Annals of Operations Research 19(1989)375 – 403

DECISION ANALYSIS WITH INDETERMINATE OR INCOHERENT PROBABILITIES*

Robert F. NAU

The Fuqua School of Business, Duke University, Durham, North Carolina 27706, USA

Abstract

This paper presents a new method of modeling indeterminate and incoherent probability judgments in decision analysis problems. The decision maker's degree of beliefs in the occurrence of an event is represented by a unimodal (in fact, concave) function on the unit interval, whose parameters are elicited in terms of lower and upper probabilities with attached confidence weights. This is shown to provide a unified framework for performing sensitivity analysis, reconciling incoherence, and combining expert judgments.

1. Introduction

In assessing subjective probabilities for the analysis of a decision problem under uncertainty, several kinds of difficulty may arise. The decision maker (DM) may be reluctant or unable to provide a sufficiently precise and detailed assessment to determine a unique distribution over the relevant states of nature. Or, at the other extreme, the distribution may be over-determined: the assessment may be internally inconsistent, a condition known as *incoherence*. Or, even if the assessment process is structured so as to enforce uniqueness and coherence, it may be felt that some of the probabilities thereby obtained are not entirely reliable, and hence some amount of sensitivity analysis will be desired. Standard Bayesian decision theory (as codified in the Savage axioms or other similar axiom systems) offers little guidance on how to deal with such problems since it does not formally acknowledge their existence. Pragmatic methods for dealing with imprecise, unreliable, or incoherent probabilities have been proposed in the decision-theoretic literature of the last few decades. Some of these have been mainly computational devices, such as the early work of Fishburn [10] on decision analysis with incomplete knowledge of probabilities, or Fishburn et al. [9] on sensitivity analysis. Others have been obtained by incorporating a theory of

*This material is based upon work supported by the National Science Foundation under grant SES-85-11675, and by the Business Associates Fund at the Fuqua School of Business.

measurement errors into the basic subjective probability model, such as the methods for reconciling incoherence by Lindley et al. [22]. More radical solutions have been based on relaxations of the standard axioms, such as the "quasi-Bayesian" interval-probability models by Smith [32], Good [16], and Giron and Rios [17]. (The belief-function models by Shafer [31] and Wolfenson and Fine [39] also use interval probabilities, but are less closely related to the Bayesian paradigm.) Descriptive models of decision making with imprecise probabilities have also been proposed by psychologists and philosophers (Levi [21], Gardenfors and Sahlin [13,14], Einhorn and Hogarth [7], Loui [24]).

This paper presents a new framework for modeling indeterminacy and incoherence in decision analysis that is based on the use of *confidence-weighted probabilities* (Nau [29]), a generalization of interval probabilities. The representation it proposes for beliefs is conceptually similar to an "epistemic reliability measure" (Gardenfors and Sahlin [13,14]) or "membership function" (Watson et al. [37], Freeling [12], Wallsten et al. [36]) describing a vague probability. The DM's degree of belief in the occurrence of an event is represented by a unimodal (in fact, concave) function on the unit interval. While the shape of this function is suggestive of a second-order probability distribution, it does not have this interpretation. Rather, the vertical scale is considered as an index for a nested sequence of probability intervals, as suggested by fig. 1. This index will be interpreted as a level of "qualification" applied to the DM's assessment, which can be given an operational interpretation in terms of betting with limited stakes. More generally, such a function is defined on the simplex of probability distributions over states, and a nested sequence of convex sets of distributions is obtained as the qualification level is varied. By observing the effects of varying the qualification level, the sensitivity of the recommended decision to different subjective inputs can be analyzed simultaneously, and incoherence can be reconciled if necessary.

The expected-utility side of Savage's [30] theory has also been criticized, and many researchers have proposed that explanations of behavior inconsistent with the principle of maximizing expected utility should be sought in weakenings of assumptions about utility instead of or in addition to those about probability. A survey of recent work in this area is given by Bell and Farquhar [2]. While the exposition in this paper will concentrate on the modeling of indeterminate probabilities, it can in principle be generalized to cover jointly indeterminate probabilities and utilities; a sketch of how to do this is offered in the last section.

We will consider the case of a decision problem in *normal* (tabular) form characterized by finite sets of states of nature and possible decisions. (The alternative representation, *extensive* form, is unsuitable in the presence of indeterminate probabilities because it may not be possible to make unequivocal decisions at intermediate nodes in the decision tree.) Let Θ denote the set of states, let its elements be indexed by a set M of integers, and let Π denote the simplex of probability distributions on Θ.

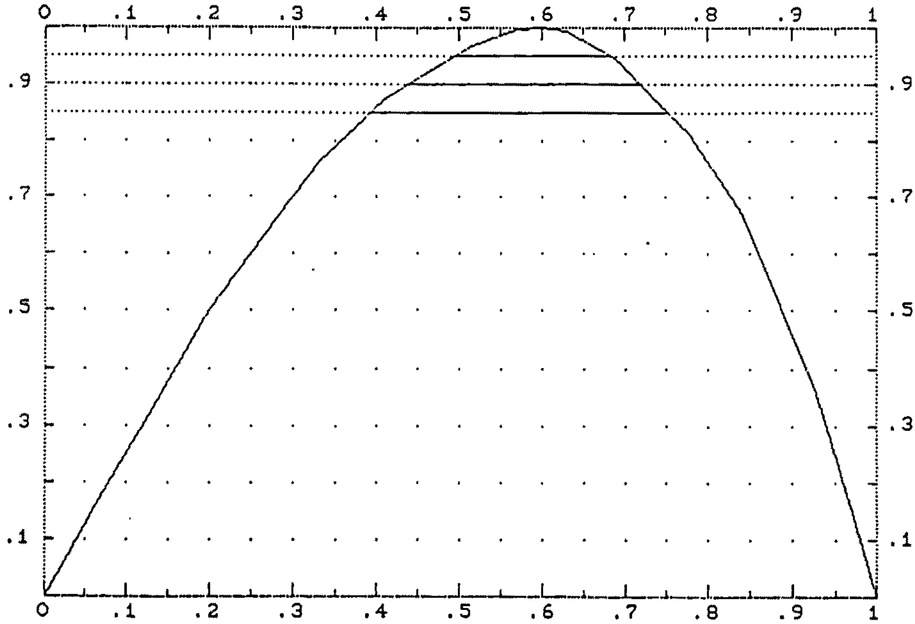

Fig. 1. Representation of an indeterminate probability by a concave function defining a sequence of nested intervals.

For any $m \in M$, θ^m denotes the mth element of Θ and π^m denotes the probability assigned to θ^m by the distribution $\pi \in \Pi$. Let E, F, E_n, F_n denote vectors of 1's and 0's defining events (subsets of Θ). The mth element of E (respectively, E_n) is denoted E^m (respectively, E_n^m), and is 1 or 0 according to whether the event E (respectively, E_n) includes or does not include the state θ^m. Finally, let D_k denote the vector of payoffs yielded by decision k, with generic element D_k^m. The event pairs $\{E_n, F_n\}$ and decisions $\{D_k\}$ will be indexed by sets of integers N and K, respectively. (For the moment, it will be assumed that payoffs are known quantities of money and that utility for money is linear. The extension to the more general case will be mentioned later.) For every distribution $\pi \in \Pi$, let $P_\pi(E)$ denote the probability of an event E, and let $P_\pi(E \mid F)$ denote the conditional probability of E given F. That is:

$$P_\pi(F) \equiv \sum_{m \in M} F^m \pi^m ,$$

$$P_\pi(E \mid F) \equiv P_\pi(EF)/P_\pi(F) \quad \text{if} \quad P_\pi(F) > 0.$$

(Here, EF denotes the elementwise product of E and F – i.e. the vector whose mth element is $E^m F^m$.) Note that if D is the payoff vector for a decision, then $P_\pi(D)$ is simply its expected value under the distribution π. Hence, a separate notation for expectation is unnecessary.

It is desired to determine which among the decisions $\{D_k\}$ is (or are) *preferred.* In a conventional decision analysis, this is determined on the basis of expected value. Let Π_j denote the set of all π for which decision j achieves the maximum expected payoff, i.e.:

$$\Pi_j \equiv \{\pi \mid P_\pi(D_j) \geqslant P_\pi(D_k) \text{ for all } k \in K\}.$$

Roughly speaking, the sets $\{\Pi_j\}$, some of which may be empty, form a partition of Π, except that their edges overlap. For example, consider the following payoff matrix for a decision problem with 6 decisions and 3 states:

	States		
Decisions	θ^1	θ^2	θ^3
D_1	4	−2	0
D_2	2	3	0
D_3	0	2	1
D_4	−1	0	2
D_5	1	4	−2
D_6	3	0	0

(This problem is an adaption of the one discussed in Fishburn et al. [9]. Two additional decisions, here numbered 5 and 6, have been added.) Figure 2 shows the partition of the probability simplex into the corresponding domains $\{\Pi_j\}$. Note that Π_6 does not appear: this set is empty, indicating that D_6 is dominated by a mixture of other decisions. For example, $0.57D_1 + 0.40D_2 + 0.03D_4$ yields the payoff vector (3.05, 0.06, 0.06).

