

# The aggregation of imprecise probabilities

Robert F. Nau\*

*Fuqua School of Business, Duke University, Durham, NC 27708-0120, USA*

Received 20 September 1999; received in revised form 18 February 2000; accepted 25 January 2001

---

## Abstract

Two methods are presented for the aggregation of imprecise probabilities elicited from a group of experts in terms of betting rates. In the first method, the experts bet with a common opponent subject to limits on their personal betting stakes, and their individual and aggregate beliefs are represented by confidence-weighted lower and upper probabilities. This method is applicable to situations in which the events in question are discrete, few in number, and ambiguous in nature, where interest centers on the degree of consensus or dissensus and on the reconciliation of incoherence among a small group of experts. In the second method, the experts bet directly with each other as a means of reconciling incoherence, and their beliefs are represented by lower and upper risk neutral probabilities—i.e., products of probabilities and relative marginal utilities for money. The latter method is applicable to situations in which many individuals jointly assess probability distributions for large numbers of events or continuous random variables. In both methods, the fundamental concept of probability is an interpersonal one, with irreducible game-theoretic elements. © 2002 Elsevier Science B.V. All rights reserved.

*MSC:* primary 60A05; Secondary 62C05; 91A99; 91B10

*Keywords:* Lower and upper previsions; Confidence weighted probabilities; Risk neutral probabilities; Expert resolution; Coherence; Fuzzy sets

---

## 1. Introduction

In statistical applications where subjectively assessed prior probabilities play a role (e.g., Bayesian models, probabilistic risk analysis, various kinds of forecasting), the probability estimates of two or more experts are often aggregated in order to gain the benefit of multiple knowledge bases and minimize the potential biasing effects of extreme opinions. Yet the question of *how* the subjective probabilities of different individuals ought to be aggregated has been controversial. Standard Bayesian theory provides little guidance: according to the Savage axioms, all subjective probabilities are purely *personal* probabilities. (See Winkler et al., 1986, for a discussion.) Various

---

\* [www.duke.edu/~rnau](http://www.duke.edu/~rnau)

*E-mail address:* robert.nau@duke.edu (R.F. Nau).

authors have, therefore, proposed additional axioms or desiderata that a rule for aggregating subjective probabilities ought to satisfy, such as (i) *unanimity-preservation*: if the experts all agree, then their aggregated probabilities should equal their common probabilities; (ii) *marginalization*: aggregation of marginal probabilities should yield the same result as marginalization of the aggregate joint distribution; and (iii) *external Bayesianity*: aggregation of prior probabilities followed by Bayesian updating (using a common likelihood function) should yield the same result as aggregation of posterior probabilities. Unfortunately, no one probability-aggregation rule satisfies all of these properties: for example, only a generalized linear aggregation rule satisfies (ii) and only a generalized logarithmic aggregation rule satisfies (iii), while supra-Bayesian methods (which treat the probabilities as data evaluated by an outside decision-maker) typically do not satisfy either (i) or (iii). (See Genest and Zidek, 1986, for a discussion.) It has often been suggested that imprecise probabilities might be more suitable for aggregation than precise probabilities, because they allow room for disagreement among individuals, they allow the experts to quantify the sharpness of their own opinions, and they provide additional parameters that might reflect the degree of consensus or dissensus (Walley, 1982; Moral and del Sagrado, 1998). However, aggregation methods for imprecise probabilities have met with theoretical and practical difficulties of their own.

This paper surveys two approaches to the aggregation of imprecise probabilities, based on the author's prior work, both of which have firm quasi-Bayesian theoretical foundations and circumvent the impossibility theorem for the aggregation of precise probabilities. They also both have implicit game-theoretic elements, which is not coincidental. I suggest that all measurements of subjective probabilities are intrinsically intersubjective—that they do not represent beliefs that exist in vacuo, but rather beliefs that exist in a strategic environment inhabited by other individuals—and it is only through modeling of the intersubjective dimension that the aggregation of probabilities can be justified and carried out in a non-arbitrary fashion.

In the first method, *confidence-weighted lower and upper probabilities* are used as the fundamental representation of uncertainty. The confidence weights permit tradeoffs to be made in a systematic way between the imprecise probability judgements of different individuals, and the confidence-weighted representation has an axiomatic foundation that supports the aggregation of opinions through a slight weakening of the property of transitivity as well as the abandonment of completeness (Nau, 1992). In the second method, the parameters of interest are not “true” subjective probabilities but rather *risk-neutral probabilities*—i.e., products of probabilities and relative marginal utilities for money. An individual's risk-neutral probabilities are the apparent probabilities revealed by his acceptance of money bets under conditions of risk aversion (non-linear utility for money and significant prior stakes in the outcomes of events). The natural way to aggregate risk-neutral probabilities is to simply let a group of individuals make bets or trade lottery tickets with each other and then record the final betting rates or ticket prices that are obtained after strategic maneuvering has taken place, reciprocal learning has occurred, risks have been hedged, and arbitrage opportunities have been

exploited. Markets for betting on securities prices, sporting events, and political races are conducted in exactly this fashion. Both of these approaches give rise to a “representative individual”, and the behavior of the representative individual is qualitatively identical to the behavior of a single individual. Hence, the representative individual is necessarily quasi-Bayesian. Furthermore, in both approaches, the imprecision in the revealed probabilities is partly due to the incompleteness of beliefs and partly due to irreducible strategic considerations.

The organization of the remainder of the paper is as follows: Sections 2 and 3 describe the aggregation methods for confidence-weighted probabilities and risk-neutral probabilities, respectively, and Section 4 presents a concluding discussion.

## 2. Aggregation of confidence-weighted probabilities

In one of the most widely-used models of imprecise probabilities (e.g., Smith, 1961; Walley, 1991), a *convex set of probability measures* is used to represent the beliefs of an individual, yielding lower and upper probabilities for events as well as lower and upper expectations for random variables. This representation of uncertainty is a natural generalization of a precise subjective probability distribution that is obtained by relaxing the axiom of completeness. But it is not obvious how aggregation should be carried out under this representation. For example, an aggregate measure of uncertainty could be defined either by the union or the intersection of the sets of probability measures of different individuals (Walley, 1991, pp. 186–188), but both of these alternatives appear problematic. The union of convex sets of probability measures generally is non-convex, and it yields too loose a representation of aggregate uncertainty. (Non-convexity of the aggregate set of measures might be considered reasonable as a reflection of dissensus among the experts, but what matters in the end is how the set is to be used for purposes of decision making. If the ultimate quantities of interest are bounds on linear functions such as expectations of uncertain quantities, then only the convex hull matters.) As more opinions are pooled, the union can only get larger, and it reflects only the least informative opinions, whereas intuitively there ought to be (at least the possibility of) an increase in precision as the pool gets larger. On the other hand, the intersection of convex sets of measures may be empty if the experts are mutually incoherent, and it generally yields too tight a representation of aggregate uncertainty. As more opinions are pooled, the intersection can only shrink, and it reflects only the most extreme among those opinions, whereas intuitively there should be some convergence to an average opinion when the pool gets sufficiently large. Moreover, neither the union nor the intersection provides an opportunity for the differential weighting of opinions, which would be desirable in cases where one individual is considered (either by herself or by an external evaluator) to be better or worse informed than another individual about a particular event under consideration. Some sort of weighted averaging over lower and upper probabilities could be performed, in principle, but the theoretical rationale for doing so is unclear. Moral and del Sagrado (1998) present several methods for

weighted aggregation of convex sets of probabilities, based on ideas from Dempster–Shafer theory and fuzzy-sets theory, but these methods are fundamentally non-Bayesian.

