

# Agents in Discord. On Preference Aggregation under Uncertainty

ULRICH METSCHL, MUNICH AND INNSBRUCK

## 1. Introduction

According to a widely held impression, consensus in pluralistic societies is more the exception than the rule. Where agreement can be found and where public issues are decided unanimously, this seems to be due to favourable circumstances rather than to a society's concord. However, hopes seem to prevail that the often intractable conflicts which beset modern societies often result from zeal and passion, so that, where conflicts are the disease, reason and rational discourse are the appropriate cure. Recent results in social choice theory indicate, however, that this hope may be overly naive.

Consider a group of individuals in a situation where a collective decision has to be made. For example, the agents might be jointly responsible for a welfare policy and its appropriate resource allocation or for the choice of an environmental action plan. Suppose, the individuals are rational in the following sense: the preferences of each person over the set of feasible outcomes of the joint decision are coherent and can be represented by a von Neumann-Morgenstern utility function (or some equivalent device), and the individuals' factual beliefs concerning the relevant states of nature are represented by individual probability functions  $p_i$  for each person  $i$ , so that each person is able to calculate her individual expected utility for each of the alternatives under consideration. If conditions are imposed to the effect that the collective decision is the 'product' of the individual assessments of the likelihood and desirability of the possible outcomes and at the same time respects unanimity and hence Pareto optimality then examples can be constructed which demonstrate that the straightforward method of aggregating the individuals' assessments by taking the average utilities and probabilities will not yield the desired result. The following illustration is a modification of Raiffa's original example, which seems to be the classical exposition of the problem of Bayesian aggregation (see Raiffa 1968, 228-230):

Alice and Bob are faced with a decision between two alternatives  $x$  and  $y$ . Two possible states of nature have to be considered,  $s_1$  and  $s_2$ . The four possible outcomes will be denoted  $x(s_1)$ ,  $y(s_1)$ ,  $x(s_2)$  and  $y(s_2)$  respectively. Alice's probabilities for  $s_1$  and  $s_2$  are 0.2 and 0.8 respectively, while Bob's are 0.7 and 0.3. Their utilities for the four possible outcomes are given in the following tables:

Alice	$S_1$	$S_2$
$x$	10	3
$y$	1	6

Bob	$S_1$	$S_2$
$x$	2	9
$y$	9	2

This yields the following expected utilities for the options  $x$  and  $y$ :  $EU_{\text{Alice}}(x) = 4.4$ ,  $EU_{\text{Alice}}(y) = 5.0$ ,  $EU_{\text{Bob}}(x) = 4.1$ , and  $EU_{\text{Bob}}(y) = 6.9$ . So, Alice and Bob both consider  $y$  as better than  $x$ , although their reasons for doing so are clearly different. If unanimous choice is to be respected, society should choose  $y$  over  $x$ . However, this choice cannot be justified on the basis of aggregated probabilities and utilities when aggregation consists in taking the average values. For the average probabilities are 0.45 for  $s_1$  and 0.55 for  $s_2$ , and the average utilities are 6 for  $x(s_1)$  and for  $x(s_2)$ , 5 for  $y(s_1)$  and 4 for  $y(s_2)$ . Thus, the expected utility for  $x$ , based on the average utility-probability values, is higher than the expected utility for  $y$ . The aggregation of the individual values stands in tension with the individuals' unanimous choice.

## 2. Unanimity and Pareto conditions

The fact, that a simple though seemingly plausible rule of aggregation fails to meet certain requirements, does not imply that the aggregation of (possibly divergent) preferences in situations of uncertainty is impossible. However, a number of results do exist which indicate that, given a suitably rigorous framework, preference aggregation for Bayesian agents is indeed impossible when the aggregation procedure is based exclusively on the individuals' assessments of the probability and desirability of the possible outcomes and is required to respect unanimous choice. These results can be summarized by the following statement (essentially due to Seidenfeld/Kadane/Shervish 1989, see also Mongin 1995).