Let $\hat{\pi}$ denote an "estimate" of the DM's probability distribution over the states. Based on this estimate, it can be inferred that D_j is *preferred* (to all other decisions) if $\hat{\pi} \in \Pi_j$. However, even if the preferred decision is D_j, there may be another decision – say, D_k – which is "close" to being preferred in the sense that $\hat{\pi}$ is near the boundary of Π_k. Fishburn et al.'s [9] method of analyzing the sensitivity of the choice of D_j to the estimate $\hat{\pi}$ is to compute the distance from $\hat{\pi}$ to the nearest set Π_k, $k \neq j$. To pursue this idea further, the distance from $\hat{\pi}$ to *each* set Π_j could be measured, and this distance would represent how "close" D_j is to being preferred. Then the alternative decisions might be *ranked* according to their closeness (in this sense) to being preferred. The latter criterion is what will be developed in more detail

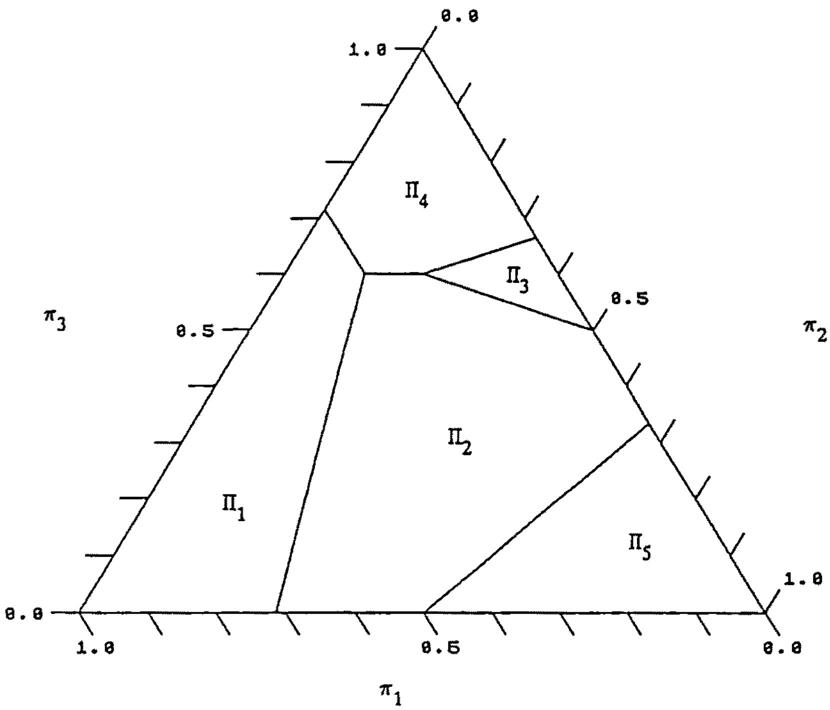

Fig. 2. Partition of the simplex into domains of optimality of different decisions.

below – the key issues are how and in what form to elicit the estimate, and how to measure the distances.

2. Confidence-weighted probabilities

An estimated distribution is usually obtained by asking the decision maker to assess the conditional or unconditional probabilities of various events that are subsets of Θ. Let A_n denote an assertion by the DM with respect to the conditional probability of an event E_n given the occurrence of another event F_n. Within the paradigm of ordinary ("sharp") probabilities, this assertion would have the form: "the conditional probability of E_n given F_n is exactly p_n", for some number p_n. Geometrically, this means the estimated distribution must lie in the intersection of a certain *linear subspace* with the simplex, namely, the set $\hat{\pi}(A_n)$ defined by:

$$\hat{\pi}(A_n) \equiv \{\pi \in \Pi \mid p_n P_\pi(F_n) - P_\pi(E_n F_n) = 0\}.$$

An *assessment* by the DM consists of a finite set of assertions $\{A_n | n \in N\}$, which will be denoted A_N.* An assessment of sharp probabilities constrains the estimated distribution to lie in the intersection of the sets $\{\hat{\pi}(A_n) | n \in N\}$, which will be denoted $\hat{\pi}(A_N)$. Ideally, $\hat{\pi}(A_N)$ is non-empty and consists of a single point. Realistically, unless artificial methods are employed to enforce this condition, $\hat{\pi}(A_N)$ is likely to contain many points or none – i.e. the assessment will either be incomplete or incoherent.

For example, the decision maker may be unable or unwilling to give precise estimates for the probabilities of all the events under consideration. In this case, it may be more natural for her☆ to merely specify lower and upper bounds on the values. Thus, statement A_n might take the weaker form: "the conditional probability of E_n given F_n is *at least* p_n." This constrains the estimated distribution to lie in the intersection of a certain *half-space* with the simplex, which is the set $\hat{\Pi}(A_n)$ defined by:

$$\hat{\Pi}(A_n) \equiv \{\pi \in \Pi \,|\, p_n P_\pi(F_n) - P_\pi(E_n F_n) \leqslant 0\}. \tag{1}$$

(Without loss of generality, we will restrict attention to statements referring to lower bounds, since a lower bound on the probability of one event determines an upper bound on the probability of its complement. In particular, if p is a lower bound on the probability of E, then $1 - p$ is an upper bound on the probability of $\bar{E}$.)

The intersection of the sets $\{\hat{\Pi}(A_n) | n \in N\}$ will be denoted $\hat{\Pi}(A_N)$ and represents our "estimate" of the decision maker's distribution based on an assessment A_N given in the form of lower probabilities. It is unlikely to consist of a single point: in general, it will be a convex polyhedron. Nonetheless, we could still rank the decisions $\{D_j\}$ according to the distances of the sets $\{\Pi_j\}$ from the set $\hat{\Pi}(A_N)$. That is, for each j we could compute the minimum distance from a point in Π_j to a point in $\hat{\Pi}(A_N)$. In practical terms, this would be an optimization problem with a quadratic objective function and linear constraints.

The decision-ranking model just described is consistent with the lower-and-upper-probability view of subjective uncertainty that was originally developed by Koopman [18], Smith [32], and Good [16, and elsewhere], and which continues to attract support. (Suppes [33], Williams [38], Giron and Rios [17], Walley [34,35], Leamer [20], Bewley [3]. See also the comments by Good, Fine, and Seidenfeld on Fishburn [11].) Under this model, a certain amount of indeterminacy is considered to be natural and inevitable, and the likelihood of encountering incoherence is reduced. For those who can live with some things undecided, this is an improvement over the sharp probability model, but it presents its own dilemma, namely, the determinancy of the lower and upper bounds. For example, if the decision maker is willing to assert that the probability of some events lies in the interval [0.6, 0.8], then she would presumable also assert – and with even more confidence – that it lies in the

*The assertions might also consist of estimates of conditional or unconditional *expectations*, and/or *differences* or *ratios* of probabilities or expectations. These would also give rise to linear constraints on the estimated distribution.

☆To avoid confusion between the two actors appearing in this paper, the DM will be referred to as "she", and the DM's betting opponent (who appears later) will be referred to as "he".

interval [0.55, 0.85]; and she might also venture – although with somewhat less confidence – that it lies in the interval [0.65, 0.75]. That is, it seems natural to envision a set of *nested* intervals, indexed by some sort of parameter representing "confidence" (reliability, caution, acceptance, membership, or whatever). This leads to questions about whether such confidence is *intercomparable* between, say, a lower and an upper probability for the same event, or between lower probabilities for different events, and if so, what sort of laws it obeys. Is confidence defined in this way *operationally measurable*, and can it be justified in terms of some set of *axioms* of rational behavior? Previous attempts to generalize the interval-probability model have come up short precisely on these points. For example, Gardenfors and Sahlin [14, p. 242] state: "We will not attempt a full description of the properties of the measure ρ . . . Nor will any attempt be made to describe *how* the values of the ρ measure are to be determined." And Freeling [12, p. 344] laments: "There is in fact a very great difficulty in asking an individual to produce fuzzy distributions. After many months working in this field, the author is still incapable of writing down a [membership] function which could accurately describe one's probability."

The generalization proposed here is that each lower (or upper) probability should be qualified by a numerical *confidence weight*, which is scaled to take on values between 0 and 1. Thus, a generic probability assertion A_n now takes the form: "the probability of E_n given F_n is at least p_n with confidence c_n." Equivalently, we will say that (p_n, c_n) is a *confidence-weighted lower probability* for $E_n | F_n$. The DM now has the option of assessing more than one lower or upper probability for the same event, at different levels of confidence, although this is not strictly necessary. For example, (0.65, 0.5), (0.6, 0.8), and (0.55, 1) could all be confidence-weighted lower probabilities for the same event. These are to be considered as sample points from a continuum of possible responses: the DM's confidence increases as the value for the lower probability decreases.

The interpretation of the confidence weights will be discussed more fully in sect. 4. For the moment, assume that the DM is capable of making such judgments. Assessments of sharp probabilities or lower and upper probabilities, *sans* confidence weights, can still be accommodated within this framework. For example, a sharp probability can be considered as a degenerate interval obtained at a confidence level near zero. An assessment of unweighted lower and upper probabilities can be interpreted as a set of bounds at the same numerical level of confidence, whose value is merely unspecified. In this context, such an assessment would not be considered as the *unique* set of intervals describing the DM's beliefs: wider or narrower bounds might in principle have been given at a greater or lesser degree of confidence.

It may be objected that the additional parameters are a nuisance or even a step in the wrong direction: why increase the number of parameters in order to decrease the precision of the representation? But, once it is admitted that probabilities may be unreliable or in error, it is natural to suppose that some are *more* unreliable or likely to be in error than others. Any attempts to quantify this will inevitably require more parameters, and such additional parameters have often appeared explicitly in the literature dealing with imprecision and error in probability assessment. For example, Fishburn et al. [9] suggest a *weighted* minimum distance method for sensitivity analysis. Lindley et al.'s [22] reconciliation method associates a *precision* factor with each elicited probability, which is interpreted as the inverse of its dispersion under an assumption of normally distributed errors. Morris [25,26] has suggested that a *credibility* factor should be associated with each subjective probability. Of more immediate relevance to this paper, the need for additional parameters has previously been noted within the literature of lower and upper probabilities. Good [16] observed that "[the] inequalities themselves have fuzziness", and advocated the use of higher-order probability distributions to quantify it. While this approach is useful as a conceptual device, it would be difficult to put into practice, and it lacks a convincing theoretical basis. The confidence-weight approach proposed here is operationally simpler, and can be defended on theoretical gounds as a natural consequence of a further relaxation of the axioms of probability.

A question might be raised as to whether an infinite-regress dilemma arises here (as it does with hierarchical probability models): should not the confidence weights themselves be considered as indeterminate, and hence described by higher-order hyperparameters, and so on? In answer to this, it should be noted that, in the progression from sharp probabilities to interval probabilities to confidence-weighted probabilities, each successive layer of modeling is qualitatively different from its predecessors and enlarges the range of behavioral phenomena that can be described. A way in which this progression might be continued to model still higher levels of indeterminacy, and its implications for behavior, are not immediately obvious. (Indeed, an axiomatically-based method of progressing even beyond interval probabilities has hitherto proved elusive.) The important consideration is whether the extension from interval probabilities to confidence-weighted probabilities confers significant practical benefits. It will be seen that what is gained is the ability to give a unified treatment of sensitivity analysis, reconciliation, and combining judgments.

