7

A Welfarist Version of Harsanyi's Aggregation Theorem

Claude d'Aspremont and Philippe Mongin

7.1 Introduction

The Aggregation Theorem is one of the main arguments used by Harsanyi in support of utilitarian ethics. It was first presented in his 1955 article and further developed in chapter 4 of his book in 1977. Since then, several authors[1] have constructed alternative proofs of this theorem in more general settings. It is generally presented as relating a "single profile" of individual utility functions $\{U_i\}$, to the utility function W of a moral observer by means of the Pareto Indifference rule. In this context, the theorem states that if all utility functions (including the moral observer's) are von Neumann–Morgenstern (VNM), then the moral observer's utility is an affine transformation of the individual utilities, that is, $W = \sum \beta_i U_i + \gamma$.

The relevance of this result in giving proper foundations to utilitarianism has been questioned on several grounds. First, the weights $\{\beta_i\}$ are not necessarily positive, and hence the welfare of some individuals might not affect, or worse, might negatively affect total welfare. This first problem can be solved quite naturally by strengthening Pareto Indifference into the

[1] See Domotor (1979), Border (1981, 1985), Fishburn (1984), Selinger (1986), Coulhon and Mongin (1989), Hammond (1992), and Weymark (1993).

This chapter was first presented at the conference on Justice, Political Liberalism, and Utilitarianism held in honour of John Harsanyi and John Rawls at the University of Caen, June 1996. We are grateful to Charles Blackorby, Vincenzo Denicolo, David Donaldson, Marc Fleurbaey, Louis Gevers, and Serge Kolm for useful discussions and comments. Special thanks are due to John Weymark for suggestions and encouragement in the writing of this final version. This text presents research results of the Belgian Program on Interuniversity Poles of Attraction (Science Policy Programming, Prime Minister's Office). Scientific responsibility is assumed by its authors.

Strong Pareto condition; the latter implies that all weights are positive.[2] A second problem is that the weights might not be uniquely defined, creating an indeterminacy. This further problem can be solved by introducing an additional condition, called Independent Prospects, which says that for every individual there exists a pair of lotteries for which that individual alone is not indifferent.[3] The third problem, which is one of the main objections formulated by Sen (1986) against the Aggregation Theorem as an axiomatisation of utilitarianism, is that the weights cannot be determined independently of the utility functions to be aggregated; indeed, if the β_i's are functions of the U_i's, the formula is different from a utilitarian rule. A related issue is how to obtain, in the context of Harsanyi's theorem, the pure classical utilitarian rule, with all weights equal to 1 (Bentham's sum rule) or to $1/n$ (as in the average utility rule). To determine the weights independently from the given utilities, and eventually to get equal weights by introducing a symmetry condition, one needs to consider a more general framework, allowing the utility profiles to vary. As suggested in Coulhon and Mongin (1989) and Mongin (1994), this can be conveniently done in Sen's (1970) framework of social welfare functionals (SWFLs). The Aggregation Theorem can then be reformulated so as to give an axiomatisation which, at least formally, relates to Utilitarianism.[4]

This chapter elaborates on this reformulation. It will not, though, start from Sen's multiprofile approach – with SWFLs defined on some universal domain – but instead from the "enlarged" single-profile approach used in Roberts (1980a) and d'Aspremont (1985), with SWFLs being defined on a restricted domain. More specifically, we will closely follow Harsanyi in assuming a *single profile of individual VNM preferences* and allow for *multiple profiles of VNM utility functions*, representing these given preferences. Following an argument in d'Aspremont (1985), this will be sufficient to obtain a VNM version of Welfarism and, thus, to introduce conditions that are usually stated in the multi-profile approach. One such condition is Anonymity, which will imply a symmetric formula. Another is Cardinality and Unit Comparability, an invariance axiom that allows for (some version of) interpersonal comparisons of utility differences. Following Mongin (1994), this cardinal condition will be shown to result from assigning VNM utility

2 See Domotor (1979), Weymark (1993), and De Meyer and Mongin (1995). This was already suggested by Harsanyi (1955).

3 See Fishburn (1984) and Coulhon and Mongin (1989). This condition was used implicitly, as a structural assumption, by Harsanyi in the proof of the Aggregation Theorem. Domotor (1979) and Border (1981) showed that it was not needed.

