Utilitarian Aggregation of Beliefs and Tastes

Itzhak Gilboa

Tel Aviv University and Cowles Foundation, Yale University

Dov Samet

Tel Aviv University

David Schmeidler

Tel Aviv University and Ohio State University

Harsanyi's utilitarianism is extended here to Savage's framework. We formulate a Pareto condition that implies that both society's utility function *and* its probability measure are linear combinations of those of the individuals. An indiscriminate Pareto condition has been shown to contradict linear aggregation of beliefs and tastes. We argue that such a condition is not compelling: Society should not necessarily endorse a unanimous choice when it is based on contradictory beliefs. Our Pareto condition is restricted to choices that involve identical beliefs only.

I. Introduction

Harsanyi (1955) offered an axiomatic justification of utilitarianism. He has assumed that all individuals in society as well as society itself are von Neumann and Morgenstern (1944) (VNM) expected utility maxi-

We wish to thank Philippe Mongin for helpful discussions and many detailed comments. Similarly, thanks are due to Pierre-André Chiappori and two anonymous referees. The research of Itzhak Gilboa and David Schmeidler has been supported by Israel Science Foundation grants 790/00 and 975/03. Dov Samet's research was supported by the Israel Institute of Business Research in the Faculty of Management at Tel Aviv University.

[*Journal of Political Economy,* 2004, vol. 112, no. 4]
 0022-3808/2004/11204-0007$10.00

mizers. In this model, a Pareto indifference condition implies that the utility function of society is a linear combination of the utility functions of individuals in society.

The VNM framework deals with a single individual facing lotteries over outcomes. The use of this framework in Harsanyi's social choice model presupposes that all individuals face the same lotteries, that is, that they agree on the probabilities of any given lottery. Thus individuals differ only in their preferences but not in their beliefs. This, however, is rather restrictive. While divergence of opinions is sometimes the result of different values assigned by different individuals to outcomes, much of it is due to differences in beliefs (i.e., different probability distributions over states of nature). Everyone wants peace and prosperity, but people have very different views about the way to achieve these common goals.

In order to examine utilitarian aggregation à la Harsanyi in the more general case in which individuals have different beliefs, a framework such as Savage's (1954) is required. In this framework, an individual has a utility function over a set of outcomes and a (subjective) probability over events. The individual is assumed to have preferences over acts, which are functions from states of nature to outcomes. These preferences can be represented by maximization of subjective expected utility, that is, the expectation of the utility function with respect to the subjective probability.

In this work we assume that each individual in society, as well as society itself, has preferences over acts and that these preferences are representable by subjective expected utility as in Savage's model, with an additional assumption guaranteeing countable additivity of the subjective probabilities. Acts represent social alternatives. We show that an appropriate generalization of Harsanyi's Pareto indifference condition to the present framework implies that society's utility function is a linear aggregation of individuals' utilities and similarly that society's belief is a linear aggregation of individuals' beliefs.

Our result stands in contrast to the conclusion of the works that have studied utilitarian aggregation in Savage-like models. Several authors have found that such an aggregation, generalizing Harsanyi's theorem, is impossible when Pareto and nondictatorial conditions are imposed (Hylland and Zeckhauser 1979; Mongin 1995). Moreover, impossibility results have also been recently obtained for more general classes of preferences (Mongin 1998; Blackorby, Donaldson, and Mongin 2000). Mongin (1995) was the first to formalize Harsanyi's utilitarianism in the countable additive version of Savage's model. We use the same model here. We argue, however, that the impossibility results are derived from a counterintuitive axiom. Further, an appropriate weakening of this axiom leads to a possibility result.

The Pareto condition that drives the impossibility of aggregation requires that if all individuals in society agree on preferences between two alternatives, so should society. Despite its apparently innocent formulation, we do not find this condition very plausible in a model incorporating subjective beliefs, as illustrated by the following example.

