The Relevance of Utilitarianism

Marc Fleurbaey,*Philippe Mongin†

December 1, 2011

Abstract

Harsanyi invested his Aggregation Theorem and Impartial Observer Theorem with deep utilitarian sense, but Sen redescribed them as "representation theorems" with little ethical import. This negative view has gained wide acquiescence in economics. Against it, we support the utilitarian interpretation by a novel argument relative to the Aggregation Theorem. We suppose that a utilitarian observer evaluates non-risky alternatives by the sum of individual utilities and investigate his von Neumann-Morgenstern (VNM) preference on risky alternatives. Adding some technical assumptions to Harsanyi's, we conclude that (i) this observer would use the utility sum as a VNM utility function, and crucially, (ii) any social observer would evaluate both risky and non-risky alternatives in terms of a weighted utility sum. Hence, *pace* Sen, VNM theory can give some interesting support to utilitarianism. The argument is conveyed by means of three theorems that encapsulate Harsanyi's original one as a particular step.

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 against 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).

Sen's critique has gained wide acquiescence among economists, especially after Weymark (1991) amplified it in a thorough discussion of the "Harsanyi-Sen debate". Those, like Mongin and d'Aspremont (1998), who maintained that the critique can be addressed, barely had a hearing, and today's prevailing view seems to be that Harsanyi's attempt at deriving utilitarianism from VNM assumptions is simply hopeless.

Running against the tide, we offer a novel argument in favour of Harsanyi's position. We regard it as being severely incomplete, but not flawed, and undertake to buttress it, somewhat paradoxically, by a further influx of "representation theorems". Those proved below complement Harsanyi's Aggregation Theorem and provide it, or so we claim, with a utilitarian interpretation that it does not have by itself.

Like Harsanyi, we suppose that the individuals' preferences satisfy the VNM axioms on risky alternatives. Unlike him, we start from a social observer who has already formed social preferences on sure alternatives according to the sum rule of classical utilitarianism, and we suppose that his preferences can be extended to lotteries so as to satisfy the VNM axioms. Under relevant technical assumptions from microeconomics, we obtain two consequences in a row. First, if the extended preferences satisfy the Pareto principle, they are represented as the expected utility of the classical utilitarian sum, i.e., this sum is also the utilitarian observer's VNM utility function (up to a positive affine transformation). Second - a result holding this time for *any* social observer - preferences on lotteries that are VNM and Paretian are represented as the expected utility of *a weighted variant of the utilitarian sum*, i.e., the variant in question is also the social observer's VNM utility function (up to a positive affine transformation). Accordingly, it is also the social rule employed to evaluate the sure alternatives.

The second statement conveys the final message: utilitarianism prevails inasmuch as the social observer must add up the same cardinal individual utility functions as our supposed utilitarian; however, it fails to the extent that the weights in the sum may not be equal. This is not classical utilitarianism all the way down, but still a good deal of it (social choice theorists, e.g., d'Aspremont and Gevers, 2002, call it *weighted* utilitarianism).

In sum, there is a sense in which Harsanyi was correct in believing that VNM theory could help support utilitarianism. He did not carry the argument to its end, and Sen was thus justified in questioning it, but the challenge is not unanswerable despite the common presumption.

In section 2, we develop Sen's objections in some formal detail. We do not review the whole controversy with Harsanyi, since part of it is self-explanatory or well covered elsewhere. In particular, we do not discuss the normative plausibility of utilitarianism

and its problematic stand towards income inequality, which have raised lively discussions among both economists and philosophers. Instead, we concentrate on the single issue of whether the Aggregation Theorem is at all relevant to utilitarianism. Section 3 sets up the formal framework with its technical assumptions. They amount to replacing Harsanyi's abstract social states by allocations of commodities between the individuals, and then imposing constraints on these more structured objects as well as their individual and social evaluations. Section 4 states three theorems that derive utilitarianism from Paretian aggregation; each follows from specific assumptions made on the allocations, preferences and utilities. The appendix explains the mathematical tools.

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 here):
"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).