4 See also Mongin and d'Aspremont (1998).

functions to the individuals and VNM preferences to Harsanyi's moral observer. Our results are closely related to the ones given by Blackorby, Donaldson, and Weymark (2008). They also investigate how the expected utility hypothesis, combined with a Paretian condition, can provide support for utilitarianism. However, their investigation is done for other domains than the one considered here.

This chapter is organized as follows. In the next section, we define a social welfare functional restricted to the domain of all VNM representations of a given single profile of VNM individual preferences, and state the corresponding Aggregation Theorem. Then, in Section 7.3, we derive Welfarism, prove the theorem, and derive the VNM characterizations of pure and generalized utilitarian rules. Finally, in the concluding section, we show that, under the VNM domain restriction adopted here, two standard cardinality notions are equivalent.

7.2 A SWFL Version of the Aggregation Theorem for a Single Profile of VNM Preferences

The social choice problem to be considered here is defined by a set of individuals $N = \{1, 2, \ldots, n\}$, a set of social states X, and a "moral observer." According to Harsanyi's approach, the moral observer is any individual, adopting a moral point of view and forming moral preferences (to be distinguished from this individual's personal preferences). But the moral rule to be finally determined should, in principle, be the same for each individual. The objective is to derive (some version of) the utilitarian rule. The set X of social states is not precisely interpreted. Following Harsanyi, who claims to be a rule-utilitarian, it could be the set of all possible rules to constrain individual behavior (including all sorts of possible amendments), or, more specifically, some given set of possible rules and all probability mixtures (i.e., lotteries) on this set. Formally, X is supposed to be a convex subset (which is not a singleton) of some vector space: for any $x, y \in X$ and any $\lambda \in [0, 1]$, the convex combination (or mixture) $[\lambda x + (1 - \lambda)y]$ is also in X.

For any set Ξ (which may be X or some other convex set that will be introduced in the sequel), a *preference ordering* R on Ξ is a reflexive, complete and transitive binary relation on Ξ. Moreover, it is a *VNM preference ordering on* Ξ if it satisfies in addition:

Continuity: $\forall a, b, c \in \Xi$, the sets $\{\lambda \in [0, 1] : c R[\lambda a + (1 - \lambda)b]\}$ and $\{\lambda \in [0, 1] : [\lambda a + (1 - \lambda)b] Rc\}$ are closed in $[0, 1]$.

Independence: $\forall a, b, c \in \Xi, \forall \lambda \in]0, 1], a Rb \Leftrightarrow [\lambda a + (1 - \lambda)c] R[\lambda b + (1 - \lambda)c]$.

A VNM preference ordering R can always be represented by a utility function υ defined on Ξ: $\forall a, b \in \Xi$, $aRb \iff \upsilon(a) \geq \upsilon(b)$. Moreover, in this framework, every utility representation of R is either *mixture-preserving*, that is,

$$\forall a, b \in \Xi, \forall \lambda \in [0, 1], \upsilon(\lambda a + (1 - \lambda)b) = \lambda \upsilon(a) + (1 - \lambda)\upsilon(b),$$

or a monotone transformation of a mixture-preserving utility function. A mixture-preserving utility representation is called a *VNM utility function.*

We start with a given *single profile* of individual preference orderings $(\bar{R}_i)_{i \in N}$. This will remain fixed throughout. Our first assumption is that each $\bar{R}_i$ is a VNM preference ordering on the set X of social states, which is *nontrivial* in the sense of being different from total indifference (for each $i \in N$, there exist $x, y \in X$ such that i strictly prefers x to y: $x \bar{P}_i y$). A *social welfare functional* (SWFL) is a function F associating to each utility profile $U = (U_1, U_2, ..., U_n)$ in some admissible domain $\mathcal{D}$, a preference ordering $F(U)$ on X. The objective here is to associate to the single profile of VNM preferences $(\bar{R}_i)_{i \in N}$ a particular SWFL $\bar{F}$, satisfying a set of conditions. The first three conditions are directly linked to Harsanyi's basic axioms: The first determines the domain of the SWFL, the second fixes its range, and the third is a strenghtening of Pareto Indifference. The last axiom is a weakening of the structural assumption used by Harsanyi.