The geometric representation of an assessment given in terms of CWPs will now be considered. Recall that the representation of an assessment of sharp probabilities is (ideally) a single point $\hat{\pi}(A_N)$ in the simplex, whereas an assessment of lower and upper probabilities is represented by a convex subset $\hat{\Pi}(A_N)$. For an assessment of CWPs, the representation takes the form of a *concave function* on the simplex, constructed in the following way. Corresponding to the assertion A_n, let a piecewise linear function $R_\pi(A_n)$ be defined on the simplex Π as:

$$R_\pi(A_n) \equiv 1 - (c_n/p_n)\max\{0, p_n P_\pi(F_n) - P_\pi(E_n F_n)\}. \tag{2}$$

Let $\hat{\Pi}(A_n)$ continue to denote the set of distributions defined by (1), irrespective of the value of c_n. Then $R_\pi(A_n) = 1$ for all $\pi \in \hat{\Pi}(A_n)$, and elsewhere $R_\pi(A_n)$ declines in proportion to c_n multiplied by the Euclidean distance from π to the nearest point in $\hat{\Pi}(A_n)$.* Hence, $1 - R_\pi(A_n)$ is a *weighted distance* from π to the set $\hat{\Pi}(A_n)$, in which the weight is proportional to the confidence c_n and also depends in more subtle ways on p_n and the structure of the events E_n and F_n.

A function $R_\pi(A_N)$ summarizing the entire assessment is now defined as the *pointwise minimum* of the functions $\{R_\pi(A_n)\}$:

$$R_\pi(A_N) \equiv \min_{n \in N} R_\pi(A_n). \tag{3}$$

This function is concave and piecewise linear, and it plays a role similar to that of Gardenfors and Sahlin's [14] measure of the "epistemic reliability" of π, or Watson et al.'s [37] and Freeling's [12] measure of the degree of "membership" of π in the DM's belief set.☆ In this paper, the function R_π will be referred to as the *Bayes risk* function summarizing the DM's assessment, since it will be seen to be interpretable as the Bayes risk (minimum expected loss as a function of probability) for a hypothetical betting opponent when the DM's confidence weights are operationally defined in terms of limits on betting stakes. This interpretation of R_π, along with a rationale for the definitions (2) and (3), will be more fully discussed in sects. 4 and 5.

*In fact, if $d_\pi(A_n)$ denotes the Euclidean distance from π to the nearest point in $\hat{\Pi}(A_n)$, then we have the following identity:

$$1 - R_\pi(A_n) = c_n d_\pi(A_n)[\sqrt{|\Theta|}\,\mathrm{STD}(F_n - E_n F_n / p_n)].$$

(Here, STD(.) denotes the sample standard deviation of the elements of the vector in parentheses computed by the "$1/n$" rule.) The term in square brackets (which does not depend on π or c_n) is a normalizing constant that exactly confines $R_\pi(A_n)$ to the interval [0, 1] in the extreme case where $c_n = 1$. Hence, $1 - R_\pi(A_n)$ is directly proportional to c_n, and also directly proportional to $d_\pi(A_n)$.

☆The model presented here is not based on the theory of "fuzzy sets" (Zadeh [41]). Yet, the formal resemblance between this model and the fuzzy-decision-analysis models by Watson et al [37] and Freeling [12], which *are* based on fuzzy sets theory, is quite strong. The functions describing probabilities obey a "calculus" based on max and min operations, and the "laws" of confidence-weighted probabilities for unconditional events are the same as those obtained by "fuzzifying" the ordinary laws of probability according to Zadeh's [41] extension principle for functions defined on fuzzy sets. It is suggestive to think of $\{R_\pi(A_n)\}$ as membership functions of fuzzy subsets of the simplex induced by the individual statements $\{A_n\}$, in which case $R_\pi(A_N)$ can be viewed as the membership function of their mutual intersection computed by Zadeh's original pointwise-minimum rule. The correspondence breaks down, though, where conditionality is present or where utility and probability are considered jointly. It might be fair to say that the present work provides a foundational justification (which has hitherto been lacking) for using fuzzy-set-like operations in the context of probability and utility modeling (and *only* in that context). However, it suggests that the extension principle needs to be modified to give a consistent treatment of conditionality. More details appear in Nau [29] and in a forthcoming paper.

The complementary function $1 - R_\pi(A_N)$ may be considered to measure the distance of the distribution π from the boundary of a "fuzzy-edged" subset of the simplex describing the DM's beliefs.* (Note that $1 - R_\pi(A_N)$ is the *maximum* of the weighted distances from π to any of the sets $\hat{\Pi}(A_n)$.) In this respect, it is similar to metrics on the simplex proposed by Fishburn et al. [9] and Lindley et al. [22]. Its potential applications to decision analysis will now be discussed.

3. The decision-ranking criterion

In the general normal-form decision problem introduced earlier, let Q_j denote the *minimum* of $1 - R_\pi(A_N)$ on the set Π_j. (Recall that Π_j is the set of distributions for which D_j is a maximal-expected-value decision.) The quantity Q_j measures, in some sense, how "far" the adoption of decision j is from being consistent with the DM's assessed beliefs. More precisely, Q_j measures the distance of decision j from being a "preferable" decision, in a sense to be discussed in sect. 6. Therefore, it is proposed that the decisions be *ranked* in ascending order of the distances $\{Q_j\}$, which are easily computed by linear programming.

The ranking process will now be illustrated for the decision problem introduced in sect. 1. Suppose the following probability assessment is obtained from the DM:

	Event	Lower prob.	Upper prob.	Confidence weight
(a)	θ^1	0.1	0.3	1.0
		0.15	0.25	0.5
(b)	θ^2	0.2	0.4	1.0
		0.25	0.35	0.5
(c)	$\theta^3 \mid \bar{\theta}^1$	0.5	0.75	1.0
		0.6	0.7	0.5

*The following canonical interpretation of "fuzziness" in metric spaces is suggested. Let a set be defined by a function whose value at any point represents the distance from that point to the nearest point in the set. At points which are definitely "in" the set, the distance function has the value zero; at all other points, its value is positive. Thus, membership and non-membership are treated asymmetrically. An ordinary ("crisp") set has the property that, for any point outside the set, there is a straight-line path (namely, the shortest path to the edge of the set) along which distance-from-the-set decreases as a *linear* function of the distance traveled. A fuzzy set, in contrast, would have the property that distance-from-the-set would in general be a *nonlinear, convex* function of the distance traveled, even along a shortest path to it. The distance function for an intersection (respectively, union) of two sets (crisp or fuzzy) would naturally be the pointwise maximum (respectively, minimum) of their individual distance functions. Defining the membership function as the complement of the distance function, the max-min rules for fuzzy-set union and intersection are obtained. However, this idea seems inapplicable to spaces lacking natural metrics, or to the description of pathological (e.g. non-compact) sets.

Figures 3(a), 3(b), and 3(c) are contour plots of the Bayes risk functions on the simplex that would be determined by parts (a), (b), and (c) of the assessment considered separately, superimposed on the partition diagram of fig. 2. (That is, fig. 3(a) is based on the directly assessed CWPs for θ^1, fig. 3(b) is based on those for θ^2, and fig. 3(c) is based on those for $\theta^3|\bar{\theta}^1$.) In all three figures, the central dark band is the set of π on which the Bayes risk achieves its maximal value of 1, namely, those π satisfying the probability bounds at the lowest level of confidence. Additional contour lines are shown at intervals of 0.1 in height. Figure 4 shows the Bayes risk for the entire assessment, which is the pointwise minimum of those in figs. 3(a), 3(b), and 3(c). Note that the darkly shaded area, where the Bayes risk equals 1, straddles both Π_2 and Π_3, indicating that decisions 2 and 3 are preferable (i.e. "within distance 0"). The distances-from-being-preferable $\{Q_j\}$ for all 6 decisons are summarized here:

j	Q_j
1:	0.1667
2:	0.0000
3:	0.0000
4:	0.0305
5:	0.3750
6:	** (dominated)

This yields a preliminary ranking: $Q_2 \approx Q_3 < Q_4 \ll Q_1 \ll Q_5$, with decision 6 eliminated on grounds of dominance.

In order to better discriminate between decisions 2 and 3, the DM may wish to sharpen her original assessment. For example, she might add the following assessments of sharp probabilities (at a low level of confidence) to her original assessment:

	Event	Lower prob.	Upper prob.	Confidence weight
(d)	θ^1	0.2	0.2	0.1
(e)	θ^2	0.3	0.3	0.1
(f)	$\theta^3|\bar{\theta}^1$	0.65	0.65	0.1

These values are merely the midpoints of the narrowest intervals previously assessed for each event, and the resulting assessment is technically incoherent: parts (d) and (e)

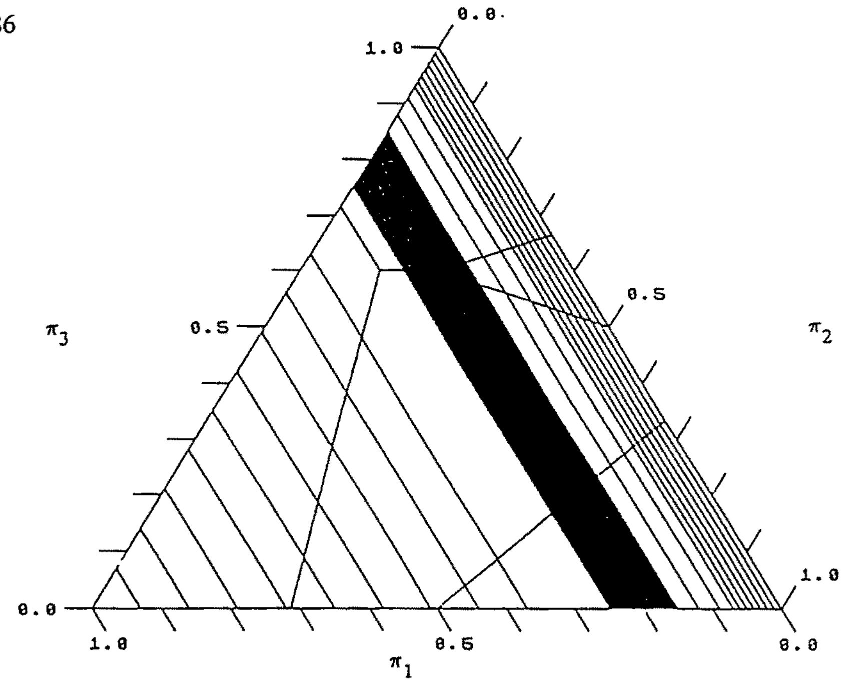

Fig. 3(a). Bayes risk function for part (a) of assessment.