It has often been pointed out that an infinite regress arises once it is admitted that subjective probabilities are imprecise: if imprecision in probability judgements is measured by lower and upper bounds, why aren't the bounds themselves somewhat imprecise? This leads to further questions about how the higher-order imprecision might be quantified and whether there would be practical or conceptual benefits in doing so. Higher-order imprecision cannot simply be represented by hierarchies of ordinary probabilities, since they can always be integrated to obtain precise first-order bounds for decision-making purposes. Models of second-order imprecision using hierarchies of lower and upper probabilities or fusions of probability and fuzzy-sets theory have been proposed by various authors (e.g., Good, 1962; Walley, 1997; de Cooman, 2001), but the behavioral implications of such higher-order imprecision are often subtle (e.g., they may refer to the behavior of another person—a modeler—as well as the original subject). A generalization of lower and upper probabilities that yields a simple behavioral interpretation of second-order imprecision was given by Nau (1989, 1992). There, so-called *confidence weights* are attached to lower and upper probabilities, and an individual is permitted to assess more than one lower or upper probability for a given event at different levels of confidence. This representation of subjective uncertainty follows in a natural way from a further relaxation of the standard axioms of coherent preferences for monetary acts, in which transitivity is weakened at the same time as completeness is abandoned. It can also be derived as a generalization of de Finetti's operational method of measuring subjective probabilities in terms of acceptable bets.

de Finetti (1937, 1974) proposed that an individual's subjective conditional probability ("prevision") for an event  $E$  given another event  $F$  should be defined as the (presumably unique) number  $p$  for which the individual is indifferent to betting on or against  $E$  at rate  $p$ , conditional on  $F$ . That is,  $p$  is the number for which he is willing to accept a bet with payoff  $\alpha(E - p)F$ , where  $\alpha$  is a small but otherwise arbitrary number, *positive or negative*, chosen at the discretion of an opponent. (Here  $E$  and  $F$  are used as indicator variables as well as names of events: the variable  $E$  takes the values 1 or 0 when the event  $E$  occurs or doesn't occur, respectively. More generally,  $E$  can be a random variable that takes on arbitrary real values, in which case its prevision is interpretable as a subjective conditional expectation. In de Finetti's view, expectation is the more fundamental concept, with probability as the special case.) This operational definition of subjective probability is readily generalized in terms of lower and upper conditional probabilities: for given events  $E$  and  $F$ , let the individual give lower and upper betting rates  $p$  and  $q$ , respectively, for  $E|F$ , meaning that he is willing to accept a bet  $\alpha(E - p)F + \beta(q - E)F$ , where  $\alpha$  and  $\beta$  are small *non-negative* numbers chosen at the discretion of an opponent. Lower and upper conditional expectations are defined similarly.

Now consider a further extension of de Finetti's elicitation method in which the individual is assumed to have a *finite total betting stake*, so that the maximum amount he is permitted to lose is bounded at some moderate value where non-linearity of

utility for money is not an issue. The finite betting stake is not to be interpreted as the individual's entire wealth, but rather as an exogenous constraint—a rule of the elicitation game—that has strategic implications. Without loss of generality, let the size of the betting stake be normalized to 1. Then the individual is able to qualify each of his lower or upper betting rates by specifying the maximum fraction of his total stake that he is willing to risk in betting at that rate, and that fraction is defined to be the *confidence weight* associated with the betting rate. For example, he might attach a confidence weight of  $c$  to a lower probability  $p$  for event  $E$ , meaning that he will accept a bet with payoff  $\alpha(c/p)(E - p)F$ , subject to the constraint that  $0 \leq \alpha \leq 1$ . (Because of the scaling factor  $c/p$ , the maximum loss is equal to  $\alpha c$ , which occurs when  $E = 0$  and  $F = 1$ . The constraint  $\alpha \leq 1$  then means the maximum loss must be less than or equal to  $c$ .) Similarly, if he attaches a confidence weight of  $c$  to an upper probability  $q$ , this means he will accept a bet with payoff  $\beta(c/(1 - q))(q - E)F$ . More generally, if he simultaneously assigns lower and upper probabilities  $p_j$  and  $q_j$  with confidence weight  $c_j$  to event  $E_j|F_j$ , for  $j = 1, \dots, J$ , this means he will accept a bet with total payoff:

$$\sum_j [\alpha_j(c_j/p_j)(E_j - p_j) + \beta_j(c_j/(1 - q_j))(q_j - E_j)]F_j,$$

where the non-negative multipliers  $\{\alpha_j\}$  and  $\{\beta_j\}$  are chosen by an opponent subject to the constraint:

$$\sum_j [\alpha_j + \beta_j] \leq 1.$$

(The events  $\{E_j\}$  and  $\{F_j\}$  in the preceding expression can have any logical relationships whatever, and need not be distinct.) The individual is free to assign more than one confidence-weighted lower or upper probability to the same event—indeed, for reasons to be explained below, he will usually wish to do so. Higher confidence is associated with looser probability bounds: the lesser [greater] of two lower [upper] probabilities will have the higher confidence.

