Towards a Unified Theory of Imprecise Probability

Peter Walley
36 Bloomfield Terrace
Lower Hutt, New Zealand

Abstract

Belief functions, possibility measures and Choquet capacities of order 2, which are special kinds of coherent upper or lower probability, are amongst the most popular mathematical models for uncertainty and partial ignorance. I give examples to show that these models are not sufficiently general to represent some common types of uncertainty. Coherent lower previsions and sets of probability measures are considerably more general but they may not be sufficiently informative for some purposes. I discuss two other models for uncertainty, involving sets of desirable gambles and partial preference orderings. These are more informative and more general than the previous models, and they may provide a suitable mathematical setting for a unified theory of imprecise probability.

Keywords. Choquet capacity, coherence, comparative probability, desirable gambles, imprecise probability, incomplete preference, lower prevision, lower probability.

1 Introduction

Can there be a unified theory of imprecise probability? At present there are numerous mathematical models, interpretations and applications of imprecise probabilities. As a general or unified theory may be expected to accommodate this variety of mathematical models, interpretations and applications, it may appear that such a theory will be difficult to attain.

My view is that a single theory of imprecise probability, as in [22], can accommodate all the kinds of uncertainty and partial ignorance that are currently being studied, including vague or qualitative judgements of uncertainty, complete ignorance and near ignorance, random sets and multivalued mappings, and partial information about an unknown probability measure. To defend this view it is necessary to examine these types of uncertainty in some detail, and this has been attempted in [22, 23, 26].

My aim in this paper is more limited: to consider what level of mathematical generality will be needed in a unified theory of imprecise probability. I will argue that none of the mathematical models that are most popular at present (numbers 1-4 in the list below) is sufficiently general, and I will suggest several other models that do seem to be sufficiently general but have received less attention than they deserve. By a 'sufficiently general' model, I mean one that can represent all the types of uncertainty and partial ignorance that are commonly encountered in applications.

The mathematical models that I consider are, in order of increasing generality,

1. possibility measures and necessity measures [4, 33]
2. belief functions and plausibility functions [2, 18]
3. Choquet capacities of order 2 [1, 3]
4. coherent upper and lower probabilities [13, 19]
5. coherent upper and lower previsions [22, 23, 24, 31]
6. sets of probability measures [9, 15]
7. sets of desirable gambles [22, 30, 32]
8. partial preference orderings [8, 22].

Another important type of model, which can be regarded as a special case of models 5-8 but which does not fit neatly into the preceding list, is

9. partial comparative probability orderings [5, 11, 12].

I think that all the models I have listed are appropriate and useful in particular types of application. Some well known examples are: (1) vague judgements of uncertainty in natural language; (2) multivalued mappings and non-specific information; (3) some types of statistical neighbourhood in robustness studies, and various economic applications; (4) personal betting rates, and upper and lower bounds for probabilities; (5) buying and selling prices for gambles, and envelopes of expert opinions; (6) partial information about an unknown probability measure; (7,8) judgements of the desirability of, or preference between, gambles; and (9) qualitative judgements of uncertainty.

If each of these models is useful in some types of application, of course it follows that a sufficiently general model should include all the listed models as special cases. This argument supports the most general of the models, 7 and 8, as the most promising candidates for a unified theory. In the rest of the paper I develop this argument by examining the mathematical models and the relationships between them in more detail, especially with regard to their mathematical generality. I have already discussed possibility measures and belief functions in [23]. In this paper I consider models 3-8 in order of increasing generality.

Almost all of the results in this paper have been discussed in my previous work, especially in [22, 23]. Nevertheless I feel that it is important to review these results at this symposium, because the choice of a mathematical model is a fundamental issue and many people continue to advocate models which have limited applicability.

My argument is based largely on examples which show that models 3-6 are not sufficiently general. To simplify the argument, I have chosen the examples to be as simple as possible and to involve only 3 or 4 possible outcomes. The phenomena illustrated in the examples are not restricted to small problems but actually occur more frequently in larger problems.

2 Choquet Capacities of Order 2

Let Ω denote the set of possibilities under consideration. Suppose that lower probabilities $\underline{P}(A)$ are defined for all $A \in \mathcal{K}$, where $\mathcal{K}$ is a collection of subsets of Ω. In this section $\mathcal{K}$ is assumed to be an algebra. For models 1-4 in the earlier list, lower probabilities determine conjugate upper probabilities through $\overline{P}(A) = 1 - \underline{P}(A^c)$, so it suffices to consider lower probabilities.

Assume that $0 \leq \underline{P}(A) \leq 1$ for all $A \in \mathcal{K}$, $\underline{P}(\emptyset) = 0$ and $\underline{P}(\Omega) = 1$. The lower probability $\underline{P}$ is said to be *2-monotone*, or a *Choquet capacity of order 2*, when it also satisfies, whenever A and B are in $\mathcal{K}$,

$$\underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B). \tag{1}$$

It is well known that probability measures, belief functions and necessity measures (the conjugates of possibility measures) are always 2-monotone lower probabilities. A simple method of constructing 2-monotone lower probabilities is to apply a convex transformation to the probability interval: if P_0 is a probability measure on $\mathcal{K}$, f is a convex function from $[0, 1]$ into $[0, 1]$ with $f(0) = 0$, and $\underline{P}$ is defined by $\underline{P}(\Omega) = 1$ and $\underline{P}(A) = f(P_0(A))$ when $A \in \mathcal{K}$ and $A \neq \Omega$, then $\underline{P}$ is a 2-monotone lower probability. Many of the neighbourhood models used in Bayesian and frequentist studies of robust statistics are of this form [25].