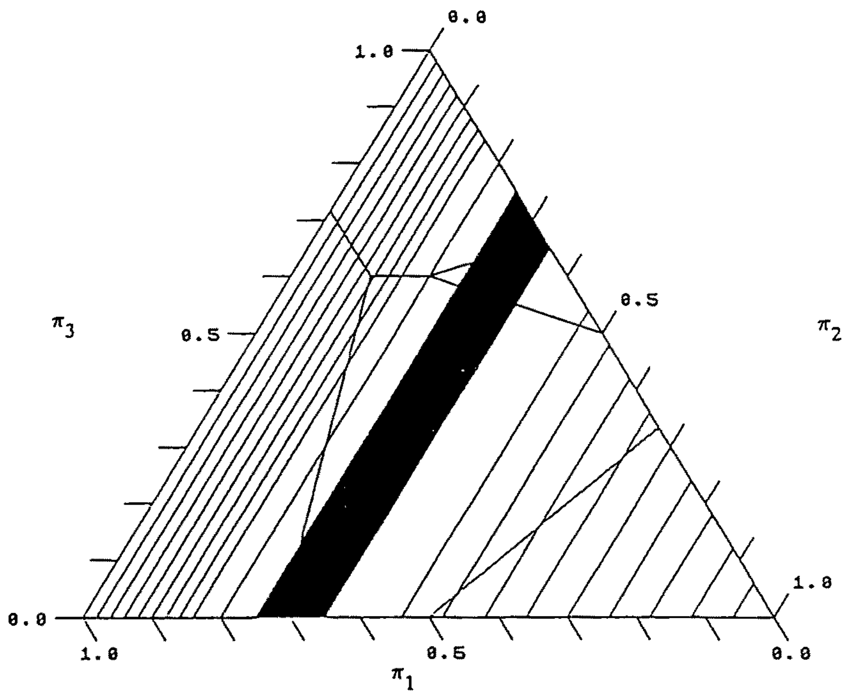

Fig. 3(b). Bayes risk function for part (b) of assessment.

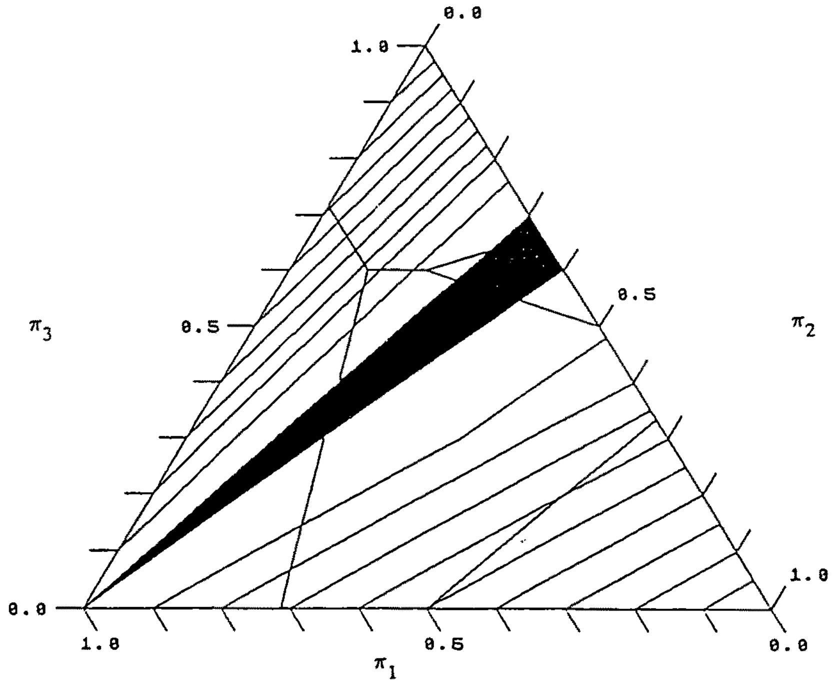

Fig. 3(c). Bayes risk function for part (c) of assessment.

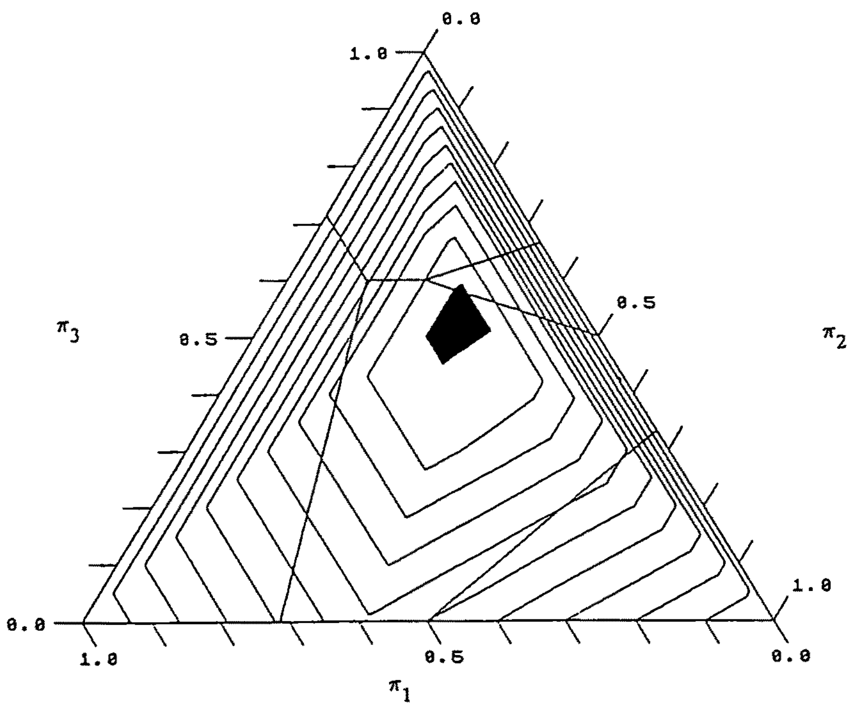

Fig. 4. Bayes risk function for entire assessment.

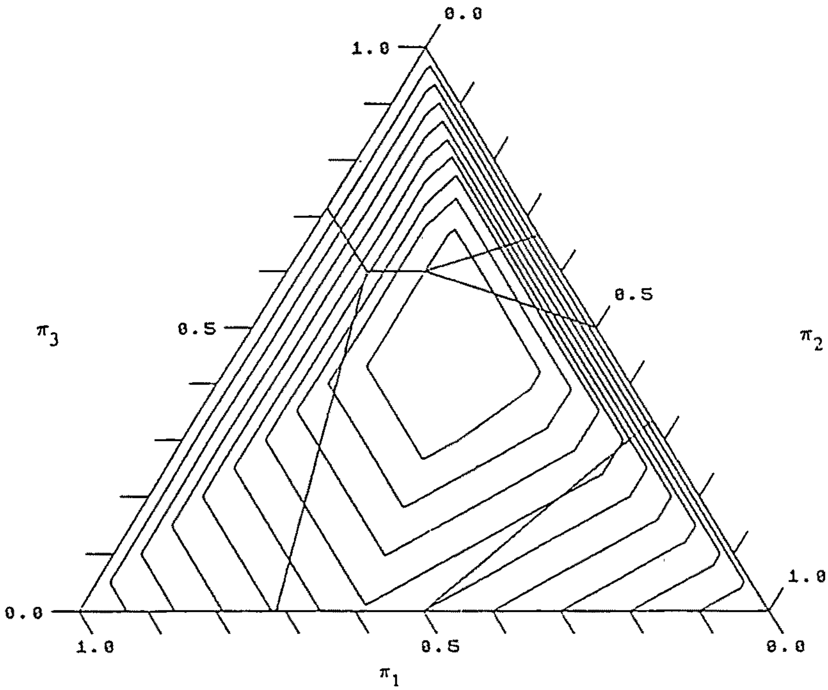

Fig. 5. Bayes risk function for sharpened (incoherent) assessment.

imply that the probability of $\theta^3 | \theta^1$ in part (f) should be 0.625, rather than 0.65. The graph of the Bayes risk function based on (a) through (f) is shown in fig. 5, and it is now seen to have a unique maximum, which occurs at $\pi = (0.1961, 0.2941, 0.5098)$ and which lies strictly inside the domains in which D_2 is optimal. The new distances for the decisions are as follows:

j	Q_j
1:	0.1667
2:	0.0020
3:	0.0167
4:	0.0305
5:	0.3750
6:	** (dominated)

Note that only Q_2 and Q_3 have changed, and decision 2 is now seen to be slightly "closer to being preferable" than decision 3 (0.002 versus 0.0167). The fact that sharpening the assessment led to incoherence presents no particular problem: this only means that the maximum height attained by the Bayes risk function is less than unity (in fact, it is 0.998), and hence *every* decision is at some positive distance from being preferable. The minimum such distance may be considered as a measure of the *relative incoherence* of the assessment, and the point at which the Bayes risk is maximized serves as a reconciled estimate of the DM's distribution. In the example above, the reconciled values of the sharp probabilities for θ_1, θ_2, and $\theta_3 | \overline{\theta}_1$ are 0.1961, 0.2941, and 0.6341, respectively. Thus, an incoherent assessment of CWPs contains, so to speak, the seeds of its own reconciliation. Of course, the DM, upon discovering her assessment to be incoherent, may wish to go back and revise some of her initial assertions, although this is not strictly necessary for the assessment to be put to use, as the above example shows. The linear programming algorithm used to evaluate the Bayes risk can help to suggest revisions by identifying subsets of assertions that are mutually contradictory.