Any function  $f$  mapping a vector  $(p_1, \dots, p_n, u_1, \dots, u_n)$  made up of individual probability-utility pairs  $(p_i, u_i)$  to a collective probability-utility pair  $(p_G, u_G)$ , such that  $f$  respects at least a weak Pareto condition, will be dictatorial in the following sense: if the individual probability functions are independent so that no  $p_i$  is a weighted average of other individual probability functions then there is a utility dictator, and if the utility functions are independent (no utility function is a mixture of other persons' utilities) then there is a probability dictator. Here, a dictator is a person whose probability (utility) function determines society's probability (utility) function, no matter what the other individuals' beliefs and preferences may be. As a dictatorial rule stands in a striking contrast to the idea of consensual decision making, the result amounts to the claim that no trivial aggregation rule for preferences exists when uncertainty (in a form that leads to diverging individual opinions) interferes.

The weak Pareto condition that is crucial in this context states that no option  $y$  can be chosen collectively that is dominated by some option  $x$  whose expected utility, according to all individual probability-utility pairs, is higher than the one for  $y$ , i.e. if  $EU_i(x) > EU_i(y)$  for all individuals  $i = 1, \dots, n$ , then:  $EU_G(x) > EU_G(y)$ . No improvement is reached when instead of the weak Pareto condition a strong Pareto is used:  $EU_G(x) > EU_G(y)$  if there is no individual  $i$  such that  $EU_i(y) > EU_i(x)$  and for at least one individual  $j$   $EU_j(x) > EU_j(y)$ .

What these results in effect say is that unless the individuals in a group are highly homogenous in their probability judgments or their preferences no consensus will be found that is faithful to those choices on which the agents, though for different reasons, agree. The example given above highlights this point. If the choice of Alice and Bob is respected in the collective decision then it is unclear on what probability-utility judgment the collective decision can be justified, unless it is either Alice's or Bob's (making one of them a 'dictator'). If, on the other hand, their individual assessments are aggregated by taking the average values then the choice based on the aggregation will stand in conflict with what Alice and Bob would choose unanimously.

The situation is troubling for the following reasons. Social choice theory began with a negative result. Arrow's Impossibility Theorem (Arrow 1951, see also Sen 1970, chap. 3) demonstrated that no non-dictatorial rule exists that allows the aggregation of arbitrary preferences such that the weak Pareto condition is obeyed and the social ranking of any two alternatives depends exclusively on the individual rankings of these alternatives and not on the availability of (or comparison with) other alternatives (Independence of Irrelevant Alternatives). An important assumption in Ar-

row's framework was, however, that the individual as well as the social preferences are represented in purely ordinal form and are not interpersonally comparable, an assumption that can be defended for the purposes of economics by the success of the Arrow-Debreu results on general equilibria (Fleurbaey and Hammond 2004). By abandoning this restriction, it became possible to mitigate Arrow's result. Harsanyi's Aggregation Theorem states that when the individual as well as the social preferences are represented by von Neumann-Morgenstern (vNM) utility functions the Pareto indifference implies that the collective utility function is a linear combination, i.e. a weighted sum of the individual utility functions. Therefore, social preferences can be ranked according to a utilitarian rule. A social alternative  $x$  is preferred over alternative  $y$  if and only if the sum of individual utilities for  $x$  is higher than for  $y$ , that is:

$$x R_G y \text{ iff } \sum u_i(x) \geq \sum u_i(y)^1$$

Pareto indifference says that when all individuals are indifferent between two alternatives  $x, y$ , then so should society be.