To show that order-2 capacities are not sufficiently general, I will give a simple example of coherent lower probabilities that are not 2-monotone. This example also shows that belief functions and necessity or possibility measures are not sufficiently general, since these models are special types of order-2 capacities. In my experience, most of the coherent lower probability models that occur in applications are not 2-monotone.

Example 1 Coin Tossing ([20] and [22], sec. 5.13.4). *Suppose that a fair coin is 'tossed' twice, in such a way that heads and tails are equally likely on each of the tosses but there can be arbitrary dependence between the tosses. For example, the coin may be tossed first in the usual way, but on the second 'toss' it may be placed to have the same outcome as the first toss, or it may be placed to have the opposite outcome from the first toss.*

Let H_1 and T_1 denote the possible outcomes (heads or tails) of the first toss, and similarly H_2 and T_2 for the second toss. To simplify the notation, denote the possible joint outcomes $(H_1H_2, H_1T_2, T_1H_2, T_1T_2)$ by (a, b, c, d). If we are completely ignorant about the interaction between the two tosses, we can model our uncertainty about the outcomes by using the set of all probability measures P that assign $P(H_1) = P(H_2) = \frac{1}{2}$. Let $\mathcal{M}$ denote this set of probability measures. The two extreme points of $\mathcal{M}$ are the probability measures that assign probabilities $(0, \frac{1}{2}, \frac{1}{2}, 0)$ and $(\frac{1}{2}, 0, 0, \frac{1}{2})$ to (a, b, c, d). These two extreme points correspond to the two possible mechanisms, mentioned in the preceding paragraph, by which the first outcome may determine the second.

We find the upper and lower probabilities that are generated by $\mathcal{M}$ by maximizing and minimizing probabilities under the two extreme points. Since $H_1 \cap H_2 = \{a\}$ and $H_1 \cup H_2 = \{a, b, c\}$, we obtain $\underline{P}(H_1) = \overline{P}(H_1) = \frac{1}{2} = \underline{P}(H_2) = \overline{P}(H_2)$, $\underline{P}(H_1 \cap H_2) = 0$, $\overline{P}(H_1 \cap H_2) = \frac{1}{2}$, $\underline{P}(H_1 \cup H_2) = \frac{1}{2}$, and $\overline{P}(H_1 \cup H_2) = 1$. These upper and lower probabilities are coherent, because they are upper and lower envelopes of a set of probability measures. But the lower probabilities are not 2-monotone, since $\underline{P}(H_1 \cup H_2) + \underline{P}(H_1 \cap H_2) = \frac{1}{2} < 1 = \underline{P}(H_1) + \underline{P}(H_2)$.

This example also illustrates that the Choquet integral is a bad way of defining lower expectations when lower probabilities are not 2-monotone. Let X denote the number of heads obtained in the two tosses, and let I denote indicator function. Since both tosses are fair, we should obtain the precise expectation $E(X) = E(I_{H_1} + I_{H_2}) = E(I_{H_1}) + E(I_{H_2}) = P(H_1) + P(H_2) = \frac{1}{2} + \frac{1}{2} = 1$.

It can be verified that the Choquet integral produces upper and lower expectations $\overline{E}(X) = \frac{3}{2}$ and $\underline{E}(X) = \frac{1}{2}$, values that are incoherent with the probabilities $P(H_1) = P(H_2) = \frac{1}{2}$. Generally the Choquet integral produces coherent upper and lower expectations if and only if the initial lower probability function is 2-monotone [20].

3 Coherent Lower Probabilities

The simplest mathematical characterization of coherent lower probabilities is that they are *lower envelopes* of a set of probability measures. That is, lower probabilities $\underline{P}$, defined on $\mathcal{K}$, are coherent if and only if there is a nonempty set of probability measures, $\mathcal{M}$, such that $\underline{P}(A) = \inf\{P(A) : P \in \mathcal{M}\}$ for all $A \in \mathcal{K}$. Another characterization in terms of positive linear combinations of desirable gambles, which shows that coherence is a normative requirement of consistency, is given in [22, 23, 24].

All 2-monotone lower probabilities are coherent [20]. Example 1 therefore shows that coherence is more general than 2-monotonicity. But coherent lower probabilities are still not sufficiently general, for the following reasons.

(a) They cannot model comparative probability judgements such as "event A is at least as probable as B" or "A is at least c times as probable as B".

(b) They do not determine unique lower (or upper) expectations, which are needed in making decisions.

(c) They do not determine unique conditional lower (or upper) probabilities, which are needed in making inferences.

(d) Even in problems where lower probabilities are an adequate model for an initial state of uncertainty, after we condition on a subset of Ω, the updated lower probabilities may no longer be adequate because they have lost relevant information [10].

Problems (a), (b) and (c) can be illustrated by a single example involving comparative probability judgements; similar examples are in [14] and [22] (sec. 2.7.3, ch. 4).