VNM-Utility Domain (VNM-D): For every $i \in N$, $\bar{R}_i$ is a nontrivial VNM preference ordering on X. The domain of $\bar{F}$ is the set $[\bar{U}]$ of all vectors of possible individual VNM utility representations of $(\bar{R}_i)_{i \in N}$; that is, $\mathcal{D} = [\bar{U}]$.

VNM-Range (VNM-R): For any $U \in [\bar{U}]$, the moral observer's preference ordering $\bar{F}(U)$ satisfies continuity (**VNM1-R**) and independence (**VNM2-R**).

These first two conditions reflect Harsanyi's commitment to the VNM preference axioms as a norm of rationality for both the personal and moral preferences. It will become clear that the restriction on the domain, as well as the restriction on the range of the SWFL, plays an important role in moving away from an ordinal noncomparable framework and in giving some ethical relevance to the rules that will be derived. Indeed, Harsanyi's choice to restrict consideration to the class of VNM (i.e., mixture-preserving) utility representations of each individual preference is used in the next section to transpose the VNM-Range condition (both continuity and independence) to the welfarist framework (obtained after the last two conditions have been introduced). Then, eventually, the welfarist version of

VNM-independence will be shown equivalent to cardinality and interpersonal unit comparability. This fact gives some foundation to Harsanyi's claim [or Vickrey's (1945)] for basing the determination of moral preferences on individual attitudes toward risk or, more precisely, on the various factors explaining these attitudes. In other terms, VNM representations of individual preferences provide cardinal information to a VNM rational moral observer.

This claim should not be interpreted as meaning that an individual's risk attitudes are not already contained in his VNM preferences (the primitive of expected utility theory) and cannot be represented by nonmixture preserving utility functions. This is justly stressed by Sen (1986), Weymark (1991), and Blackorby, Donaldson, and Weymark (2008). For example, taking a single-dimension outcome space and any one particular VNM utility representation of some individual's VNM preference relation, we can compute the corresponding Arrow-Pratt (absolute or relative) measure of risk aversion. It is a consequence of the VNM theorem that this piece of information about this individual's risk attitude can be recovered from any other utility representation of his VNM preferences because any such representation has to be a monotone transformation of the VNM utility used in computing the measure. However, it is only when this other representation is itself VNM (i.e., taking the transformation to be positive affine) that one can recompute directly (without making some preliminary ordinal retransformation) the Arrow-Pratt measure and get the same number. In other words, the Arrow-Pratt measure is only invariant to positive affine transformations. It is this particular invariance property, holding within the class of all VNM representations, that is exploited in Harsanyi's approach (as described for instance by our four axioms), the objective being not simply to get an evaluation of each single individual's risk attitude but to allow for interpersonal comparisons of risk attitudes, that is, to make sense of statements such as "individual i is more risk averse than individual j, in the Arrow-Pratt sense."

What we want to stress here, though, is that such an interpersonal cardinalization is not a consequence of just the domain restriction but of a combination of this restriction and of the one limiting the range of the SWFL to VNM preference orderings on X. Moreover, we need the other two conditions.

The third one replaces Harsanyi's Pareto Indifference. Having adopted an *enlarged* single-profile approach, Pareto Indifference needs to be strengthened. It is replaced by a neutrality condition, restricted to the set of VNM utility representations.

Relative Neutrality (RN): For any $U, U' \in [\bar{U}]$, any two pairs $\{x, y\}$ and $\{x', y'\}$, if $U(x) = U'(x')$ and $U(y) = U'(y')$, then $x\bar{R}y \Leftrightarrow x'\bar{R}'y'$, with $\bar{R} = \bar{F}(U)$ and $\bar{R}' = \bar{F}(U')$.

To see that RN implies Pareto Indifference, it is enough to put $U' = U$, $x = y'$, and $x' = y$. One needs, finally, a structural assumption not directly imposed on $\bar{F}$ but on the set X and on the given single profile of individual preferences. It ensures that any three vectors in the n-dimensional utility space (the real Euclidean space indexed by the names of the individuals), denoted by E^N, is attainable.[5]

Relative Attainability (RA): For any $u, v, w \in E^N$, there are $x, y, z \in X$ and $U \in [\bar{U}]$ such that $U(x) = u$, $U(y) = v$, and $U(z) = w$.