Two gentlemen agree to fight a duel at dawn, although either can back down. The result of a duel is that one wins and the other loses (fatally). The gentlemen have opposite rankings of the three outcomes: (1 wins and 2 loses), (no duel), and (2 wins and 1 loses). Assume, for example, that the cardinal rankings of these outcomes are $(1, 0, -5)$ for the first gentleman and $(-5, 0, 1)$ for the second. The fact that both prefer a duel to no duel, in spite of their opposite taste for the outcomes, is possible only because they hold contradictory beliefs. Each of them believes, with a probability of at least 85 percent, that he will win the duel. The combination of very different utility functions and very different subjective probabilities yields the same preferences. If they had similar beliefs, a duel would not take place. Yet, a straightforward adaptation of Pareto's condition implies that society should rank duel above no duel, since both prefer it.

In this example, unanimity of preferences results from the fact that disagreement over tastes and disagreement over beliefs "cancel out." Such spurious unanimity[1] of preference does not appear to be a compelling reason for society to adopt these preferences. By contrast, the Pareto condition appears more convincing when all individuals have identical beliefs or, dually, when they have identical tastes.

Specifically, an alternative is said to be a *lottery* if all individuals agree on the distribution of outcomes induced by this alternative. We require that, if all individuals are indifferent between two alternatives, which are lotteries in our sense, then society too is indifferent between them. Note that if such an alternative has finitely many outcomes, it corresponds to a classical VNM lottery. Thus, by Harsanyi's theorem, our Pareto condition implies that society's utility function is a linear combination of the utility functions of the individuals. More surprisingly, this condition also implies that the probability measure of society is a linear combination of those of the individuals. Thus the dual condition that society approves unanimity on the basis of identical tastes is superfluous since it is implied by our condition. Conversely, our Pareto condition is satisfied whenever society's utility is a linear combination of the individuals' utilities and society's probability measure is a linear combination of those of the individuals.

Our Pareto condition seems to be the natural extension of Harsanyi's

[1] The term *spurious unanimity* appears in Mongin (1997) in the same context and meaning as here.

utilitarianism to the model of *subjective* expected utility. In fact, our Pareto condition *is* Harsanyi's condition. As pointed out above, lotteries with finitely many outcomes constitute a subset of alternatives. We require Harsanyi's condition on lotteries, as he did.

Coming back to our example, observe that we do not argue that a duel *should not* take place. We only argue that the unanimity among individuals should not be used by society to justify a duel. Our duel example is similar to an example in Raiffa (1968), suggesting the rejection of the Pareto condition in the face of contradictory beliefs. Raiffa deals with aggregation of the opinions of different experts. These experts are consulted about both the desirability of outcomes and the plausibility of scenarios. Thus each expert provides a utility function and a probability measure. Raiffa notes that it might well be the case that all experts will prefer alternative a over b, whereas b will be preferred to a if one adopts the average utility function and the average probability measure. The latter procedure seems the correct one to Raiffa.[2]

To the extent that theoretical results may serve as rhetorical devices, we find the present result encouraging. The alternative, namely the theoretical conclusion that aggregating tastes and beliefs is impossible, is troubling. If there is, indeed, no way to aggregate preferences of all individuals, then a ruling party or a president may feel exempted from seeking to represent society in its entirety even if elected by an incidental majority. This seems to contradict our moral intuition on this issue, which demands that a majority should not disregard opinions and desires of minorities. Moreover, ignoring minorities may lead to instability and inefficiency in a society in which governments shift frequently between opposing parties. However, we argue that the impossibility results cannot be cited as an indirect justification of ignoring minority views, because they rely on a counterintuitive assumption. By contrast, a more intuitive version of this assumption necessitates aggregation of preferences.[3]

Section II presents the main result, and the Appendix contains the proof.

II. The Main Result

Let (S, Σ) be a σ-measurable space, where S is a set of *states* (of nature), and Σ is a σ-algebra of *events*. Denote by X a set of *outcomes* endowed with a σ-algebra. The set $A = \{a | a : S \to X,\ a \text{ is } \Sigma\text{-measurable}\}$ is the set

[2] Postlewaite and Schmeidler (1987) argue against welfare comparisons in a general equilibrium model with differential information but without a common prior.

