The Utilitarian Relevance of the Aggregation Theorem

Marc Fleurbaey,*Philippe Mongin†

October 17, 2012

Abstract

Harsanyi invested his Aggregation Theorem and Impartial Observer Theorem with utilitarian sense, but Sen redescribed them as "representation theorems" with little ethical import. This negative view has gained wide acquiescence among economists. Against it, we support the utilitarian interpretation by a novel argument relative to the Aggregation Theorem. We suppose that an exogeneously defined utilitarian observer evaluates social states by the sum of individual utilities and we apply the assumptions of the Aggregation Theorem to this observer. Adding technical conditions from microeconomics, we conclude that *any* social observer who is subjected to the assumptions of the Aggregation Theorem evaluates social states in terms of a weighted variant of the utilitarian sum. Hence, *pace* Sen, utilitarianism and the Aggregation Theorem are mutually relevant. The argument is conveyed by means of a main theorem, an algebraic refinement of this theorem, and a variant in which the utilitarian sum is unconventionally defined on lotteries. Each result encapsulates Harsanyi's original one as a particular step.

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

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

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

JEL Classification: D63, D71, D81.

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 novel arguments 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, i.e., lotteries. Unlike him, we assume that there exists a social observer who has formed social preferences on sure alternatives according to the sum rule of classical utilitarianism, that this observer's preferences can be extended to lotteries so as to satisfy the VNM axioms, and that the VNM extension in question satisfies the Pareto principle with respect to individual preferences. The last two conditions are those which the Aggregation Theorem imposes on *any* social preferences, and the novelty here is to suppose that they apply to classical utilitarian preferences. We move to Harsanyi's step only after exploring the consequences of this supposition. Under added conditions borrowed from microeconomics, it turns out to deliver significant information on the VNM indexes of the individuals and the utilitarian observer, and it follows by a surprising connection that, if Harsanyi's non-descript observer is now required to satisfy the same conditions of being VNM and Paretian on lotteries, *he is weighted utilitarian*, in the sense of evaluating sure alternatives by means of a weighted variant of the given classical utilitarian sum.

This new theorem, the proof of which encapsulates Harsanyi's as a mere lemma, is exemplified in several ways. Our salient result, Theorem 1, assumes for simplicity that the utilitarian sum is defined on a full-dimensional individual utility set. Since this requires assumptions about the diversity of individuals, we propose the more technical Theorem 2, which relies on a weaker dimensionality assumption. Both of these results follow the tradition of defining classical utilitarianism on sure alternatives, by considering extensions of the doctrine to lotteries only as convenient demonstrative tools. Moreover, they specialize the sure alternatives by taking them to be allocations of commodities to the individuals —this justifies introducing microeconomic conditions. However, under the influence of social choice theory, 20th century social ethics often envisages alternatives more abstractly, and Harsanyi himself leant in this direction. Thus, we also propose the variant Theorem 3, in which the lottery set is basic and lotteries become the primary objects of social evaluation. We doubt that classical utilitarians, in the Benthamite and Millian way, would be at ease with this analytical framework, but some 20th century utilitarian followers of Harsanyi clearly are (see in particular Hammond, 1979, and Mirrlees, 1979; more review in Mongin and d'Aspremont, 1998) .

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

True, we only conclude that social observers add up the cardinal utility functions fixed by the utilitarian observer, not that they give them equal weight as he does. However, *weighted utilitarianism*, as d'Aspremont and Gevers (2002) label it, goes a long way towards utilitarianism proper.[1] We emphasize the utilitarian intent of our formulas, because weighted sums can be used to evaluate social states in a non-utilitarian way, and this has precisely been offered as a solution to the conundrum raised by Harsanyi's theorems. Each time in a different way, Karni (1998), Dhillon and Mertens (1999), and Segal (2000) illustrates this alternative line of analysis.