4. Axiomatic basis of confidence-weighted probabilities

This section presents an axiomatic basis and an operational measurement scheme for confidence-weighted probabilities to support the decision-making model presented above. It is essentially a synopsis of results in Nau [29].

De Finetti [4,5] observes that there are two ways to characterize rational behavior under uncertainty that both lead to the construction of a subjective probability measure. One way is to take as primitive the judgment of comparative likelihood between pairs of events (e.g. "E is as likely as F"), and impose on such comparisons the usual axioms of completeness, transitivity, and independence that lead to a complete ordering among events. The other way is to take as primitive the determination of an acceptable price p for a unit lottery ticket whose payoff is 1 if event E occurs and 0 otherwise, conditional on the occurrence of event F. Thus, in the terms introduced in sect. 1, the DM is assumed able to determine a number p for which she will accept a monetary gamble whose payoff vector is $\beta(E-p)F$, where β is "arbitrary (positive or negative) and at the choice of an opponent". The opponent selects β after p is announced but before the outcomes of E and F are observed, and the payoff to the DM is then $\beta(E^m - p)F^m$ when state θ^m obtains. In this setting, the price p constitutes a direct measurement of the numerical probability of E given F, and hence de Finetti refers to this as the "operational" way of defining probability. Actually, E may be considered as an arbitrary *uncertain quantity* rather than an event; that is, the elements of the vector E may be payoffs taking on values other than 0 or 1. In this case, p is interpreted as the subjective conditional *expectation* of E given the occurrence of F. Conditional expectation is actually the primitive concept, with conditional probability as a special case.

De Finetti presents the operational method somewhat informally, but the principal "axioms" of rational gamble-acceptance behavior on which is rests are:

(i) every lottery has a fair price, at which it can be indifferently bought or sold;

(ii) any non-negative linear combination of acceptable gambles is acceptable (a "gamble" meaning the purchase or sale of a lottery); and

(iii) a strictly negative gamble – a "Dutch book" in the parlance of odds-making – is not acceptable.

De Finetti's celebrated result, the "Dutch book theorem", is that (iii) is compatible with (i) and (ii) if and only if there exists a probability measure under which the *price* of every unit lottery is the *probability* of the corresponding event. More precisely, if p_n is the price of the ticket that yields E_n conditional on F_n, for $n \in N$, then (iii) is consistent with (i) and (ii) if and only if there exists $\pi \in \Pi$ such that, for every $n \in N$, either $p_n = P_\pi(E_n | F_n)$ or else $P_\pi(F_n) = 0$. This follows from a separating-hyperplane theorem that is a variant of Farkas' lemma, the basis of the duality theorem of linear programming.

The generalization from sharp probabilities to interval probabilities can be obtained by relaxing the fair-price assumption (i), allowing the DM to quote possibly-distinct buying and selling prices for a lottery. The buying and selling prices are direct numerical measurements of lower and upper probabilities. The corresponding generalization of the Dutch book theorem (Smith [32], Williams [38], Nau [27]) states that the expert avoids exposure to certain loss if and only if there exists some $\pi \in \Pi$ such that, for every n, either $P_\pi(E_n|F_n) \geqslant p_n$ or else $P_\pi(F_n) = 0$. Such a π was referred to by Smith as "medial odds". The set of all medial odds is in general a convex subset of Π, and was introduced in sect. 2 as $\hat{\Pi}(A_N)$. Note that if p and q are lower and upper probabilities, respectively, for the same *unconditional* event, they must satisfy $p \leqslant q$ in order to avoid a trivial Dutch book.

The generalization from lower and upper probabilities to confidence-weighted probabilities is achieved by weakening assumption (ii) to state that only *convex* combinations (rather than non-negative linear combinations) of acceptable gambles are required to be acceptable. The effect of this is to allow the DM to limit the number of lottery tickets she will buy or sell on a given event at a given price. Thus, she can set different limits on the sizes of the stakes associated with different lower and upper probabilities. Assume the DM has a limited total stake for purposes of betting on the outcomes of events in Θ. Then, for each event $E_n | F_n$, let her give a lower probability p_n and an associated *confidence weight* c_n, with the interpretation that she will accept a non-negative multiple of the gamble $(E_n - p_n)F_n$ in which she stands to lose not more than a fraction c_n of her total stake. This may be taken as the operational definition of (p_n, c_n) as a confidence-weighted lower probability for $E_n | F_n$. If units of currency are normalized so the size of the DM's stake is 1, her assertion of (p_n, c_n) as a CWP for $E_n | F_n$ means that she will accept precisely the gamble whose payoff vector is $(c_n/p_n)(E_n - p_n)F_n$, which yields her a payoff of $(c_n/p_n)(E_n^m - p_n)F_n^m$ when state θ^m obtains.

The Dutch book theorems for sharp probabilities and interval probabilities are essentially linear programming duality results. The *information* conveyed by the DM's assessment may be identified with the *statistical decision problem* it presents to her opponent, and such a problem has both a primal representation and a dual representation. The primal representation is a set of payoff vectors representing gambles the DM has agreed explicitly or implicitly to accept; the dual representation is a subset of the probability simplex Π on which all the acceptable gambles have non-negative expected value. In the sharp-probability case, the primal representation is a half-space of gambles, and the dual representation is a single point in Π; in the interval-probability case, the primal representation is a convex cone of gambles, and the dual representation is a convex polyhedral subset of Π. The Dutch book theorem relates a property of the primal representation (that it should not contain a strictly negative vector) to a property of the dual representation (that it should be non-empty), which can be proved constructively by linear programming.

The same considerations apply to an assessment given in the form of CWPs. The primal representation of the assessment is now a *convex polyhedron* of acceptable payoff vectors; the dual representation is a *concave function on* Π, namely, the "Bayes risk" function for the decision problem (DeGroot [6], p. 123 ff.). This is the *opponent's* minimum achievable expected loss expressed as a function of his own hypothetical probability distribution π, where "loss" is conventionally defined as maximum possible gain minus actual gain so as to have a minimum achievable value of zero. If units of currency are normalized so that the DM's total stake is equal to 1, then the opponent's maximum gain is 1. His loss may therefore be defined as 1 minus his gain; and his optimal expected loss is never greater than 1 since he can always achieve that value by choosing the zero gamble. With this scaling convention, the Bayes risk function is bounded below by 0 and above by 1. The Dutch book theorem for CWPs (Nau [29]) states that the DM avoids exposure to certain loss if and only if there is some π in Π at which the Bayes risk function attains the value 1, which is to say, the set $\Pi(A_N)$ defined by (1) is non-empty. An opponent with such a π would decline all the DM's gambles.

To illustrate the construction of the Bayes risk function, consider the case in which the assessment refers to a single unconditional event E. Here, $\Theta = \{E, \bar{E}\}$, and Π may be taken to be the unit interval, with π representing a hypothetical value for the opponent's probability of E. Suppose that the DM has assessed lower probabilities p_1 and p_2 for E with confidence weights c_1 and c_2, and upper probabilities p_3 and p_4 with confidence weights c_3 and c_4, where $p_1 < p_2 < p_3 < p_4$. (That is, $1 - p_3$ and $1 - p_4$ are lower probabilities for $\bar{E}$ with confidence weights c_3 and c_4.) Presumably, $c_1 > c_2$ and $c_3 < c_4$. The graph of the Bayes risk function in this case is the point-wise minimum of:

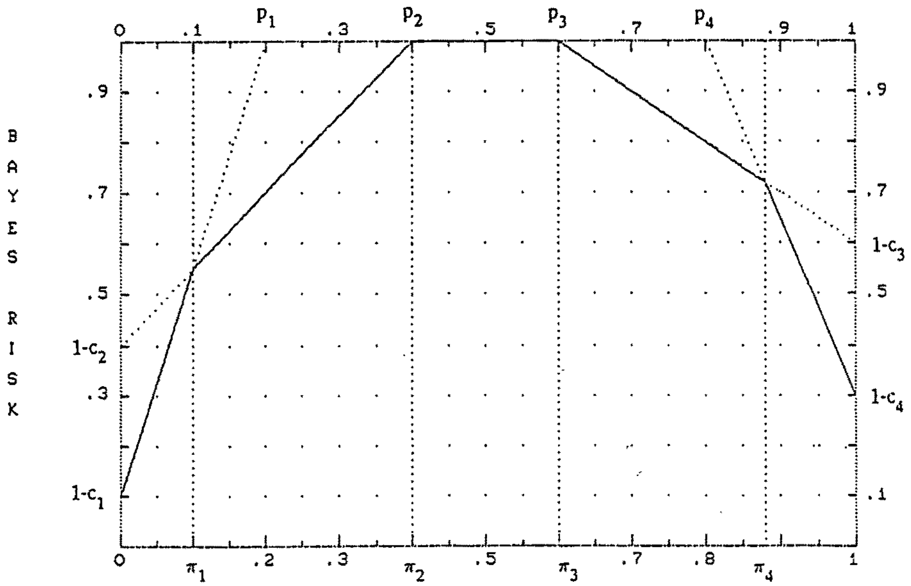

Fig. 6. Construction of (marginal) Bayes risk function for a single event.

the straight line through the points $(0, 1 - c_1)$ and $(p_1, 1)$,
the line through $(0, 1 - c_2)$ and $(p_2, 1)$,
the line through $(1, 1 - c_3)$ and $(p_3, 1)$,
the line through $(1, 1 - c_4)$ and $(p_4, 1)$, and
the horizontal line $y = 1$,

as shown in fig. 6. Each of these lines describes the opponent's expected loss as a function of π for a certain pure strategy, namely, taking the maximum allowed multiple of the gamble defined by a single CWP, or else (for the $y = 1$ line) the zero gamble. For any π, there is always a pure strategy that achieves the minimum expected loss; hence, for practical purposes, there is only a finite number of responses a "rational" opponent need consider, namely, $\beta_n > 0$ for at most one value of n. In the example of fig. 6, the opponent will choose $\beta_1 > 0$ if his probability π lies in the range $(0, \pi_1)$, he will choose $\beta_2 > 0$ if π is in the range (π_1, π_2), and so on.