[3] This result does not preclude dictatorial aggregation rules. These rules are excluded by more restrictive conditions that imply the positivity of all coefficients (see the end of Sec. II).

of *alternatives* (or *acts* in Savage's terminology). Society is a set of individuals $N = \{1, \ldots, n\}$. Individual $i \in N$ has preferences $\succsim_i \subset A \times A$, whereas society's preferences are denoted by $\succsim_0 \subset A \times A$. For $0 \leq i \leq n$, the relations $\sim_i$ and $\succ_i$ are defined as the symmetric and asymmetric parts of $\succsim_i$, as usual. We assume that each of the preference relations is represented by expected utility maximization; that is, for $0 \leq i \leq n$, there are a measurable and bounded utility function $u_i : X \to \mathbb{R}$ and a probability measure λ_i on Σ such that, for every $a, b \in A$, $a \succsim_i b$ iff $\int_S u_i(a(s))d\lambda_i \geq \int_S u_i(b(s))d\lambda_i$. We assume that, for each $0 \leq i \leq n$, λ_i is countably additive and nonatomic and that u_i is not constant.[4]

Let $\Lambda = \{E \in \Sigma | \text{for all } 1 \leq i, j \leq n, \lambda_i(E) = \lambda_j(E)\}$. Thus an event E is in Λ when all individuals agree on its probability.

An alternative a is a *lottery* if for each measurable subset of outcomes Y, $a^{-1}(Y) \in \Lambda$. Thus, in a lottery, all individuals agree on the probability of the events that are involved in the definition of the lottery. Note the formal difference between lotteries as defined here and VNM lotteries over X. The latter are probability distributions over X (with finite support), whereas the former are measurable functions. Nevertheless, it is easy to show that finitely valued lotteries, as defined here, can be identified with VNM lotteries over X. This is done in claim 4 in the proof of the theorem.

The Restricted Pareto Condition. For all lotteries a and b, if, for every $i \in N$, $a \sim_i b$, then $a \sim_0 b$.

Theorem 1. The restricted Pareto condition holds iff λ_0 is a linear combination of $\{\lambda_i\}_{i=1}^n$ and u_0 is a linear combination of $\{u_i\}_{i=1}^n$.

This theorem does not restrict the coefficients used in the linear combinations that define beliefs and tastes of society, except that the sum of the coefficients of λ_0 is one since $\lambda_i(S) = 1$ for $0 \leq i \leq n$. One may wish to augment our condition to obtain results according to which λ_0 and u_0 are convex combinations of $\{\lambda_i\}_{i=1}^n$ and of $\{u_i\}_{i=1}^n$, respectively, or even convex combinations with positive coefficients.

Harsanyi (1955) introduced conditions guaranteeing that society's utility function is a convex combination of individuals' utility functions or a convex combination with strictly positive coefficients. Similarly, Mongin (1995) introduced conditions that imply that society's probability measure is a convex function of individuals' probability measures, or a convex combination with strictly positive coefficients. Harsanyi's and Mongin's conditions can be directly adapted to our framework, with the same implications. For brevity's sake, we do not spell out these conditions here.

[4] Conditions on preferences guaranteeing representation by a subjective countably additive probability measure and a bounded utility function are well known. See Savage (1954) and Villegas (1964) or Arrow (1965).

Appendix

Proofs

Proof of the Theorem, Part 1: The Restricted Pareto Condition Is Sufficient

Assume that this condition holds. We prove, first, that the condition implies that λ_0 is a linear combination of $\{\lambda_i\}_{i=1}^n$, with coefficients summing to one. Next we show that this implies also that u_0 is a linear combination of $\{u_i\}_{i=1}^n$.

Denote $\boldsymbol{\lambda} = (\lambda_i)_{i=1}^n$ and $\hat{\boldsymbol{\lambda}} = (\lambda_i)_{i=0}^n$. Let $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ be the ranges of the vector measures $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}$ correspondingly. Note that $\mathbf{Z}$ is the projection of $\hat{\mathbf{Z}}$, and for any $\hat{\mathbf{z}} \in \hat{\mathbf{Z}}$, $\hat{\boldsymbol{\lambda}}(S) - \hat{\mathbf{z}} \in \hat{\mathbf{Z}}$. Owing to the Lyapunov theorem, both $\mathbf{Z}$ and $\hat{\mathbf{Z}}$ are convex.