Section 2 develops 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 explore the normative plausibility of utilitarianism and its prob-

[1] Weighted utilitarianism has also been explored in connection with Harsanyi's other theorem (see in particular Mongin, 2001, and Grant, Kajii, Polak and Safra, 2010).

lematic 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 has utilitarian relevance. Section 3 sets up the formal framework with its microeconomic assumptions. Section 4 states the main theorem, and section 5 the two complementary results. 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).

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. 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 $v = \varphi \circ (\sum_i \varphi_i^{-1} \circ v_i)$, 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 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 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 $\mathcal{U}$ or $\mathcal{U}'$ rather than $\mathcal{F}$ or $\mathcal{F}'$. This will be referred to as *Weymark's point.*

We are also concerned with the social rule of a utilitarian observer, and we define it as follows. In this 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 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 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. 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 does not deliver cardinal equivalence with u_i; that is, one gets $u_i^* \in \mathcal{F}_i$, and not $u_i^* \in \mathcal{U}_i$, as establishing a connection with utilitarianism would require; 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 is the most transparent criticism in Sen's quotations, henceforth *Sen's point*.

In sum, two distinct problems stand in the way of Harsanyi's utilitarian interpretation of his results. 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 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, a cardinal preference is a relation on *pairs* of sure alternatives, i.e., $(x, y) \succsim_i^* (z, w)$, and the axioms for $\succsim_i^*$ embody the 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 these comparisons, despite the difference between lotteries and sure alternatives. This axiom will lead to the conclusion 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 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 as an argument for this position, it is question-begging, because the axioms on $\succsim_i^*$ are merely a 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, and we use a non-obvious argument instead. As will be seen, it also takes care of Sen's and Weymark's points at the same time.

3 The framework and assumptions

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. 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 case of private goods, it may be inappropriate if X takes the availability of resources into account. We require connectedness, which is less restrictive than the convexity assumption of standard microeconomics.

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

The VNM apparatus can now be introduced formally. When it applies to the social observer, all 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 these systems do not 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 sets of X, 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. 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 Theorems 2 and 3 (1972, p. 48-49) apply to X and $\Delta(X)$ 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 $\Delta(X)$. 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 part 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 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 $\mathbb{R}^n$ and is such that $U^*(X) \subseteq \overline{U^*(X)^\circ}$, i.e., it 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 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 regularity assumption. In another respect, it requires $U^*(X)$ to have full affine dimension n.[2] This is not demanding under standard microeconomic conditions. If there are private consumption goods, if each individual is concerned only with how much he consumes, and if 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 (for instance, no two individuals can be alike in the 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

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

would in essence fall back on the equivalence $(*)$ of the 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 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.*

We complete the groundwork for the next section by stating a functional equation theorem. For $k = 2$, it follows from Rado and Baker's (1987) Theorem 1 (see also their Corollary1). No proof being available for $k \geq 2$, we provide one here.[3] 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*

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

for all $(z_1, ..., z_n) \in T$. Suppose that f is continuous on T_+. 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.

[3] When extending Harsanyi's Aggregation Theorem to a framework of state-contingent alternatives, Blackorby, Donaldson and Weymark (1999) implicitly use the result proved here.

4 A theorem on 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 we adopt here a more assertive form based on Strong Pareto.[4]

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.

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

The following assumption is the cornerstone of our conceptual and mathematical argument. It amounts to applying to the *given* utilitarian preference the two assumptions that Harsanyi wanted immediately to apply to *any* social observer.

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 $\succsim_i$.

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 its PAT. 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:

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

As the left-hand side is increasing in every u_i^*, each of them being non-constant (see fn 2), necessarily $b_i > 0$ for all i.

As a metric connected space, X is path-connected, hence 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.

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

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

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 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 $x \in X$,

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

From these results, 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. 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. $Eu' = \sum\limits_{i=1}^{n} a_i E u_i^* + b$, 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^* (see equation ($2''$)). 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 last section amounts to assuming what is proved here in a roundabout way. 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.