This assumption is weaker than Harsanyi's own structural assumption, Independent Prospects. The latter could be rephrased as saying that, for each VNM utility profile, the range of that profile has full dimension, hence that any n-tuple of vectors in E^N can be attained from $[\bar{U}]$.

The following is now a version of the Aggregation Theorem, adapted to the present enlarged single-profile approach.

Theorem 7.1: *If the SWFL $\bar{F}$ satisfies conditions VNM-D, VNM-R, RN, and RA, then there is a real vector of weights $(\beta_1, \ldots, \beta_n)$, unique up to a positive scale factor, such that for all $U \in [\bar{U}]$, for all $x, y \in X$, and $\bar{R} = \bar{F}(U)$,*

$$x\bar{R}y \Leftrightarrow \sum_{i=1}^{n} \beta_i U_i(x) \geq \sum_{i=1}^{n} \beta_i U_i(y).$$

The proof is delayed until the next section. There, it will mainly be argued that Harsanyi's theorem is best seen as a welfarist result. Under Welfarism, another version of the Aggregation Theorem will be stated and proved. This version will have the advantage of making clear the cardinal content of the theorem. Also, this further version will incorporate two additional assumptions directly stated in welfarist terms, Strict Pareto and Anonymity, to ensure, respectively, that the weights (β_i) are all positive and that they are all equal.

[5] In d'Aspremont (1985), this condition is called "Unrestricted Individual Utility Profile" (UP).

7.3 A Welfarist Version of the Aggregation Theorem

The first result in this section shows that the SWFL $\bar{F}$ restricted to $[\bar{U}]$, as in Theorem 7.1, can be used to derive *welfarism*. This property means that the preference ordering of the moral observer $\bar{F}(U)$ on X can be translated, whatever $U \in [\bar{U}]$, into a *social welfare ordering* (SWO), that is, a preference ordering R^* defined on the n-dimensional utility space E^N. This moral observer is then truly consequentialist in the sense of taking into account only the utility consequences of all social states and not the social states themselves. In addition, he will be a VNM decision maker because R^* will be shown to satisfy:

VNM-Social Welfare Ordering (VNM-R*): The moral observer's preference ordering R^* defined on E^N satisfies continuity (**VNM1***) and independence (**VNM2***).

The following lemma combines results from d'Aspremont (1985) and Mongin (1994).

Lemma 7.1 (VNM Welfarism): *If the SWFL $\bar{F}$ satisfies VNM-D, RN, and RA, then there exists a SWO R^* defined on E^N such that: For any $x, y \in X$, for any $U \in [\bar{U}]$, and $\bar{R} = \bar{F}(U)$, if $U(x) = u$ and $U(y) = v$, then $uR^*v \Leftrightarrow x\bar{R}y$. Moreover, if $\bar{F}$ also satisfies VNM-R, then R^* satisfies VNM-R*.*

Proof: The first step in the proof consists in constructing a binary relation R^* on E^N. By RA, we may take, for every $u, v \in E^N$, some $x, y \in X$ and $U \in [\bar{U}]$ such that $U(x) = u$ and $U(y) = v$ and put: $uR^*v \Leftrightarrow x\bar{F}(U)y$. The relation R^* is well-defined by RN: for any other profile $U' \in [\bar{U}]$ and pair $\{x', y'\} \subset X$ such that $U'(x') = u$ and $U'(y') = v$, $x'\bar{F}(U')y' \Leftrightarrow x\bar{F}(U)y$. It is complete and transitive because of, respectively, the completeness and the transitivity of $\bar{F}(U)$ for any $U \in [\bar{U}]$.