It is intuitively reasonable that, the more favorable the betting rate to himself, the higher the stakes an individual should be willing to play for, but it may not be immediately apparent why he should want to adopt this particular formula for constrained betting at the discretion of an opponent. The rationale for this method of attaching a confidence weight to a lower or upper probability is that it directly addresses a shortcoming that de Finetti acknowledged in his original definition of probability, namely “the possibility that people accepting bets against our individual have better information than he has... [which] would bring us to game-theoretic situations” [footnote to 1964 translation by Kyburg]. The constrained betting system described above has the useful property that *it allows the individual to perform a reciprocal measurement on the subjective probability of his opponent and to adjust his own betting rate accordingly in advance*. To see this, note that the individual's assessment of confidence weighted probabilities presents a well-defined statistical decision problem to a Bayesian opponent (“she”). For simplicity, consider the case of an unconditional probability assessment

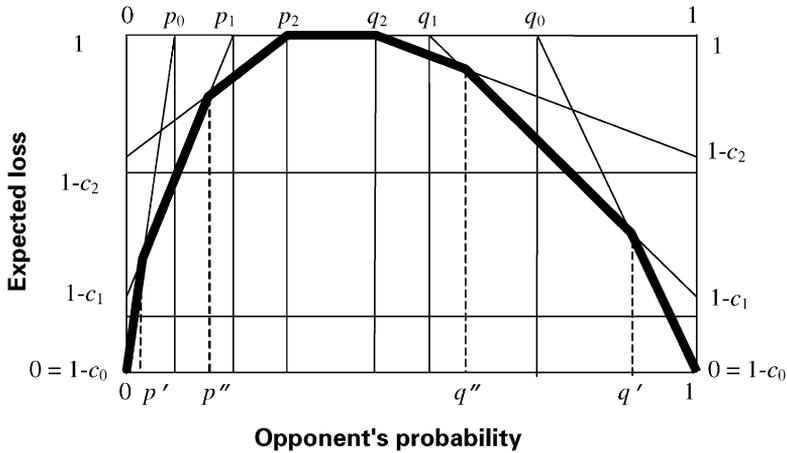


Fig. 1. Construction of the Bayes risk function representing an assessment of confidence-weighted probabilities.

for a single event  $E$ , and suppose that three lower probabilities ( $p_0, p_1, p_2$ ) and three upper probabilities ( $q_0, q_1, q_2$ ) are assessed with distinct confidence weights ( $c_0, c_1, c_2$ ), where  $1 = c_0 > c_1 > c_2$ . Correspondingly,  $p_0 < p_1 < p_2$  and  $q_0 > q_1 > q_2$ . In other words, the individual is “100% confident” that the probability lies in the interval  $[p_0, q_0]$ . He is somewhat less confident that it lies in the narrower interval  $[p_1, q_1]$ , and still less confident that it lies in the still-narrower interval  $[p_2, q_2]$ . The statistical decision problem that this probability assessment presents to an opponent is described by a *Bayes risk function* (DeGroot, 1970) which specifies the opponent’s minimum expected loss as a function of her own hypothetical probability distribution. The construction of the Bayes risk function is shown in Fig. 1.

Here, the horizontal interval  $[0, 1]$  represents the set of possible values for the opponent’s probability of  $E$ . The vertical interval  $[0, 1]$  measures the opponent’s expected loss, where “loss” is defined as maximum possible gain minus actual gain. Because the opponent’s maximum possible gain is 1 (the size of the probability assessor’s total stake), her expected loss is 1 minus her expected gain from the bets she places. Hence her loss is equal to 1 if she places no bets at all, and her loss cannot be less than zero.

Each of the sloping construction lines in the figure is an upper bound on the opponent’s Bayes risk function determined by a single confidence-weighted lower or upper probability. In particular, the assessment of a lower probability  $p$  with confidence  $c$  determines a construction line through the points  $(0, 1 - c)$  and  $(p, 1)$ . If the opponent takes this bet alone (i.e., assigns it a multiplier  $\alpha = 1$ ), then her expected loss is the linear function of her probability whose graph is this line. (If her probability is 0, then she is certain  $E$  will not occur, and she expects to win the amount  $c$ , in which case her loss is equal to  $1 - c$ . On the other hand, if her probability for  $E$  is exactly  $p$ , she thinks the bet is fair, and her expected gain from it is zero, in which case her loss is equal to 1. Expected losses for other probability values follow by linear interpolation.)

Similarly, the assessment of a single upper probability  $q$  with confidence  $c$  determines a construction line through the points  $(1, 1 - c)$  and  $(q, 1)$ . If the opponent takes *none* of the bets, her actual and expected loss is equal to 1 regardless of her probability for  $E$ , which corresponds to the horizontal line at  $y = 1$  in the figure. The opponent's optimal decision is to take the *single* bet, if any, that yields the *minimum expected loss*, and consequently her Bayes risk function is the envelope (pointwise minimum) of the aforementioned construction lines, which is the heavy line shown in the figure.

Note that the Bayes risk function partitions the probability interval into subintervals in which the same bet is always taken by the opponent. For example, if the opponent's probability lies in the subinterval  $[p', p'']$ , she will take the bet determined by lower probability  $p_1$  with confidence  $c_1$ , while if her probability lies in the subinterval  $[p'', p_2]$ , she will take the bet determined by lower probability  $p_2$  with confidence  $c_2$ . (Here  $p'$  is the probability value at which the construction lines corresponding to  $p_0$  and  $p_1$  intersect, and similarly for  $p'', q'$ , and  $q''$ .) If her probability is in the subinterval  $[p_2, q_2]$ , she will not bet at all. If her probability is in the subinterval  $[q_2, q'']$ , she will take the bet determined by upper probability  $q_2$  with confidence  $c_2$ , and so on. Notice that this optimal strategy for the opponent enables the probability assessor to effectively adjust his betting rate in response to what he may learn about the opponent's probability. If the opponent's probability is between  $q_2$  and  $p_2$  (the assessor's greatest lower and least upper probability with non-zero confidence), he will not bet at all with the opponent: he cannot discern a profitable difference of opinion. If the opponent's probability is less than  $p_2$  but greater than  $p''$ , the assessor is willing to bet on  $E$  at the relatively high rate  $p_2$ . But if the opponent's probability is revealed to be less than  $p''$ , though greater than  $p'$ , the assessor revises his betting rate downward from  $p_2$  to  $p_1$ , and he revises it even further downward to  $p_0$  if the opponent's probability is revealed to be less than  $p'$ . Similarly, on the upper end, the assessor is willing to bet against  $E$  at the relatively low rate  $q_2$  if the opponent's probability is between  $q_2$  and  $q''$ , but he revises the rate upward to  $q_1$  if the opponent's probability is revealed to be greater than  $q''$ , though less than  $q'$ , and so on.