A major claim can be read between these lines: 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, the sum rule of classical utilitarianism presupposes that individual utility functions are cardinally comparable, 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. Weymark brings it out clearly:

"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 the 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 $v = \varphi(\sum_i \varphi_i^{-1} \circ v_i)$, 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 also leads to the conclusion that Harsanyi proved no more than representation theorems (see

Weymark, 1991, p. 305).

Some definitions and notation, which anticipate on the framework of next section, will help make the two objections formally. 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 V 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 $\mathcal{U}'$ rather than $\mathcal{F}'$. This is the point well emphasized by Weymark.

We are also concerned with the social rule of a utilitarian observer, and we define it as follows. In the observer's eyes, the individuals $i = 1, ..., n$ are associated with welfare

indexes u_i^* on the outcome set X that meaningfully add up, i.e., the u_i^* must be *both cardinally measurable and comparable*. Accordingly, this social rule may be any element of the set:

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

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

With the definitions just made, 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. This is the most transparent claim in Sen's above quotations: VNM utility values do not have to measure welfare, and if Harsanyi proves that the social utility function is a sum of individual VNM utility values, this does not by itself speak in favour of 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 does not deliver cardinal equivalence with u_i; that is, one gets $u_i^* \in \mathcal{F}_i$, the set of i's representations on X at large, and not $u_i^* \in \mathcal{U}_i$, as utilitarianism would require; and similarly, the assumption does not make $\sum_i u_i$ a member of the set $\mathcal{C}$.

To take stock, two problems stand in the way of Harsanyi's utilitarian interpretation of his results. One - call it Sen's point - is that VNM theory has a cardinalization of its own; the other - call it Weymark's point - is that this cardinalization is artificial given the ordinal nature of that theory. There is a convenient joint answer, which is *to ground the utility functions* u_i *and* u_i^* *on a common basis of cardinal preference*. If the utilitarian cardinalization rests on a preference basis and the VNM cardinalization can be reduced to that basis, the two coincide (*pace* Sen) and the former escape irrelevance (*pace* Weymark). Technically, a cardinal preference is a relation on *pairs* of sure alternatives, i.e., $(x, y) \succsim_i^* (z, w)$, and the axioms for $\succsim^*$ embody the utilitarian tenet that coherent comparisons

can be made of intraindividual preference differences. There will be a connecting axiom to ensure that the VNM cardinalization is rooted in the same comparisons. It will entail the consequence that for all $x, y, w, z \in X$,

$$(1/2)u(x) + (1/2)u(y) \geq (1/2)u(z) + (1/2)u(w) \text{ iff}$$

$$u^*(x) - u^*(z) \geq u^*(w) - u^*(y). \quad ((*))$$

Mongin (2002) develops this strategy, which was already suggested, but without axiomatic detail, in Weymark (1991, p. 308) and Mongin and d'Aspremont (1998, p. 435). It is helpful in making Harsanyi's position logically consistent, but it is question-begging, because the axioms are only a roundabout way of getting the problematic equivalence $(*)$. Moreover, for most economists, preference is an ordinal concept by definition, and it may even be so for Harsanyi himself. In this paper, we dispense with $(*)$ or its axiomatic counterpart in terms of cardinal preference, and we use an indirect argument instead. It will be seen that it takes care of the two critical points at the same time.

3 The framework and assumptions

We assume that there is *a feasible set* X, the elements of which are allocations of commodities to the $n \geq 2$ individuals. As a typical application, $X \subseteq \mathbb{R}^{mn}$, where m is the number of commodities. Unlike basic consumer theory, which takes $X = \mathbb{R}^{mn}_+$, we do not require X to be a Cartesian product, and indeed, this structure becomes 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 standard case of private goods, it may be an inappropriate structure if X takes the availability of resources into account. We require connectedness, a weakening of the convexity assumption that the basic theory often makes. We can afford to be general in another direction

by allowing X to be a subset of a function space, and not simply of the Euclidean space. All that matters is the following domain assumption.

Assumption 1: X is a metric connected space.