The second step is to show that VNM-R implies VNM-R*. Consider VNM-independence first. We have to get the conclusion that VNM-2* holds, namely, that $\forall u, v, w \in E^N, \forall \lambda \in [0, 1]$,

$$uR^*v \Leftrightarrow [\lambda u + (1-\lambda)w]\, R^*[\lambda v + (1-\lambda)w].$$

By RA (or by the definition of R^*), there are x, y, and z in X and $U \in [\bar{U}]$ such that $U(x) = u$, $U(y) = v$, and $U(z) = w$, and using VNM1-R,

$$x\bar{F}(U)y \Leftrightarrow (\lambda x + (1-\lambda)z)\bar{F}(U)(\lambda y + (1-\lambda)z).$$

So, by the definition of R^*, we get

$$U(x)\,R^*\,U(y) \Leftrightarrow U(\lambda x + (1-\lambda)z)\,R^*\,U(\lambda y + (1-\lambda)z).$$

Since VNM-D holds, U is mixture-preserving, that is,

$$U(\lambda x + (1-\lambda)z) = \lambda U(x) + (1-\lambda)U(z),$$
$$U(\lambda y + (1-\lambda)z) = \lambda U(y) + (1-\lambda)U(z).$$

The conclusion follows. To derive VNM-1*, a similar argument can be used. □

From now on, we may as well assume that the moral observer's preferences are given by a VNM social welfare ordering R^* on E^N (which amounts to assuming VNM-R, VNM-D, RA, and RN) and introduce additional axioms directly on this R^*. But first, let us prove Theorem 7.1.

Proof of Theorem 7.1: Because the preference ordering R^*, defined on the convex set E^N, satisfies VNM-R*, it has a VNM utility representation W. This mixture-preserving function is affine on E^N, that is, for all $u \in E^N$, $W(u) = \sum_{i \in N} \beta_i u_i + \gamma$, for some vector $(\beta_1, \ldots, \beta_n)$ and some scalar γ (for the equivalence of mixture-preserving and affine functions on convex sets, see, e.g., Coulhon and Mongin, 1989). Moreover, any other VNM representation, with weights $(\beta'_1, \ldots, \beta'_n)$, must be such that $\beta'_i = \lambda \beta_i$, for some $\lambda > 0$ and all $i \in N$. □

To understand better the ethical relevance of this result, another observation is in order. This is the logical equivalence between the independence axiom (VNM2*) and a well-known invariance property of the SWO R^*, stating the minimal kinds of measurability (cardinality) and interpersonal comparability (unit comparability), which are compatible with utilitarianism.

Cardinality and Unit Comparability (CU*): For any $u, v \in E^N$, any vector $(\alpha_1, \ldots, \alpha_n)$, and any $\beta > 0$, if $u'_i = \alpha_i + \beta u_i$ and $v'_i = \alpha_i + \beta v_i$ for all $i \in N$, then $uR^*v \Leftrightarrow u'R^*v'$.

The following argument, given in Mongin (1994), is close to the one used by Harsanyi to show the linear homogeneity of the function W representing moral preferences (1977, chapter 4, Lemma 4).

Lemma 7.2 (VNM cardinality): *A SWO R^* on E^N satisfies CU* if and only if it satisfies VNM2*.*

Proof: Suppose first that CU* holds. We want to show that VNM2* holds, that is, that, for $u, v \in E^N$, uR^*v if and only if for any $\lambda \in]0, 1]$ and $w \in E^N$, $[\lambda u + (1-\lambda)w] R^* [\lambda v + (1-\lambda)w]$. Taking $\alpha_i = (1-\lambda)w_i$, for every i, and $\beta = \lambda$, this equivalence immediately follows from CU*. Second, to prove the converse, suppose that VNM2* holds and take any vector $\alpha = (\alpha_1, \ldots, \alpha_n)$ and $\beta > 0$. If $\beta < 1$, we can simply put $w = \alpha/(1-\beta)$ and $\lambda = \beta$, then apply VNM2*. If $\beta > 1$, clearly $uR^*v \Leftrightarrow \frac{1}{2\beta}(2\beta u) R^* \frac{1}{2\beta}(2\beta v)$, which by VNM2* is equivalent to $(2\beta u) R^* (2\beta v)$ [letting $w \equiv 0$ and $\lambda = 1/(2\beta)$]. To get the conclusion, it is then enough to let $\lambda = 1/2$ and $w = 2\alpha$, and apply VNM2* again. □

This shows that VNM2* (hence, granting welfarism, VNM2-R) implies the axiom that traditionally formalizes the possibility of making interpersonal comparisons of utility differences. If two utility vectors $u, v \in E^N$ are transformed into two vectors $u', v' \in E^N$ according to CU*, then for any $i, j \in N$

$$u_i - v_i \geq u_j - v_j \Leftrightarrow u'_i - v'_i \geq u'_j - v'_j.$$

This invariance property is clearly important from a moral point of view. However, it might be objected that this property is here only a necessary condition, not a sufficient one. We shall come back to this objection (which is relevant to both VNM2* and CU*) in the next section. We now pursue the task of deriving an improved version of the Aggregation Theorem from an ethical point of view.