CLAIM 1. If $(z_0, \frac{1}{2}\boldsymbol{\lambda}(S)) \in \hat{\mathbf{Z}}$, then $z_0 = \frac{1}{2}$.

Proof. Assume the contrary. Then, without loss of generality, for some event E, $\lambda_0(E) > \frac{1}{2}$ whereas $\boldsymbol{\lambda}(E) = \frac{1}{2}\boldsymbol{\lambda}(S)$. Choose $x, y \in X$ with $u_0(x) > u_0(y)$. Consider alternatives a and b such that a is x on E and y on E^c, whereas b is y on E and x on E^c. Then $a \sim_i b$ for each $i \in N$ but $a \succ_0 b$, contrary to our assumption. Q.E.D.

Proposition 3 in Mongin (1995) states that, under the conclusion of claim 1, λ_0 is a linear combination of $\{\lambda_i\}_{i=1}^n$ with coefficients summing to one. The following two claims constitute a shorter proof of this fact and are given here for the sake of completeness.

CLAIM 2. For every $\mathbf{z} \in \mathbf{Z}$ there exists a unique $z_0 = z_0(\mathbf{z})$ such that $(z_0, \mathbf{z}) \in \hat{\mathbf{Z}}$.

Proof. Assume that $\hat{\mathbf{z}} = (z_0, \mathbf{z})$ and $\hat{\mathbf{w}} = (w_0, \mathbf{z})$ are in $\hat{\mathbf{Z}}$, with $z_0 < w_0$. Since $\hat{\boldsymbol{\lambda}}(S) - \hat{\mathbf{w}} \in \hat{\mathbf{Z}}$, it follows from the convexity of $\hat{\mathbf{Z}}$ that $\frac{1}{2}\hat{\mathbf{z}} + \frac{1}{2}[\hat{\boldsymbol{\lambda}}(S) - \hat{\mathbf{w}}] = (\frac{1}{2}z_0 + \frac{1}{2}(1 - w_0), \frac{1}{2}\boldsymbol{\lambda}(S)) \in \hat{\mathbf{Z}}$. This contradicts claim 1, since the first coordinate of this point is less than $\frac{1}{2}$. Q.E.D.

CLAIM 3. For every $\mathbf{z}$, $\mathbf{w} \in \mathbf{Z}$ and every $0 \le \beta \le 1$, $z_0(\beta\mathbf{z} + (1 - \beta)\mathbf{w}) = \beta z_0(\mathbf{z}) + (1 - \beta)z_0(\mathbf{w})$.

Proof. Let $\hat{\mathbf{z}} = (z_0(\mathbf{z}), \mathbf{z})$ and $\hat{\mathbf{w}} = (z_0(\mathbf{w}), \mathbf{w})$. By the convexity of $\hat{\mathbf{Z}}$, $\beta\hat{\mathbf{z}} + (1 - \beta)\hat{\mathbf{w}} \in \hat{\mathbf{Z}}$. The first coordinate of this point is $\beta z_0(\mathbf{z}) + (1 - \beta)z_0(\mathbf{w})$. The last n coordinates are $\beta\mathbf{z} + (1 - \beta)\mathbf{w}$, and thus the result follows by claim 2. Q.E.D.

By claim 3, $z_0(\mathbf{z})$ is a linear function on $\mathbf{Z}$. Linearity together with the fact that $z_0(0) = 0$ imply the existence of $(\theta_i)_{i\in N}$ in $\mathbb{R}^N$ such that $z_0(\mathbf{z}) = \sum_{i\in N}\theta_i z_i$. Hence, for each event E, $\lambda_0(E) = \sum_{i\in N}\theta_i\lambda_i(E)$. Substituting S for E in the last equality, we conclude that $\sum_{i\in N}\theta_i = 1$.

Now we show that u_0 is a linear combination of $\{u_i\}_{i=1}^n$. We first recall a well-known conclusion of the Lyapunov theorem (proved by induction).