Example 2 A Football Game [22, 23]. *Consider a football game with three possible outcomes for the home team, labeled as W (win), D (draw) and L (loss). Suppose that a subject makes three qualitative judgements of his uncertainty concerning the outcome: (i) not win is at least as probable as win; (ii) win is at least as probable as draw; and (iii) draw is at least as probable as loss.*

Judgement (i) can be represented in terms of upper or lower probabilities by $\overline{P}(W) \leq \frac{1}{2}$ or $\underline{P}(D \cup L) \geq \frac{1}{2}$. The other two judgements cannot be represented adequately in terms of upper or lower probabilities: for example, the translation $\underline{P}(W) \geq \overline{P}(D)$ of judgement (ii) is too strong. Instead, the three judgements should be regarded as constraints on a coherent lower prevision, or equivalently as constraints on a probability measure P, of the form: (i) $P(W) \leq \frac{1}{2}$; (ii) $P(W) \geq P(D)$; and (iii) $P(D) \geq P(L)$.

Let $\mathcal{M}$ denote the set of all probability measures that satisfy these three constraints. Then $\mathcal{M}$ is a closed convex polyhedron (in this case a triangle) whose three extreme points assign probabilities $(\frac{1}{2}, \frac{1}{2}, 0), (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$ and $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ to the outcomes (W, D, L). The set $\mathcal{M}$ generates upper probabilities $\frac{1}{2}, \frac{1}{2}, \frac{1}{3}$ and lower probabilities $\frac{1}{3}, \frac{1}{4}, 0$ for W, D, L respectively, by maximizing and minimizing probabilities under the extreme points.

Now suppose that the subject is permitted to express his uncertainty only by making judgements of upper and lower probabilities, and he assesses the upper and lower probabilities that were just specified. Let $\mathcal{M}^*$ be the set of all probability measures that lie between the upper and lower probabilities, i.e., that satisfy $\underline{P}(A) \leq P(A) \leq \overline{P}(A)$ for $A = W, D, L$. We find that $\mathcal{M}^*$ is a closed convex polyhedron with five extreme points: in addition to the three extreme points of $\mathcal{M}$, $\mathcal{M}^*$ has the extreme points $(\frac{1}{3}, \frac{1}{2}, \frac{1}{6})$ and $(\frac{5}{12}, \frac{1}{4}, \frac{1}{3})$. Because $\mathcal{M}$ is strictly contained in $\mathcal{M}^*$, the new model $\mathcal{M}^*$ is less informative than the previous model $\mathcal{M}$. In other words, the upper and lower probabilities are less informative than the qualitative judgements (i)-(iii) which generated them. Information may be lost when uncertainty is modeled only in terms of upper and lower probabilities.

To see that the lost information is relevant in calculating lower expectations, consider the gamble $X = I_W - I_D$, which takes the value 1 if W occurs, -1 if D, and 0 otherwise. By minimizing expectations with respect to the extreme points, we find that the lower expectation is $\underline{E}(X) = 0$ under the first model ($\mathcal{M}$), but $\underline{E}(X) = -\frac{1}{6}$ under the second model ($\mathcal{M}^*$).

To see that the lost information is relevant in calculating conditional lower probabilities, consider conditioning on the observation (B) that the outcome is not L. By conditioning the extreme points and minimizing, we find that $\underline{P}(W|B) = \frac{1}{2}$ under $\mathcal{M}$ but $\underline{P}(W|B) = \frac{2}{5}$ under $\mathcal{M}^*$.

The sets $\mathcal{M}$ and $\mathcal{M}^*$ generate the same unconditional upper and lower probabilities. Given only these upper and lower probabilities, we cannot tell whether the underlying model for uncertainty is $\mathcal{M}$ or $\mathcal{M}^*$ or some other set. Consequently we cannot tell whether the upper and lower expectations and conditional probabilities should be those generated by $\mathcal{M}$ or by $\mathcal{M}^*$ or by some other model.

The lower probability function in this example is not a belief function, but it is 2-monotone, simply because the possibility space is so small: all coherent lower probabilities on a 3-point space are 2-monotone. Suppose that we extend the space to 4 possible outcomes by distinguishing scoring draws from non-scoring draws, and we add one judgement to the earlier ones (i-iii): the event that there is a scoring draw or a loss is at least as probable as a win. Then the resulting lower probabilities are not 2-monotone. When the lower probabilities are generated by the kinds of qualitative judgements in this example, 2-monotonicity becomes less likely as the possibility space gets larger.

Example 3 Coin Tossing. *To illustrate problem (d), consider the set of probability measures, $\mathcal{M}$, defined in Example 1, which models two coin tosses with unknown interaction. It can be verified that $\mathcal{M}$ is the set of all probability measures that lie between the upper and lower probabilities stated in Example 1, and in fact $\mathcal{M}$ is the unique closed convex set of probability measures that generates these upper and lower probabilities as its upper and lower envelopes. In this case the upper and lower probabilities are adequate models for uncertainty, because they uniquely determine $\mathcal{M}$.*