Pareto Indifference: if  $x I_i y$  for all individuals  $i$  then  $x I_G y$

Society, this is to say, should not impose preferences where these are not based on the judgments of its members. But although vNM utility functions represent preferences over lotteries, uncertainty in a genuine sense is beyond Harsanyi's approach. The probabilities for the chancy prospects of the lotteries are, in the von Neumann-Morgenstern setting, objective (statistical) probabilities; the lotteries are, in terms of the Anscombe-Aumann framework, roulette lotteries, in contrast to horse lotteries (which formally are defined as functions from states of nature to roulette lotteries and which can be associated with subjective probabilities). Thus, no disagreement between equally legitimate assessments of the facts pertinent to the problem of choice under consideration will arise, because rational agents are bound in their judgments by the objective probabilities.

Note, that Harsanyi even may feel justified in ignoring subjective probabilities and their effect of divergent beliefs. For Harsanyi maintains explicitly that differences in beliefs, where beliefs are represented by probability assignments, must be explained in terms of differences in information, in the sense that rational agents, when given the same items of information, will come to hold identical beliefs. This claim, known as the Harsanyi doctrine, is supported by Aumann's result that Bayesian agents with the same prior probabilities and whose posterior probabilities are common

---

<sup>1</sup> Strictly speaking, Harsanyi's utilitarianism must attach weights to individual utilities when unanimity is expressed by Pareto indifference. However, stronger versions of the Pareto requirement allow to consider sums of utilities without weights, see Weymark 1993.

knowledge (i.e. everybody knows them and everybody knows that everybody knows them and everybody knows that ... that everybody knows them), must agree in their posterior probabilities, irrespective of how these were gained. In combination with what is known as ‘washing out of priors’ (or ‘merging of opinions’) by Bayesian up-dating, this means that any differences in subjective beliefs will disappear in the due course of time when existing information is exchanged freely among the agents and new information is processed according to the Bayesian recommending.

However, not always can a decision be delayed until all disagreement concerning the relevant situations is removed. Often enough, the information that would be needed for considered judgment is not available, and differences in belief will persist.

Situations of choice under uncertainty thus seem to point to a serious limitation of Harsanyi’s axiomatic defence of utilitarianism, in particular when it is presented as a rational method for establishing welfare decisions. But the impossibility of preference aggregation under uncertainty is not only a challenge to Harsanyi’s specific position in the theory of social choice. It throws doubt on our understanding of the norms that regulate collective and in particular public decision making in general, norms which Harsanyi’s version of neo-utilitarianism tried to capture in the tradition of welfare economics. Given that ignorance and uncertainty in matters of public concern is our everyday situation, how should consensual decisions, conforming to minimal requirements for democratic and liberal societies, be possible when even purely rational agents cannot expect to specify an acceptable aggregation rule?

However, closer scrutiny reveals that the predicament of the impossibility of finding a consensus reflects the specific constraints that were imposed on the aggregation procedure. Although these constraints may seem to express only minimal requirements, their actual content depends to some extent on the particular context of their application. This is true in particular for the unanimity requirement and the corresponding Pareto condition.

Raiffa’s example of a panel of Bayesian experts whose judgments are to inform a neutral decision maker illustrates the central dilemma for preference aggregation under uncertainty (Raiffa 1968, 228). If the experts agree in their preferences concerning two risky options  $x$ ,  $y$ , although they disagree in their probabilities and their utilities for the prospects of  $x$  and  $y$ , then the decision maker whose judgment is the result of the aggregation of the individual judgments (with separate aggregation of probabilities and utilities) can respect the experts’ unanimous choice of  $x$  over  $y$  only if the decision maker’s probability-utility pair coincides with that of one of the experts. However, where one of the expert opinions is dictatorial, the idea

of a consensus is thwarted and the disagreement in the individual assessments is left unresolved. Unless there are independent reasons for ignoring all the other experts, the decision maker's choice to accept one of the experts' opinions seems arbitrary and, at any rate, leaves the issue of finding a compromise unaddressed. If, on the other hand, the decision maker sets for a compromise by aggregating the individual probability and utility judgments, then no non-trivial aggregation rule will respect unanimous choice under all situations. Respecting unanimity in all cases therefore amounts to a neglect of the potential disagreement in individual judgments. But to ignore disagreement falls short of the aim of consensual decision making.