The first part of the proof delivers information on the utilitarian observer that we single out in a separate statement. Prima facie, the preference ordering on L that is represented by $E\sum_i u_i^*$ is merely one among the extensions $\succsim^{*ext}$. That it is the only one follows as a conclusion (see ($1''$)). Another result of interest is that each Eu_i^* turns out to be a VNM utility function for $\succsim_i$ on L (see ($2''$)). That is to say, under the present assumptions, not only are individual VNM utility functions on L cardinally meaningful by themselves, but they are also cardinally comparable, and the classical utilitarian formula can be extended from X to L.

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.*

5 Other forms of the theorem

As a technical refinement of Theorem 1, we show how Assumption 4 can be weakened so as to accommodate a less than full dimensional utility set $U^*(X)$. Take a vector $\varphi = (\varphi_1, ..., \varphi_k)$ of real-valued functions on some domain Y, and denote by $\overline{\succsim}_i$ the preference ordering on Y represented by φ_i. We say that *Independent Prospects* (IP) holds with respect to $\overline{\succsim}_1, ..., \overline{\succsim}_k$ if, for all $i = 1, ..., k$, there exist $y^i, z^i \in Y$ such that $y^i \succ_i z^i$ and $y^i \sim_j z^i$, $j \neq i$. This entails that the φ_i are affinely independent, but the converse does not hold in general.[5] However, it does when the φ_i are the utility functions of this paper, and we will often replace affine independence by this more interpretable condition.

Proposition 2 *Suppose that $\varphi = (\varphi_1, ..., \varphi_k)$ represent $(\overline{\succsim}_1, ..., \overline{\succsim}_k)$ on Y, and that either*

[5] Recall that $\varphi_1, ..., \varphi_j, ..., \varphi_k$ are *affinely independent* if no φ_j can be written as an affine function of the others.

(i) $\varphi(Y)$ has a non-empty interior in $\mathbb{R}^k$, or (ii) Y is a lottery set, the orderings $\overline{\succsim}_i$ are VNM preferences, and the functions φ_i are affinely independent VNM representations of these preferences. Then, IP holds.

Proof. See appendix. ■

Now, consider the effect of weakening Assumption 4 by requiring that the utility set $U^*(X)$ 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 Lemma 4 applied to $Y = X$ and the k-dimensional subspace of $\mathbb{R}^n$, IP holds of the orderings represented by $u_1^*, ..., u_k^*$, and since $X \subset L$, Assumptions 2, 3 and 5 entail that IP also holds of $\succsim_1, ..., \succsim_k$. 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 $\{v_1, ..., v_n\}$, that is to say in preference terms, IP can be realized on L for more individuals than it is on X. The only conclusion is that $\{v_1, ..., v_k\}$ can be enlarged to a set $\{v_1, ..., v_{k'}\}$, with $k' \geq k'$, so as to deliver an affine basis for $\{v_1, ..., v_n\}$. This analysis motivates the new 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 IP applies to the orderings represented by $u_{j_1}^*, ..., u_{j_k}^*$, and moreover, for any strict superset $\{j_1, ..., j_{k'}\}$, IP does not 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 $k = n$, $\{u_1^*, ..., u_n^*\}$ has full affine dimension, and Assumption 4' reduces to Assumption 4. Thus, we have managed to weaken this assumption, though not at the highest possible level of generality. We are now ready for a stronger form of Theorem 1.

Theorem 2 *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 also hold.*

Proof. See appendix. ■

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^*, and this may be criticized as lacking generality. Our assumptions 1-4 (or 4') define a restricted domain by the standards of social ethics, where rules are often discussed in terms of abstract social states. Sen's point refers to classical utilitarianism, which often makes special economic assumptions, but Weymark's point does not, and in its original version, Harsanyi's Aggregation Theorem hinges only on the structure of $\Delta(X)$, regardless of what X may be. This motivates devising a variant theorem in which the main assumptions directly concern $\Delta(X)$ 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. The earlier analysis used the continuity of u_i^* and the dimensionality of $U^*(X)$, and we will now need these properties to hold on v_i^* and $V^*(L)$. Grandmont's (1972) theorems will be invoked again, and since the underlying X does not matter here, we may as well take up his abstract assumption for this set.