In the more general case of multiple, conditional events that are subsets of a finite state space, the Bayes risk function is defined on the probability simplex in  $\mathfrak{R}^m$ , where  $m$  is the number of states. The Bayes risk function on the simplex is typically a piecewise linear concave function—i.e., its graph is a convex polytope. Details are given in Nau (1989, 1992). The Bayes risk function on the simplex can then be marginalized to obtain concave, piecewise linear Bayes risk functions for conditional or unconditional events, leading to natural generalizations of the usual probability laws, including Bayes' theorem. (Posterior confidence-weighted probabilities can be deduced from priors and likelihoods by forming appropriate linear combinations of bets, exactly as de Finetti did for precise probabilities.) Bayes risk functions behave very much like fuzzy-set membership functions—in particular, they obey the max–min rules of union and intersection—although they are not derived from the assumptions of fuzzy-set theory. Rather, these results establish *as a theorem* that fuzzy-set theory is applicable to the modeling of imprecise subjective probabilities, as previously suggested by Freiling

(1980), Watson et al. (1979), and Dubois and Prade (1989). The Bayes risk function also has the same qualitative properties as the *epistemic reliability function* proposed by Gärdenfors and Sahlin (1982, 1983), insofar as it serves to index nested convex sets of probability measures.

The use of confidence-weighted probabilities in decision analysis is discussed by Nau (1989). Briefly, the second-order imprecision in probabilities leads naturally to second-order imprecision in expected values of decisions, and the Bayes risk function provides a metric in terms of which alternative decisions can be ranked according to their “distance” from the set of potentially optimal (expected-value-maximizing) decisions. This decision-ranking criterion is similar in principle to methods of sensitivity analysis developed by Rios Insua (1990).

To determine the implications of the confidence-weighted probability model for the aggregation of opinions of different experts, consider a roomful of experts, each of whom has a finite total betting stake. In principle, the experts’ stakes could be either exogenously or endogenously determined. For example, the organizer of the elicitation game could assign the experts equal or unequal stakes, according to an outside assessment of their relative expertise or credibility. Or, the experts could choose their own stakes through some process of negotiation in which they self-assess their all-around confidence relative to each other. Let  $w_i$  denote the betting stake of expert  $i$ , and without loss of generality assume that  $\sum_i w_i = 1$ . Let each expert assess his own subjective uncertainty for various (possibly conditional) events in terms of confidence-weighted probabilities, where the confidence weights are the fractions of his own betting stake that he is willing to risk in betting at the corresponding lower or upper probabilities. (The experts need not all contemplate the same events, although it is assumed that the events considered by different experts are somehow logically related—i.e., they are subsets of a common state space.) Now consider the statistical decision problem that is perceived by an opponent who bets against all of the experts simultaneously. Because the opponent’s gain is naturally the sum of her gains in the bets against the different experts, it immediately follows that her Bayes risk is a weighted average of the Bayes risks of the different experts, with weights equal to their respective betting stakes  $w_i$ . In other words, *the Bayes risk function of the combined experts is a linear pool of their individual Bayes risk functions, and the experts are weighted in proportion to their total betting stakes*. Because the concavity and piecewise linearity and 0–1 range of the Bayes risk function are preserved under weighted averages, it follows that the Bayes risk function of the combined experts has exactly the same qualitative properties as the Bayes risk function of a single expert. Hence, if all she observes are the combined bets that are available, as summarized by the aggregate Bayes risk function, the opponent cannot really tell whether there is a single expert or multiple experts on the other side of the door. The linear pool of Bayes risk functions also satisfies the belief-aggregation desiderata that are incompatible when beliefs are represented by ordinary subjective probabilities. It obviously preserves unanimity, and it also satisfies generalizations of the marginalization and external Bayesianity properties. The result of pooling-before-marginalizing is

always consistent with the result of pooling-after-marginalizing but slightly more informative in the sense that the Bayes risk functions are nested. Similarly, the result of pooling-after-Bayesian-updating is always consistent with, but slightly more informative than, the result of pooling-before-updating. (See Nau, 1990, for further details.)

When confidence-weighted probabilities are aggregated as described above, it is quite possible that the aggregate assessment will be *incoherent*. This will happen, for example, if one expert assesses a lower probability (at some positive level of confidence) which exceeds an upper probability for the same event (at some positive level of confidence) assessed by another expert. However, incoherence is not catastrophic in the framework of confidence-weighted probabilities: the aggregate set of probability measures is not just empty, as it would be in the case of incoherent precise probabilities or incoherent interval probabilities. Rather, the [fuzzy] set of aggregate probability measures merely has a *subnormal* Bayes risk [membership] function—i.e., a Bayes risk function whose maximum value is less than 1. The common betting opponent will perceive an arbitrage opportunity, but the magnitude of the arbitrage profit—which measures the *relative incoherence* of the assessment—will be bounded at some fraction of the maximum betting stake. In particular, the arbitrage profit is precisely the amount by which the maximum value of the aggregate Bayes risk is less than 1 (Nau, 1989). This measure of relative incoherence is a generalization of one of the two measures of incoherence independently proposed by Schervish et al. (1998): their model is a special case of confidence-weighted probabilities in which a point probability (i.e., a lower and upper probability that coincide) is assessed for every event with a confidence weight of 1.

The aggregation of confidence-weighted probabilities of two equally-weighted experts for the same unconditional event, in terms of their Bayes risk functions, is illustrated by Figs. 2 and 3. Note that the combined Bayes risk function (the heavy line in each figure) is simply an average of the Bayes risk functions of the individual experts, and as such it typically has a smoother appearance. Fig. 2 illustrates the case in which the experts are mutually coherent: the combined Bayes risk achieves a normal maximum height of 1. Fig. 3 illustrates the case in which the experts are mutually incoherent. Here, one expert has asserted that 0.4 is an upper probability while the other has asserted that 0.5 is a lower probability, both with positive confidence. The Bayes risk function is, therefore, subnormal: its maximum height is 0.95, indicating a relative incoherence of 5%.

In the special case of  $n$  equally-weighted experts who assess mutually coherent (overlapping) probability intervals with full confidence—i.e., lower and upper probabilities with confidence 1—it is simple to characterize the aggregate assessment. Let their respective intervals be  $[p_i, q_i]$ , for  $i = 1, \dots, n$ . Let  $p^* = \max\{p_i\}$  and  $q^* = \min\{q_i\}$ , and assume  $p^* < q^*$ . Let  $p^{**}$  denote the harmonic mean of the quantities  $\{p_i\}$  and let  $q^{**}$  denote one minus the harmonic mean of the quantities  $\{1 - q_i\}$ . Then, in the aggregate assessment,  $p^*$  and  $q^*$  are lower and upper probabilities with confidence  $1/n$ —the greatest lower and least upper probabilities with non-zero confidence—while  $p^{**}$  and  $q^{**}$  are lower and upper probabilities with confidence 1. (With more than two experts,

