Preference Aggregation After Harsanyi

Matthias Hild
Christ's College, Cambridge

Richard Jeffrey
Department of Philosophy, Princeton University

Mathias Risse
Department of Philosophy, Princeton University

August 3, 1998

Justice, Political Liberalism, and Utilitarianism: Proceedings of the Caen Conference in Honor of John Harsanyi and John Rawls

Edited by Maurice Salles and John A. Weymark

1 Introduction

Consider a group of people whose preferences satisfy the axioms of one of the current versions of utility theory, such as von Neumann-Morgenstern (1944), Savage (1954), or Bolker-Jeffrey (1965). There are political and economic contexts in which it is of interest to find ways of aggregating these individual preferences into a group preference ranking. The question then arises of whether methods of aggregation exist in which the group's preferences also satisfy the axioms of the chosen utility theory, and in which at the same time the aggregation process satisfies certain plausible conditions (e.g., the Pareto conditions below).

The answer to this question is sensitive to details of the chosen utility theory and method of aggregation. Much depends on whether uncertainty, expressed in terms of probabilities, is present in the framework and, if so, on how the probabilities are aggregated. The goal of this paper is (a) to provide a conceptual map of the field of preference aggregation—with special emphasis, prompted by the occasion, on Harsanyi's aggregation result and its relations to other results—and (b) to present a new problem ("Flipping") which we see as leading to a new impossibility result.

The story begins with some bad news, roughly 50 years old, about "purely ordinal" frameworks, in which probabilities play no role.[1]

> **Arrow's General Possibility Theorem (1950, 1951, 1963)** No universally applicable non-dictatorial method of aggregating individual preferences into group preferences can satisfy both the Pareto Preference condition (Unanimous individual preferences are group preferences) and the condition of Independence of Irrelevant Alternatives (Group preference between two prospects depends only on individual preferences between those same prospects).

But for nearly as long we have had some good news about the *vN–M* (von Neumann-Morgenstern) framework, in which probabilities play an essential role:[2]

> **Harsanyi's Representation Theorem (1955)** If individual and group preferences all satisfy the *vN–M* axioms, if ("Pareto Indifference") the group is indifferent whenever all individuals are, and if ("Strong Pareto") group preference agrees with that of an individual whenever no individual has the opposite preference, then group utility is a linear function W of individual utilities.

[1] Sen (1970) chapter 3 provides an excellent exposition.

[2] Here and in sec. 2 we draw on Weymark's (1991) reconstruction of the Harsanyi theorem.

Both news items are accurate. Their differences stem from differences in the requirements they place on utility functions that count as representing a given preference ordering. Arrow's framework was "purely ordinal" in the sense that for a utility function to count as a representation of a preference ordering he only required the numerical ordering of utilities to agree with the given preference ordering of prospects. But in the von Neumann-Morgenstern framework, where the agent is assumed to have preferences between lotteries that yield particular outcomes with particular numerical probabilities, there is a second requirement: The place of a lottery in the preference ranking must correspond to the *eu* (the expected utility, the probability–weighted sum) of the utilities of its possible outcomes. In the *vN–M* framework utilities of outcomes and *eu's* of lotteries are uniquely determined by the preference ranking once a zero and a unit have been chosen.

Actual personal probabilities play no part in Harsanyi's aggregation process: even though individuals may have personal probabilities and use them to solve their own decision problems, the process does not aggregate these into group probabilities; it is only personal utilities for outcomes that are aggregated. These will determine social *eu's* for chancy prospects in which outcomes are assigned definite numerical probabilities. Harsanyi's result will be our main concern in section 2.

In various other frameworks, e.g., Savage's (1954), and Bolker and Jeffrey's (1965), personal probabilities as well as utilities are deducible from preferences. If both group and individual preferences are to be placed in these frameworks we need to decide how to use personal probabilities as well as personal utilities in the aggregation process—a decision that does not arise in the von Neumann-Morgenstern framework. There are two ways to go: "*ex ante*" and "*ex post*". (Harsanyi's own method of aggregation falls into neither of these categories, since personal probabilities have no place in his *vN–M* framework.) Both methods of aggregation face serious problems.

In *ex ante* aggregation (sec. 3) group *eu* is a function—say, W—of individual *eu's.* Here the question arises: under what conditions is the aggregate $W(eu_1, \ldots, eu_I)$ of individual *eu's* itself an *eu*? The answer is bad news for those who hope to use aggregation as a way of arriving at compromises among conflicting judgments of fact or value:[3]

> **Generic ex ante Impossibility Theorem.** In general, *ex ante* aggregation is possible only for groups that are highly homogeneous in their probability judgments or in their value judgments.

In *ex post* aggregation (sec. 4) individual *eu's* are first disintegrated into utilities and probabilities. These are then aggregated separately into group

[3] Among the bearers of bad tidings have been Broome (1987), (1990), Seidenfeld *et al.* (1989), and Mongin (1995).

utilities and group probabilities, which are finally reintegrated into group *eu's*. This blocks the difficulty that led to the generic *ex ante* possibility theorem. But later in sec. 4 we announce some new bad news for the *ex post* approach:

> **Flipping.** In *ex post* aggregation utility and probability profiles for individuals exist relative to which group preference between some pair of options reverses repeatedly or even endlessly as the analysis is refined, although individual preferences remain constant throughout these analyses.

Finally, we note that Harsanyi's good news is not vitiated by the flipping phenomenon, and we suggest a connection between that fact and a certain sort of individualism.

2 Harsanyi's Utilitarianism

In "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility" (1955) Harsanyi challenged Arrow's (1951, p. 9) thesis "that interpersonal comparison of utilities has no meaning and, in fact, that there is no meaning relevant to welfare comparisons in the measurability of individual utility." Both saw themselves as responding to Bergson's (1938, 1948) challenge "to establish an ordering of social states which is based on indifference maps of individuals".[4] But their responses were radically different, with Arrow reaffirming the ordinalism of the 1930's, and Harsanyi rejecting it in favor of von Neumann and Morgenstern's revived cardinalism, which he applied to social as well as individual preferences.