It seems also important that all individuals be given positive weights. This is ensured by adding the following condition.

Strict Pareto (S-P*): If $u, v \in E^N$ are such that $u_i \geq v_i$, for all $i \in N$, and $u_j > v_j$, for some $j \in N$, then uP^*v.

In conjunction with Pareto Indifference (which is satisfied by construction in a welfarist framework), this principle is equivalent to the usual Strong Pareto principle.

To give positive weight to each individual might even be considered as insufficient. It is an advantage of our welfarist approach – as opposed to the initial Harsanyi approach where only a single profile of individual preferences was considered – to make it possible to formulate an anonymity axiom.

This axiom will make the chosen weights definitely independent from the single profile fixed at the outset.

Anonymity (A*): For all $u \in E^N$, and any permutation σ of N,

$$uI^*(u_{\sigma_1}, \ldots, u_{\sigma_n}).$$

We may, finally, state the two welfarist versions of Harsanyi's Aggregation Theorem characterizing utilitarian rules, one of which is anonymous and the other not. These theorems can be seen as alternative versions of already known welfarist characterizations of utilitarianism.

Theorem 7.2 (Pure Utilitarianism): *If the SWO R^* satisfies S-P*, A*, and VNM2*, then for all $u, v \in E^N$,*

$$uR^*v \Leftrightarrow \sum_{i=1}^{n} u_i \geq \sum_{i=1}^{n} v_i.$$

Several proofs of this theorem are available, bearing in mind that in a context of cardinal comparisons, Anonymity implies the suitable notion of continuity for the SWO. More precisely, in the presence of CU* (or, equivalently, VNM2*), A* implies VNM1*. One proof relies on Theorem 7.1 (as in Mongin, 1994). Another uses the equivalence between VNM2* and CU*, as well as the characterization of the pure utilitarian rule in terms of the latter condition (see d'Aspremont and Gevers, 1977). In either case, axiom A* is to be part of the conditions.

Theorem 7.3 (Generalized Utilitarianism): *If the SWO R^* satisfies S-P* and VNM-R*, then there is a real vector of positive weights $(\beta_1, \ldots, \beta_n)$, such that for all $u, v \in E^N$,*

$$uR^*v \Leftrightarrow \sum_{i=1}^{n} \beta_i u_i \geq \sum_{i=1}^{n} \beta_i v_i.$$

Again, this result can be seen as a corollary to Theorem 7.1. Alternatively, the proof can use a theorem characterizing "weak utilitarianism" (i.e., $\sum_{i=1}^{n} \beta_i u_i > \sum_{i=1}^{n} \beta_i v_i \Rightarrow uR^*v$ for some positive weights $\beta_1, \ldots, \beta_n$) in terms of S-P*, VNM1*, and CU*; see, e.g., Blackwell and Girschik (1954), Roberts (1980b), and d'Aspremont (1985). Using this last reference (Theorem 3.3.5), it is easy to get generalized utilitarianism by showing that the continuity condition VNM1* implies the following: for any $i, j \in N$, there exist $u, v \in E^N$ such that $u \neq v$, $u_h = v_h$ for $i \neq h \neq j$, and vI^*u. To show this property (called *Weak Anonymity*), it is enough to pick $u, a, b \in E^N$, with $a_h = b_h = u_h$, for $i \neq h \neq j$, such that u is not a convex combination of a and b, but aR^*uR^*b. Then, by VNM1*, for some $\lambda \in [0, 1]$ and $v = [\lambda a + (1 - \lambda)b]$, we get vI^*u.