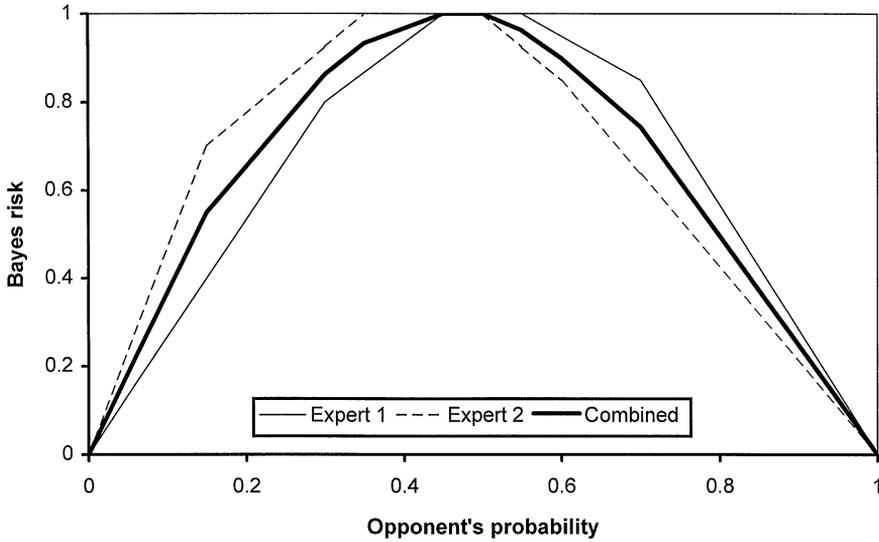


Fig. 2. Aggregation of coherent experts.

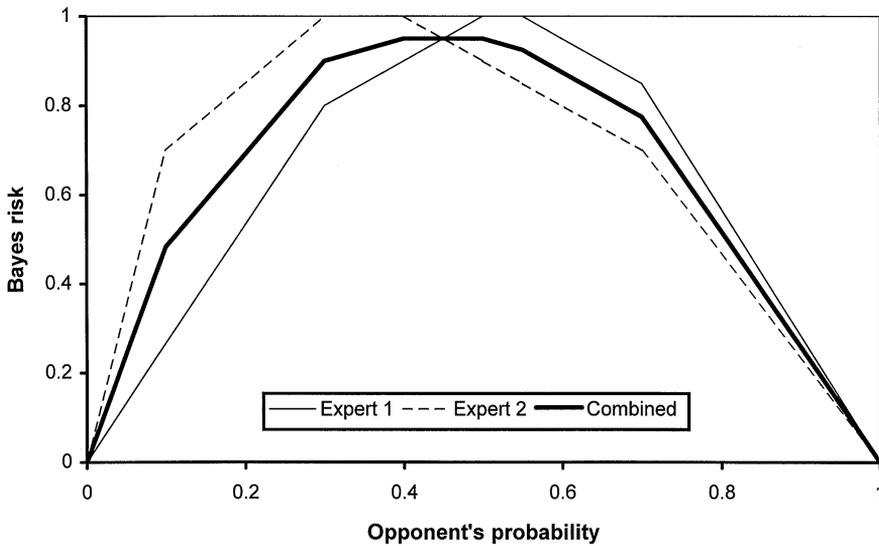


Fig. 3. Aggregation of incoherent experts.

there are also intermediate lower and upper probabilities with intermediate confidence levels.)

The aggregation method illustrated in the preceding examples can be implemented via linear programming. Suppose that  $n$  experts assess confidence-weighted probabilities for the same unconditional event. In particular, suppose that expert  $i$  assesses lower

and upper probabilities  $p_{ij}$  and  $q_{ij}$  with confidence  $c_{ij}$ , for  $j = 1, \dots, J_i$ . Consider the following system of constraints (1), where  $i$  ranges from 1 to  $n$  and  $j$  ranges from 1 to  $J_i$  as appropriate:

$$\begin{aligned} r &= \sum_{i=1}^n w_i z_i, \\ z_i &\leq 1, \\ z_i &\leq 1 - c_{ij} + (c_{ij}/p_{ij})x, \\ z_i &\leq 1 + c_{ij}q_{ij}/(1 - q_{ij}) - (c_{ij}/(1 - q_{ij}))x, \\ x &\geq 0, \quad x \leq 1, \quad z_i \geq 0. \end{aligned} \tag{1}$$

Next consider the following three linear programs incorporating these constraints:

*LP1*: maximize  $r$  over all  $\{r, z_i, x\}$  subject to (1).

*LP2*: for a fixed  $x$  in  $[0, 1]$ , maximize  $r$  over all  $\{r, z_i\}$  subject to (1).

*LP3*: for a fixed  $\varepsilon \geq 0$ , minimize [maximize]  $x$  over all  $\{r, z_i, x\}$  subject to (1) and  $r \geq 1 - \varepsilon$ .

Let  $r^{**}$  denote the optimal objective value for LP1, let  $r^*(x)$  denote the optimal objective value for LP2, and let  $x_*(\varepsilon)$  and  $x^*(\varepsilon)$  denote the optimal min and max objective values for LP3. Then  $r^{**}$  is the maximum value of the Bayes risk function on the interval  $[0, 1]$ ,  $r^*(x)$  is the value of the Bayes risk function at  $x$ , and  $x_*(\varepsilon)$  and  $x^*(\varepsilon)$  are the endpoints of the level set obtained by cutting the graph of the Bayes risk function at a height of  $1 - \varepsilon$ . The quantity  $1 - r^{**}$  is the relative degree of incoherence of the experts, and  $x_*(\varepsilon)$  and  $x^*(\varepsilon)$  are the greatest lower and least upper probabilities at a distance  $\varepsilon$  from the fuzzy set of aggregate probabilities. (A more efficient and more general LP algorithm applicable to the case of multiple, conditional events is given in Nau, 1990, 1992.)

There are several characteristic features of this approach to aggregating imprecise probabilities. First, it uses a measure of the *second-order imprecision* in the experts' probability judgements as a basis for making tradeoffs among the lower and upper bounds offered by different experts. The confidence weights provide the necessary measure of second-order imprecision. Second, the assumption of a *limited total betting stake*, and the interpretation of confidence weights as fractions of the total stake allocated to different bets, turns out to be the key to modeling the second-order imprecision and adapting de Finetti's probability elicitation method to the case of multiple experts. Under the usual extension of de Finetti's elicitation method to the case of lower and upper probabilities, there is no explicit limit on the size of any of the bets—they are just assumed to be small. If an opponent bets with multiple experts under such conditions, she will only bet against the expert offering the most favorable rate on any given event: she will have no incentive to bet with anyone offering a less attractive rate. Hence the opponent will only see the greatest lower and least upper probability bounds among all the experts, and the aggregate set of probability measures will naturally be

the intersection of their individual sets of probability measures. Here, by comparison, the fact that each expert's stake is limited means that the opponent will wish to bet some amount against every expert whose least confident probability bounds are disjoint from her own probabilities. The possibility that different experts might be endowed with different total betting stakes also provides a basis for differential weighting.