Needing cardinal utilities for game theory, von Neumann and Morgenstern had turned the tables on ordinalists who had argued that the significance of a numerical utility function for prospects $X, Y, \ldots$ is exhausted by the corresponding order relation ($\succeq$) of preference–or–indifference on those prospects:

$$(1) \qquad u(X) \geq u(Y) \text{ iff } X \succeq Y$$

By replacing the old prospects $X, Y, \ldots$ by the set $\mathcal{G}$ of all gambles among them[5] and replacing the utilities $u(X), u(Y), \ldots$ by expected utilities $eu(P), eu(Q), \ldots$ relative to $P, Q, \ldots \in \mathcal{G}$ they obtained a preference relation with definite cardinal significance:

$$(2) \qquad eu(P) \geq eu(Q) \text{ iff } P \succeq Q$$

[4] The words are Arrow's (1951, p. 9).

[5] In these gambles the probabilities of outcomes must be specified explicitly in numerical form, e.g., "Victory with probability .1, defeat with probability .9". The contrast is with specifications in terms of events for which different individuals might have different probabilities, e.g., "Victory if Ruritania joins us, defeat if it does not".

Here $eu(P) = u(X)P(X) + u(Y)P(Y) + \ldots$, and similarly for $eu(Q)$. In the presence of (2) the full set of monotone increasing transformations of u under which (1) is preserved shrinks to its positive affine subset.

Note that it is not eu's (or their ratios, or differences) that are invariant, but ratios of differences, ratios of "preference intensities":[6]

$$\frac{eu(P) - eu(Q)}{eu(R) - eu(S)} = \frac{\text{intensity of preference for P over Q}}{\text{intensity of preference for R over S}} \tag{3}$$

Harsanyi used Marschak's formulation of the *vN–M* theory. In Marschak's framework the outcomes $X, Y, \ldots$ of gambles are off stage; it is only members of the set $\mathcal{G}$ that appear on stage. But each outcome off stage is represented on stage by the member of $\mathcal{G}$ that assigns probability 1 to it and 0 to all the others.

Marschak's Postulates[7]

For $P, Q, R, S \in \mathcal{G}$ and $x, \tilde{x} \in [0, 1]$ where $\tilde{x} = 1 - x$:

M_1 $\succeq$ is a complete, transitive relation on $\mathcal{G}$.

M_2 If $P \succ Q \succ R$ then $xP + \tilde{x}R \approx Q$ for some x.

M_3 $P \succ Q \succ R \succ S$ for some P, Q, R, S.

M_4 If $Q \approx R$ then $xP + \tilde{x}Q \approx xP + \tilde{x}R$ for all P, x.

Representation Theorem

Given $M_1 - M_4$ there exist functions eu satisfying (2). These are unique up to a positive affine transformation.

InTRApersonal comparison of preference intensities

To compare i's preference intensity for P_1 over P_2 with that for P_3 over P_4, select suitable test–gambles P_{14}, P_{23} from $\mathcal{G}$, i.e.,

$$P_{14} = \frac{1}{2}P_1 + \frac{1}{2}P_4, \qquad P_{23} = \frac{1}{2}P_2 + \frac{1}{2}P_3, \tag{4}$$

and note their relative positions in i's preference ranking. It will turn out that $eu_i(P_1) - eu_i(P_2) \geq eu_i(P_3) - eu_i(P_4)$ iff $P_{14} \succeq_i P_{23}$, for by (2), the three conditions (5) are equivalent:

$$\frac{eu_i(P_1) - eu_i(P_2)}{eu_i(P_3) - eu_i(P_4)} \geq 1, \ \frac{eu_i(\frac{1}{2}P_1 + \frac{1}{2}P_4)}{eu_i(\frac{1}{2}P_2 + \frac{1}{2}P_3)} \geq 1, \ P_{14} \succeq P_{23} \tag{5}$$

In a single episode of group decision making, the group (e.g., perhaps, a legislature) will choose from a small set of pairwise incompatible options (perhaps, bills for combinations of taxation and public expenditure). The set $\mathcal{G}$ of all probability distributions over those options is the common field of

[6] See remark (3) at the end of this section.

[7] $\succeq, \succ$, and $\approx$ are the relations of weak preference, strong preference, and indifference.

the group preference ranking $\succeq_0$ and the individual preference rankings $\succeq_i$ of group options. In Harsanyi's postulates the number 0 represents a group and the numbers $1, \ldots, I$ represent the individuals who make it up.

Harsanyi's postulates. For $i, j = 1, \ldots, I$ and $P, P_i, Q \in \mathcal{G}$:
H_1 All individuals' rankings $\succeq_i$ satisfy $M_1 - M_4$.
H_2 So does the group's ranking, $\succeq_0$.
H_3 $Functionality$: $P \approx_0 Q$ if $P \approx_i Q$ for all i.
H_4 $Uniqueness$: $\exists Q\ \forall i\ \exists P\ \forall j \neq i\ (P \succ_i Q$ but $P \approx_j Q)$.[8]
H_5 $Positivity$: $P \succeq_0 Q$ if $P \succeq_i Q$ for all i and $\succ_i$ for some i.

Harsanyi's Aggregation (= Representation) Theorem: Postulates $H_1 - H_5$ imply the existence of *eu*'s for the preferences $\succeq_0, \succeq_i$ that satisfy the condition $eu_0 = \sum_i eu_i$. These are unique up to a positive affine transformation.

For an accessible explanation of the axioms and a proof of a somewhat stronger form of this theorem, see Weymark (1991) sec. 3.[9]

When is individual i's preference intensity for P_1 over P_2 greater than (or less than, or equal to) individual j's for P_3 over P_4? This is the form that questions of interpersonal comparison of utilities take when individual and group preferences determine only ratios of differences of utilities as in (3) above. These may well be substantive questions, which people do sometimes manage to answer correctly by various devices appropriate to particular persons and their situation.[10] Answers to such questions guide the synthesis of group preferences out of individual ones.

But here we work backwards, from a group preference ranking that all find acceptable as an even–handed aggregation of their various preferences to the interpersonal comparison of individual utility differences which that ranking presupposes. Whether or not the individuals have accurately answered the substantive questions, their group ranking can be analyzed so as to discover what are in effect common judgments, right or wrong, of form "$r =$ the ratio of i's preference intensity for P_1 over P_2 to j's for P_3 over P_4".