A smooth concave function such as shown in fig. 1 represents the limiting case of a Bayes risk function constructed in the manner of fig. 6 when intervals (lower and upper probabilities) have been assessed at many different levels of confidence. Note, however, that an interval between lower and upper probabilities having the same confidence is *not* an interval obtained by cutting the graph with a *horizontal* *line*, as in fig. 1. Rather, the lower and upper probabilities with confidence c are the $y = 1$ intercepts of *tangent* lines to the graph which intersect the $x = 0$ and $x = 1$ lines at a height of $1 - c$. The interval defined by a horizontal cut will be interpreted as the range of probability values considered under a given level of *qualification* of the entire assessment, which will be discussed in sect. 6. Thus, a given Bayes risk function defines nested sequences of intervals in two different ways, one obtained by varying a threshold level of confidence, and another obtained by varying a threshold level of qualification.

In announcing more than one lower or upper probability for the same event at different levels of confidence, the DM is offering to bet successively larger stakes at successively more favorable odds. Such behavior is not to be attributed to risk aversion (i.e. concave utility), but rather to *ambiguity* in her perception of the probability, in the sense of Ellsberg's [8] paradoxes. Ambiguity may have many intrinsic sources, such as lack of information, limits of cognition, ill-defined events, and so on, but an intriguing canonical definition is that it is the manifestation of an "uncertainty principle" – an interference of the instrument with the object of measurement – in the realm of probability measurements. Here, the measuring instrument is the opponent who may accept or decline the gambles offered by the DM. If the opponent's acceptance or rejection of gambles is potentially informative to the DM, then she will prefer to maintain a spread in the odds by giving her assessment in terms of lower and upper probabilities rather than sharp probabilities. The generalization from lower and upper probabilities to confidence-weighted probabilities allows an even richer description of this interaction between the DM and her betting opponent. The Bayes risk function in fig. 6 can be interpreted as follows: if the opponent's probability lies in the range $(0, \pi_1)$, then he will take the gamble defined by (p_1, c_1). For this to be agreeable to the DM, then p_1 must be a lower bound on her probability *conditional* on knowing the opponent's probability to be in the range $(0, \pi_1)$. Similarly, p_2 is a lower bound on her probability conditional on knowing his probability to lie in the range (π_1, π_2), and so on. Thus, by offering more than one lower or upper probability for an event, the DM performs a more detailed reciprocal measurement on her betting opponent, and thereby reveals more details about the relative firmness of her own beliefs.

A connection may be noted with the theory of *regret* (Bell [1], Loomes and Sugden [23]) which has been used to explain certain kinds of anomalous decision-making behavior. The size of the confidence weight has been defined as the relative amount of loss to which the DM is willing to expose herself in a gamble on a given event at given odds. This is precisely the *relative regret* to which she is exposed as a result of the decision to offer the gamble, suggesting that the extent to which regret is an important consideration in the DM's choices is related to the degree of ambiguity in her beliefs.

5. Inference with confidence-weighted probabilities

In application of subjective probability, it is often desired to infer the probability of some "target" event from probabilities that have been assessed for other, related events. The interential process can be discussed in terms of the primal representation (acceptable gambles) or the dual representation (subsets of the simplex). For example, suppose that within the sharp-probability paradigm the DM asserts that the probability of E_1 is p_1 and the probability of E_2 is p_2, where E_1 and E_2 are disjoint events. In the primal representation of the DM's beliefs, this means that the gambles $E_1 - p_1$ and $E_2 - p_2$ and any linear combination thereof are acceptable. Now let E_3 be defined as the *union* of E_1 and E_2, and suppose we now wish to infer the DM's probability for E_3. Since E_1 and E_2 are disjoint, we can write: $E_3 = E_1 + E_2$. The inferred probability of E_3 is the value of p_3 for which a gamble of the form $E_3 - p_3$ can be constructed as a linear combination of the first two gambles. Observing that $(E_1 - p_1) + (E_2 - p_2) = E_3 - (p_1 + p_2)$ is a gamble of this form, it follows immediately that $p_3 = p_1 + p_2$. In the dual representation, without loss of generality, let E_1 be composed of the states $\{\theta^1, \ldots, \theta^r\}$, and let E_2 be composed of $\{\theta^{r+1}, \ldots, \theta^s\}$, for some $s > r > 1$, whence E_3 is composed of $\{\theta^1, \ldots, \theta^s\}$. Then, $P_\pi(E_1) = \pi^1 + \ldots + \pi^r$, $P_\pi(E_2) = \pi^{r+1} + \ldots + \pi^s$, and $P_\pi(E_3) = \pi^1 + \ldots + \pi^s = P_\pi(E_1) + P_\pi(E_2)$. The assertions of p_1 and p_2 as probabilities for E_1 and E_2 imply that the DM's belief set consists of π satisfying $P_\pi(E_1) = p_1$ and $P_\pi(E_2) = p_2$, whence it follows that $P_\pi(E_3) = p_3 = p_1 + p_2$ as before.

Rules of inference for CWPs can be derived analogously. For example, suppose that in the CWP paradigm the DM asserts that (p_1, c_1) and (p_2, c_2) are confidence-weighted lower probabilities for disjoint events E_1 and E_2. In the primal representation, this means that the gambles $(c_1/p_1)(E_1 - p_1)$ and $(c_2/p_2)(E_2 - p_2)$ are both acceptable. If we now wish to infer a CWP for $E_3 = E_1 + E_2$, we must find p_3 and c_3 such that a gamble of the form $(c_3/p_3)(E_3 - p_3)$ can be constructed as a *convex* combination of the preceding two gambles. Letting $\beta = (c_2/p_2)/((c_1/p_1) + (c_2/p_2))$, we find:

$$(c_3/p_3)(E_3 - p_3) = \beta(c_1/p_1)(E_1 - p_1) + (1 - \beta)(c_2/p_2)(E_2 - p_2),$$

where p_3 and c_3 are determined by $p_3 = p_1 + p_2$ and $p_3/c_3 = (p_2/c_1) + (p_1/c_2)$. This result might be termed the "additive law of CWPs" for disjoint events. (Note that in the special case where $c_1 = c_2 = c$, it follows also that $c_3 = c$.) Similar analogs can be developed for the other laws of ordinary (sharp) probabilities, including Bayes' theorem. However, unlike with sharp probabilities, a complex inferential calculation with CWPs cannot generally be decomposed into a sequence of calculations in which such laws are applied to combinations of two or three events at a time. That is, such laws do not form an alternative set of "axioms" for CWPs. Rather, in the most general case, linear programming must be used to simultaneously bring to bear *all* of the information in the assessment upon the target event.

The general inference procedure for CWPs is most easily discussed in terms of the dual representation, wherein the assessment A_N is summaried by the Bayes risk function on the simplex. The formulas for constructing the Bayes risk function introduced in sect. 2 will now be derived, and its role in performing inference with CWPs will be discussed. As before, let A_n stand for the DM's assertion of (p_n, c_n) as a confidence-weighted lower probability for $E_n|F_n$, and let A_N denote the entire assessment $\{A_n | n \in N\}$. The DM is committed to accept any gamble whose payoff vector is

$$\sum_{n \in N} \beta_n (c_n/p_n)(E_n - p_n)F_n$$

for non-negative $\{\beta_n\}$ chosen by her opponent such that $\Sigma\beta_n \leqslant 1$. With respect to A_n by itself, the opponent's optimal betting strategy when his own distribution is π is to choose $\beta_n = 1$ (the largest allowable value) if $P_\pi(E_n|F_n) < p_n$, and $\beta_n = 0$ otherwise. This yields a conditional expected gain of $\max\{0, (c_n/p_n)(p_n - P_\pi(E_n|F_n))\}$ given the occurrence of F_n, which occurs with probability $P_\pi(F_n)$, and a conditional gain of zero given the non-occurrence of F_n. Since the opponent's expected loss is 1 minus his expected gain, his Bayes risk function determined by A_n alone is therefore:

$$R_\pi(A_n) \equiv 1 - P_\pi(F_n)\ \max\{0, (c_n/p_n)(p_n - P_\pi(E_n|F_n))\}$$

$$= 1 - (c_n/p_n) \max\{0, p_n P_\pi(F_n) - P_\pi(E_n F_n)\},$$

as given in (2). The opponent's overall problem is to choose a *mixture* of the gambles offered by the DM. However, in mixed decision problems, as is well known, there is an optimal *pure* strategy for any π – i.e. an optimal strategy in which at most one of the $\{\beta_n\}$ is nonzero. It follows that the Bayes risk for the entire assessment $R_\pi(A_N)$ is merely the pointwise minimum of the Bayes risk functions $\{R_\pi(A_n)\}$, as given by (3).

In performing inference with sharp or interval probabilities within their respective dual representations, the image of the DM's belief set (a subset of the simplex) is marginalized or projected onto the subspace of the target event, yielding an inference in the form of a subset of the unit interval – i.e. a point or interval representing the inferred probability. A similar procedure is followed with CWPs: the image of the *function* on the simplex describing the DM's beliefs is projected onto the subspace of the target event, yielding an inference in the form of a *function* defined on the unit interval. From this function, individual probability bounds and their confidence weights can then be extracted. In particular, the function $R_\pi(A_N)$ is "marginalized" with respect to an arbitrary target event $E|F$ to obtain a function $r_x(E|F; A_N)$, defined for $x \in [0, 1]$ as follows:

$$r_x(E|F; A_N) = \sup_{\pi: P_\pi(E|F) = x,\, P_\pi(F) > 0} 1 - (1 - R_\pi(A_N))/P_\pi(F).$$

This is called the *marginal Bayes risk* (MBR) function for $E|F$ determined by the assessment A_N, and it has the following interpretation: $r_x(E|F; A_N)$ is the minimum conditional expected loss, given the occurrence of F, that is achievable by the opponent when x is his conditional probability for $E|F$. (In the case where the set Θ consists of a single event, say E, and its complement, the Bayes risk function and the MBR function for E are identical, and are constructed in the manner of fig. 6.)