The VNM apparatus can now be introduced formally. When we apply it to the social observer, all standard construals of VNM theory in expectational form work; take any one of the axiomatizations in Fishburn (1982). However, concerning the individuals, we need *continuous* VNM utility functions, a property which the usual systems do not provide, so we turn to Grandmont's (1972) version, which is 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 sets, and take the set $\Delta(X)$ 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 in its own right. Now, a *continuous VNM preference relation* $\succsim$ on $\Delta(X)$ 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 $p \in \Delta(X)$, the sets

$$\{p' \in \Delta(X) : p' \succsim p\} \text{ and } \{p' \in \Delta(X) : p \succsim p'\}$$

are closed in $\Delta(X)$.

(*Independence*) For all $p, q, r \in \Delta(X)$ and all $\lambda \in [0, 1]$, $p \sim_i q$ iff $\lambda p + (1 - \lambda)r \sim_i \lambda q + (1 - \lambda)r$.

Grandmont's Theorem 3 (1972, p. 49) ensures 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 $\Delta(X)$. 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 of PAT of u.

We retain Grandmont's definition of the lottery set, letting $L = \Delta(X)$, and apply his preference apparatus to the individuals.

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

By contrast, the utilitarian half of the construction relies on primitives directly expressed in terms of utility functions. 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 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 nonempty connected interior $U^*(X)^\circ$ in $\mathbb{R}^n$ and is such that $U^*(X) \subseteq \overline{U^*(X)^\circ}$, i.e., is included in the closure of its interior.

It would be equivalent to impose these assumptions on any collection of PAT $\varphi_i \circ u_i^*$ (with the same α for all i), so that they make good utilitarian sense. Assumption 3 is mild and standard, but Assumption 4 is less so. In one respect, it simply complements Assumptions 1 and 3, which entail that $U^*(X)$ is connected, by excluding unusual ways in which connectedness applies to this set. In another respect, it requires $U^*(X)$ to have full dimension n, hence $u_1^*, ..., u_n^*$ to be linearly independent functions. This is not demanding under standard microeconomic conditions. If there are private consumption goods, each individual is concerned only with how much of these goods he consumes, and free disposal is allowed, then throwing away some of the individual'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 (for instance, no two individuals can be alike in the utilitarian observer's eyes).

In a variant of our analysis, we shift the dimensionality requirement from the util-

itarian to the VNM side. This variant requires the continuity of Eu_i^* instead of u_i^* and dispenses with connectedness assumptions. Accordingly, 1, 3 and 4 are modified thus.

Assumption 1': X is a metric space.

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

Assumption 4': For each $i = 1, ..., n$, there are $p^i, q^i \in L$ such that $p^i \succ_i q^i$ and $p^i \sim_j q^i$, $j \neq i$.

The last property is sometimes called *Independent Prospects* (IP) in the literature, and it is provably equivalent to the following: for any vector $V = (v_1, ..., v_n)$ of VNM utility functions for $\succsim_1, ..., \succsim_n$ on L, the v_i are affinely independent. (That IP entails linear, hence affine independence, is obvious and not limited to VNM functions; however, the converse is non-trivial and specific to such functions.)

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 and the strategy that was criticized as being question-begging.

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, more can be said on f_i in view of the previous assumptions.

Lemma 1 *Suppose that g and h are continuous real-valued functions defined on a path-connected set X and f is a real-valued function defined on $g(X)$ such that $h = f \circ g$; then, f is also continuous. It follows from Assumptions 1–3 that f_i is continuous for $i = 1, ..., n$.*

To complete the mathematical groundwork for the next section, we state a functional equation theorem that drives its mathematical analysis. Rado and Baker (1987) proved it for $k = 2$, but it generalizes to $k \geq 2$. Blackorby, Donaldson and Weymark (1999)

assumed this more general form to derive Harsanyi's Aggregation Theorem on a domain of state-contingent alternatives.

Take Z and E to be normed vector spaces, and T to be an open connected subset of Z^k, $k \geq 2$. For any set $S \subset Z$, we put $S_+ = \left\{ \sum_{i=1}^{k} z_i \mid (z_1, ..., z_k) \in S \right\}$ and $S_i = proj_i S$, i.e., the projection of S on the i-th factor of Z^k.