The idea is adequately illustrated in the case of a two–person group. Suppose that, somehow or other, individuals 1 and 2 have come to regard a particular preference ranking $\succeq_0$, satisfying $H_1 - H_5$ for the group constituted by the two of them, as an even–handed aggregation of their individual

[8] Harsanyi (1955) does not state H_4 as an axiom, but presupposes it in the first sentence of the proof of his Theorem V. Note that in H_4, P depends on i but Q does not.

[9] In his treatment, Weymark (1991, p. 272) permutes the first two quantifiers in H_4 to obtain a weaker axiom ("Independent Prospects") in which both P and Q depend on i, and which still yields uniqueness.

[10] See Harsanyi (1955, 1990) and Weymark's (1991) counterarguments. See also Jeffrey (1992), chapter 10.

preference rankings, $\succeq_1$ and $\succeq_2$. Then any function eu_0 representing $\succeq_0$ can be used to determine whether or not given functions eu_1, eu_2 representing the personal rankings are interval–commensurate:

> **Interval Commensuration Revealed Retrospectively.** If $H_1 - H_5$ hold with $I = 2$, then by H_4 there are $P_1, P_2, Q \in \mathcal{G}$ satisfying (a) and (b).
>
> $$\text{(a) } P_1 \succ_1 Q \approx_1 P_2 \qquad \text{(b) } P_2 \succ_2 Q \approx_2 P_1$$
>
> Representations eu_1, eu_2 of $\succeq_1, \succeq_2$ will be called "interval commensurate" iff some (and, so, every) representation eu_0 of $\succeq_0$ satisfies
>
> $$\text{(6)} \qquad \frac{eu_1(P_1) - eu_1(Q)}{eu_2(P_2) - eu_2(Q)} = \frac{eu_0(P_1) - eu_0(Q)}{eu_0(P_2) - eu_0(Q)}$$
>
> Given conditions (c) and (d), formula (6) follows from conditions (a) and (b):[11]
>
> (c) $eu_0(P) = eu_1(P) + eu_2(P)$
> (d) eu_0, eu_1, eu_2 represent $\succeq_0, \succeq_1, \succeq_2$

Note that differences of form $eu_j(P) - eu_j(Q)$ are not uniquely determined by the corresponding relation $\succ_j$, but ratios of such differences *are*—e.g., as on the right–hand side of (6).[12] Then in view of (6) the ratio of differences for $j = 1, 2$ (i.e. a ratio of interval commensurate preference intensities) is fixed by certain group preference intensities, and thus, in view of Marschak's representation theorem, by the group's preference ranking.[13]

We conclude this section with three remarks:

(1) Of course questions of interpersonal comparison are idle if Harsanyi's aggregation theorem is vitiated by an *ex ante* impossibility theorem, as some would seem to think;[14] but it is not so. On the contrary, Harsanyi's method

[11] *Proof.* By (a), (b), (d) the denominator on the left of (6) is non–null. Now operate on the right: First apply (c) to the four eu_0 terms; by (a) and (b) we may now substitute $eu_1(Q)$ for $eu_1(P_2)$ and $eu_2(Q)$ for $eu_2(P_1)$; after cancelling the $\pm eu_2(Q)$ terms in the numerator and the $\pm eu_1(Q)$ terms in the denominator, equation (6) becomes an identity.

[12] The social preference ranking determines eu_0 uniquely up to an affine transformation $eu_0 \mapsto a \cdot eu_0 + b$ with $a > 0$, and the value of the right–hand side of (6) is unaffected by any such transformation because we can drop $b - b$ from the numerator and the denominator, after which the a's in the numerator cancel those in the denominator.

[13] By confining this commensuration technique to consecutive pairs $(1, 2), \ldots (I-1, I)$ of individuals, Harsanyi's aggregation result might be obtained with H_4 weakened to this: $\forall i = 1, \ldots, I-1[\exists P\, \exists Q(P \succ_i Q \text{ but } P \approx_{i+1} Q) \text{ and } \exists P\, \exists Q(P \succ_{i+1} Q \text{ but } P \approx_i Q)]$.

[14] Broome (1991, pp. 160, 201) *seems* to be saying that Harsanyi's scheme is vitiated in that way, but this impression is created by his broad use of the term "Harsanyi's theorem" not only for Harsany's own aggregation theorem (above), but for variants of it in which the *vN–M* framework is replaced by frameworks like those of Savage and Bolker–Jeffrey, in which personal probabilities figure alongside utilities.

of utility aggregation is immune to *ex ante* impossibility theorems simply because, as we have observed, it is neither *ex ante* nor *ex post.*

(2) The object of the *vN–M* and Marschak axiomatic treatments of preference was to counter the view that game theory's cardinal concept of utility was metaphysical nonsense. Since there were no such qualms about the long–run frequency view of cardinal *probability*, von Neumann and Morgenstern adopted that view in their exposition (p. 19):

> "Probability has often been visualized as a subjective concept, more or less in the nature of an estimation. Since we propose to use it in constructing an individual, numerical estimation of utility, the above view of probability would not serve our purpose. The simplest procedure is, therefore, to insist upon the alternative, perfectly well founded interpretation of probability as frequency in long runs. This gives directly the necessary numerical foothold.[2]
>
> [2] If one objects to the frequency interpretation of probability then the two concepts (probability and preference) can be axiomatized together. This too leads to a satisfactory numerical concept of utility which will be discussed on another occasion."

But what made Harsanyi adopt the vN–M framework was no commitment to a long run frequency view of probability; rather, it was his view of probability as (in von Neumann and Morgenstern's words, above) "a subjective concept, more or less in the nature of an estimation." Harsanyi was that sort of subjectivist well before Savage showed how personal probabilities of events can be recovered from personal *eu*'s—i.e., ultimately, from personal preferences among gambles on those events. From the start, Harsanyi took it for granted that your expectations concerning random variables would be represented by probability–weighted means in which the probabilities are "subjective", representing your own uncertain judgments.[15] He could use the *vN–M* utility theory without the sorts of qualms mentioned in the unkept promise made in their footnote 2, above—a promise that Savage later made good.[16] The *vN–M* theory provided Harsanyi with a random variable u that could be combined with personal probabilities, exogenous to that theory, to yield exogenous personal *eu's*. It was Ramsey (1931) and Savage (1954) who provided decision theories with endogenous personal probabilities as well as utilities.