Third, the second-order imprecision has an explicit *intersubjective* dimension: it encodes the way in which the probability assessor wishes to respond to information that is revealed about the probabilities held by his betting opponent. It is precisely this intersubjective quality that sets the stage for a rational aggregation method, because the probability assessments of different experts are all referenced to a common betting opponent.

### 3. Aggregation of risk-neutral probabilities

In the preceding method of aggregating subjective probabilities, every probability assessor was able to adjust his betting rate in response to the actions of a betting opponent, but the details of the adjustment process had to be specified in advance, encoded in the assignment of appropriate confidence weights to different lower and upper probabilities. The aggregation of such probabilities was imagined to be carried out by assembling a group of experts in the same room for purposes of betting against a common opponent, but the experts did not interact directly with each other. The possibility was admitted that the combined experts could be incoherent—i.e., that a lower probability of one expert might be greater than the upper probability of another expert for the same event, at some positive level of confidence—but the effects of incoherence were not disastrous, since betting stakes were a priori limited. Indeed, a useful measure of the relative incoherence of the experts was obtained.

An alternative, though conceptually related, approach, is to let the different experts in the same room bet *with each other* prior to interacting with a common betting opponent. Specifically, assume that upon entering the room, the experts announce to each other their lower and upper betting rates on events—i.e., their bid–ask spreads for monetary lottery tickets pegged to those events—and they then proceed to make sequences of small bets with each other at the quoted rates. Presumably their betting rates will change somewhat over time due to a combination of reciprocal learning, mutual hedging of risks, and exploitation of arbitrage opportunities. Under these conditions, an expert's bid–ask spread for a given lottery ticket at any given time may reflect not only the intrinsic imprecision in his beliefs but also his desires (if any) to conceal or misrepresent his beliefs, to hedge himself against the possibility that his opponents have superior information, or to profit from the possibility that his opponents have inferior information. Note that if the incremental bets are small relative to the experts' total betting stakes, and if the experts have the opportunity to adjust their rates continuously over time, it is unnecessary for them to use the device of confidence weights described in the preceding section: the only betting rates that matter at any given moment are their greatest lower and least upper betting rates with positive confidence.

In this setting (or indeed, any setting in which material rewards are used to elicit beliefs), the agents' lower and upper betting rates on events are not necessarily measures of pure belief. Rather, they are measures of belief *compounded with possibly-state-dependent marginal utilities for money*. For, on the assumption that an individual is an expected-utility-maximizer, his betting rate on an event ought to be proportional to the product of his true probability and his relative marginal utility for money given the occurrence of that event, notwithstanding any strategic hedging he may wish to undertake. This product of true probability and relative marginal utility for money is known as a *risk neutral probability* in the literature of finance: it is the probability we would infer from the individual's betting behavior on the (possibly incorrect) assumption that he is risk neutral. If the individuals in the room all have decreasing marginal utility for money in every state, then as they accumulate bets with each other, their state-dependent marginal utilities will shift so that their risk neutral probabilities will converge, even if their true probabilities remain relatively constant (Kadane and Winkler, 1988). Their true probabilities might also converge to some extent as the experts learn from their interactions with each other, but the precise extent of that convergence will remain unknown because their true probabilities and their marginal utilities will be inseparable in each other's eyes. Indeed, the unobservability of their true probabilities explains how different individuals can "agree to disagree" even when their betting rates are common knowledge (Nau, 1995b).

The inseparability of subjective probabilities and utilities is a fundamental problem that afflicts *any* method of eliciting probabilities that uses material rewards to motivate or discipline the experts. It cannot be circumvented by substituting lottery-ticket currency for ordinary money, for if the lottery tickets offer chances at some material prize, the utility of the prize need not be state-independent, and if it is not, the probabilities inferred from gambles expressed in lottery-ticket currency will be distorted by the state-dependent utilities. Of course, this problem arises only when the events of interest are important to the experts who are assessing the probabilities, in the sense that the experts have significant personal stakes in them. But many events have this character—e.g., events that refer to the general state of the economy, the outcomes of divisive political contests, global calamities, or the vindication of scientific theories. (See Aumann, 1971. Further discussion of the inseparability of probability and utility is given by Shafer, 1986; Schervish et al., 1990; Karni and Safra, 1995; Karni and Mongin, 2000.) Fortunately, the distortion of subjective probabilities by state-dependent utilities is not catastrophic for purposes of decision analysis. It is possible in principle to perform decision analysis using risk neutral probabilities (Nau, 1995a), and risk neutral probabilities also provide the basis for a useful integration of decision analysis and options pricing methods of project valuation (Smith and Nau, 1995). Wakker and Zank (1999) show that Bayesian analysis remains possible even if the axiom that guarantees the existence of true prior probabilities (Savage's postulate P4) is dropped.

To return to the proposed aggregation method, as the experts in the room continue to bet with each other, updating their beliefs and marginal utilities and strategic

positions, an equilibrium will eventually be reached in which their betting rates converge on stable values, and those stable values will necessarily be coherent if all arbitrage opportunities have been rationally exploited. A consensus of the experts is thereby achieved, but the consensus is with respect to risk neutral probabilities rather than true subjective probabilities of events. In the final equilibrium, every expert has his own lower and upper risk neutral probabilities for events (or more generally, bid–ask spreads for arbitrary lotteries pegged to those events), defining a *convex set of risk-neutral probability measures*. From the perspective of an observer who enters the room at this point, only the greatest lower and least upper betting rates (or bid–ask prices) are of interest. Hence, there appears to be a single “representative agent” whose convex set of risk-neutral probability measures is the *intersection* of the final sets of risk-neutral probability measures of the separate experts. Thus, we find support for the representation of the experts’ aggregate opinion by the intersection of their respective convex sets of probability measures, but the probability measures are risk-neutral rather than true measures, and the intersection is taken only *after* the experts have had the opportunity to bet with each other, strategize, learn, and exploit any arbitrage opportunities that may be discovered. The problem of incoherence (i.e., an empty intersection) therefore does not arise at all, and the problem that only the most extreme beliefs are represented is mitigated by the fact that the experts have had the opportunity to learn from each other and revise their beliefs.