Now suppose that we learn partial information about the outcomes of the two tosses: we learn that at least one outcome was heads. How does this change our uncertainty? We should update $\mathcal{M}$ to $\mathcal{M}'$ by conditioning the two extreme points of $\mathcal{M}$ on $\{a, b, c\}$, using Bayes' rule ([22], sec. 6.4). *Hence the two extreme points of $\mathcal{M}'$ assign probabilities $(0, \frac{1}{2}, \frac{1}{2})$ and $(1, 0, 0)$ to (a, b, c). The updated set $\mathcal{M}'$ generates upper probabilities $\overline{P}(\{a\}) = 1$ and $\overline{P}(\{b\}) = \overline{P}(\{c\}) = \frac{1}{2}$, and lower probability 0 for each possible outcome. But now these upper and lower probabilities are* not *an adequate model for the updated uncertainty, because they are not sufficiently informative to determine $\mathcal{M}'$. The set of all probability measures that lie between the upper and lower probabilities, $\mathcal{M}''$, has four extreme points: $(0, \frac{1}{2}, \frac{1}{2})$, $(1, 0, 0)$, $(\frac{1}{2}, \frac{1}{2}, 0)$, and $(\frac{1}{2}, 0, \frac{1}{2})$. If we replace $\mathcal{M}'$ by the upper and lower probabilities or by $\mathcal{M}''$, we lose information that might be needed in making decisions.*

In view of these inadequacies of upper and lower probabilities, why have they received so much attention in the literature on imprecise probability? I think that this is due largely to an uncritical acceptance of the traditional approach of probability theory. A precise probability measure P does determine unique expectations, through the formula $E_P(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$ if Ω is finite, and E is the unique linear expectation operator whose restriction to events is P. Thus there is a one-to-one correspondence between probability measures and linear expectations, and no information is lost when uncertainty is specified in terms of a probability measure. Also a probability measure P determines unique conditional probabilities through the formula $P(A|B) = P(A \cap B)/P(B)$ (Bayes' rule), provided that $P(B) > 0$. This explains why probability theory can be formulated in terms of unconditional probabilities. Nevertheless, there are some advantages in formulating probability theory in terms of expectations or previsions, as in [6, 29], and, as de Finetti [6] recognized, the usual formulation is inadequate for dealing with conditioning events that have probability zero. The case $P(B) = 0$ is discussed in later sections.

It is clear from (b) and (c) above that these properties of probability measures do not generalize to lower probabilities. Lower probabilities are not sufficiently informative to determine unique lower expectations or unique conditional lower probabilities. In my experience, lower probabilities are inadequate models in many applications, including most applications in which imprecise probability models are constructed from sets of probability measures, as in Example 1, or from qualitative judgements of uncertainty, as in Example 2.

Upper and lower probability models which are not necessarily coherent, including *fuzzy measures* [27] (also known as Choquet capacities of order 1), are mathematically more general than coherent upper and lower probabilities but they are inadequate for the same reasons: as set functions, they are not sufficiently informative about upper and lower expectations and conditional probabilities.

4 Coherent Lower Previsions

A bounded mapping from Ω to $\mathbb{R}$ (the real numbers) is called a *gamble*. Let $\mathcal{K}$ be a nonempty set of gambles. A mapping $\underline{P} : \mathcal{K} \to \mathbb{R}$ is called a *lower prevision* or *lower expectation*. A lower prevision is said to be *coherent* when it is the lower envelope of some set of linear expectations, i.e., when there is a nonempty set of probability measures, $\mathcal{M}$, such that $\underline{P}(X) = \inf \{E_P(X) : P \in \mathcal{M}\}$ for all $X \in \mathcal{K}$, where $E_P(X)$ denotes the expectation of X with respect to P. The conjugate upper prevision is determined by $\overline{P}(X) = -\underline{P}(-X)$.

Example 4 Coin Tossing. *The set of probability measures $\mathcal{M}$ in Example 1 generates a coherent lower prevision $\underline{P}$ through $\underline{P}(X) = \inf \{E_P(X) : P \in \mathcal{M}\}$. Using the fact that the infimum is achieved by one of the two extreme points specified in Example 1, we find that*

$$\underline{P}(X) = \frac{1}{2} \min \{X(b) + X(c), X(a) + X(d)\} \qquad (2)$$

for every gamble X. The upper prevision $\overline{P}(X)$ is obtained by replacing min *by* max *in this formula.*

When $\mathcal{K}$ is a linear space of gambles, coherence is equivalent to the three simple axioms (for all $X, Y \in \mathcal{K}$):

A1. $\underline{P}(X) \geq \inf \{X(\omega) : \omega \in \Omega\}$

A2. $\underline{P}(cX) = c\underline{P}(X)$ whenever $c > 0$

A3. $\underline{P}(X + Y) \geq \underline{P}(X) + \underline{P}(Y)$.

Coherent lower probabilities can be regarded as a special type of coherent lower prevision, by taking $\mathcal{K}$ to be a set of indicator functions of subsets of Ω and identifying the lower probability of a subset with the lower prevision of its indicator function. For that reason, it is convenient to adopt de Finetti's convention of using the same symbol A to denote both a subset of Ω and its indicator function.

Coherent lower previsions avoid most of the defects of lower probabilities that were discussed in Section 3:

(a) Lower previsions can model the comparative probability judgement "A is at least as probable as B" by $\underline{P}(A-B) \geq 0$, and "$A$ is at least c times as probable as B" by $\underline{P}(A-cB) \geq 0$. (Here A and B denote indicator functions.) In Example 2, for instance, the three judgements would be modeled through the constraints $\underline{P}(D+L-W) \geq 0$, $\underline{P}(W-D) \geq 0$ and $\underline{P}(D-L) \geq 0$.

