Utilitarian Aggregation of Beliefs and Tastes*

Itzhak Gilboa, Dov Samet, and David Schmeidler

March 2001

Abstract

Several authors have indicated a contradiction between consistent aggregation of subjective beliefs and tastes, and a Pareto condition. We argue that the Pareto condition that implies the contradiction is not compelling. Society should not necessarily endorse a unanimous choice when it is based on contradictory beliefs. Restricting the Pareto condition to choices that only involve identical beliefs allows a utilitarian aggregation: both society's utility function *and* its probability measure are linear combinations of those of the individuals.

1 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-Morgenstern (vNM) expected utility maximizers (von Neumann and Morgenstern (1944)). In this model, one Pareto indifference condition implies that the utility function of society is a convex 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, they agree on the

*We wish to thank Philippe Mongin for helpful discussion and many detailed comments.

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 of 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 Savage's (1954) model there is a separation between preferences represented by a cardinal (i.e., vNM) utility and information represented by a (subjective) probability. This model permits the examination of Harsanyi's utilitarian aggregation in the more general case where individuals have different beliefs. Indeed, more recent works have studied utilitarian aggregation in Savage-like models, aggregating individuals' utilities into society's utility and individuals' probabilities into society's probability. 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) and Blackorby et al. (2000)).

These impossibility results are 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 where governments shift frequently between opposing parties. In this note we argue, however, that the impossibility results are derived from a counter-intuitive axiom. Further, an appropriate weakening of this axiom leads to a possibility result. Thus, we conclude that the impossibility result cannot be cited as an indirect justification of ignoring minority views.

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 the following example shows.

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 consequences: (1 wins and 2 loses), (no duel), and (2 wins and 1 loses). Assume, for example, that the cardinal rankings of these consequences are $(1, 0, -5)$ for the first gentleman and $(-5, 0, 1)$ for the second. The fact that both prefer duel to no duel, in spite of their opposite taste for the consequences, is possible only because they hold contradictory beliefs. Each of them believes, with a probability of more than 90%, 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.

A straightforward adaptation of Pareto's condition implies that society should rank duel above no duel, as both prefer it. Indeed, so do the Pareto axioms used to obtain impossibility results. But we claim that the Pareto condition cannot serve as an argument for society to prefer that a duel take place.

Observe that we do not argue that a duel *should not* take place. A liberal approach might suggest that society should have no preferences at all over issues such as the duel. Moreover, even if one takes a more paternalistic approach, according to which society does have a complete preference relation, society might favor a duel for a variety of reasons. We only argue that the Pareto condition cannot be one of them.

Consider, first, the following example. Mary has to choose between two cups containing hot drinks. She believes that the first contains coffee and the second – tea, and, given her strong preference for coffee, she prefers the first cup. No other individual is affected by this choice. Hence all other individuals are indifferent between Mary's two choices. But everyone else is convinced that Mary is in fact wrong, and that the contents of the cups are reversed. Society may refrain from expressing preferences over Mary's two choices. But if society does have such preferences, we find that basing them on a formal application of the Pareto condition is unjustified.

By a similar token, the duel is preferred by both gentlemen only because they differ in their beliefs. Society cannot, perhaps, judge who is more accurate in his probability assessments. But it is evident that at least one of them is wrong: every probability assessment would induce at least one of the gentlemen to back down. In this situation, we argue, the Pareto condition cannot serve as a reason for society to prefer that a duel take place.

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, while b will be preferred to a if one adopts the average utility function and the average probability measure. Thus, Raiffa (1968) argues against the Pareto condition whenever it contradicst expected utility maximization with respect to average utility and average probability. By contrast, we argue that the Pareto condition is unpalatable without reference to a particular method of aggregation. [1] Moreover, whereas Raiffa deals with a single decision maker, who is unsure about the measurement of utility and of probability, our focus is on social welfare, where conflict of interest is inherent. Thus, in Raiffa's problem one may be "wrong" both about probabilities and

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

about utilities, whereas in the social welfare set-up only probabilities admit a notion of correctness.

