f_i' and f_i entail that for all $x \in X$,

$$(2'')\ b_i u_i(x) = a u_i^*(x) + c_i.$$

These equations show that the sets of VNM utility functions for $\succsim^{*ext}$ and $\succsim_i$ on X are the sets of PAT of $\sum_{i=1}^n u_i^*$ and u_i^*, respectively.

Now, take $\succsim$ as specified and fix a VNM utility function Eu' for $\succsim$ on L. Lemma 3 can be applied to Eu', and for each i, some choice of VNM utility function for $\succsim_i$ on L. As the last paragraph has shown, this utility function must be a PAT of Eu_i^*. Hence there are real numbers $a_i, i = 1, ..., n$, and b s.t. $Eu' = \sum_{i=1}^{n} a_i Eu_i^* + b$, and by restriction to X,

$$u' = \sum_{i=1}^{n} a_i u_i^* + b.$$

It follows that the set of VNM utility functions for $\succsim$ on X is the set of PAT of $\sum_{i=1}^{n} a_i u_i^*$. The a_i are unique because the u_i^* are affinely independent by assumption. If $\succsim$ satisfies the Strong Pareto condition, Lemma 3 entails that the a_i are positive. ■

Up to the penultimate paragraph, the proof consists in establishing that for each individual preference $\succsim_i$, the set of its VNM representations is the set of PAT of u_i^*. The key point is that, by Assumption H and Lemma 3, the utilitarian ordering must be linear in the VNM utilities, while, by assumption, it is linear in the u_i^*. Therefore the two additive representations of the utilitarian ordering on X must be related in such a way that each u_i^* is itself a VNM utility. Using this fact, the last paragraph easily connects the social preference $\succsim$ with the u_i^* in the desired weighted utilitarian way.

The equivalence statement (*) of the previous section amounted to assuming what is proved here indirectly. Unlike (*), Theorem 1 is not question-begging because none of its assumptions by itself entails cardinal relevance for individual VNM utility functions; indeed, this follows only from putting all assumptions together. Furthermore, unlike the argument based on (*), the assumptions eschew the notion of a cardinal preference relation, which has no place in traditional economics.

The following proposition collects information on the utilitarian observer and the individuals that the statement of Theorem 1 does not mention; see equations ($1''$) and ($2''$), respectively. Notice that the last equations, when considered jointly, also say that the individuals' expected utilities are cardinally comparable.

Proposition 1 *Let Assumptions 1–5 and H hold. Then, the set of VNM representations of $\succsim^{*ext}$ is the set of PAT of $E\sum_i u_i^*$. Furthermore, every Eu_i^* is a VNM utility function for $\succsim_i$ on L.*

The previous results depend on putting economic structure on the set of alternatives X, the individual preferences $\succsim_i$ and the individual utility functions u_i^*, while social ethics and social choice theory often discuss normative rules in terms of abstract domains and preference properties. Sen refers to classical utilitarianism, which makes specific assumptions in the style of economic theory, but Weymark does not, and in its original version, Harsanyi's Aggregation Theorem hinges only on the lottery structure of $L = \Delta(X)$, regardless of what X may be. This motivates devising a variant theorem in which the main assumptions directly concern L and utility representations on this set.

Consistently, this variant must shift the utilitarian benchmark to the lottery side. The individuals $i = 1, ..., n$ are now associated with a vector $V^* = (v_1^*, ..., v_n^*)$ of utility functions on L that meaningfully add up, and the utilitarian preference ordering $\succsim^*$ is now defined on L by

$$p \succsim^* q \text{ iff } \sum_{i=1}^{n} v_i^*(p) \geq \sum_{i=1}^{n} v_i^*(q).$$

The v_i^* function will have to represent $\succsim_i$ on L, in the same way as u_i^* earlier represented $\succsim_i$ on X, but we will not assume that it is VNM, for this is precisely one of the things to prove. Grandmont's (1972) axioms for continuous VNM preference orderings are still suitable for our purpose, and as his assumption for the set X is very general, we simply reproduce it below. The previous analysis used the continuity of the u_i^*, and the full dimensionality of $U^*(X)$. Here we will similarly require the v_i^* to be continuous, but shift the dimensionality property from the observer's utility set to the set of individual preferences. Formally, we say that *Independent Prospects* (IP) holds with respect to $\succeq_1, ..., \succeq_n$ if, for all $i = 1, ..., n$, there exist $p^i, q^i \in L$ such that $p^i \succ_i q^i$ and $p^i \sim_j q^i$, $j \neq i$. This property is a *preference* rendering of individual diversity, unlike algebraic independence conditions put on utility representations.[7] The convex structure of L permits reducing Assumption 4 to (IP) alone.