### 3. Considered unanimity

As Amartya Sen once aptly remarked, "the rejection of the Pareto principle cannot be a source of great joy" (Sen 1970, 85). The problem of preference aggregation under uncertainty, however, can be seen as evidence that the unanimity condition is a substantial and possibly controversial requirement. As a constraint on pure preference aggregation the Pareto principle may still be persuasive. After all, unanimity in preferences (as expression of 'values' or 'tastes'<sup>2</sup>) is unanimity along one dimension, and we may therefore presume that it does not conceal any conflict or disagreement.<sup>3</sup> But where the evaluation of options is based simultaneously on value judgments and on judgments concerning facts, unanimity may indeed be the result of sheer coincidence, thereby concealing profound differences in the assessments on which unanimous choice is based.

Reservations concerning Pareto unanimity have led several authors to attempt improvements in the formulation of principles of preference aggregation under uncertainty. Gilboa, Samet and Schmeidler are among those who argue for a rejection of the unqualified Pareto requirement (Gilboa, Samet and Schmeidler 2004). They claim that the Pareto condition is

---

<sup>2</sup> The question what it is that preferences are the expression of, is beyond the scope of the present paper. It should be kept in mind, however, that the precise nature of preferences is far from clear; for a thorough discussion of this point see Griffin 1986. Still instructive for the difference between values and tastes are Arrow's remarks in 1951.

<sup>3</sup> Even this is not quite true. It may happen, for example, that agents agree on the ranking of two options, because they believe that the preferred option serves best their individual interests, although their aims are clearly different. This phenomenon, known under the heading 'politics makes strange bed fellows', is nicely illustrated by Isaac Levi's example of a catholic and a communist who agree that a national health insurance plan prohibiting abortion is better than no health insurance plan, see Levi 1990. The art of logrolling consists, of course, precisely in bringing enemies to bed.

implausible when subjective beliefs are involved. Their misgivings are illustrated by the following example of two gentlemen who agree to meet for a duel. Their common preference for duel over no duel is a result of the fact that each of them is pretty sure, i.e. believes with a probability of 0.9 that he will win the duel. Gilboa, Samet and Schmeidler maintain that, despite the gentlemen's agreement, it is not evident that society should respect their choice. The argument is not that society should opt against duel. All that is claimed is that the agents' unanimous choice in this case, based on opposite utility and probability judgments, is no compelling reason for the society to respect that choice and prefer duel over no duel. In suspending judgement on this issue, society may point out that it is obvious that not both gentlemen can be right in their beliefs to win the duel. Hence, one of them will have his preference based on a false assumption.

Gilboa, Samet and Schmeidler argue that the duel example corresponds to Raiffa's example of the panel of experts. This is, to some extent, doubtful because society may come, in the case of the duel example, to a decision independent of the gentlemen's choice, whereas in Raiffa's example of the panel of experts the decision maker has, by assumption, no opinion of his own, independently of the aggregated opinions of the experts. But even if the Pareto condition were less objectionable in Raiffa's example, the duel example does indicate that this condition cannot be applied uncritically. Gilboa, Samet and Schmeidler suggest the following restriction: let  $x$  be an option on whose prospects the agents agree, i.e.  $x$  is an alternative with outcomes  $x(s_1), \dots, x(s_m)$  for the possible states of nature  $s_1, \dots, s_m$  such that for all individuals  $i, j$   $p_i(s_k) = p_j(s_k)$  (for  $1 \leq k \leq m$ ). Such an option will be called a lottery. The probabilities for the outcomes of an alternative  $x$  that is a lottery in the Gilboa/Samet/Schmeidler sense thus are the same for all individuals, and therefore their notion of a lottery corresponds roughly to a lottery in the von Neumann-Morgenstern sense (modulo the question whether the probabilities result from a chance mechanism or represent subjective ignorance). The restriction mentioned above can now be specified as follows:

**Restricted Pareto Indifference:** For all lotteries  $x, y$ : if  $x I_i y$  for all individuals  $i$  then  $x I_G y$

The restricted Pareto condition crowds out unanimity considerations for all those options for which the agents hold different beliefs with regard to the outcomes. In fact, it may be said that it copies Harsanyi's approach to preference aggregation in the wider context of preference aggregation under uncertainty. The result is remarkable:

**Theorem (Gilboa/Samet/Schmeidler):** The restricted Pareto condition is satisfied if and only if the collective probability function  $p_G$  is an affine

combination of the individual probability functions  $p_i$  and the collective utility function  $u_G$  is a linear combination of the individual utility functions  $u_i$ .

Gilboa, Samet and Schmeidler seem to understand their result as salvage to Harsanyi's Aggregation Theorem. Note, however, that the price is a radical treatment of unanimity in situations of uncertainty. Unanimity is resolutely ignored, unless it can be taken for granted that differences in subjective belief play no role. To some extent, this position may be defended by considerations that initially motivated Pareto requirements in the context of social choice theory. Collective (consensual) decisions in liberal, democratic societies should be sensitive to the aims and ambitions of its members and they should not impose collective judgments where these are not founded in the individuals' judgments. Societies, that is to say, should pay respect to the individuals' values, but it is less clear that the same should hold for the individuals' beliefs. A social planner intending to make a welfare decision should be committed to the individuals' interests, aims and values, but (s)he might well feel exempted from taking the individual judgments on the factual situation into account.

Nevertheless, in situations like those that are illustrated in Raiffa's example of a panel of experts preferences over alternatives are intrinsically preferences over uncertain prospects and their evaluation may depend crucially on the combined assessment of their desirability and their likelihood. The rejection of the Pareto condition is a move that may result in an aggregation theorem in the tradition of Harsanyi's. It does not answer the question under which circumstances a unanimous choice should be respected and Pareto requirements are appropriate.

Doubts concerning the unqualified Pareto condition motivate also Isaac Levi's proposal of robust Pareto unanimity (Levi 1990). Robust Pareto unanimity demands that the options under consideration are compared independently from the probabilities in the following sense: for any outcome  $x(s_j)$  individual  $i$  is supposed to consider not only his or her weighted utility  $p(s_i)u(x(s_i))$ , but also the weighted utilities that would result from adopting another person's probability function  $p_j$ . That is, each person is asked to make an evaluation of the options under consideration not only on the basis of his or her own beliefs but also from the perspective of the other agents. This leads to hypothetical assessments in which the agents keep their utilities for the outcomes fixed but hypothetically take into account other agents' probabilities. A preference  $xRy$  for  $x$  over  $y$  is then called robust if the expected utility of  $x$  is at least as great as the expected utility for  $y$ , for all probability functions  $p_i$  ( $1 \leq i \leq n$ ), i.e. if  $x$  has higher expected utility under all controversial probability assignments. To illustrate, in the

duel example above the preference of the gentlemen for duel is not robust, as none of the gentlemen would prefer a duel given the supposition that he would lose. Identifying the robust preferences makes sense from the perspective of a social decision maker, whereas the individuals might see no reason to adopt a perspective of whose falsehood they are convinced. Robust preferences, however, represent unanimity in a strong sense: agreement in robust preferences holds irrespective of individual probability judgments. Levi's qualified Pareto condition therefore demands that not all unanimous preferences are respected for the social preference ordering but only those that are robust.

But of course it may happen that the set of robust preferences is empty. In that case a social utility function will be a linear combination of the individual utilities, just as it is under application of the restricted Pareto condition proposed by Gilboa, Samet and Schmeidler. Levi's proposal may seem less drastic than the restricted Pareto condition, but it still implies a dismissal of unanimity as a value in itself. From the perspective of the individual agents, however, such a move may come untoward. They might insist on their unanimous choice, even when they are aware that their evaluations result from radically different beliefs and values. It is not clear to what extent a social planner may feel exempted from respecting unanimity by restricting the Pareto condition to robust preferences.