The MBR function, like the Bayes risk function, is piecewise linear and concave, and every tangent to its graph corresponds to a nontrivial CWP that can be inferred for $E|F$ from the assessment A_N. That is, each tangent defines a gamble of the form $(c/p)(E - p)F$ that is acceptable for the DM and admissible (undominated) for the opponent. Those tangents that are *facets* of the graph represent irreducible CWPs in terms of which all the inferences for $E|F$ can be summarized. Thus, from an assessment given in the form of a finite number of CWPs, the inferences that can be obtained for any event can also be summarized by a finite number of CWPs. In practice, this can be accomplished by parametric linear programming, and involves not much more computational effort than the manipulation of interval probabilities.

6. Derivation of the decision-ranking criterion

The decison-ranking criterion developed in sect. 3 refers to sets Π_j on which different decisions yield the maximum expected payoff. Thus, it may appear that the principle of maximizing the DM's expected payoff has been assumed *a priori*. In this section, a more fundamental derivation of the decision-ranking criterion will be presented which does not assume the principle of maximizing expected payoff, but instead derives it from an operational definition of *preference*. The natural way to define preference operationally is to say that decision j is preferred to decision k if the DM is willing to *exchange* D_k for D_j. This is equivalent to saying that the DM would pay (at least) \$0 for the lottery whose payoff is $D_j - D_k$, which by definition means that 0 is a *lower expectation* for $D_j - D_k$.

In the CWP context, every inference is qualified by a confidence weight; hence, the notion of preference must also be confidence-weighted. Therefore, suppose that the DM has to option to buy, sell, or trade "shares" of each decision. A share in a decision is some multiple of the corresponding payoff vector. Then, given two decisions D_j and D_k, let $A_{jk}(c)$ stand for the assertion that "D_j is *preferred* to D_k with confidence c", and define this to mean that $(0, c)$ is a *confidence-weighted lower expectation* for the lottery $D_j - D_k$. That is, the DM is willing to exchange a positive multiple of D_k for the same multiple of D_j such that the maximum loss to which she might be exposed due to the exchange is not more than c. More precisely, letting $D_{jk}^{\min}$ denote the minimum element of $D_j - D_k$, the assertion $A_{jk}(c)$ is equivalent to acceptance of the gamble whose payoff vector is $(c/|D_{jk}^{\min}|)(D_j - D_k)$. Bayes risk function determined by this assertion is:

$$R_\pi(A_{jk}(c)) = 1 - (c/|D_{jk}^{\min}|)\max\{0, P_\pi(D_k) - P_\pi(D_j)\}.$$

If $c > 0$, then $R_\pi(A_{jk}(c))$ is equal to 1 for all π such that $P_\pi(D_j) \geqslant P_\pi(D_k)$. Suppose that, for some j, the DM can assert that D_j is preferred to *all* other decisions at some positive level of confidence. Such a decision will be called simply a *preferred* decision Let $A_{jK}(c)$ stand for the conjunction of assertions $\{A_{jk}(c)\}$ for all $k \in K$. Then the corresponding Bayes risk is the pointwise minimum of separate Bayes risk functions:

$$R_\pi(A_{jK}(c)) = \min_{k \in K} R_\pi(A_{jk}(c)).$$

The set of points for which this function is equal to 1 is precisely the set of points for which $P_\pi(D_j) \geqslant P_\pi(D_k)$ for all $k \neq j$, which we previously denoted by Π_j. Thus:

$$\Pi_j = \{\pi \mid R_\pi(A_{jK}(c)) = 1 \text{ for some } c > 0\}.$$

Now return to a consideration of the DM's assessment A_N, and suppose that for some $c > 0$ and some j we have:

$$R_\pi(A_N) \geqslant R_\pi(A_{jK}(c)) \text{ for all } \pi \in \Pi.$$

This means that the set of assertions $A_{jK}(c)$ can be *inferred* from the assessment A_N, which is to say, it can be inferred that j is a preferred decision. In this case, we have $\Pi(A_N) \subseteq \Pi_j$: the *only* distributions satisfying all the probability bounds are those under which decision j is optimal.

In general, such unequivocal preference may not emerge. However, weaker forms of support can also be distinguished. To aid in the formulation, two additional devices will be introduced: the "sharpening" and "qualifying" of an assessment. An Assessment is *sharpened* by conjoining assertions stronger than, but not inconsistent with, those already made. More precisely, A_{N^*} is defined to be a sharpening of A_N if both are coherent and A_N can be inferred from A_{N^*} but not vice versa. (Geometrically, this means the Bayes risk function for A_{N^*} can be obtained by "whittling down" that of A_N, as in the sharpening of a pencil.) Now, if decision j cannot be determined to be preferred on the basis of the assessment A_N, the next strongest support for j would be that there exist some *sharpening* of A_N under which j is preferred. Such a decision will be called a *preferable* ("able to be preferred") decision. Thus, a preferable decision is one that *could* subsequently be determined to be preferred if suitable information were added to the existing assessment. Or, to put it another way, a preferable decision is one that could subsequently be asserted to be preferred (at some positive level of confidence) without contradicting previous testimony.

Formally, decision j is preferable if and only if the conjunction of A_N and a set of assertions of the form $A_{jK}(c)$, for some $c > 0$, is not incoherent. Since:

$$R_\pi(A_N, A_{jK}(c)) = \min\{R_\pi(A_N), R_\pi(A_{jK}(c))\},$$

it follows from the Dutch book theorem (cited in sect. 4) that the conjunction of A_N and $A_{jK}(c)$ is coherent if and only if there is some π for which $R_\pi(A_N) = R_\pi(A_{jK}(c)) = 1$. This, in turn, holds if and only if the sets Π_j and $\hat{\Pi}(A_N)$ have a non-empty intersection, which implies $Q_j = 0$. (Recall that the "distance" Q was defined as the minimum of $1 - R_\pi(A_N)$, on the set Π_j.) Hence, the set of preferable decisions is precisely the set of decisions $\{D_j\}$ for which $Q_j = 0$.

A decision that cannot even be inferred to be preferable (much less preferred) is still not to be entirely disregarded. The distribution of values $\{Q_j\}$ provides the basis for a *sensitivity analysis* of the set of preferable decisions with respect to global features of the assessment A_N. Suppose that, upon reconsideration, the DM wishes to *qualify* her assessment – that is, to retract information in a generalized way. This can be operationalized as follows: let the DM be allowed to demand from the opponent a side payment of size Q (e.g. an an "entry fee") prior to taking up any of the bets that have been offered. The parameter Q, which takes on values between 0 and 1, will be called the *level of qualification* of the assessment, and the qualified assessment will be denoted $A_N(Q)$. The effect of the qualification is that, for every gamble other than the zero gamble, the opponent's expected loss is increased by Q. Therefore, the Bayes risk for the qualified assessment is:

$$R_\pi(A_N(Q)) \equiv \min\{1, Q + R_\pi(A_N)\}.$$

A decision D_j is preferable under the qualified assessment if and only if Π_j has a non-empty intersection with $\hat{\Pi}(A_N(Q))$, which is the set of all π satisfying $R_\pi(A_N(Q)) = 1$, or equivalently $R_\pi(A_N) \geqslant 1 - Q$. This is the set of π obtained by cutting the graph of $R_\pi(A_N)$ with a horizontal line (or hyperplane) at height $1 - Q$, in the manner of fig. 1.

The quantity Q_j introduced in sect. 3 can now be interpreted as *the minimum level of qualification needed in order for decision j to be preferable.* That is, Q_j is the least value of Q for which there exists a sharpening of $A_N(Q)$ under which j is preferred. In this operational sense, Q_j measures the amount by which the assessment must be "stretched" in order to accommodate the selection of decision j as a preferred decision. Thus, the qualification level provides a single parameter through which a sensitivity analysis of the decision model can be performed *jointly* with respect to all the assessed probabilities. As a "metric" with which to perform sensitivity analysis or reconcile incoherence, it is endogenous to the theory, and has a direct operational significance for the DM.

The following guidelines for decision analysis are therefore offered: starting from an assessment A_N, choose a reference value Q for the qualification level and restrict attention to those decisions which are preferable under $A_N(Q)$ – i.e. those $\{j\}$ for which $Q_j \leqslant Q$. (The reference level might be chosen after looking at the distribution of $\{Q_j\}$.) The qualification level here plays a rôle analogous to that of the "desired level of epistemic reliability" in Gardenfors and Sahlin's [14] model, the "precision level" in Watson et al.'s [37] model, or the "membership threshold" in Wallsten et al.'s [36] model. However, no universal prescription (such as Gardenfors and Sahlin's maximin criterion) will be offered here for how to select a unique decision from the restricted set. Rather, some or all of the following holistic strategies might be employed: (i) additional information could be extracted from the existing assessment – e.g. by computing and displaying MBR functions for the expectations of each of the remaining decisions, or for pairwise differences in their expectations; (ii) an attempt could be made to sharpen the original assessment by further introspection; (iii) additional attributes of decision consequences, not included in the original analysis, might be brought into play.

An objection might be raised to the fact that the decision analysis procedure does not guarantee a unique recommended decision: the prescription appears "incomplete" in this sense. Yet, once it is acknowledged that the DM's beliefs may be partly indeterminate, it naturally follows that her preferred decision may also be partly indeterminate. (For an eloquent defense of this view, see Bewley [3].) Any mathematical attempt to force a unique solution on the problem in such a case will be guilty of arbitrariness. The procedure suggested here is actually not qualitatively different from what is already considered good practice: if sensitivity analysis of a decision model reveals that small perturbations of subjectively-assessed parameters lead to different solutions, then a reasonable conclusion (in the absence of additional information) is that the best decision cannot be unequivocally specified. What the CWP-based method offers is a way to systematically and simultaneously explore the sensitivity of the model to all of its subjective inputs.

7. Discussion

A theoretical framework has been presented for modeling ambiguous beliefs in more detail than is possible with interval probabilities, integrating sensitivity analysis into decision analysis, and reconciling incoherence. This has been characterized as a generalization of the Bayesian paradigm, at least as it applies to problems involving finite numbers of states and decisions. However, the CWP model can be also viewed pragmatically as a technology for aiding decision analysis in the presence of imprecise or incoherent judgments even if probabilities are still held to be "theoretically" determinate. In this respect, the key features of the model are (i) the initial association of a *confidence weight* with each elicited probability, similar in spirit to weighting schemes proposed by Fishburn et al. [9], Lindley et al. [22], Nau [27], Morris [25,26], and others; and (ii) the construction of a *piecewise linear metric* incorporating these weights, with which to measure relative distances on the probability simplex. The calculations can be carried out by straightforward linear programming. This framework can be employed no matter whether the probability of a typical event is elicited as a single number, an interval, or nested intervals.