Lemma 2 *Suppose that $f : T_+ \to E$ and $f_i : T_i \to E$, $i = 1, ..., k$ satisfy the equation*

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

for all $(z_1, ..., z_k) \in T$. Suppose that one of the f, f_i is continuous. Then, there exist a linear function $A : Z \to E$ and scalars b_1,...,b_k such that the functions on Z defined by

$$F(z) = A(z) + \sum_{i=1}^{k} b_i,$$

$$F_i(z) = A(z) + b_i, i = 1, ..., k,$$

extend f and f_i, $i = 1, ..., k$, respectively, and such that

$$F(\sum_{i=1}^{k} z_i) = \sum_{i=1}^{k} F_i(z_i)$$

holds for all $(z_1, ..., z_k) \in Z^k$. There are no other functions than F and the F_i just defined that extend f and the f_i while satisfying this equation.

4 Theorems on utilitarianism from Paretian aggregation

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, and this needs spelling out. 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 we adopt here a more assertive form based on Strong Pareto. 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.

Now to our results. In each of them, we suppose that the utilitarian social preference $\succsim^*$ on X can be *extended to a VNM preference* $\succsim^{*ext}$ *on* L, which calls for a technical comment. By definition, for all $x, y \in X$,

$$x \succsim^{*ext} y \text{ iff} \sum\nolimits_i u_i^*(x) \geq \sum\nolimits_i u_i^*(y).$$

Clearly, the following preference $\succsim^{**}$ satisfies the requisites: for all $p, q \in L$,

$$p \succsim^{**} q \text{ iff } E_p \sum\nolimits_i u_i^*(x) \geq E_q \sum\nolimits_i u_i^*(y).$$

But clearly also, there are other VNM extensions $\succsim^{*ext}$ of $\succsim^*$ to L. They involve cardinalizations different from the utilitarian one — another occurrence of Sen's point in section 2. The force of our theorem is precisely to exclude these other extensions, using Paretian assumptions as further constraints on $\succsim^{*ext}$. Once uniqueness of $\succsim^{*ext}$ is secured, much more follows, in particular a strong claim concerning any arbitrary social observer.

Theorem 1 *Let assumptions 1–5 hold. Consider an extension $\succsim^{*ext}$ to L of the utilitarian $\succsim^*$ on X that satisfies the VNM axioms, and suppose that Pareto indifference holds between $\succsim^{*ext}$ and the $\succsim_i$. Then, the set of VNM utility functions for $\succsim^{*ext}$ on X is the set of PAT of $\sum_{i=1}^n u_i^*$. Furthermore, 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 set of VNM utility functions for $\succsim$ on X is the set of PAT of $\sum_i a_i u_i^*$. If $\succsim$ satisfies the Strong Pareto condition, the a_i are positive.*

Proof. Let Eu_i and Eu be VNM utility functions for $\succsim_i$ and $\succsim^{*ext}$, respectively, on the lottery set on L. By Lemma 3 applied to these functions, there are constants b_i, $i = 1, ..., n$ and d s.t.

$$Eu = \sum_{i=1}^{n} b_i Eu_i + d.$$

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

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

The increasing property of f makes u increasing in each u_i^*, and therefore in each u_i, as f_i^{-1} is increasing, which leads to $b_i > 0$, $i = 1, ..., n$. Also, by Lemma 1, each f_i is continuous, and so is f because $u = \sum_{i=1} b_i u_i + d$ and $\sum_{i=1}^n u_i^*$ are continuous functions. Defining $f_i' = b_i f_i + d_i$ for arbitrary choices of d_i s.t. $\sum_{i=1}^n d_i = d$, 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 one of the f, f_i' is continuous (all are), so we can apply Lemma 2 to the functional equation by restricting it to T. It follows that there exist constants $A > 0$ and c_1,...,c_n s.t.

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

where T_+, T_i are defined like S_+, S_i before Lemma 2.

A stronger result actually holds:

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

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 initial definition of f entail that, for all $x \in X$,

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

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 $x \in X$,

$$b_i u_i(x) + d_i = Au_i^*(x) + c_i,$$

hence, given $b_i > 0$, that u_i is a PAT of u_i^*. It follows 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. But the last paragraph has shown that this utility function must be a PAT of Eu_i^*. It follows that there are real numbers $a_i, i = 1, ..., n$, and b s.t.

$$Eu' = \sum_{i=1}^{n} a_i Eu_i^* + b,$$

hence s.t.