Assumption 1': X is a separable metric space.

[7] (IP) clearly entails that the utility representations of the $\succeq_i$ are affinely independent. If these representations are VNM, (IP) is morever equivalent to that property; see, e.g., Weymark (1991, p. 272).

Assumption 2 is unchanged.

Assumption 3': For each $i = 1, ..., n$, v_i^* is continuous on L.

Assumption 4': *Independent Prospects* holds of $\succsim_1, ..., \succsim_n$.

Assumption 5': For each $i = 1, ..., n$, v_i^* is a utility function for $\succsim_i$ on L.

Our main assumption becomes:

Assumption (H'): The utilitarian social preference $\succsim^*$ on L satisfies the VNM axioms as well as Pareto indifference with respect to the $\succsim_i$.

Theorem 2 *Let Assumptions 1',2,3',4',5' and H' hold. Then, for any preference relation $\succsim$ on L satisfying the VNM axioms, if Pareto indifference holds between $\succsim$ and the $\succsim_i$, there are unique constants a_i, $i = 1, ..., n$, such that the VNM utility functions for $\succsim$ on L are $\sum_i a_i v_i^*$ and its PAT. If $\succsim$ satisfies the Strong Pareto condition, the a_i are positive.*

Proof. See the appendix. ∎

5 Conclusion

The normative import of our analysis lies with the conclusion, obtained in two ways, that a social observer whose preferences on lotteries meet the two conditions of the Aggregation Theorem must follow a weighted utilitarian sum rule $\sum_i a_i u_i^*$ or $\sum_i a_i v_i^*$. That the a_i may be unequal is a weakness from the perspective of utilitarianism. However, the weighted utilitarian variant has been defended for itself by some, and the measurement stage is anyhow the decisive one on the road to the standard form of the doctrine.

By introducing a utilitarian observer, we have followed Sen's suggestion that utilitarianism had to be defined *independently*, or else the results would bear no connection with this doctrine. The exogenously given sums $\sum_i u_i^*$ or $\sum_i v_i^*$ provide the benchmark for comparisons with the preference orderings studied in the theorems. Sen's point appears to be answered by the mathematical consequences of making the very addition he recommends.

Perhaps less obviously, Weymark's reminder about VNM utility is also taken care of. Our assumptions on the u_i or v_i take them to be ordinal representations, but the theorems invest them with a relevant cardinal meaning. Technically, non-affine φ_i drop out from the social observer's additively separable criterion $\sum_i \varphi_i^{-1} \circ u_i$ (or $\sum_i \varphi_i^{-1} \circ v_i$) that the Aggregation Theorem by itself only delivers. This crucial simplification occurs in the proof as a result of Assumptions H or H', when the two conditions of the Aggregation Theorem are applied to the extension of $\sum_i u_i^*$ (or $\sum_i v_i^*$). See in particular Proposition 1.

To sum up, when the theorem is reconstructed, it provides an argument for its ethical relevance. There seem to be only two ways to escape this conclusion. One is to reject Assumptions H or H', as if they were irrelevant to utilitarian social ethics. We have argued that this rejection would be an inconsistent move for Harsanyi, and we do not see any other utilitarian theorist who would defend it. The other, of course much more significant move is to reject H or H' because the VNM and Pareto conditions are not appealing in and of themselves. It has been argued that the VNM conditions are questionable for a social observer (Diamond, 1967), and that the Pareto principle becomes dubious in the risk context (Fleurbaey 2010), not to mention the uncertainty context, where this point is now well taken. Weighty as these objections are, they come into play only if one has disposed of the claim that the Aggregation Theorem had nothing to do with utilitarianism as an ethical doctrine.

Acknowledgement. Many thanks to the participants in the Conference in Honour of Peter Hammond (University of Warwick), the D-TEA Conference (HEC Paris), the Conference in Honor of Claude d'Aspremont and Jean-François Mertens (CORE), a LSE

Choice Group seminar and a Yale workshop, where this paper was given. Special thanks are due to John Duggan, Michael Mandler, Marcus Pivato, John Roemer, Uzi Segal, two anonymous referees and the editor of this journal for their comments on previous versions. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

6 References

Aczel J. 1987, *A Short Course on Functional Equations*, Dordrecht: Reidel.

d'Aspremont, C. and L. Gevers (2002), "Social welfare functionals and interpersonal comparability", in K. J. Arrow, A. K. Sen and K. Suzumura (eds), *Handbook of Social Choice and Welfare*, New York, Elsevier, vol. 1, ch. 10, p. 459–541.