(b) Lower expectations (i.e., lower previsions) are uniquely determined for all gambles in $\mathcal{K}$.

(c) Provided that $\underline{P}(B) > 0$ and the gamble $B[X - \underline{P}(X|B)]$ is in $\mathcal{K}$, lower previsions determine conditional lower previsions $\underline{P}(\cdot|B)$ uniquely, through the *generalized Bayes rule*

$$\underline{P}(B[X - \underline{P}(X|B)]) = 0. \tag{3}$$

This equation, like Bayes' rule, is necessary for coherence of conditional and unconditional previsions ([22], sec. 6.4). The equation has a unique solution $\underline{P}(X|B)$ because $\underline{P}(B[X-c])$ is strictly decreasing in c if $\underline{P}(B) > 0$. Conditional lower probabilities $\underline{P}(A|B)$ are determined by taking X in (3) to be the indicator function of A.

Example 5 Coin Tossing. *In Example 1, suppose we want to condition on the event $B = \{a, b, c\}$, that at least one outcome was heads. Using formula (2) for lower previsions, we can solve (3) to obtain the conditional lower previsions $\underline{P}(X|B) = \min\{X(a), \frac{1}{2}X(b) + \frac{1}{2}X(c)\}$ (for all gambles X). This is the lower envelope of the updated set of probability measures, $\mathcal{M}'$, in Example 3.*

(d) Because there is a one-to-one correspondence between coherent lower previsions and closed convex sets of probability measures, coherent lower previsions also solve the problem of missing information that was illustrated in Example 3.

So coherent lower previsions are more general and more informative than coherent lower probabilities. However, there remain two respects in which coherent lower previsions may not be sufficiently informative:

(e) When $\underline{P}(B) = 0$, coherent lower previsions do not determine conditional lower previsions $\underline{P}(\cdot|B)$ ([22], sec. 6.10). This is important when we need to update lower previsions after observing B.

Example 6 Coin Tossing. *In Example 1, suppose we learn the information, denoted by S, that the tosses produced the same outcome: either both heads (H) or both tails (T). What are the updated lower probabilities $\underline{P}(H|S)$ and $\underline{P}(T|S)$? Because $\underline{P}(S) = 0$, the generalized Bayes rule (3) does not have a unique solution. It implies only that $\underline{P}(H|S) \leq \frac{1}{2}$ and $\underline{P}(T|S) \leq \frac{1}{2}$. The vacuous conditional probabilities $\underline{P}(H|S) = \underline{P}(T|S) = 0$ are coherent with the initial model, but so are the precise conditional probabilities $P(H|S) = P(T|S) = \frac{1}{2}$, and so are any conditional lower probabilities that lie between these two extremes. A modified conditioning rule which produces the precise conditional probabilities as the unique solution was studied in* [22], Appendix J; *this is equivalent to removing the extreme point $(0, \frac{1}{2}, \frac{1}{2}, 0)$ from the set of probability measures $\mathcal{M}$ in Example 1.*

(f) Lower previsions cannot distinguish preference from weak preference. For example, if X and Y are two gambles such that $\underline{P}(X-Y) = 0$, then X is weakly preferred to Y ($X \succeq Y$), but we do not know whether X is preferred to Y ($X \succ Y$), and it is possible that Y is preferred to X. This may be important in decision problems, where X and Y represent the utility functions that are associated with two feasible actions and we must decide which action to select.

Problems (e) and (f) are both caused by the inadequacy of the real-number scale. The same problems occur for precise probabilities, which of course are a special case of lower probability or lower prevision. Conditioning on events of probability zero causes real difficulties in Kolmogorov's theory of probability (Borel's paradox is a well known example): if $P(B) = 0$ then the conditional probability measure $P(\cdot|B)$ is completely indeterminate. Also, if two gambles X and Y have the same expectation $E(X) = E(Y)$ then preferences between them are indeterminate: we may have $X \succ Y$, $Y \succ X$, or $X \approx Y$ (we are *indifferent* between X and Y). It may seem that we should always be indifferent between X and Y when $E(X) = E(Y)$, but if $X \geq Y$ and $X(\omega) > Y(\omega)$ for some possible outcomes ω then we would surely prefer X to Y. De Finetti [6] considered using infinitesimals (nonstandard real numbers) to provide a richer scale for probability.

It is arguable that problems (e) and (f) are unimportant, because they concern infinitesimal differences in unconditional expected utility. In the Kolmogorov approach, it is often claimed that events of probability zero are negligible. That may be true before the conditioning event is observed, but after observing an event of probability zero, differences that were previously negligible may become important. In statistical problems with a continuous sample space, it is usual that all possible observations have (upper) probability zero, and then posterior probabilities based on the observation are indeterminate. Also, it is more common for an event to have lower probability zero than to have precise probability zero: on an epistemic interpretation, $\underline{P}(B) = 0$ means only that there is no evidence at the present time to support the occurrence of B, not that it has no chance of occurring. In Example 6, for instance, we have $\underline{P}(S) = 0$ but $\overline{P}(S) > 0$.