As a simple example, suppose there are two experts and a single event  $E$  whose probability is of interest. Suppose the experts have initial lower and upper risk neutral probabilities  $[p_1, q_1]$  and  $[p_2, q_2]$  respectively. If the intervals overlap—say, if  $q_1 > p_2$ —then the experts will not bet at all with each other, and trivially their aggregated risk neutral probabilities will be the intersection of the intervals, namely  $[p_2, q_1]$ . In the more interesting case where their intervals are initially disjoint—say,  $q_1 < p_2$ —they will bet with each other and adjust their risk neutral probabilities until some overlap is achieved. Suppose, for example, that their initial intervals are  $[0.2, 0.3]$  and  $[0.4, 0.5]$ , respectively. Then for any betting rate between 0.3 and 0.4, expert 1 is initially willing to bet against  $E$  while expert 2 is willing to bet on  $E$ . The actual rate at which betting occurs may depend on which agent has more patience, stubbornness, or savvy than the other. A canonical way to model the outcome of the betting is to imagine that an arbitrageur bets with each expert at the least favorable rate the expert is willing to accept, so that the expert’s total expected utility remains unchanged. Thus, initially, the arbitrageur will bet with expert 1 at the rate 0.3 (with the expert betting *against*  $E$ ) and with expert 2 at the rate 0.4 (with the expert betting *on*  $E$ ). As the stakes accumulate, the betting rate of expert 1 will rise while that of expert 2 falls, until they converge on a common value that represents their aggregate beliefs. The arbitrageur will reap a positive profit during this process, and the total arbitrage profit can later be redistributed (or not) between the experts in an arbitrary manner, which may (or may not) trigger further rounds of betting. If the arbitrage profits are *not* redistributed, the solution is generally unique (under suitable regularity conditions on the utility functions) and can be found by solving a nonlinear programming problem

in which the arbitrageur's minimum profit across states is maximized while holding the experts' expected utilities constant.

To complete the analysis in this example, suppose that the experts have exponential utility functions  $u(x) = -\exp(-x/t)$  where  $x$  denotes wealth and  $t$  is the expert's *risk tolerance* parameter. (The exponential utility function is widely used in applied decision analysis and financial modeling. Roughly speaking, an exponential utility maximizer with risk tolerance  $t$  is just willing to accept a bet that offers equal chances of winning  $t$  or losing  $t/2$ .) Let the experts have risk tolerances of \$10,000 and \$20,000, respectively. A convenient property of exponential utility functions is that they exhibit constant absolute risk aversion, which implies that lump-sum changes in wealth have no effects on risk neutral probabilities. Hence, if the experts have exponential utility functions, they will converge on a unique risk neutral distribution regardless of whether or how they share the arbitrage profits. The exponential utility function is unique in leading to the construction of an unambiguous aggregate probability distribution in this manner, as shown by Wilson (1968) and Rubinstein (1974). In the present example, the experts will converge uniquely on an aggregate risk neutral probability of 0.3652, which is closer to the initial risk neutral probability of expert 2 than expert 1 because expert 2 has a higher risk tolerance. (This value is obtained by solving a nonlinear program to find a mutually acceptable bet that is Pareto optimal.) Under more general utility functions, such as log or power functions, the solution would depend on how the profits were shared. In this example it has been assumed that no learning or strategic maneuvering takes place, so that changes in the experts' risk neutral probabilities are due *only* to changes in relative marginal utilities as betting stakes accumulate. If their true probabilities also changed because they learned from each other while betting, the convergence process would be even harder to analyze.

The convergence of risk neutral probabilities among agents who are free to bet with each other, leading to the construction of a representative agent, is discussed by Kadane and Winkler (1988), Nau and McCardle (1991), and Nau (1995b), and it is also well known to be a necessary condition for competitive equilibrium in a contingent claims market (Drèze, 1970). (In the economics literature, risk-neutral probabilities are also referred to as normalized *state prices*.) This market-based approach to aggregating risk-neutral probabilities is widely employed in practice, most notably in markets for derivative securities (e.g., stock options and futures) and for betting on sporting events (e.g., parimutuel oddsmaking) and political races (e.g., the University of Iowa political markets). The efficiency and calibration of such markets are well known.

We see once again that, in order to aggregate subjective probabilities in a rational and non-arbitrary fashion, it is necessary to begin with a definition of probability that is intrinsically intersubjective and allows for varying degrees of precision in the assessed values. In this case, the intersubjective element is modeled by allowing the experts to bet with each other during the elicitation process, so that the revealed bounds on their (risk-neutral) probabilities are not merely measures of private belief but rather are measures of public willingness to bet against each other. The final risk-neutral probabilities to which the experts will converge under these conditions are not predictable

a priori from their hypothetical true prior probabilities and utilities, even in principle, except under very restrictive conditions on utility functions, learning algorithms, strategic behavior, and trading mechanisms. There may be many possible final allocations of state-contingent wealth that are mutually preferred to the initial allocation, and correspondingly many possible final distributions of risk-neutral probabilities. The one to which the experts converge may depend on the vagaries of the sequence of trades, on their relative bargaining powers with respect to each other, on the way they learn from their interactions with each other, and on other psychological factors or environmental contingencies. The final equilibrium, therefore, is not uniquely determined by initial conditions that are subject to independent measurement. Rather, the observation of the final equilibrium is a *fundamental measurement* of aggregate belief.

#### 4. Conclusions

The aggregation of probability assessments, if it is to be more than an exercise in descriptive statistics applied to psychometric data, requires an interpersonal concept of subjective probability that allows the beliefs of different individuals to react naturally with each other in the presence of a physical measuring instrument. Two such concepts of probability have been surveyed here, leading to two aggregation methods which are conceptually related but quite different in their strengths and weaknesses and potential areas of application. The first method, involving confidence-weighted probabilities, is best suited for situations in which the events in question are discrete, few in number, and ambiguous in nature, where interest centers on the degree of consensus or dissensus and on the reconciliation of incoherence among a small group of experts. The strength of the method is that it allows each expert to articulate his second-order uncertainty in a detailed and operationally meaningful way that has a nice graphical representation; the weakness is that this potentially requires the elicitation of more than just one or two parameters per event. On a purely theoretical level, it provides some support for the concept of a linear opinion pool (although the appropriate objects to pool are Bayes risk functions, not point probabilities) and also for the use of harmonic means to pool lower and upper probabilities.