7.4 Concluding Remarks: More on SWFLs and Cardinality

Once stated in an appropriate framework, that is, the welfarist framework, the Aggregation Theorem performs no worse and no better, from an ethical point of view, than existing characterizations of the utilitarian rule. It offers an alternative but equivalent axiomatization. This results from the equivalence between the VNM-independence axiom (VNM2*) and the cardinality-with-unit-comparability axiom (CU*), as imposed on the social welfare ordering. In our presentation, VNM2* was taken to be the welfarist translation of VNM2-R, the VNM-independence axiom imposed on the SWFL $\bar{F}$. Knowing now this equivalence, VNM2* can as well be viewed as the translation of an axiom of interpersonal utility comparison, which would be imposed from the start on $\bar{F}$. More formally, under RN and RA, the condition CU* is equivalent to the following:

Cardinality and Unit Comparability (CU): For any $U \in [\bar{U}]$, any vector $\alpha = (\alpha_1, \ldots, \alpha_n)$, and any $\beta > 0$, if $U' = \beta U + \alpha$, then $\bar{F}(U) = \bar{F}(U')$.

This is cardinality in a specific sense, to be compared with cardinality in the larger and more meaningful sense of preserving interpersonal utility differences. This other axiom is (see Bossert and Weymark, 1997):

Interpersonal Difference Comparability (IRDC): For any $U, U' \in [\bar{U}]$, if, for all $x, y, x', y' \in X$ and all $i, j \in N$,

$$U_i(x) - U_i(y) \geq U_j(x') - U_j(y') \Leftrightarrow U'_i(x) - U'_i(y) \geq U'_j(x') - U'_j(y'),$$

then $\bar{F}(U) = \bar{F}(U')$.

In general, conditions on the individual utility functions are needed to get the equivalence between these two cardinality principles. An interesting fact, in the context of Harsanyi's Aggregation Theorem, is that one such sufficient condition is VNM-D.

Lemma 7.3 (Cardinality): *If the SWFL $\bar{F}$ satisfies VNM-D, then CU is equivalent to IRDC.*

Proof: That IRDC implies CU is easily verified. For any $U \in [\bar{U}]$, any vector $\alpha = (\alpha_1, \ldots, \alpha_n)$, and any $\beta > 0$, if $U' = \beta U + \alpha$, then obviously, for all $x, y, x', y' \in X$ and all $i, j \in N$, differences orderings are preserved, that is,

$$U_i(x) - U_i(y) \geq U_j(x') - U_j(y') \Leftrightarrow U_i'(x) - U_i'(y) \geq U_j'(x') - U_j'(y'),$$

so that $\bar{F}(U) = \bar{F}(U')$ by IRDC.

For the reverse implication, select any $U, U' \in [\bar{U}]$ preserving all differences orderings. By VNM-D, U and U' are nontrivial and there are $x^i, y^i, z^i \in X$ and $\lambda^i \in [0, 1]$, for every $i \in N$, such that

$$z^i = \lambda^i x^i + (1 - \lambda^i) y^i,$$
$$U_i(x^i) - U_i(z^i) = U_j(x^j) - U_j(z^j) > 0,$$

hence,

$$U_i'(x^i) - U_i'(z^i) = U_j'(x^j) - U_j'(z^j) > 0,$$

for all $j \in N$. But, by VNM-D again, for every $i \in N$, we must have $U_i' = \alpha_i + \beta_i U_i$ for some α_i and some $\beta_i > 0$. Using the above equalities, we obtain $\beta_i = \beta_j$, for all $i, j \in N$. By CU, it implies $\bar{F}(U) = \bar{F}(U')$. □

In other words, to restrict individual utility functions, as Harsanyi does, to nontrivial VNM representations entails equivalence between the two definitions of cardinality. This conclusion holds more generally in a multiprofile approach, for a SWFL F defined on a domain $\mathcal{D}$ of profiles of mixture-preserving individual utility functions (not all trivial). The conditions CU and IRDC have simply to be rephrased by substituting F for $\bar{F}$ and $\mathcal{D}$ for $[\bar{U}]$.