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

is a VNM utility function for $\succsim$ on X. Thus, 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. ■

We now provide a variant of the previous result that shifts the dimensionality requirement from the utilitarian vector $U^* = (u_1^*, ..., u_n^*)$ to the vector of VNM representations $V = (v_1, ..., v_n)$ of $\succsim_1, ..., \succsim_n$. This is perhaps not so intuitive, but the technical advantage is that the connectedness assumptions can be dropped. The role played earlier by the connected sets X and $U^*(X)^\circ$ is now fulfilled by the convex sets L and $V(L)$.

Theorem 2 *Assumptions 1',2,3',4',5 hold. Then, the conclusions of the previous theorem follow.*

Proof. Let v_i and v be VNM utility functions for $\succsim_i$ and $\succsim^{*ext}$ on L. By Lemma 3, these satisfy

$$v = \sum_{i=1}^{n} b_i v_i + d,$$

where $b_i > 0$ for all i. There are increasing functions f_i on $(Eu_i^*)(L)$, and f on $(E\sum_{i=1}^n u_i^*)(L) = (\sum_{i=1}^n Eu_i^*)(L)$, s.t. $v_i = f_i \circ Eu_i^*$ and $v = f \circ \sum_{i=1}^n Eu_i^*$. Putting

$$w_i = b_i v_i + d_i,\ i = 1, ..., n$$

for arbitrary d_i s.t. $d = \sum_{i=1}^n d_i$, we rewrite the equation as:

$$\sum_{i=1}^{n} f_i^{-1} \circ \frac{w_i - d_i}{b_i} = f^{-1} \circ \sum_{i=1}^{n} w_i.$$

or

$$\sum_{i=1}^{n} g_i \circ w_i = f^{-1} \circ \sum_{i=1}^{n} w_i$$

after suitably relabelling the functions on the LHS. Now, L is a convex, hence connected set, and the weak topology ensures that it is path-connected. The functions w_i and $\sum_{i=1}^n w_i$ are continuous on L, as are the composed functions $g_i \circ w_i = Eu_i^*$ and $f^{-1} \circ \sum_{i=1}^n w_i = \sum_i Eu_i^*$; hence, by Lemma 1, f^{-1} and the g_i are continuous on their respective domains. The last equation can be restated as

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

for all $(z_1, ..., z_n) \in V(L) \subseteq \mathbb{R}^n$. As $V(L) = (v_1(L), ..., v_n(L))$ is a convex subset of full dimension, it has non-empty interior $T = V(X)^\circ$ in $\mathbb{R}^n$. Furthermore, T is itself convex, hence connected in $\mathbb{R}^n$, so Lemma 2 can be applied to the functional equation as restricted to T.

The proof now parallels the previous one. It entails that v is a PAT of $E\sum_i u_i^*$, that, for $i = 1, ..., n$, v_i is a PAT of Eu_i^*, and that there are $a_i, i = 1, ..., n$ and b s.t. $\succsim$ is represented by

$$\sum_{i=1}^{n} a_i E u_i^* + b,$$

on L, hence by

$$\sum_{i=1}^{n} a_i u_i^* + b,$$

on X. These facts entail the two claims on the sets of VNM functions made by the theorem.

Since the v_i are affinely independent by assumption, the Eu_i^* also are, and the coefficients a_i are unique in the representation of $\succsim$ on L. Given the linearity of E, they are also unique in the representation of R on X. ■

As a last variant, we dispense with the assumption that either $U^*(X)$ or $V(L)$ is full-dimensional, but require that these two utility sets have the same affine dimension and realize their common dimension on the same subset of individuals. Formally, there is $\{j_1, ..., j_k\} \subseteq \{1, ..., n\}$ such that

$$\dim \left\{u_{j_1}^*, ..., u_{j_k}^*\right\} = \dim U^* = \dim \left\{v_{j_1}, ..., v_{j_k}\right\} = \dim V.$$

For $k \geq 2$, we call the set $\{j_1, ..., j_k\}$ a *common basis* for U^* and V. Clearly, this algebraic assumption is neither weaker nor stronger than the preceding ones. It can be made part of either the first or the second set of assumptions. Specifically, take the first one, and replace Assumption 4 by:

Assumption 4". The image set $U^*(X)$ has nonempty connected relative interior $U^*(X)^\circ$ and is such that $U^*(X) \subseteq \overline{U^*(X)^\circ}$. Furthermore, there is a common basis B for U^* and V, where V is any vector of individual VNM representations on L.