#### **4. Concluding remarks**

In the context of collective choice under uncertainty, the unqualified unanimity requirement is challenged by severe objections. In combination with the expected utility hypothesis the Pareto condition implies the impossibility of preference aggregation under uncertainty. Still, the deeper meaning of Pareto conditions, namely to found collective decisions in the individuals' evaluations of the options under consideration, thereby making any 'improvement' by switching from social alternative  $x$  to social alternative  $y$  subject to the individual's consent, is not confined to social choice problems free of uncertainty. The Pareto condition may not be plausible in all choice situations. But its rejection needs careful argumentation. In particular, a proper understanding of the norms that govern collective or even public decision-making requires a better understanding of the interaction of different conditions where each of them may claim initial plausibility but which jointly lead to difficulties. A closer analysis of unanimity conditions is only a first step in this endeavour.

## REFERENCES

- Arrow, K. 1951 *Social Choice and Individual Values*, New Haven and London: Yale University Press, 2<sup>nd</sup> edition 1963.
- Blackorby, C., Donaldson, D. and Mongin, P. 2000 “Social Aggregation without the Expected Utility Hypothesis”, working paper.
- Fleurbaey, M. and Hammond, P. 2004 “Interpersonally Comparable Utility”, in S. Barberà, P. Hammond, and C. Seidl (eds.), *Handbook of Utility Theory, Vol. 2*, Amsterdam: Kluwer Academic Publishers.
- Fleurbaey, M., Suzumura, K., and Tadenuma, K. 2003 “The Informational Basis of the Theory of Fair Allocation”, working paper.
- Gajdos, T., Tallon, J.-M., and Vergnaud, J.-C. 2005 “On the Impossibility of Preference Aggregation under Uncertainty”, working paper.
- Gilboa, I., Samet, D., and Schmeidler, D. 2004 “Utilitarian Aggregation of Beliefs and Tastes”, *Journal of Political Economy*, 112, 932-938.
- Griffin, J. 1986 *Well-Being. Its Meaning, Measurement and Moral Importance*, Oxford: Clarendon Press.
- Harsanyi, J. 1955 “Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility”, *Journal of Political Economy*, 63, 309-321.
- Hild, M. 2001 “The Instability of Robust Aggregation”, Social Science Working Paper 1113, Division of the Humanities and Social Sciences, California Institute of Technology.
- Hild, M., Jeffrey, R. and Risse, M. 2001 “Preference Aggregation after Harsanyi”, in M. Salles and J. Weymark (eds.), *Justice, Political Liberalism, and Utilitarianism. Themes from Harsanyi and Rawls*, Cambridge: Cambridge University Press, forthcoming.
- Hylland, A. and Zeckhauser, R. 1979 “The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values”, *Econometrica*, 47, 1321-1336.
- Levi, I. 1990 “Pareto unanimity and Consensus”, *Journal of Philosophy*, 87, 481-492.
- Mongin, P. 1995 “Consistent Bayesian Aggregation”, *Journal of Economic Theory*, 66, 313-351.
- Raiffa, H. 1968 *Decision Analysis. Introductory Lectures on Choices under Uncertainty*, New York: McGraw-Hill.
- Seidenfeld, T., Kadane, J., Shervish, M. 1989 “On the Shared Preferences of two Bayesian Decision Makers”, *Journal of Philosophy*, 86, 225-244
- Sen, A. 1970 *Collective Choice and Social Welfare*, San Francisco: Holden.
- Weymark, J. 1993 “A reconsideration of the Harsanyi-Sen debate on utilitarianism”, in J. Elster and J. Roemer (eds.), *Interpersonal Comparisons of Well-Being*, Cambridge: Cambridge University Press, 255-321.