In this note we suggest a weaker Pareto condition, which is more plausible in a model that separates tastes and beliefs. With this condition a nondictatorial separate aggregation of tastes and of beliefs, a la Harsanyi, is possible. Our Pareto condition is restricted to alternatives in which individuals may differ only in their tastes and not in their beliefs. More 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 consequences, 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 an affine combination of those of the individuals. Conversely, our Pareto condition is satisfied whenever society's utility is a linear combination of the individuals' utilities and society's probability measure is an affine combination of those of the individuals.

In conclusion, we argue that the impossibility theorems result from a Pareto condition that is far from compelling. It is based on the requirement that society approves unanimous preferences of individuals, even when unanimity is the result of conflicting tastes and conflicting beliefs. Our Pareto condition restricts social approval of unanimity only to those cases where all individuals have identical beliefs. This seems to be the natural extension of Harsanyi's utilitarianism for the expected utility model to that of subjective expected utility.

The next section presents the main result, while Section 3 contains the proof.

2 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 $A = \{a \mid a\colon S \rightarrow X,\, a \text{ is } \Sigma - \text{measurable}\}$ is the set of *alternatives* (or *acts* in Savage's terminology). Society is a set of individuals $N = \{1, ..., n\}$. Individual $i \in N$ has preferences $\succsim_i \subset A \times A$, whereas society's preferences are denoted $\succsim_0 \subset A \times A$. 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 preference relations is represented by expected utility maximization, that is, for $0 \leq i \leq n$ there are a measurable and bounded utility function $u_i : X \rightarrow \mathbb{R}$ and a probability measure μ_i on Σ such that, for every $a, b \in A$, $a \succsim_i b$ iff $\int_S u_i(a(s))d\mu_i \geq \int_S u_i(b(s))d\mu_i$. We assume that, for each $0 \leq i \leq n$, μ_i is countably additive and non-atomic, and that u_i is not constant.[2]

Let $\Lambda = \{E \in \Sigma \,|\, \text{for all } 1 \leq i, j \leq n,\ \mu_i(E) = \mu_j(E)\}$. 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 outcome, 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) while the first 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 $a \sim_0 b$.

[2] 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).

Theorem *The Restricted Pareto Condition holds iff μ_0 is an affine combination of $\{\mu_i\}_{i=1}^n$ and u_0 is a linear combination of $\{u_i\}_{i=1}^n$.*

This theorem does not restrict the coefficients used in the affine and linear combinations that define beliefs and tastes of society. One may wish to augment our condition to obtain results according to which μ_0 and u_0 are convex combinations of $\{\mu_i\}_{i=1}^n$ and of $\{u_i\}_{i=1}^n$, respectively, with strictly positive coefficients, and perhaps also with equal coefficients, at least in the case of beliefs (where interpersonal comparisons are natural). Such derivations are beyond the scope of this note.

3 Proof of the Theorem

The restricted Pareto condition is sufficient. Assume that this condition holds. We prove, first, that the condition implies that μ_0 is an affine combination of $\{\mu_i\}_{i=1}^n$. Next, we show that this implies also that u_0 is a linear combination of $\{u_i\}_{i=1}^n$.

Denote $\mu = (\mu_i)_{i=1}^n$ and $\hat{\mu} = (\mu_i)_{i=0}^n$. Let Z and $\hat{Z}$ be the ranges of the vector measures μ and $\hat{\mu}$ correspondingly. Note, that Z is the projection of $\hat{Z}$, and for any $\hat{z} \in \hat{Z}$, $\hat{\mu}(S) - \hat{z} \in \hat{Z}$. Due to the Lyapunov Theorem, both Z and $\hat{Z}$ are convex. ∎

Claim 1: If $(z_0, \frac{1}{2}\mu(S)) \in \hat{Z}$, then $z_0 = \frac{1}{2}$.

Proof: Assume the contrary. Then, without loss of generality, for some event E, $\mu_0(E) > \frac{1}{2}$ while $\mu(E) = \frac{1}{2}\mu(S)$. Choose $x, y \in X$ with $u_0(x) > u_0(y)$. Consider alternatives a, b such that a is x on E and y on E^c, while 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. ∎

Proposition 3 in Mongin (1995) states that, under the conclusion of Claim 1, μ_0 is an affine combination of $\{\mu_i\}_{i=1}^n$. The following two claims constitute a shorter proof of this fact, and are given here for the sake of completeness.