Assumption 1': X is a separable metric space.

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 3 *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 appendix. ■

Theorem 3 is related to Theorem 1 by its full dimensionality assumption, but a stronger variant can be devised to parallel Theorem 2, in which no such assumption is made. Theorem 3 is more economical than Theorem 1 in terms of the other assumptions, but this is only because $\Delta(X)$ has a preexisting structure, whereas conditions must be postulated on X; by and large, the convexity of $\Delta(X)$ plays the role of Assumption 3. In the end, Theorem 3 carries less information, since there is no analogue of Proposition 1 (but each v_i^* turns out to be a VNM function for $\succsim_i$on L, a relevant finding).

Conceptually, it seems to us questionable to define utilitarianism directly as a sum of utilities over lotteries. The classical version never took into consideration the uncertainty context, and one may wonder if our strategy of fixing an exogeneous benchmark for the doctrine remains plausible in this case. Theorem 3 *is devised for the sake of formal generality* and in order to respond to the social choice tradition of minimizing substantial assumptions.[6]

6 Conclusion

The normative import of our analysis lies with the conclusion, obtained each time, that a social observer whose preferences on lotteries meet the two conditions of the Aggregation Theorem must follow a weighted sum rule $\sum_i a_i u_i^*$. That the a_i may be unequal or even (in one theorem) nonpositive is a weakness from the perspective of classical utilitarianism. However, weighted utilitarianism has some theoretical standing by itself, and the measurement stage is anyhow the decisive one on the road to full Benthamism. To rely on cardinally comparable u_i^* is more momentous than to give each of them the same role.

By introducing a utilitarian observer at the outset, we followed Sen's suggestion that utilitarianism had to be defined *independently* in the analytic framework, or else the results would bear no connection with this doctrine. The exogenously given sum $\sum_i u_i^*$ (or $\sum_i v_i^*$ in the variant) provides the desired basis for a comparison with the preference orderings studied in the theorems. Sen's point appears to be fully answered precisely by making this addition and drawing the full mathematical consequences.

Perhaps less obviously, Weymark's point is also answered. Our assumptions about individual utility functions u_i (or v_i) take them to be representations of ordinal VNM preferences, but the theorems invest them with a cardinal meaning. Technically put, 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. As the proof goes, this crucial simplification follows from applying the two conditions of the Aggregation Theorem to the given utilitarian preference before applying them to Harsanyi's nondescript social

[6] Segal (2000) also takes the view that allocations of goods, rather than abstract social states, should be evaluated.

preference.

To sum up, when the theorem is properly reconstructed, utilitarianism becomes relevant to it, and conversely, it provides an argument for the ethical relevance of that doctrine. There seem to be only two ways to block this conclusion. One is to reject Assumptions **H** or **H'**, to the effect that a *utilitarian* preference over lotteries should satisfy the VNM and Pareto conditions. Although one may abstractly think of a utilitarian stance that would reject H or H', we are not aware of any theorist to defend it. For Harsanyi and his predecessor Vickrey (1945), the VNM and Pareto conditions had universal normative appeal, and they would apply them as a matter of course to our classical utilitarian observer. Today's economists who pushed Harsanyi's work further either are on the same line, or are not full-fledged utilitarian.[7] 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). 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.

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. Special thanks are due to Marcus Pivato and Uzi Segal for their comments on a previous version.

[7] Hammond (1982, 1996) is perhaps exceptional in illustrating the first attitude. In various ways, Weymark (1993), Mongin and d'Aspremont (1998), Blackorby, Donaldson and Weymark (1999), Mongin (2001), Grant, Kajii, Polak and Safra (2010) exemplify the second attitude. Karni (1998), Dhillon and Mertens (1999), Segal (2000) distance themselves even more from the utilitarian tradition.

7 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.

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

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.