[15] In this sense of the term, Carnap (1945, 1950, 1962) was also a subjectivist. Like Carnap, Harsanyi took the legitimate source of the differences between different people's "subjective" probability judgments to be differences in the data on which those judgments are based.

[16] Savage (1954) points out that Ramsey (1931) had made the promise good decades earlier.

(3) We *form* our preference ranking of acts under uncertainty by judging the probabilities and utilities of the possible outcomes of those acts as best we can. From this constructive point of view it is our probability and utility judgments that determine our *eu's*, and our *eu's* that determine our preferences. This way of forming preferences has been tuned up over the past three centuries and more. A high–tech version can be found in Raiffa's 1968 "How to Think" book for MBA's. And a low–tech version had the place of honor at the end Arnauld's 1662 "How to Think" book for the innumerate:

> "To judge what one must do to obtain a good or avoid an evil, it is necessary to consider not only the good and the evil in itself, but also the probability that it happens or does not happen; and to view geometrically the proportion that all these things have together."

Representation theorems are analytical, not constructive: given a fully formed preference ranking that satisfies the axioms, they assure us of the existence of *eu* functions that represent the ranking, and of the uniqueness of those representations up to a positive linear transformation. But of course we do not have fully formed preference rankings over all the prospects that interest us. (If we did, we could simply read the solutions to our decision problems off them.) The problem in decision making is the constructive one of forming or discovering preferences we can live with. From the analytical point of view taken in representation theorems it is true enough that an *eu* function is a mere representation of a given preference ranking. But from the point of view of decision makers it is their preference rankings that merely represent their *eu* functions, which in turn merely reflect their probabilities and utilities.

3 Aggregation ex ante

We now turn to frameworks for preference in which actual personal probabilities play a role—in particular, the Savage framework in the present section, and the Bolker–Jeffrey framework in sec. 4. In the *vN–M* framework numerical probabilities of lottery outcomes are specified explicitly, and actual personal probabilities play no role. In the new frameworks personal probabilities play a central role, and are recoverable from the given preference ranking if it satisfies the relevant axioms. Here are thumbnail sketches of the two frameworks:

Savage. Preference is a relation between "acts". Acts are represented by functions f, each of which assigns to each possible "state of nature" s a definite "consequence" $f(s)$. If the act is betting \$10 on Bluebell to win, then we have

$$f(s) = \text{ "be \$10 richer" if Bluebell wins in state } s$$

$$f(s) = \text{“be \$10 poorer” if Bluebell does not win in state } s$$

The expected utility $eu(f)$ of an act f is the mean value of $u(f(s))$ for all states of nature s, weighted with the individual's personal probability distribution P over the states of nature. Savage's representation theorem guarantees the existence of functions u and P which together represent the preference ranking in the sense that act f is preferred to act g if and only if $eu(f)$ is greater than $eu(g)$.

Bolker–Jeffrey. Here preference is a relation between “events” A (i.e., between the same things to which probabilities are attributed), and utilities $u(s)$ are attributed to states of nature s. Performing an act is a matter of making some particular event true, e.g., the event of betting $10 on Bluebell to win. Given a utility function u and a probability function P, the “desirability” $des(A)$ of an event A is defined as the mean value of $u(s)$ for all states of nature s, weighted with the conditional probability distribution $P(-|A)$. According to Bolker's representation theorem truth of event A is preferred to truth of event B if and only if the desirability of A is greater than that of B.[17]

Desirability can be defined as conditional expectation of utility,[18] $des(A) = E(u|A) = \int_A u\, dP(-|A)$. In the discrete case, where the set S of states of nature is finite or countably infinite, the integral becomes a sum:

$$des(A) = \sum_{s \in A} u(s) P(\{s\}|A) \tag{7}$$

Example: “Dessert?” Consider Alice's problem of deciding whether to say “Yes” or “No” in answer to this question. She is sure that dessert would turn out to be chocolate ice cream (c), vanilla ice cream (v) or pie (p), i.e., $Dessert = \{c, v, p\}$—but she does not know which.

Data: For these possibilities her probabilities conditionally on $Dessert$ are $P_{Alice}(\{c\}|\{c, v, p\}) = P_{Alice}(\{v\}|\{c, v, p\}) = \frac{1}{8}$ and $P_{Alice}(\{p\}|\{c, v, p\}) = \frac{3}{4}$, and her utilities are $u_{Alice}(c) = 68$, $u_{Alice}(v) = -100$, $u_{Alice}(p) = 16$. For the remaining possibility, *None (“n”),* her utility is $u_{Alice}(n) = 0$.

[17] For accessible overviews of the theory see Bolker (1967), Jeffrey (1983), and Broome (1990). For important modifications of the theory see Joyce (1992) and Bradley (1997).

[18] Bolker's (1965, 1966, 1967) representation theorem guarantees existence of a function des representing preference between elements of a Boolean algebra—but on assumptions under which the algebra cannot be a field of sets (of “states”). Under those assumptions the function des is not the conditional expectation of any function $u(s)$. But of course existence of such a representation when those assumptions hold does not imply non–existence when they do not. Jeffrey (1992, chapter 15) recasts Bolker's theorem in a form applicable to Boolean algebras of sets of states—algebras on which $des(A)$ can be defined as $E(u|A)$ after all. (The gimmick is like the one Kolmogorov [1948, 1995] uses to transform fields of sets on which probability measures exist into Boolean algebras of the sort postulated in Bolker's theorem.)