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

Proof: Assume that $\hat{z} = (z_0, z)$ and $\hat{w} = (w_0, z)$ are in $\hat{Z}$, with $z_0 < w_0$. Since $\hat{\mu}(S) - \hat{w} \in \hat{Z}$, it follows from the convexity of $\hat{Z}$ that $\frac{1}{2}\hat{z} + \frac{1}{2}(\hat{\mu}(S) - \hat{w}) = (\frac{1}{2}z_0 + \frac{1}{2}(1 - w_0), \frac{1}{2}\mu(S)) \in \hat{Z}$. This contradicts Claim 1, as the first coordinate of this point is less than $\frac{1}{2}$. ∎

Claim 3: For every $z, w \in Z$ and every $0 \leq \beta \leq 1$, $z_0(\beta z + (1 - \beta)w) = \beta z_0(z) + (1 - \beta)z_0(w)$.

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

By Claim 3, $z_0(z)$ is an affine function on Z, that is, there are $(\lambda_i)_{i \in N} \in \mathbb{R}^N$ such that $z_0(z) = \sum_{i \in N} \lambda_i z_i$. Hence, for each event E, $\mu_0(E) = \sum_{i \in N} \lambda_i \mu_i(E)$. Substituting S for E in the last equality we conclude that $\sum_{i \in N} \lambda_i = 1$.

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 $p_1, ...p_m$ are non-negative numbers whose sum is 1. Then there is a partition of S, $(E_1, ...E_m)$ such that, for all $1 \leq j \leq m$ $\mu(E_j) = p_j \mu(S)$.

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 $0 \leq i \leq n$, all such lotteries a are $\sim_i$-equivalent since they have the expected utility. Conversely, any finitely-valued lottery $a \in A$ defines a distribution over X, which is a vNM lottery.

It follows that, restricting $\{\succsim_i\}_{0 \leq i \leq n}$ 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, DeMeyer and Mongin (1995) is recommended.)

∎

The restricted Pareto condition is necessary. Let a, b be lotteries. Since μ_0 is an affine combination of $\{\mu_i\}_{i=1}^n$, $\mu_0(E) = \mu_i(E)$ for all $i \in N$ and all $E \in \Lambda$. Therefore, for every lottery c, $\int_S u_i(c) d\mu_i = \int_S u_i(c) d\mu_0$. For all $i \in N$, the condition $a \sim_i b$ implies that $\int_S (u_i(a) - u_i(b)) d\mu_i = 0$. Hence, for all $i \in N$, $\int_S (u_i(a) - u_i(b)) d\mu_0 = 0$. Since u_0 is a linear combination of $\{u_i\}_{i=1}^n$, $\int_S (u_0(a) - u_0(b)) d\mu_0 = 0$ also follows, and $a \sim_0 b$ is proved. ∎

References

Arrow, K.J. (1965), *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjo Jahnssonin. (Reprinted in *Essays in the Theory of Risk Bearing* (1970), North Holland.)

Blackorby, C., D. Donaldson, and P. Mongin (2000), Social Aggregation Without the Expected Utility Hypothesis, A working paper.

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

Harsanyi, J.C. (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility, *Journal of Political Economics*, 63, 309-321.

Hylland A. and R. Zeckhauser (1979), The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values, *Econometrica*, 6, 1321-1336.

Mongin, P. (1995), Consistent Bayesian Aggregation, *Journal of Economic Theory*, 66, 313-351.

Mongin, P. (1998), The Paradox of the Bayesian Experts and State-Dependent Utility Theory, *Journal of Mathematical Economics*, 29, 331-361.

Postlewaite, A. and D. Schmeidler (1987), Differential Information and Strategic Behavior in Economic Environments: A General Equilibrium Approach, *Information Incentives and Economic Mechanisms, Essays in Honor of Leonid Hurwicz*, T. Groves, R. Radner and S. Reiter (eds.), University of Minnesota Press, 330-48.

Raiffa, H. (1968), *Decision Analysis,* Addison-Wesley, 1968.

Savage, L.J. (1954), *The Foundation of Statistics,* Wiley and Sons (second edition 1972, Dover).

Villegas, C. (1964), On Quantitative Probability Σ-Algebras, *Annals of Mathematical Statistics,* 35, 1787-1796.

von Neumann, J. and O. Morgenstern (1944) *Theory of Games and Economic Behavior,* Princeton University Press.