CLAIM 4. Assume that $p_1, \ldots, p_m$ are nonnegative numbers whose sum is one. Then there is a partition of S $(E_1, \ldots, E_m)$ such that, for all $1 \le j \le m$, $\boldsymbol{\lambda}(E_j) = p_j\boldsymbol{\lambda}(S)$.

Using claim 4, we can identify finitely valued lotteries with VNM lotteries over X. Specifically, given a VNM lottery L with a finite support over X, one may use claim 4 to construct a lottery $a \in A$ such that L is the distribution on X defined by a. Moreover, for all $0 \le i \le n$, all such lotteries a are $\sim_i$-equivalent since they have the same expected utility. Conversely, any finitely valued lottery $a \in A$ defines a distribution over X, which is a VNM lottery.

It follows that, restricting $\{\succsim_i\}_{0\le i\le n}$ to lotteries in A, one may apply Harsanyi's (1955) theorem to conclude that u_0 is a linear combination of $\{u_i\}_{i=1}^n$. (For a recent proof of Harsanyi's theorem, De Meyer and Mongin [1995] is recommended.) Q.E.D.

Proof of the Theorem, Part 2: The Restricted Pareto Condition Is Necessary

Let a and b be lotteries. Since λ_0 is a linear combination of $\{\lambda_i\}_{i=1}^n$, $\lambda_0(E) = \lambda_i(E)$ for all $i \in N$ and all $E \in \Lambda$. Therefore, for every lottery c, $\int_S u_i(c)d\lambda_i = \int_S u_i(c)d\lambda_0$. For all $i \in N$, the condition $a \sim_i b$ implies that $\int_S [u_i(a) - u_i(b)]d\lambda_i = 0$. Hence, for all $i \in N$, $\int_S [u_i(a) - u_i(b)]d\lambda_0 = 0$. Since u_0 is a linear combination of $\{u_i\}_{i=1}^n$, $\int_S [u_0(a) - u_0(b)]d\lambda_0 = 0$ also follows, and $a \sim_0 b$ is proved. Q.E.D.

References

Arrow, Kenneth J. 1965. *Aspects of the Theory of Risk-Bearing.* Helsinki: Yrjö Jahnsson Säätiö. Reprinted in *Essays in the Theory of Risk-Bearing.* Amsterdam: North-Holland, 1971.

Blackorby, Charles, David Donaldson, and Philippe Mongin. 2000. "Social Aggregation without the Expected Utility Hypothesis." Discussion paper. Vancouver: Univ. British Columbia, Dept. Econ.

De Meyer, Bernard, and Philippe Mongin. 1995. "A Note on Affine Aggregation." *Econ. Letters* 47 (February): 177–83.

Harsanyi, John C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *J.P.E.* 63 (August): 309–21.

Hylland, Aanund, and Richard J. Zeckhauser. 1979. "The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values." *Econometrica* 47 (November): 1321–36.

Mongin, Philippe. 1995. "Consistent Bayesian Aggregation." *J. Econ. Theory* 66 (August): 313–51.

———. 1997. "Spurious Unanimity and the Pareto Principle." Working paper. Cergy-Pontoise, France: Univ. Cergy-Pontoise, Théorie Economique, Modélisation et Applications and Centre Nacional Recherche Scientifique.

———. 1998. "The Paradox of the Bayesian Experts and State-Dependent Utility Theory." *J. Math. Econ.* 29 (April): 331–61.

Postlewaite, Andrew, and David Schmeidler. 1987. "Differential Information and Strategic Behavior in Economic Environments: A General Equilibrium Approach." In *Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz*, edited by Theodore Groves, Roy Radner, and Stanley Reiter. Minneapolis: Univ. Minnesota Press.

Raiffa, Howard. 1968. *Decision Analysis: Introductory Lectures on Choices under Uncertainty.* Reading, Mass.: Addison-Wesley.

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York: Wiley. 2d ed. New York: Dover, 1972.

Villegas, C. 1964. "On Qualitative Probability σ-Algebras." *Ann. Math. Statis.* 35 (December): 1787–96.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior.* Princeton, N.J.: Princeton Univ. Press.