Solution: As Alice sees it, the states of nature form the set $S = \{c, v, p, n\}$ and the event *Dessert* has desirability $des_{Alice}(\{c, v, p\}) =$

$$\textstyle\sum_{s\in\{c,v,p\}} u_{Alice}(s)P_{Alice}(\{s\}|\{c, v, p\}) = 68(\frac{1}{8}) - 100(\frac{1}{8}) + 16(\frac{3}{4}) = 8.$$

Then since $des_{Alice}(\{n\}) = u_{Alice}(n) = 0 < 18$, Alice does want dessert: $\{c, v, p\} \succ_{Alice} \{n\}$. Similar calculations show that she prefers pie to ice cream: $des_{Alice}(\{p\}) = u_{Alice}(p) = 16 > des_{Alice}(\{c, v\}) = -16$. Note that until she makes her decision, Alice's probability for dessert will be strictly between 0 and 1, e.g., $P_{Alice}(Dessert)$ might be 1/2, or 7/10, or whatever. But the actual value makes no difference to her decision, since the probabilities of interest are all conditional on *Dessert*, and we suppose (see Jeffrey 1996) that those remain constant as the unconditional probability of *Dessert* varies.

Where Savage assigns probabilities to events independently of what act is being performed, Bolker and Jeffrey assign conditional probabilities to events given acts. (Since acts are not events for Savage, these conditional probabilities make no sense for him.) The Bolker–Jeffrey framework allows probabilities to be updated either by observation or by decision: the updated unconditional probability will be the prior conditional probability given the event observed or chosen. But in the Savage framework choice of an option cannot affect probabilities. Note, too, that Savage's treatment is problematical in cases where it is important to consider players' probabilities for other players' performing various acts, as in interactive decision theory (= game theory).

Two Dismal Possibility Theorems. Here we note two specifications of the generic *ex ante* possibility theorem indicated in sec. 1. The species is Mongin's (1995) modification of the Savage framework—a modification in which an additional postulate assures σ-additivity of the probability measure.

Let $\succeq_i$, u_i, and P_i be individual i's preference relation, utility function, and probability function. Mongin adopts analogs of Harsanyi's "Pareto" conditions H_3 (functionality) and H_5 (positivity). To give these postulates material to work on he adds an assumption of diversity (linear independence) of the various individuals' probabilities or utilities. Either assumption implies the following condition, which is an analog of H_4:

> **Independence.** Each individual i has some preference $f \succ_i g$ where all others are indifferent: $f \succ_i g$, but $f \approx_j g$ if $j \neq i$.

Finally, Mongin postulates a minimal *Agreement* condition:

> **Agreement.** There exist consequences c_1, c_0 such that all individuals i assign higher utility to the former: $u_i(c_1) > u_i(c_0)$.

Mongin uses the term "overall dictator" for an individual whose probabilities and preference intensities are the same as Society's. Of course, such individuals need not really be dictators—e.g., they might be immensely public-spirited citizens, or ones whose personal attitudes are somehow formed by the same causes as the group's; or the "dictator" might be chosen by lot, or by vote; or the coincidence might be the result of blind chance. As Hylland and Zeckhauser (1979) point out, real dictatorship would be a property of the preference aggregation scheme, W, i.e., the property of assigning a particular individual's preferences to society regardless of what probabilities and utilities the others may have. But anyway it would be a very restrictive possibility theorem that implied the existence of Mongin's "dictators".

Below, Mgn_1 and Mgn_2 are weaker consequences of Mongin's main possibility results.[19] "Positivity" is the analog of H_5 (i.e., the group prefers f to g if some member does and none prefer g to f), and "Functionality" is the analog of H_3. In Mgn_2 we use the terms "diverse" and "clone" as follows:

Probability clones are individuals with identical probability functions.

Utility clones are individuals with affine equivalent utility functions.

Diversity of the individuals' probability functions means that all are distinct and none are weighted averages of others.

> $\mathbf{Mgn_1}$: In the modified Savage framework with functionality and positivity there will be an overall "Dictator" if no individual probability or utility function is a linear combination of others.

> $\mathbf{Mgn_2}$: In the modified Savage framework, Positivity and Agreement together imply (1) and (2):
> (1) If the probability functions are diverse, all are utility clones.
> (2) If not all are probability clones, some are utility clones.

Politics makes strange bedfellows. Results like Mgn_1 and Mgn_2 may seem less disturbing—only to be expected—in the light of the well known fact that unanimity about the relative ranking of two options may be based on quite incompatible assessments of probability or utility. Raiffa (1968, p. 230) offers a simple, striking example, with two options (a_1, a_2), two states of nature (θ_1, θ_2), and a pair of experts, Alice and Bob, who are indifferent between the options for very different reasons: Alice assigns probabilities $.2, .8$ to θ_1, θ_2 and utilities $1, 0, .5, 1$ to $a_1\theta_1, a_2\theta_1, a_1\theta_2, a_2, \theta_2$, while Bob assigns probabilities $.8, .2$ and utilities $.5, 1, 1, 0$ to the same states and act-state pairs. These experts have the same expected utilities ($.6$ for a_1, $.8$ for a_2) but for precisely opposite reasons. As Raiffa argues, such examples cast doubt on the seemingly ineluctable functionality principle, H_3. This idea is pushed further in the next section, under "flipping".

[19] See Mongin's (1995) observation 1 on p. 341, and proposition 7 on pp. 343-4.

4 Aggregation ex post

The strange bedfellows phenomenon may be seen as a warning against muddling judgments of fact and value, and as a call to take the *ex post* stance, in which members' *eu*'s are not directly aggregated, but are first analyzed into probabilities and utilities—which are aggregated separately into group probabilities and group utilities, and only then recombined into group expected utilities.

It should be noted that this stance, with its rationale, was forcefully enunciated by Raiffa (1968) 30 years ago in sections 12 and 13 of his classical text, *Decision Analysis*—e.g, on "The Problem of the Panel of Experts" (pp. 232-233):

> "If I were solely responsible as the decision maker, I should want to probe the opinions of my experts to assess my own utility and probability structure. I should try to keep my assessments for utilities separate from my assessments for probabilities, and I should try to exploit such common agreements as independence.[20] Wherever possible, I should want to decompose issues to get at basic sources of agreement and disagreement. I should compromise at the primitive levels of disagreement and adopt points of common agreement as my own, so long as these common agreements were not compensating aggregates of disagreements. I should do so knowing full well that I might end up choosing an action which my experts would say is not as good as an available alternative. Throughout this discussion, of course, I am assuming that I do not have to worry about the viability of my organization, its morale, and so on."