Problem (e) could be solved by taking *conditional* lower prevision to be the fundamental concept, and specifying $\underline{P}(\cdot|B)$ directly, when necessary, rather than attempting to define it in terms of unconditional lower previsions. That approach was followed in [6] for prevision and in [22] for lower prevision. But it does not solve problem (f).

5 Sets of Probability Measures

Can these problems be solved by using a set of probability measures as the mathematical model for uncertainty? It is immediately clear that the answer is no. In the special case of precise probability, the set of probability measures reduces to a single measure and the inadequacies of the real-number scale remain. More generally, there is a one-to-one correspondence between coherent lower previsions (defined on the set of all gambles) and nonempty closed convex sets of probability measures: the closed convex set is the set of all probability measures whose expectations dominate the lower prevision, and the lower prevision is the lower envelope of this set of expectations ([22], Thm. 3.6.1). Examples of closed convex sets of probability measures have been given in Examples 1-3. If we restrict attention to sets of probability measures $\mathcal{M}$ that are closed and convex, they are exactly as general as coherent lower previsions.

Greater generality might be achieved by dropping the requirement of convexity, but convexity of $\mathcal{M}$ does not appear to have any behavioural or practical significance, at least when the behaviour is generated by $\mathcal{M}$ alone. (This can change when we combine $\mathcal{M}$ with other sets of probability measures.) Any set $\mathcal{M}$ has exactly the same behavioural implications as its convex hull: both sets generate the same lower previsions and preference orderings. In Example 1, for instance, it makes no difference to preferences whether we are completely ignorant about the interaction between the two tosses, which produces the convex set $\mathcal{M}$ in Example 1 as the model for uncertainty, or we know that the second outcome is completely determined by the first through one of the two possible deterministic mechanisms, which produces the 2-point set containing the two extreme points of $\mathcal{M}$.

A little more generality can be achieved by dropping the closure requirement. For example, if the three qualitative judgements in the football example are modified by replacing ‘at least as probable as’ by ‘more probable than’, the judgements determine an open set which is the interior of the set $\mathcal{M}$ in Example 2. The open set models a preference for W over D, since all the probability measures in it satisfy $P(W) > P(D)$, whereas the closed set $\mathcal{M}$ contains probability measures with $P(W) = P(D)$ and models only a weak preference for W over D. Distinguishing between open and closed sets of probability measures can therefore solve problem (f) in some cases.

Similarly, problem (e) can be avoided in some examples by using an open set of probability measures which does not assign probability zero to any conditioning event B but may have lower envelope $\underline{P}(B) = 0$. Then conditional probabilities and lower probabilities are uniquely determined through Bayes’ rule. In the coin-tossing Example 6, if we modify the set $\mathcal{M}$ by removing the extreme point $(0, \frac{1}{2}, \frac{1}{2}, 0)$, then all probability measures in the modified set assign positive probability to S, and we obtain the unique conditional probabilities $P(H|S) = P(T|S) = \frac{1}{2}$.

Sets of probability measures can be a little more informative than coherent lower previsions, but they are still not sufficiently informative to avoid problems (e) and (f) in general. Problem (e) remains whenever $\overline{P}(B) = 0$, as in statistical problems with a continuous sample space, since then every probability measure in the set must assign $P(B) = 0$ and conditional probabilities are completely indeterminate. Problem (f) remains whenever $\underline{P}(X - Y) = \overline{P}(X - Y) = 0$.

6 Sets of Desirable Gambles and Partial Preference Orderings

Let $\mathcal{L}$ denote the set of all gambles (bounded mappings $\Omega \to \mathbb{R}$). For $X, Y \in \mathcal{L}$, write $X \geq Y$ to mean that $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$, and write $X > Y$ to mean that $X \geq Y$ and $X(\omega) > Y(\omega)$ for some $\omega \in \Omega$. A *set of desirable gambles*, denoted by $\mathcal{D}$, is a subset of $\mathcal{L}$. A set of desirable gambles is said to be *coherent* when it satisfies the four axioms [22, 30]:

D1. if $X \in \mathcal{L}$ and $0 > X$ then $X \notin \mathcal{D}$

D2. if $X \in \mathcal{L}$ and $X > 0$ then $X \in \mathcal{D}$

D3. if $X \in \mathcal{D}$ and $c \in \mathbb{R}^+$ then $cX \in \mathcal{D}$

D4. if $X \in \mathcal{D}$ and $Y \in \mathcal{D}$ then $X + Y \in \mathcal{D}$.

Thus a coherent set of desirable gambles is a convex cone of gambles that contains all positive gambles ($X > 0$) but no negative gambles ($X < 0$). An additional conglomerability axiom, which implies stronger properties of coherence, was required in [22].

A *partial preference ordering* $\succ$ is a partial ordering of the gambles in $\mathcal{L}$. $X \succ Y$ is read as ‘gamble X is preferred to gamble Y’. Coherent partial preference orderings can be characterized through a set of axioms that are closely related to D1-D4 ([22], Appendix F).

There is a one-to-one correspondence between coherent sets of desirable gambles and coherent partial preference orderings, defined by $X \succ Y$ if and only if $X - Y \in \mathcal{D}$. (See [22], p. 153, for justification.) With this correspondence, the two models are equally general. As mathematical objects, coherent sets of desirable gambles are simpler than coherent partial preference orderings because they eliminate some of the redundancy in the ordering. Here I concentrate on sets of desirable gambles, but all of the following discussion applies to partial preference orderings through the one-to-one correspondence.