This chapter has shown that an "enlarged" single-profile approach leads to reformulating Harsanyi's Aggregation Theorem in welfarist terms and thus turns it into an alternative characterization of utilitarianism, along standard lines in social choice theory. The theorem may now include an anonymity condition and seems compatible with meaningful comparisons of cardinal utility functions. Whatever ethical content it has depends essentially on the following three assumptions: to consider only VNM representations of the individual preferences, to strengthen Pareto Indifference so as to get welfarism, and to impose VNM-independence on the moral observer's preferences. These three conditions appear to constitute the proper content of Harsanyi's particular approach to utilitarianism.

References

Blackorby, C., Donaldson, D., and Weymark, J. A. 2008. Social aggregation and the expected utility hypothesis. In *Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls*, ed. M. Fleurbaey, M. Salles, and J. A. Weymark. Cambridge University Press, Cambridge, pp. 136–183.

Blackwell, D., and Girshick, M. A. 1954. *Theory of Games and Statistical Decisions.* Wiley, New York.

Border, K. 1981. Notes on von Neumann-Morgenstern social welfare functions. Division of the Humanities and Social Sciences, California Institute of Technology (unpublished).

Border, K. 1985. More on Harsanyi's utilitarian cardinal welfare function. *Social Choice and Welfare* 1, 279–281.

Bossert, W., and Weymark, J. A. 2004. Utility in social choice. In *Handbook of Utility Theory,* Vol. 2: *Extensions*, ed. S. Barberà, P. J. Hammond, and C. Seidl. Kluwer Academic, Boston, pp. 1099–1177.

Coulhon, T., and Mongin, P. 1989. Social choice theory in the case of von Neumann-Morgenstern utilities. *Social Choice and Welfare* 6, 175–187.

d'Aspremont, C. 1985. Axioms for social welfare orderings. In *Social Goals and Social Organizations: Essays in Memory of Elisha Pazner*, ed. L. Hurwicz, D. Schmeidler, and H. Sonnenschein. Cambridge University Press, Cambridge, pp. 19–76.

d'Aspremont, C., and Gevers, L. 1977. Equity and the informational basis of collective choice. *Review of Economic Studies* 44, 199–209.

De Meyer, B., and Mongin, P. 1995. A note on affine aggregation. *Economics Letters* 47, 177–183.

Domotor, Z. 1979. Ordered sum and tensor product of linear utility structures. *Theory and Decision* 11, 375–399.

Fishburn, P. C. 1984. On Harsanyi's utilitarian cardinal welfare theorem. *Theory and Decision* 17, 21–28.

Hammond, P. J. 1992. Harsanyi's utilitarian theorem: A simpler proof and some ethical considerations. In *Rational Interactions: Essays in Honor of John C. Harsanyi*, ed. R. Selten. Springer-Verlag, Berlin, pp. 305–319.

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

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

Mongin, P. 1994. Harsanyi's aggregation theorem: Multi-profile version and unsettled questions. *Social Choice and Welfare* 11, 331–354.

Mongin, P., and d'Aspremont, C. 1998. Utility theory and ethics. In *Handbook of Utility Theory,* Vol. 1: *Principles*, ed. S. Barberà, P. J. Hammond, and C. Seidl. Kluwer Academic, Boston, pp. 371–481.

Roberts, K. W. S. 1980a. Social choice theory: The single-profile and multiprofile approaches. *Review of Economic Studies* 47, 441–450.

Roberts, K. W. S. 1980b. Interpersonal comparability and social choice theory. *Review of Economic Studies* 47, 421–439.

Selinger, S. 1986. Harsanyi's aggregation theorem without selfish preferences. *Theory and Decision* 20, 53–62.

Sen, A. K. 1970. *Collective Choice and Social Welfare.* Holden-Day, San Francisco.
Sen, A. K. 1986. Social choice theory, in *Handbook of Mathematical Economics,* Vol. 3, eds. K. J. Arrow and M. D. Intriligator. North-Holland, Amsterdam, pp. 1073–1181.
Weymark, J. A. 1993. Harsanyi's social aggregtion theorem and the weak Pareto principle. *Social Choice and Welfare* 10, 209–221.
Weymark, J. A. 1994. Harsanyi's social aggregation theorem with alternative Pareto principles. In *Models and Measurement of Welfare and Inequality*, ed. W. Eichhorn. Springer-Verlag, Berlin, pp. 869–887.