There, too, he reports a result of Zeckhauser's that would be published 11 years later (Hylland and Zeckhauser 1979) in a somewhat different version:

> "Richard Zeckhauser has proved a mathematical theorem that states this result:
> "No matter what procedure you use for combining the utility functions and for combining the probability functions, so long as you keep these separate and do not single out one individual to dictate the group utility and probability assignments, then you can concoct an example in which your experts agree on which act to choose but in which you are led to a different conclusion."
> (Raiffa 1968, p. 230)

[20] Convex combinations of i.i.d. distributions are not generally i.i.d., so averaging such distributions would not be a way of preserving common agreement on independence. (To preserve independence one could form the average of the individuals' i. i. d. distributions and use that as the 1–shot probability of an i. i. d. group distribution.)

Raiffa (1968, pp. 233-237) explores the tension this theorem reveals between the following two conditions.

> **Reification**: "the group members should consider themselves as constituting a panel of experts who advise the organizational entity: they should imagine the existence of a higher decision–making unit, the organization incarnate, so to speak, and ask what *it* should do. Just as it made sense to give up Pareto optimality in the problem of the panel of experts, it likewise seems to make sense in the group decision problem." (Raiffa, pp. 233-234)
> **Pareto Optimality**: The group prefers one prospect to another if some members do and none have the opposite preference.

We now introduce a new problem for *ex post* aggregation:[21]

> **Flipping:** In *ex post* aggregation, utility and probability profiles for individuals exist relative to which group preference between some pair of options reverses endlessly as the analysis is refined, even though all individual preferences remain unchanged.

Note how this relates to the result of Hylland and Zeckhauser. They use the *ex ante* Pareto condition in an *ex post* framework, i.e., a framework in which probabilities and utilities are aggregated separately. We use the *ex post* Pareto condition (i.e., on utilities, not expected utilities) in an *ex post* framework. Thus fliping is a problem inherent in the *ex post* approach: Unlike the Hylland–Zeckhauser (1979) result, it does not depend on the tension between *ex ante* standards and *ex post* aggregation.[22]

The flipping phenomenon is illustrated by the following example, which we formulate here in the Bolker–Jeffrey framework sketched in sec. 3 above.[23] In the example, initial group desirabilities $12, 0$ of two options (*dessert*, *none*) change to $-8, 0$ upon closer examination of the first option, and change back to $12, 0$ upon still closer examination. The group desirabilities can flip because the individuals have opposed probabilities and differently opposed utilities, somewhat as in the "politics makes strange bedfellows" example, but here with the opposed tendencies overbalancing in opposite directions at each stage of refinement.

[21] Here we illustrate the problem for a particular aggregation rule, i.e., straightforward averaging of probabilities and summing of utilities. But the problem can arise for any *ex post* Pareto optimal aggregation rule.

[22] See Hylland–Zeckhauser (1979, pp. 1325–6). Their axioms 2 and 3 stipulate *ex post* aggregation of individual probabilities p^k and utilities u^k. Their axiom 5 is a weak *ex ante* Pareto optimality condition: "If $E(a_m|p^k, u^k) > E(a_i|p^k, u^k)$ for all k, then a_i is not an element of the choice set."

[23] I.e. the simplest framework for the purpose. A treatment in a modification of the Savage framework will be published elsewhere.

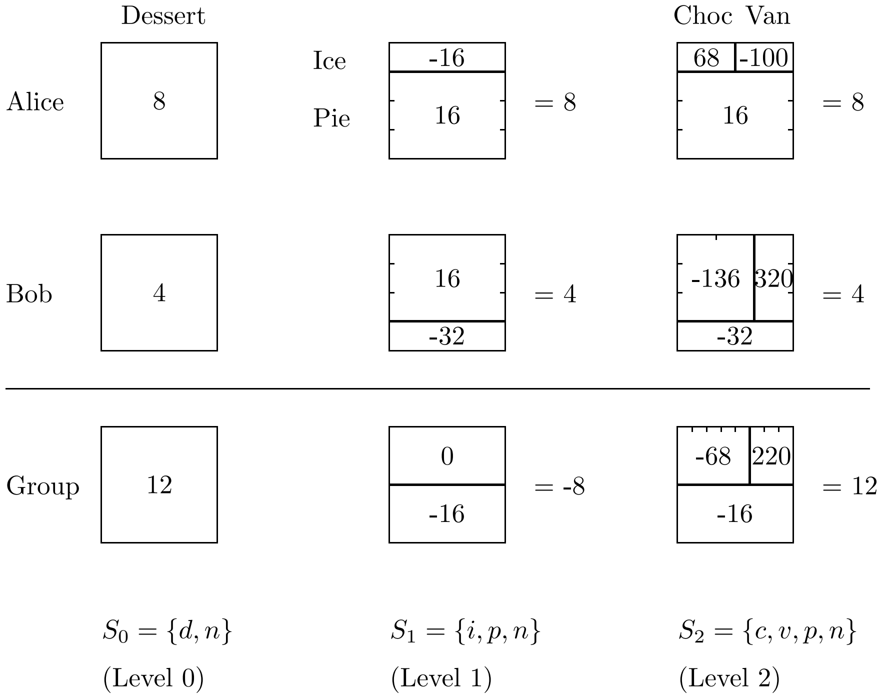

Figure 1: Flipping illustrated by refinements of the “Dessert” option.