Blackorby, C., D. Donaldson, and Weymark, J. (1999), "Harsanyi's Social Aggregation Theorem for State-Contingent Alternatives", *Journal of Mathematical Economics*, p. 365–387.

De Meyer, B. and P. Mongin (1995), "A Note on Affine Aggregation", *Economics Letters,* 47, p. 177–183.

Dhillon, A. and J.F. Mertens (1999), "Relative Utilitarianism", *Econometrica*, 67, p. 471-498

Diamond, P.A. (1967), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment", *Journal of Political Economy*, 75, p. 765–766.

Eichhorn, W. (1978), *Functional Equations in Economics*, Reading, Mass., Addison Wesley.

Fishburn, P. (1982), *The Foundations of Expected Utility*, Dordrecht, Reidel.

Fleurbaey, M. (2010), "Assessing Risky Social Situations", *Journal of Political Economy* 118: 649–680.

Grandmont, J.M. (1972), "Continuity Properties of a von Neumann-Morgenstern Utility", *Journal of Economic Theory*, 4, p. 45–57.

Grant, S., A. Kajii, B. Polak and Z. Safra, 2010. "Generalized Utilitarianism and Harsanyi's Impartial Observer Theorem", *Econometrica*, 78, p. 1939-1971.

Hammond, P.J. (1982), "Utilitarianism, Uncertainty and Information", ch. 4 in A. Sen and B. Williams (eds), *Utilitarianism and Beyond*, Cambridge, Cambridge University Press, p. 85-102.

Hammond P. (1996), "Consequentialist Decision Theory and Utilitarian Ethics", in F. Farina, F. Hahn, and S. Vannucci (eds.), *Ethics, Rationality, and Economic Behaviour*, Oxford: Clarendon Press.

Harsanyi, J. C. (1955), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility", *Journal of Political Economy*, 63, p. 309–321.

Harsanyi, J. C. (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge, Cambridge University Press.

Harvey, C.M. (1999), "Aggregation of Individuals' Preference Intensities into Social Preference Intensity", *Social Choice and Welfare*, 16, p. 65-79.

Karni, E. (1998), "Impartiality: Definition and Representation", *Econometrica*, 66, p. 1405-1415.

Mongin, P. (2001), "The Impartial Observer Theorem of Social Ethics", *Economics and Philosophy*, 17, p. 147-179.

Mongin, P. (2002), "Impartiality, Utilitarian Ethics, and Collective Bayesianism", Cahier de recherche du Laboratoire d'économétrie, Ecole Polytechnique.

Mongin, P. and C. d'Aspremont (1998), "Utility Theory and Ethics", in S. Barberà, P. Hammond and C. Seidl (eds), *Handbook of Utility Theory*, vol. 1, Dordrecht, Kluwer, ch. 10, p. 371–481.

Rado, F. and J. Baker (1987), "Pexider's Equation and Aggregation of Allocations", *Aequationes Mathematicae*, 32, p. 227–239.

Segal, U. (2000), "Let's Agree That All Dictatorships Are Equally Bad", *Journal of Political Economy*, 108, p. 569-589.

Sen, A.K. (1974), "Rawls versus Bentham: An Axiomatic Examination of the Pure Distribution. Problem", *Theory and Decision*, 4, p. 301–309

Sen, A.K. (1977), "Nonlinear social welfare functions: A reply to Professor Harsanyi", in R. Butts and J. Hintikka (eds.), *Foundational Problems in the Special Sciences*, Dordrecht, Reidel.

Sen, A. K. (1986), "Social Choice Theory", in K.J. Arrow and M.D. Intriligator (eds), *Handbook of Mathematical Economics*, Amsterdam, North Holland, vol. 3, p. 1073–1181.

Weymark, J. (1991), "A Reconsideration of the Harsanyi-Sen Debate on Utilitarianism", in J. Elster and J.E. Roemer (eds), *Interpersonal Comparisons of Well-Being*, Cambridge, Cambridge University Press, p. 255–320.

Weymark, J. (1993), "Harsanyi's Social Aggregation Theorem and the Weak Pareto Principle", *Social Choice and Welfare*, 10, p. 209–221.