Theorem 3 *Assumptions 1,2,3,4",5 hold. Then, the conclusions of the initial theorem*

follow, except that the coefficient a_i may not be unique and even under Strong Pareto may be of any sign.

Proof. Assume w.l.g. that the common basis B consists of the first k individuals. Since the Pareto indifference condition can be restated in terms of the individuals in B, Lemma 3 leads to

$$Eu = \sum_{i=1}^{k} b_i Eu_i + d.$$

Utility functions of individuals outside B can be reexpressed as

$$u_j^* = \sum_{i=1}^{k} \mu_{ji} u_i^* + \nu_j,\ j = k+1, ..., n,$$

and the equation in the proof of the initial theorem becomes

$$f \circ \left(\sum_{i=1}^{k} \left(1 + \sum_{j=k+1}^{n} \mu_{ji} \right) u_i^* + \sum_{j=k+1}^{n} \nu_j \right) = \sum_{i=1}^{k} b_i^* f_i \circ u_i^* + c^*,$$

or

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

for suitably defined functions $\widetilde{u_i^*}$,$\widetilde{f_i}$.

This leads to a functional equation on $\widetilde{T} = (\widetilde{u_i^*}(X))_{i=1,...,k}$ that can be solved like the equation on $T = (u_i^*(X))_{i=1,...,n}$ in the first proof. ($\widetilde{T}^\circ$ can be viewed as a full-dimensional subset of $\mathbb{R}^k$ satisfying the conditions of Lemma 2, and the affine solution on $\widetilde{T}^\circ$ can be extended by continuity to $\widetilde{T}$.) It follows that u is a PAT of $\sum_{i=1}^{k} \widetilde{u_i^*} = \sum_{i=1}^{n} u_i^*$, which establishes the claim on the set of VNM utility functions for $\succsim^{*ext}$ on X. It also follows that, for $i = 1, ..., k$, u_i is a PAT of u_i^*.

We may now apply Lemma 3 to $\succsim$ and $\succsim_i$, $i = 1, ..., k$, and conclude that there are coefficients a_i', $i = 1, ..., k$ s.t.

$$u' = \sum_{i=1}^{k} a_i' u_i^*$$

is a VNM utility function for $\succsim$ on X. This can be rewritten as

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

for appropriate coefficients that will not be unique if $k < n$. This establishes the claim on the set of VNM utility functions for $\succsim$. 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 $u' = a_1 u_1^* + a_2 u_2^* + a_3 u_3^*$, we see that the coefficients a_i can be chosen to be negative, e.g.,

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

■

5 Conclusion

The three theorems of the paper differ only by the technical assumptions they make, and their ethical import lies with the conclusion, obtained each time, that a social observer whose preferences on lotteries meet the conditions of the Aggregation Theorem must follow a weighted sum rule $\sum_i a_i u_i^*$. That the coefficients may be unequal or (in the last theorem) nonpositive is a weakness from the perspective of classical utilitarians like Bentham. However, weighted utilitarianism has some theoretical standing, and the measurement stage is anyhow the decisive one on the road to Benthamism. What matters most are the u_i^* appearing in the formula.

By introducing a utilitarian observer at the outset, we make a salient addition to Harsanyi, in a way that is fully consistent with Sen's point that utilitarianism has to be defined independently within the Harsanyi framework, or else the results will bear no connection with this doctrine. The exogenously given u_i^* provide the desired basis of cardinal utility measurement. Thus, we take for granted that *nonrisky alternatives* are evaluated in the cardinal utilitarian way, the point of the theorems being to disseminate this evaluation to *lotteries*. Eventually, there will be a unique cardinalization for both items, which Harsanyi claimed without adducing any reason. If one accepts our conceptual addition of the utilitarian observer — as well as, of course, the added technical restrictions — Sen's point appears to be fully answered.