Dessert makes strange bedfellows. Alice and Bob are being given a dinner for two in which they must make the same choice from the menu, course by course. Having agreed on all courses so far, they are trying to decide whether or not to have dessert, which the menu lists with no details. Suppose that in fact they both prefer the event *Dessert* to the event *None*, and that on personal desirability scales des_{Alice}, des_{Bob} which they regard as interpersonally commensurate, their desirabilities for *Dessert* are 8 and 4 as shown in Figure 1 (level 0, above the line), and their desirabilities for $None$ (not shown in Figure 1) are both 0. Suppose they are sure that dessert will turn out to be *Ice* cream or *Pie*, concerning which their respective commensurate desirabilities are $-16, 16$ for Alice, and $16, -32$ for Bob, as shown above the line at level 1 of Figure 1. Suppose that $P_{Alice}(Ice|Dessert) = 1/4, P_{Alice}(Pie|Dessert) = 3/4$, and that the values for P_{Bob} are just the reverse. These conditional probabilities are represented by the areas of the respective compartments, on a scale where the whole square has area 1. Since $des(A) = E(u|A)$, the desirability of the union of two events that are judged to be incompatible is a weighted average of their separate desirabilities: *If* $P(A \cap B) = 0$, *then*

$$des(A \cup B) = des(A)P(A|A \cup B) + des(B)P(B|A \cup B) \tag{8}$$

It is easy to verify that with $Dessert = Ice \cup Pie = A \cup B$ this equation, applied to Alice's and Bob's level 1 desirabilities and probabilities, yields their level 0 desirabilities, 8 and 4. And similarly, if both are convinced that Ice would turn out to be $Choc$(olate) or Van(illa), equation (8) delivers their ±16 level 1 desirabilities for Ice when their probabilities and desirabilities for $Choc$ and Van are as shown at level 2. Then above the line, the three levels of analysis of Alice's attitudes depicted in Figure 1 are mutually consistent, as are the three levels of Bob's.

But *ex post* aggregation of Alice's and Bob's desirabilities by applying the following formula to the numbers shown in Figure 1 yields mutually inconsistent group desirabilities, for the results, shown below the line, exhibit the flipping phenomenon: group desirabilities for *Dessert* flip from 12 to −8 and back again as the aggregation process is applied to finer analyses of the individuals' probabilities and desirabilities.

$$des_{Group}(A) = des_{Alice}(A) + des_{Bob}(A) \tag{9}$$

And it would be straightforward to devise probabilities and utilities for a further stage (say, with $Pie = Apple \cup Banana$) at which group desirability flips back from 12 at stage 2 to −8 at a new stage 3; and one can give an algorithm for continuing the refinements of consistent individual probabilities and utilities so as to carry the $12, -8, 12, -8, \ldots$ flipping process as far as you like—even, endlessly.

The flipping problem has another aspect, i.e., inconsistency of group probabilities and desirabilities with formula (8) when group probabilities conditionally on an act-event D (e.g., the event that we have dessert) are obtained by averaging:

$$P_{Group}(A|D) = \frac{1}{2}P_{Alice}(A|D) + \frac{1}{2}P_{Bob}(A|D) \tag{10}$$

Thus, the desirability of Ice at level 1, obtained via equation (9) as the simple sum of Alice's and Bob's level 1 desirabilities for Ice, is inconsistent with the value obtained via equation (8) as the probability–weighted average of the group's level 2 desirabilities for $Choc$ and Van:

$$des_{Group}(Ice) = -16 + 16 = 0 \text{ from } (9)$$

$$des_{Group}(Ice) = \frac{5}{8}(-68) + \frac{3}{8}(220) = 40 \text{ from } (8)$$

But is formula (9) a correct description of *ex post* aggregation? By definition, *ex post* aggregation adds *utilities*, not desirabilities, so that in genuine *ex post* aggregation formula (9) would be replaced by the corresponding formula for utilities:

$$(11) \qquad u_{Group}(s) = u_{Alice}(s) + u_{Bob}(s)$$

Can the effect of applying formula (11) be the same as that of applying formula (9) to the desirabilities of the smallest compartments in Figure 1? The answer is "Yes" if we represent the refinement process as applying primarily to the set S of states of nature, and only derivatively to the events, the subsets of S. Thus, at level 0 there are just two states of nature, the state d in which Alice and Bob have dessert, and the state n in which they have none: at level 0 the set of states of nature is $S_0 = \{d, n\}$ as indicated in Figure 1. The set S_1 of states at level 1 is obtained by replacing d by two states: a state i in which the waiter brings ice cream, and a state p in which he brings pie. And similarly S_2 comes from S_1 by replacing i by c (he brings chocalate ice cream) and v (he brings vanilla).

Here we have three Boolean algebras $\mathcal{A}_k$ of subsets of S_k, with $k = 0, 1, 2$. The algebra $\mathcal{A}_k$ contains $2^{(2^{k+1})}$ events, e.g., $\mathcal{A}_0 = \{\emptyset, \{d\}, \{n\}, S_0\}$. In these, *Dessert* is represented by three different sets: by $\{d\}$ at level 0, by $\{i, p\}$ at level 1, and by $\{c, v, p\}$ at level 2. We shall say that these three are "associated" with each other, in order to indicate that they are all representations of what is informally seen as one and the same event, *Dessert*. In general, any $A \in \mathcal{A}_k$ for $k = 0, 1$ is associated with an $A' \in \mathcal{A}_{k+1}$ defined as follows, where $\{s\} = S_k - S_{k+1}$ and $\{s', s''\} = S_{k+1} - S_k$:[24]

$$(12) \qquad A' = (A - \{s\}) \cup \{s', s''\} \text{ if } s \in A, \text{ else } A' = A$$

As an ideal beyond human powers of attainment, one could think of continuing this process of refinement endlessly, specifying not only the ways in which *Dessert* and *None* might turn out, but also possibilities about other things one might care about, e.g., the weather tomorrow (and tomorrow, and tomorrow, ...), various people's states of health, and births, deaths, wars, football scores—whatever. The *ultimate* states of nature are the maximal consistent sets of such specifications. From this idealized point of view the elements of S_k for finite k will be pseudo–states, events (sets of ultimate states) masquerading as states.

Where "s" ranges over ultimate states, aggregation via equations (10) and (11) is immune to the flipping phenomenon illustrated by Figure 1, e.g., because the putative utilities $u_{Alice}(p) = 16$, $u_{Bob}(p) = -32$ at level 1 must really be seen as desirabilities $des_{Alice}(Pie) = 16$, $des_{Bob}(Pie) = -32$ of an event *Pie*; and formula (11) is no warrant for summing desirabilities. But

[24] I.e., s is the element of S_k that is split into two elements s', s'' to produce S_{k+1}.

application of formula (11) to utilities of ultimate states is beyond human powers: this way out "in principle" leaves *ex post* aggregation impossible in practice. One way or the other, *ex post* aggregation looks like a pipe dream.