The second method, involving risk-neutral probabilities, is best suited for situations in which many individuals jointly assess probability distributions for large numbers of events or continuous random variables. The latter method is already deeply rooted in our economic institutions: it is just a familiar market for betting or trading lottery tickets (contingent claims) on events. Its strong point is that it is psychologically transparent, easy to implement, and well known to be an efficient method of aggregating opinions. Its weak point—if this can be considered one at all—is that it does not register pure belief: it measures, for each individual, the product of his or her relative probabilities for events and relative marginal utilities for money should those events occur. In cases where the individuals do not have significant prior stakes in the events (or otherwise state-dependent marginal utilities), the distortions of probabilities are not likely to be significant, but for personally or societally important events,

the distortions could conceivably be severe. However, this is not necessarily problematic for statistical theory—nor for decision or game theory. In a Bayesian statistical model, the contaminating effects of state-dependent marginal utilities are otherwise indistinguishable from the effects of unexplained prior probabilities. And in a decision or game-theoretic model, it is possible to perform money-backed utility measurements that interlock naturally with the risk-neutral probabilities to permit the determination of optimal strategies (Nau, 1995a). In any case, risk neutral probabilities are more easily understood and more readily translated into action by denizens of market economies than pure subjective probabilities would be, even if the latter could reliably be measured. de Finetti (1937) suggested that the definition of subjective probability in terms of money-backed gambles might have practical benefits that would outweigh the seeming impurity of the measurements. We find considerable prescience in that view.

### Acknowledgements

This research was supported by the National Science Foundation under Grant 98-09225, the Fuqua School of Business, and INSEAD.

### References

- Aumann, R., 1971. Letter to L.J. Savage. Reprinted in: Drèze, J., 1987. *Essays on Economic Decision under Uncertainty*. Cambridge University Press, London.
- de Cooman, G., 2001. Precision–imprecision equivalence in a broad class of imprecise hierarchical uncertainty models. *J. Statist. Plann. Inference* 105, 175–198.
- de Finetti, B., 1937. La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincaré* 7, 1–68. (Translation in: Kyburg, H.E., Smokler, H.E. (Eds.), 1980. *Studies in Subjective Probability*, 2nd edition. Robert Krieger, New York.)
- de Finetti, B., 1974. *Theory of Probability*, Vol. 1. Wiley, New York.
- DeGroot, M., 1970. *Optimal Statistical Decisions*. McGraw-Hill, New York.
- Drèze, J., 1970. Market allocation under uncertainty. *European Econom. Rev.* 2, 133–165.
- Dubois, D., Prade, H., 1989. Fuzzy sets, probability, and measurement. *European J. Oper. Res.* 40, 135–154.
- Freeling, A.N.S., 1980. Fuzzy sets and decision analysis. *IEEE Trans. Systems Man Cybernet.* SMC-10, 341–354.
- Gärdenfors, P., Sahlin, N., 1982. Unreliable probabilities, risk-taking, and decision making. *Synthese* 53, 361–386.
- Gärdenfors, P., Sahlin, N., 1983. Decision making with unreliable probabilities. *British J. Math. Statist. Psych.* 36, 240–251.
- Genest, C., Zidek, J.V., 1986. Combining probability distributions: a critique and annotated bibliography. *Statist. Sci.* 1, 114–148.
- Good, I.J., 1962. Subjective probability as the measure of a non-measurable set. In: Nagel, E., Suppes, P., Tarski, A. (Eds.), *Logic, Methodology, and Philosophy of Science*. Stanford University Press, Stanford, CA. (Reprinted in: Kyburg, H.E., Smokler, H.E. (Eds.), 1980. *Studies in Subjective Probability*, 2nd edition. Robert Krieger, New York.)
- Kadane, J.B., Winkler, R.L., 1988. Separating probability elicitation from utilities. *J. Amer. Statist. Assoc.* 83, 357–363.
- Karni, E., Safra, Z., 1995. The impossibility of experimental elicitation of subjective probabilities. *Theory and Decision* 38, 313–320.

- Karni, E., Mongin, P., 2000. On the determination of subjective probabilities by choices. *Management Sci.*, 46, 233–248.
- Moral, S., del Sagrado, J., 1998. Aggregation of imprecise probabilities. In: Bouchon-Meunier, B. (Ed.), *Aggregation and Fusion of Imperfect Information*. Physica-Verlag, New York, pp. 162–188.
- Nau, R.F., 1989. Decision analysis with indeterminate or incoherent probabilities. *Ann. Oper. Res.* 19, 375–403.
- Nau, R.F., 1990. Indeterminate probabilities and utilities on finite sets. Working paper, Fuqua School of Business, Duke University.
- Nau, R.F., 1992. Indeterminate probabilities on finite sets. *Ann. Statist.* 20, 1737–1767.
- Nau, R.F., 1995a. Coherent decision analysis with inseparable probabilities and utilities. *J. Risk Uncertainty* 10, 71–91.
- Nau, R.F., 1995b. The incoherence of agreeing to disagree. *Theory and Decision* 39, 219–239.
- Nau, R.F., McCardle, K.F., 1991. Arbitrage, Rationality, and Equilibrium. *Theory and Decision* 31, 199–240.
- Rios Insua, D., 1990. *Sensitivity Analysis in Multi-objective Decision Making*. Springer-Verlag, Berlin.
- Rubinstein, M., 1974. An aggregation theorem for securities markets. *J. Financial Econom.* 1, 225–244.
- Schervish, M.J., Seidenfeld, T., Kadane, J.B., 1990. State-dependent utilities. *J. Amer. Statist. Assoc.* 85, 840–847.
- Schervish, M.J., Seidenfeld, T., Kadane, J.B., 1998. two measures of incoherence: how not to gamble if you must. Technical Report #660, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213.
- Shafer, G., 1986. Savage revisited (with discussion). *Statist. Sci.* 1, 463–501.
- Smith, C.A.B., 1961. Consistency in statistical inference and decision. *J. Roy. Statist. Soc. B* 23, 1–25.
- Smith, J.E., Nau, R.F., 1995. Valuing risky projects: option pricing theory and decision analysis. *Management Sci.* 41, 795–816.
- Wakker, P., Zank, H., 1999. State dependent expected utility for Savage's state space. *Math. Oper. Res.* 24, 8–34.
- Walley, P., 1982. The elicitation and aggregation of beliefs. Technical Report, University of Warwick, Coventry, UK.
- Walley, P., 1991. *Statistical Reasoning with Imprecise Probabilities*. Chapman & Hall, London.
- Walley, P., 1997. Statistical inferences based on a second-order possibility distribution. *Internat. J. Gen. Systems* 26, 337–383.
- Watson, S.R., Weiss, J.J., Donnell, M.L., 1979. Fuzzy decision analysis. *IEEE Trans. Systems Man Cybernet.* SMC-9, 1–9.
- Wilson, R., 1968. The theory of syndicates. *Econometrica* 36, 119–132.
- Winkler, R.L., Lindley, D.V., Schervish, M.J., Clemen, R.T., French, S., Morris, P.A., 1986. Expert resolution. *Management Sci.* 32, 298–328.