A set of desirable gambles can retain all the information in the earlier models, and it can supply some additional information by specifying which of the gambles on the boundary of the set are desirable ([22], sec. 3.8.6 and App. F). This additional information is exactly what is needed to condition on events of probability zero and to distinguish preference from weak preference.

To see that all the information in the earlier models can be represented in terms of a set of desirable gambles, suppose that a coherent lower prevision $\underline{P}$, defined on a set of gambles $\mathcal{K}$, is given. Define

$$\begin{aligned}\mathcal{D} = \quad & \{X \in \mathcal{L} : X > \textstyle\sum_{i=1}^{n} c_i[X_i - \underline{P}(X_i) + \varepsilon] \\ & \text{for some } n \geq 0, c_i \geq 0, \varepsilon > 0, X_i \in \mathcal{K}\}. \quad (4)\end{aligned}$$

Then $\mathcal{D}$ is a coherent set of desirable gambles, and $\underline{P}$ can be recovered from $\mathcal{D}$ by, for all $X \in \mathcal{K}$,

$$\underline{P}(X) = \sup\{c : X - c \in \mathcal{D}\}. \quad (5)$$

Coherent lower probabilities are a special case of coherent lower previsions and so they can be recovered from $\mathcal{D}$ by $\underline{P}(A) = \sup\{c : A - c \in \mathcal{D}\}$.

Similarly, given a closed convex set of probability measures, $\mathcal{M}$, define

$$\mathcal{D} = \{X \in \mathcal{L} : X > 0, \text{ or } E_P(X) > 0, \forall P \in \mathcal{M}\}. \quad (6)$$

Then $\mathcal{D}$ is coherent and $\mathcal{M}$ can be recovered from it by

$$\mathcal{M} = \{P : E_P(X) \geq 0, \forall X \in \mathcal{D}\}. \quad (7)$$

Example 7 Coin Tossing. *The set of probability measures, $\mathcal{M}$, in Example 1 generates a coherent set of desirable gambles $\mathcal{D}_1$ through (6). Using the fact that $E_P(X) > 0$ for all $P \in \mathcal{M}$ if and only if $E_P(X) > 0$ for both extreme points of $\mathcal{M}$, which are the probability distributions $(0, \frac{1}{2}, \frac{1}{2}, 0)$ and $(\frac{1}{2}, 0, 0, \frac{1}{2})$, we find that $\mathcal{D}_1 = \{X \in \mathcal{L} : X > 0$, or $X(b) + X(c) > 0$ and $X(a) + X(d) > 0\}$. This is the set of all gambles that must be judged desirable, given $\mathcal{M}$.*

Gambles on the boundary of $\mathcal{D}_1$, those which satisfy $X(b) + X(c) \geq 0$ and $X(a) + X(d) \geq 0$ with at least one equality, are not included in $\mathcal{D}_1$ because they may or may not be desirable. By classifying the desirability of these gambles we can obtain a larger and more informative set $\mathcal{D}$. For example, consider the coherent set $\mathcal{D}_2 = \{X \in \mathcal{L} : X > 0$, or $X(b) + X(c) \geq 0$ and $X(a) + X(d) > 0\}$, which strictly contains $\mathcal{D}_1$. Both $\mathcal{D}_1$ and $\mathcal{D}_2$ generate the same set $\mathcal{M}$, through (7), and the same lower previsions, through (5), but we shall see that they produce different conditional lower probabilities and different preferences.

The model $\mathcal{D}$ can be sufficiently informative to overcome problems (e) and (f):

(e) Conditional lower previsions are uniquely determined by $\mathcal{D}$ through the formula

$$\underline{P}(X|B) = \sup\{c : B(X - c) \in \mathcal{D}\}. \quad (8)$$

Hence conditional lower probabilities $\underline{P}(A|B)$ are determined by taking X to be the indicator function of A. There is no special difficulty when $\underline{P}(B) = 0$, because $\mathcal{D}$ can provide sufficient information to discriminate between sets of probability zero ([22], App. F). An example is given below (Example 8).

(f) There is a preference for X over Y if and only if $X - Y \in \mathcal{D}$. There is a weak preference for X over Y if and only if $X - Y + \varepsilon \in \mathcal{D}$ for all $\varepsilon > 0$. Thus the model can distinguish between preference and weak preference. In Example 7, let $X = 2\{a\}$ and $Y = \{d\}$. Since the gamble $X - Y = (2, 0, 0, -1)$ is in $\mathcal{D}_2$ but not in $\mathcal{D}_1$, X is preferred to Y under model 2, but X is only weakly preferred to Y under model 1. So $\mathcal{D}_2$ is more informative than $\mathcal{D}_1$ about preferences.

Similarly, a comparative probability judgement "A is more probable than B" can be modeled by requiring that $A - B \in \mathcal{D}$, and "A is at least as probable as B" by requiring $A - B + \varepsilon \in \mathcal{D}$ for all $\varepsilon > 0$. Sets of desirable gambles are therefore more general than *partial comparative probability orderings*, which are a special type of partial preference ordering in which preferences are specified only between indicator functions of events. Partial comparative probability orderings are not sufficiently general because usually they do not determine lower probabilities, lower previsions and preferences between other gambles.