7 Appendix

Proof. (Lemma 2) Assume w.l.g. that if f is not bounded on an interval, then f_1 is. Denote by T_{ij} the projections of T on the i-th and j-th factors of $\mathbb{R}^n$. Fix $z^* \in T$, and consider the open subset of T_{12}:

$$T_{12}(z^*) = \{(z_1', z_2') \in T_{12} \mid (z_1', z_2', z_3^*, ..., z_n^*) \in T\}.$$

Define the set:

$$T_{12}(z^*)_+ = \{z_1' + z_2' \mid (z_1', z_2') \in T_{12}(z^*)\},$$

and the function on $T_{12}(z^*)_+$:

$$h : z_1' + z_2' \mapsto f(z_1' + z_2' + z_3^* + ... + z_n^*) - f_3(z_3^*) - ... - f_n(z_n^*).$$

Consider some $\overline{z} \neq z^*$, such that $T_{12}(z^*)$ and $T_{12}(\overline{z})$have non-empty intersection. One checks that h is defined identically on this intersection, as the following holds for all $(z_1', z_2') \in T_{12}(z^*) \cap T_{12}(\overline{z})$:

$$f(z_1'+z_2'+z_3^*+...+z_n^*)-f_3(z_3^*)-...-f(z_n^*) = f_1(z_1')+f_2(z_2') = f(z_1'+z_2'+z_3^{**}+...+z_n^{**})-f_3(z_3^{**})-...-f(z_n^{**}).$$

Using the fact that T_{12} is path-connected (for a similar step, see Rado and Baker, 1987, p. 232), we conclude that h is uniquely defined on T_{12+} and that the following functional equation holds on the whole domain T_{12}: for all $(z_1', z_2') \in B(z_1, z_2)$,

$$h(z_1' + z_2') = f_1(z_1') + f_2(z_2').$$

Since T_{12} is open and connected, and either h or f_1 is bounded on an interval, the Rado-Baker theorem (more precisely, their Corollary 3) for $n = 2$ applies. Hence, the functions f, f_1, f_2 are affine (with the same multiplicative constant).

The same argument can be reproduced for the domains $T_{1,3}, ..., T_{1,n}$, implying that every f_i is affine (with the same multiplicative constant). ■

Proof. (Theorem 2) Let $V = (v_1, ..., v_n)$ a vector of VNM functions representing the $\succsim_i$ orderings on L. Each v_i can be taken to be continuous by section 3, and there is no loss of generality in assuming that for some $\overline{p} \in L$, $\sum_i v_i^*(\overline{p}) = 0 = v_i(\overline{p})$, $i = 1, ..., n$. By Lemma 3, there are constants b_i, $i = 1, ..., n$, s.t.

$$\sum_{i=1}^{n} v_i^* = \sum_{i=1}^{n} b_i v_i.$$

Applying *Independent Prospects* (IP) leads to

$$v_i^*(p^i) - v^*(q^i) = b_i(v_i(p^i) - v(q^i)),$$

hence to $b_i > 0$, $i = 1, ..., n$. There are increasing functions g_i on $v_i(L)$ s.t. $v_i^* = g_i \circ b_i v_i$. Lemma 1 can be applied because v_i^* and $b_i v_i$ are continuous on L, which is convex, hence path-connected; thus the g_i are continuous.

Define $V' = (b_1 v_1, ..., b_n v_n)$. The first equation becomes

$$\sum_{i=1}^{n} g_i(z_i) = \sum_{i=1}^{n} z_i,$$

for all $(z_1, ..., z_n) \in V'(L) = Z$. The set Z is is convex and has an non-empty interior in $\mathbb{R}^n$. We can use Lemma 2 on $T = Z^\circ$, noticing that none of the g_i is constant by (IP). Since the f of this lemma is the identity function, $g_i(z) = z + c_i$ for all $z \in T_i$ and for all $i = 1, ..., n$, with $\sum_{i=1}^{n} c_i = 0$. The function g_i is continuous on Z_i, the projection of Z on its ith component, and $Z_i \subseteq \overline{T_i^\circ}$, so an extension argument leads to $g_i(z) = z + c_i$ for all $z \in Z_i$ and all $i = 1, ..., n$. Returning to the initial functions, we see that, up to an additive constant, $v_i = v_i^*/b_i$ for $i = 1, ..., n$. Hence the set of VNM utility functions for $\succsim_i$ on L is the set of PAT of v_i^*.