If the *ex post* approch is ruled out in this way, the *ex ante* approach has its own severe difficulties. In particular, the *ex ante* possibility theorems rule out any version of liberalism that satisfies the following two conditions. (1) Unanimous individual preferences are preserved as group preferences. (2) Diversity is tolerated as part of political reality, or even cherished, as in Mill's *On Liberty.* (By excluding all linear independence of probability measures and of utility functions, the *ex ante* possibility theorems exclude such diversity.) Liberalism that mets these two conditions violates Bayesian rationality of individuals or the group: it requires irrational people or an irrational society.

In closing we recall that flipping does not arise in Harsanyi's aggregation scheme, for the vN–M or Marschak framework attributes no judgmental probabilities to groups or to individuals.[25] From a certain individualistic point of view this opportunity to deny that groups have beliefs (i.e., judgmental probabilities) is most welcome. On that view we may perhaps speak of groups as agents, and even as having aggregate preferences, but on that view groups are not the sorts of things to which beliefs are to be attributed, and so groups are not to be thought of as rational or irrational.

References

Arnauld, A. (1662), *La logique, ou l'art de penser.* Paris. Translation (1964), *The Art of Thinking*, Indianapolis: Bobbs–Merrill.

Arrow, K. (1950), A Difficulty in the Concept of Social Welfare. *Journal of Political Economy* **58**, 328-346; reprinted in Arrow and Scitovsky

Arrow, K. (1951, 1963), *Social Choice and Individual Values.* New York: Wiley.

Arrow, K. and Scitovsky, T., eds. (1969), *Readings in Welfare Economics.* Allen and Unwin, London.

Bergson, A. (1938), A Reformulation of Certain Aspects of Welfare Economics. *Quarterly Journal of Economics* **52**, 310-334; reprinted in Arrow and Scitovsky (1969).

Bergson, A. (1948), Socialist Economics. *A Survey of Contemporary Welfare*

[25] It does arise in other schemes that are neither *ex ante* nor *ex post,* e.g., that of Levi (1997, chapter 9), and the pseudo *ex post* scheme illustrated in Figure 1 above.

Economics, H. S. Ellis (ed.), pp. 412-448: Blakiston, Philadelphia.

Bolker, E. (1965), *Functions Resembling Quotients of Measures.* Ph.D. Dissertation, Harvard University.

Bolker, E. (1966), Functions Resembling Quotients of Measures. *Transactions of the American Mathematical Society* **124**, 293-312.

Bolker, E. (1967), A Simultaneous Axiomatization of Subjective Probability and Utility. *Philosophy of Science* **34**, 333-340.

Bradley, R. (1997). *The Representation of Beliefs and Desires within Decision Theory.* Ph. D. Dissertation, University of Chicago.

Broome, J. (1987), Utilitarianism and Expected Utility. *Journal of Philosophy* **84**, 402–422.

Broome, J. (1990), Bolker-Jeffrey Expected Utility Theory and Axiomatic Utilitarianism. *Review of Economic Studies* **57**, 477-503

Broome, J. (1991), *Weighing Goods.* Oxford: Basil Blackwell

Carnap, R. (1945), On Inductive Logic. *Philosophy of Science* **12**, 72-97.

Carnap, R. (1950, 1962), *Logical Foundations of Probability.* Chicago: University of Chicago Press.

Harsanyi, J. (1955), Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility. *Journal of Political Economy* 63: 309-321; reprinted in Arrow and Scitovsky (1969)

Harsanyi, J. (1990), Interpersonal Utility Comparisons. *Utility and Probability*, J. Eatwell, M. Milgate, and P. Newman (eds.), New York: Norton

Hylland, A. and R. Zeckhauser (1979), The Impossibility of Bayesian Group Decision Making with Separate Aggregation of Beliefs and Values. *Econometrica* 47: 1321-1336

Jeffrey, R. (1965, 1983), *The Logic of Decision.* 1st ed., New York: McGraw Hill. 2nd ed., Chicago: University of Chicago Press

Jeffrey, R. (1992), *Probability and the Art of Judgment.* Cambridge: Cambridge University Press.

Jeffrey, R. (1996), Decision Kinematics. *The Rational Foundations of Eco-*

nomic Behaviour, K. J. Arrow, E. Colombatto, and M. Perlman (eds.), Macmillan (GB) and St. Martin's (USA).

Joyce, J. M. (1992), *The Foundations of Causal Decision Theory.* Ph. D. dissertation, University of Michigan, Ann Arbor.

Kolmogorov, A. N. (1948, 1995), Algèbres de Boole métriques complètes, *IV Zjadz Matemayików Poslkich*, Warsaw (1948) pp. 21-30. Translation: Complete Metric Boolean Algebras, *Philosophical Studies* **77** (1995) 57-66.

Levi, I. (1997), *The Covenant of Reason.* Cambridge: Cambridge University Press.

Marschak, J. (1950), Rational Behavior, Uncertain Prospects, and Measurable Utility. *Econometrica* **18**, 111-141

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

Raiffa, Howard (1968), *Decision Analysis.* Reading, Mass.: Addison–Wesley

Ramsey, F. P. (1931), Truth and probability, in *The Foundations of Mathematics and other Logical Essays*, R. B. Braithwaite (ed.), Kegan Paul. Reprinted in Ramsey (1990).

Ramsey, F. P. (1990), *Philosophical Papers*, D. H. Mellor (ed.), Cambridge University Press.

Savage, L. J. (1954), *Foundations of Statistics.* New York: Wiley

Seidenfeld, T., Kadane, J. B., and Schervish, M. J. (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–Day.

von Neumann, J. and O. Morgenstern (1944, 1947, 1953), *Theory of Games and Economic Behavior.* Princeton: Princeton University Press.

Weymark, J. A. (1991), A Reconsideration of the Harsanyi–Sen Debate on Utilitarianism, in *Interpersonal Comparisons of Well–Being*, J. Elster and J. Roemer (eds.), Cambridge: Cambridge University Press.