Perhaps less obviously, Weymark's point is also answered. Our assumptions about individual utility functions u_i take them to be representations of ordinal VNM preferences, but the theorems invest them with a cardinal meaning, which is furthermore utilitarian in the sense just said. Briefly put, non-affine φ_i drop out from the social observer's criterion $\sum_i \varphi_i^{-1} \circ u_i^*$. As the proof goes, this follows because the utilitarian observer has evaluated lotteries by $E \sum_i u_i^*$, not by some more general $E \sum_i \varphi_i \circ u_i^*$, which in turn follows because the two conditions of the Aggregation Theorem have been applied to a utilitarian preference over lotteries that extends the utilitarian preference over sure outcomes.

From this summary, there remain only two ways of opposing the conclusion that the theorem confers ethical relevance to utilitarianism. One is to deny that a utilitarian preference over lotteries necessarily satisfies the stated conditions - the VNM one being more problematic than the Pareto principle. Not to mention Harsanyi himself, those economists, like Hammond (1996) and Mongin and d'Aspremont (1998), who updated Bentham's doctrine, do endorse the conditions. While we are not aware of utilitarian theories that would violate them, such theories can be conceived of. The other move is of

course to deny that the conditions are appealing in and of themselves. It has been argued that the VNM conditions are questionable for a social observer (e.g., Diamond, 1967), and that the Pareto principle is not compelling in a lottery context (e.g., Fleurbaey 2010). All these objections are in some sense secondary to the pivotal claim that the Aggregation Theorem has nothing to do with utilitarianism as an ethical doctrine. The aim of this paper was to debunk this claim.

Acknowledgements. Many thanks to the participants to the Conference in Honour of Peter Hammond (Department of Economics, University of Warwick, March 2010), to the D-TEA Conference (HEC Paris, June 2010), and to the Conference in Honor of Claude d'Aspremont and Jean-François Mertens (CORE, June 2011), where this paper was given.

6 References

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.

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.

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.

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.

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 1) Let (g_n) be a sequence in $g(X)$ s.t. $g_n \to g_0 \in g(X)$ as $n \to \infty$. We prove by reductio that $f(g_n) \to f(g_0)$ as $n \to \infty$. Suppose not; one can then find $\varepsilon > 0$ and a subsequence (g_m) s.t. $|f(g_m) - f(g_0)| > \varepsilon$ for all m. In this subsequence, we select a weakly monotonic subsequence (g_p) s.t. $g_p \to g_0$ as $p \to \infty$; necessarily, $g_0 \neq g_p$ for all p. Take $x_0, x_1 \in X$ satisfying $g(x_0) = g_0$ and $g(x_1) = g_1$. The set X is path-connected, so there is a continuous function $t : [0,1] \to X$ s.t. $t(0) = x_0$ and $t(1) = x_1$. By continuity of $g \circ t$ and the properties of (g_p), the set $g \circ t([0,1])$ is a compact nondegenerate interval having endpoints $g \circ t(1)$ and $g \circ t(0)$, and this interval contains every g_p. The intermediate value theorem for $g \circ t$ ensures that, for every $p \geq 1$, there is $a_p \in [0,1]$ s.t. $g(t(a_p)) = g_p$. Let $x_p = t(a_p)$ for all p. The sequence so constructed is contained in the compact set $t([0,1]) \subset X$, so it has a subsequence (x_q) converging to some $x^* \in X$, and by continuity of g, $g(x_q) \to g(x^*)$ as $q \to \infty$. But $(g(x_q))$ is a subsequence of (g_p), which has been said to converge to g_0, whence $g(x^*) = g_0$. By continuity of h, $h(x_q) = f((g(x_q)) \to h(x^*) = f(g_0)$ as $q \to \infty$. This is a contradiction, because no subsequence of $(f(g_m))$ can converge to $f(g_0)$. ■

Note: Lemma 1 also holds under the assumption that X is a compact metric set, and this is shown by a related mathematical argument.

Proof. (Lemma 2) A proof for $k \geq 2$ from Rado and Baker's (1987) statement for $k = 2$ is available on request. ■

Note: A significant advantage of Lemma 2 is that it does not impose a Cartesian product domain on the functions f and f_i; nor does it impose more than a mild continuity restriction. Compare with the less general results surveyed in Eichhorn (1978).