Kannapan P. 2009, *Functional Equations and Inequalities with Applications*, Dordrecht, Springer.

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.

Rockafellar, R.T. (1970), *Convex Analysis*, Princeton, Princeton University Press.

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.

Vickrey, W. (1945), "Measuring Marginal Utility By Reaction to Risk", Econometrica, 13, p. 319-333.

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.

8 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) For a given $z \in T$, there is $\varepsilon > 0$ such that the set $B_\varepsilon =]-\varepsilon, +\varepsilon[^n$ satisfy the two conditions that $z + B_\varepsilon \subset T$ and $z + (B_\varepsilon)_+ \subset T_+$. Thus, we may define new functions, g on $(B_\varepsilon)_+$ and g_i on $(B_\varepsilon)_i =]-\varepsilon, +\varepsilon[$ as follows: for all $y \in B_\varepsilon$,

$$\begin{aligned} g(\sum_{i=1}^n y_i) &= f\left(\sum_{i=1}^n y_i + \sum_{i=1}^n z_i\right) \\ \text{and } g_i(y_i) &= f_i(y_i + z_i), i = 1, ..., n. \end{aligned}$$

The functional equation on f, f_i entails the following one on g, g_i: for all $y \in B_\varepsilon$,

$$(*) \quad g\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n g_i(y_i).$$

Now, for $y \in B_\varepsilon$ and $i = 1, ..., n$, $(0, ..., y^i, ..., 0) \in B_\varepsilon$, so by (*),

$$g(y_i) = g_i(y_i) + \sum_{j \neq i} g_j(0),$$

Thus (*) becomes

$$(**) \quad g\left(\sum_{i=1}^n y_i\right) = \sum_{i=1}^n \left[g(y_i) - \sum_{j \neq i} g_j(0)\right].$$

Define $c_i = g_i(0)$ and the function φ such that $\phi(x) = g(x) - \sum_{i=1}^n c_i$ for all $x \in (B_\varepsilon)_+$. We may rewrite (**) as

$$\begin{aligned} \phi\left(\sum_{i=1}^n y_i\right) &= \sum_{i=1}^n g_i(y_i) - \sum_{i=1}^n c_i \\ &= \sum_{i=1}^n \left[g(y_i) - \sum_{j \neq i} c_j\right] - \sum_{i=1}^n c_i \\ &= \sum_{i=1}^n \left[g(y_i) - \sum_{i=1}^n c_i\right] = \sum_{i=1}^n \phi(y_i). \end{aligned}$$

This shows that ϕ satisfies a Cauchy equation $\varphi(x+x') = \varphi(x) + \varphi(x')$ for all $x, x', x+x' \in \left]-\varepsilon, +\varepsilon\right[$. Since φ is continuous like g and f, we may conclude that ϕ is linear on the relevant domain, i.e., there is $a \in \mathbb{R}$ such that $\phi(x) = ax$ for all $x \in \left]-n\varepsilon, +n\varepsilon\right[$.[8] Returning to the ancestor functions, we conclude that for all $y \in B_\varepsilon$,

$$\begin{aligned} f(\sum_{i=1}^{n}(y_i+z_i)) &= a^z\sum_{i=1}^{n}y_i + f(\sum_{i=1}^{n}z_i) \\ \text{and } f_i(y_i+z_i) &= a^z y_i + f_i(z_i), i=1,...,n, \end{aligned}$$

where the indexing a^z expresses that this constant depends on the chosen $z \in T$. Using the equality

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

we may restate the equations as follows: for all $w \in z + B_\varepsilon$,

$$\begin{aligned} (***) \quad f(\sum_{i=1}^{n}w_i) &= a^z\sum_{i=1}^{n}w_i + \sum_{i=1}^{n}b_i^z \\ \text{and } f_i(w_i) &= -a^z w_i + f_i(z_i), i=1,...,n, \end{aligned}$$

where $b_i^z = -a^z z_i + f(z_i)$, $i = 1, ..., n$ are constants depending on z.