In general, the consequences of decisions may include non-monetary and/or imprecisely known payoffs, and utility for money may be nonlinear. The model described in this paper is not directly applicable in such cases. However, it can be extended to include a joint treatment of probability and utility along the lines developed in Nau [29] if the set of possible consequences is finite. The DM would then be asked to assess *confidence-weighted utilities for consequences* as well as confidence-weighted probabilities for events. The joint assessment of probabilities and utilities would be summarized by a Bayes risk function defined over a product-set Ω of probabilities and utilities, and the analog of the sets $\{\Pi_j\}$ would be subsets $\{\Omega_j\}$ of Ω. The qualification level would then allow a sensitivity analysis to be carried out jointly with respect to probabilities and utilities.

In addition to indeterminacy and incoherence, the CWP model addresses another problem that has been troublesome for Bayesian decision theory: the combination of expert judgments ("expert resolution"). There is a natural method for pooling probability (or utility) judgments obtained in the form of CWPs from different experts, namely, *a linear pool of their Bayes risk functions.* This pooling method has a direct operational significance, and can be shown to have the desirable properties of "marginalizability" and "external Bayesianity", which are known to be incompatible in the context of sharp probabilities.* Details of this appear in Nau [28]. For recent discussions of the expert resolution problem, see Genest and Zidek's [15] survey paper and the collection of papers by Winkler, Lindley, Schervish, Clemen, French and Morris (Management Science 32, 3(1986) pp. 298–328).

It has been shown that confidence-weighted probabilities, like sharp probabilities or interval probabilities, can be given an operational basis in terms of betting. Still, further work is needed on developing guidelines for a DM to follow in articulating assessments in this form, as well as on providing a more detailed prescription for decision analysis. Behavioral research may be undertaken to determine the extent to which actual decision making is described by the model. The CWP model "predicts" (or rather, allows) that decision-making behavior may depend on (i) the intrinsic uncertainty *and ambiguity* characterizing the DM's beliefs, which can be quantified in the form of confidence-weighted probabilities, and (ii) a reference level of *qualification* adopted by the DM for purposes of making a decision in a given context. The

*Thus, "impossibility theorems" concerning the aggregation of probabilities or utilities do not necessarily hold in the CWP context. Of course, the price paid for a representation that allows aggregation is that it does not necessarily lead to a unique optimal decision.

latter might be considered to express the DM's "attitude toward ambiguity" in that context. Together, the assessment of CWPs and the qualification level determine a set of *preferable* decisions. If more than one decision is preferable, and if the assessment cannot be "sharpened" further, the model does not dictate which among the preferable decisions should be chosen. This leaves open the possibility that the final choice may be partly influenced by contextual factors (e.g. status quo attraction or other framing effects and heuristics) that lie outside the scope of the model. In this respect, the CWP paradigm allows room for a rapprochement between normative and descriptive models of choice under uncertainty.

The CWP model does have qualitative features in common with a number of descriptive models that have already found support in behavioral studies. For example, Einhorn and Hogarth's [7] "beta-theta" model of ambiguity postulates that the subjective probability of an event is initially anchored on some reference value (which might be suggested by frequency data) and then adjusted upward or downward for decision-making purposes in a manner that depends on the intrinsic ambiguity surrounding the event and the individual's attitude toward ambiguity in the context of that decision. This model is summarized by graphs describing how a subject's response will vary as the anchoring point is moved (e.g. by considering similar events with different frequency probabilities) while the intrinsic ambiguity and the nature of the decision are held fixed. The CWP model might be considered as a cross-sectional view of the same process: it describes how responses to the same event may vary as features of the decision (gains versus losses, high stakes versus low stakes) are manipulated. Wallsten et al. [36] have tested a model of linguistic uncertainty in which unimodal functions are used to describe vague subjective probabilities. (These are referred to as "membership" functions, although fuzzy-set theory is not otherwise invoked.) In their model, it is postulated that the individual selects a threshold level of membership which depends on features of the decision and then bases his decision on a probability value randomly selected from the interval on which the function exceeds that threshold. This is suggestive of the "level of qualification" used in the CWP decision model. The CWP model does not provide a normative basis for extracting a single probability value from the resulting interval (by randomization or otherwise), but apparent randomization in behavioral experiments might be considered to arise from uncontrolled contextual factors.

References

[1] D.E. Bell, Regret in decision making under uncertainty, Oper. Res. 30(1982)961.

[2] D.E. Bell and P.H. Farquhar, Perspectives on utility theory, Oper. Res. 34(1986)179.

[3] T.F. Bewley, Knightian decision theory, part I, Cowles Foundation Discussion Paper No. 807, Yale University (1986).

[4] B. de Finetti, Foresight: Its logical laws, its subjective sources (1937), translation reprinted in: Kyburg and Smokler (1980)53.

[5] B. de Finetti, *Theory of Probability*, Vol. 1 (Wiley, New York, 1974).
[6] M.H. DeGroot, *Optimal Statistical Decisions* (McGraw-Hill, New York, 1970).
[7] H.J. Einhorn and R.M. Hogarth, Ambiguity and uncertainty in probabilistic inference, Psych. Rev. 92(1985)433.
[8] D. Ellsberg, Risk, ambiguity, and the Savage axioms, Quarterly Journal of Economics 75(1961)643.
[9] P.C. Fishburn, A.H. Murphy and H.H. Isaacs, Sensitivity of decisions to probability estimation errors: A reexamination, Oper. Res. 16(1968)254.
[10] P.C. Fishburn, Analysis of decisions with incomplete knowledge of probabilities, Oper. Res. 13(1965)217.
[11] P.C. Fishburn, The axioms of subjective probability (with comments), Statistical Science 1(1986)335.
[12] A.N.S. Freeling, Fuzzy sets and decision analysis, IEEE Trans. Sys. Man & Cyb. SMC-10:7 (1980)341.
[13] P. Gardenfors and N. Sahlin, Unreliable probabilities, risk-taking, and decision making, Synthese 53(1982)361.
[14] P. Gardenfors and N. Sahlin, Decision making with unreliable probabilities, Brit. J. Math. Stat. Psych. 36(1983)240.
[15] C. Genest and J.V. Zidek, Combining probability distributions: A critique and annotated bibliography, Statistical Science 1(1986)114.
[16] I.J. Good, Subjective probability as the measure of a non-measurable set, in: *Logic, Methodology, and Philosophy of Science*, ed. E. Nagel, P. Suppes and A. Tarski (Stanford University Press (1962), reprinted in Kyburg and Smokler (1980) pp. 134 – 146.
[17] F.J. Giron and F. Rios, Quasi-Bayesian behavior: A more realistic approach to decision-making? in: *Bayesian Statistics II*, ed. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (University Press, Valencia, 1980) pp. 17 – 38.
[18] B.O. Koopman, Axioms and algebra of intuitive probability, Ann. Math. 41(1940)269.
[19] H.E. Kyburg and H.E. Smokler, *Studies in Subjective Probability*, 2nd ed. (Robert Krieger, New York, 1980).
[20] E.A. Leamer, Bid-ask spreads for subjective probabilities, in: *Bayesian Inference and Decision Techniques*, ed. P. Goel and A. Zellner (Elsevier, 1986) pp. 217 – 232.
[21] I. Levi, *The Enterprise of Knowledge* (MIT Press, Cambridge, MA, 1980).
[22] D.V. Lindley, A. Tversky and R.V. Brown, On the reconciliation of probability assessments, J. Roy. Stat. Soc. A142, Part 2 (1979)146.
[23] G. Loomes and R. Sugden, Regret theory: An alternative theory of choice under uncertainty, Econ. J. 92(1983)805.
[24] R.P. Loui, Decisions with indeterminate probabilities, Theory and Decision 21(1986)283.
[25] P.A. Morris, The credibility of decision models, Presentation at ORSA/TIMS Chicago Meeting (1983).
[26] P.A. Morris, The credibility of probabilities, Presentation at ORSA/TIMS Miami Meeting, and unpublished manuscripts (1986).
[27] R.F. Nau, Coherent assessment of subjective probability, ORC 81-5, Operations Research Center, Berkeley (1981).
[28] R.F. Nau, Combining and reconciling expert judgments with confidence-weighted probabilities, Working Paper No. 8610, Fuqua School of Business, Duke University (1986).
[29] R.F. Nau, A new theory of indeterminate probabilities and utilities, Working Paper No. 8609, Fuqua School of Business, Duke University (1988).
[30] L.J. Savage, *The Foundations of Statistics* (Wiley, New York, 1954).
[31] G. Shafer, *A Mathematical Theory of Evidence* (Princeton University Press, Princeton, 1976).

[32] C.A.B. Smith, Consistency in statistical inference and decision, J. Roy. Stat. Soc. B23 (1961)1.
[33] P. Suppes, The measurement of belief, J. Roy. Stat. Soc. B36(1974)160.
[34] P. Walley, The elicitation and aggregation of beliefs, Statistics Research Report, University Warwick (1982).
[35] P. Walley, *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall, London, 1987), in preparation.
[36] T.S. Wallsten, D.V. Budescu and I. Erev, Understanding and using linguistic uncertainty, Presented at SPUDM-11, Cambridge, U.K. (1987).
[37] S.R. Watson, J.J. Weiss and M.L. Donnell, Fuzzy decision analysis, IEEE Trans. Sys. Man & Cyb. SMC-9:1(1979)1.
[38] P.M. Williams, Indeterminate probabilities, in: *Formal Methods in the Methodology of Empirical Sciences*, ed. M. Przlecki, K. Szaniawki and R. Wojciki (Ossolineum and Reidel, 1976) pp. 229–246.
[39] M. Wolfenson and T.L. Fine, Bayes-like decision making with upper and lower probabilities. J. Amer. Stat. Assoc. 77(1982)80.
[40] L.A. Zadeh, Fuzzy sets, Information and Control 8(1965)338.
[41] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning – I. Information Sciences 8(1975)199.