The rest of the proof makes use of this finding when Lemma 3 is applied to $\succsim$ on L, following the pattern already used for Theorem 1. ■

If we had postulated Assumption 4 instead of Assumption 4', we would have derived the (IP) property on $\succeq_1, ..., \succeq_n$ by the following argument. From Assumption 4, there is a vector $\overline{U}^* = (\overline{u}_1^*, ..., \overline{u}_n^*) \in U^*(X)$ and there are numbers $\varepsilon_i > 0$, $i = 1, ..., n$, such that for all i,

$$\overline{U}_{\varepsilon_i}^* = (\overline{u}_1^*, ..., \overline{u}_i^* + \varepsilon_i, ..., \overline{u}_n^*) \in U^*(X).$$

Take $x \in X$ such that $U^*(x) = \overline{U}^*$, and for each i, $x^i \in X$ such that $U^*(x^i) = \overline{U}_{\varepsilon_i}^*$. Then, by Assumption 5 and the fact that $X \subset L$, the definition of (IP) holds with $p^i = x^i$ and $q^i = x$, $i = 1, ..., n$. The role of Assumption 4' is only to save the generality of Theorem 2 by assuming nothing on the set of sure alternatives X.

We now discuss how Theorem 1 can be revised to accommodate a less than full dimensional utility set $U^*(X)$. Consider the effect of weakening Assumption 4 by requiring that $U^*(X)$ have affine dimension k, with $2 \leq k \leq n$. By reindexing if necessary, we may suppose that $\{u_1^*, ..., u_k^*\}$ is an *affine basis* —i.e., a maximal affinely independent subset— for $\{u_1^*, ..., u_n^*\}$. Then, by the argument of last paragraph, $\succsim_1, ..., \succsim_k$satisfy (IP). It follows that, for any choice $v_1, ..., v_n$ of VNM representations of $\succsim_1, ..., \succsim_n$ on L, the set $\{v_1, ..., v_k\}$ is affinely independent. However, $\{v_1, ..., v_k\}$ may not be an affine basis for $\{v_1, ..., v_n\}$, and we can only conclude that the former set is *included* in such a basis $\{v_1, ..., v_{k'}\}$, where $k' \geq k$. This analysis motivates a more complex form of Assumption 4, which in effect imposes that $k' = k$.

Assumption 4".1: There is a subset of individuals, $\{j_1, ..., j_k\} \subseteq \{1, ..., n\}$, with $k \geq 2$, such that $\left\{u_{j_1}^*, ..., u_{j_k}^*\right\}$ is an affine basis for $\{u_1^*, ..., u_n^*\}$, and moreover, for no strict superset $\{j_1, ..., j_{k'}\}$ does (IP) apply to $\succsim_{j_1}, ..., \succsim_{j_{k'}}$.

Assumption 4".2: $U^*(X)$ has a nonempty connected relative interior $U^*(X)^\circ$ and is such that $U^*(X) \subseteq \overline{U^*(X)^\circ}$.

In the limiting case $k = n$, Assumption 4" reduces to Assumption 4. We are now ready for a more general form of Theorem 1.

Theorem 3 *Let Assumptions 1,2,3,4",5, and H hold. Then, the conclusions of Theorem 1 hold, except that the coefficient a_i may not be unique, and even under Strong Pareto, may be of any sign. The conclusions of Proposition 1 still hold.*

Proof. The proof follows that of Theorem 1 in outline; the Working Paper version has the full details. To show that even with Strong Pareto a_i may be nonpositive, take $n = 3$ and $B = \{1, 2\}$ with

$$u_1 = u_1^*, u_2 = u_2^*, u_3 = u_1 + u_2 \text{ and } u_3^* = 2u_1^* + 2u_2^*,$$

with u_1^* and u_2^* being unrestricted. Define $\succsim$ on L from the representation $E(u_1+u_2+u_3)$. By construction, $\succsim$ satisfies Strong Pareto and has a VNM utility function $u' = 2u_1^*+2u_2^* = u_3^*$ on X. Now, if we put $u' = a_1u_1^* + a_2u_2^* + a_3u_3^*$, we see that the coefficients a_i can be chosen to be negative, e.g.,

$$u' = 4u_1^* + 4u_2^* - u_3^* = -u_1^* - u_2^* + 1.5u_3^* \ .$$

■

A more general variant of Theorem 2 can be devised along similar lines.