For any $z' \in T$, $z' \neq z$, there is a set $B_{\varepsilon'}$ like B_ε above, and the previous reasoning delivers solutions for f, f_i on $z' + B_{\varepsilon'}$ that obey the form of equations (***) with constant $a^{z'}$and $b_i^{z'}$, $i = 1, ..., n$. If $z + B_\varepsilon$ and $z' + B_{\varepsilon'}$ have non-empty intersection, the conclusion readily follows that $a^z = a^{z'}$ and $b_i^z = b_i^{z'}$, $i = 1, ..., n$. This conclusion extends to the case in which $z + B_\varepsilon$ and $z' + B_\varepsilon$ do not overlap by a classic argument based on path-connectedness (as in the final step of Rado and Baker, 1987, p. 232). Hence dependency on z is cancelled, which completes the proof. ■

[8] The steps are as follows: 1) ϕ satisfies the Cauchy equation for $x, x', x+x' \in \left]-\varepsilon, +\varepsilon\right[$; 2) therefore ϕ has a unique extension to $\mathbb{R}$ that satisfies the Cauchy equation (Kannapan 2009, p.57); 3) any continuous solution to the Cauchy equation on $\mathbb{R}$ is linear (Kannapan 2009, p. 3).

Note: A significant advantage of Lemma 2 is that it does not impose a Cartesian product domain on the functions f and f_i; compare with the less general results surveyed in Eichhorn (1978).

Proof. (Theorem 2) We fix $V = (Eu_1, ..., Eu_n)$ and Eu as in the previous proof, assume w.l.g. that the set K of first k individuals defines an affine basis for both U^* and V, and using this fact, observe that the Pareto indifference condition can be stated in terms of K alone. Thus, Lemma 3 leads to

$$Eu = \sum_{i=1}^{k} b_i E u_i.$$

Utility functions of individuals outside K 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 Theorem 1 becomes

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

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 earlier equation on $T = (u_i^*(X))_{i=1,...,n}$. ($\widetilde{T}^\circ$can be viewed as a full-dimensional subset of $\mathbb{R}^k$ satisfying the conditions of Lemma 2, and the affine solutions on $\widetilde{T}^\circ$ can be extended by continuity to $\widetilde{T}$.) It follows that u is a PAT of $\sum_{i=1}^{k} \widetilde{u_i^*} = \sum_{i=1}^{n} u_i^*$, and that, for $i = 1, ..., k$, the u_i are PAT of $u_i^*$with the same positive factor, which establishes the claims made in Proposition 1.

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

$$u' = \sum_{i=1}^{k} a'_i u^*_i + b'$$

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

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

for appropriate coefficients that will not be unique if $k < n$. So the conclusion of Theorem 1 holds with the proviso said in the statement.

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.,

$$u' = 4u^*_1 + 4u^*_2 - u^*_3 = -u^*_1 - u^*_2 + 1.5u^*_3\ .$$

■

Proof. (**Proposition 2**) By assumption (i), there are $\varphi^* \in \varphi(Y)^\circ$ and $\varepsilon_i > 0$, $i = 1, ..., k$, s.t.

$$\varphi^{*i} = (\varphi^*_1, ..., \varphi^*_i + \varepsilon_i, ..., \varphi^*_k) \in \varphi(Y).$$

Take $y \in Y$ s.t. $\varphi(y) = \varphi^*$, and for each $i = 1, ..., k$, $y^i \in Y$ s.t. $\varphi(y^i) = \varphi^{*i}$. Then, IP holds with $z^i = y$, $i = 1, ..., k$.

By (ii), $\varphi(Y)$ is a convex set as the image of a vector of VNM functions, and it has full affine dimension, hence it has non-empty interior in $\mathbb{R}^k$ (see, e.g., Rockafellar, 1970, section 6), so that the previous argument applies. ■

Proof. **(Theorem 3)** 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 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 connected and (since it is metric) 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 Theorems 1 and 2. ■