Coherent sets of desirable gambles, or (equivalently) coherent partial preference orderings, appear to be sufficiently general and sufficiently informative to model the common types of uncertainty and the most important aspects of uncertainty. Of course coherence is a normative (consistency) requirement and it is unlikely to be an accurate description of people's intuitive reasoning. Sets of desirable gambles or partial preference orderings which satisfy weaker properties than coherence, such as 'avoiding sure loss' or 'n-coherence' [22], may be more useful as descriptive models.

Although sets of desirable gambles are more general than the previous models, they simplify the mathematical theory of coherence and natural extension [22]. For example, the generalized Bayes rule (3) can be expressed in the following simple form: if we observe a subset B of Ω, we should update the initial set of desirable gambles, $\mathcal{D}$, to $\mathcal{D}' = \{X \in \mathcal{L} : BX \in \mathcal{D}\}$ ([22], sec. 6.1.6). More generally, if we obtain a statistical observation that generates a bounded likelihood function L on Ω, we should update $\mathcal{D}$ to $\mathcal{D}' = \{X \in \mathcal{L} : LX \in \mathcal{D}\}$. More generally still, if we observe upper and lower likelihood functions U and L ([22], sec. 8.5.3), then we should update $\mathcal{D}$ to $\mathcal{D}' = \{X \in \mathcal{L} : LX^+ + UX^- \in \mathcal{D}\}$, where X^+ and X^- denote the positive and negative parts of X. These simple rules apply even when conditioning on events of upper or lower probability zero.

Example 8 Coin Tossing. *Consider the coin tossing example, and suppose we learn that $S = \{a, d\}$ has occurred, as in Example 6. Two sets of desirable gambles, $\mathcal{D}_1$ and $\mathcal{D}_2$, were defined in Example 7. After observing S, these are updated to $\mathcal{D}_1' = \{X \in \mathcal{L} : SX \in \mathcal{D}_1\} = \{X \in \mathcal{L} : (X(a), X(d)) > (0,0)\}$, and $\mathcal{D}_2' = \{X \in \mathcal{L} : SX \in \mathcal{D}_2\} = \{X \in \mathcal{L} : X(a) + X(d) > 0\}$. Using (8) or (5), $\mathcal{D}_1$ generates the vacuous conditional probabilities and $\mathcal{D}_2$ generates the precise conditional probabilities given in Example 6. The extra information in $\mathcal{D}_2$ determines conditional probabilities precisely.*

A central idea of the theory in [22] is the idea of natural extension. Suppose we judge all the gambles in a set $\mathcal{D}_0$ to be desirable, where $\mathcal{D}_0$ is a subset of some coherent set but is not necessarily coherent. Then the *natural extension* of $\mathcal{D}_0$, denoted by $\mathcal{D}$, is defined to be the smallest coherent set of desirable gambles that contains $\mathcal{D}_0$. So $\mathcal{D}$ is the smallest convex cone that contains $\mathcal{D}_0$ and all positive gambles, and it can be generated from $\mathcal{D}_0$ by applying the rules D2-D4 [32]. The coherent set $\mathcal{D}$ fully expresses the implications of the desirability judgements in $\mathcal{D}_0$.

An important special case is that in which both Ω and $\mathcal{D}_0$ are finite sets. In that case the model $\mathcal{D}$ is said to be *finitely generated* ([22], sec. 4.2). Finitely generated models occur frequently in practice, when the modeling or elicitation process produces a finite set of basic judgements which can be translated into judgements that particular gambles are desirable. A finitely generated set $\mathcal{D}$ produces, through (7), a closed convex set of probability measures, $\mathcal{M}$, that has finitely many extreme points.

Example 9 Football. *Suppose that we take the judgements in Example 2 to be the strict comparative probability judgements $D \cup L \succ W \succ D \succ L$, which is equivalent to taking $\mathcal{D}_0 = \{D + L - W, W - D, D - L\}$. The natural extension of these judgements is the coherent set of desirable gambles $\mathcal{D} = \{X \in \mathcal{L} : X \geq c_1(D + L - W) + c_2(W - D) + c_3(D - L)$ for some $c_1, c_2, c_3 \geq 0$, and $X \neq 0\}$. This $\mathcal{D}$ generates, through (7), the set of probability measures $\mathcal{M}$ that was defined in Example 2, and, through (5), the upper and lower probabilities in Example 2.*

For finitely generated models, it is often convenient to calculate inferences directly from $\mathcal{D}_0$, rather than to first calculate the extreme points of $\mathcal{M}$. For example, upper and lower previsions, defined by (5), can be computed directly from $\mathcal{D}_0$ by using linear programming techniques. Also the generalized Bayes rule can be applied directly to $\mathcal{D}_0$. Suppose that we obtain a statistical observation which generates a likelihood function L on Ω and $L(\omega) > 0$ for all $\omega \in \Omega$. Then we simply update $\mathcal{D}_0$ to the finite set $\mathcal{D}_0' = \{X/L : X \in \mathcal{D}_0\}$. Inferences can be calculated directly from $\mathcal{D}_0'$ because the updated uncertainty model $\mathcal{D}' = \{X \in \mathcal{L} : LX \in \mathcal{D}\}$ is the natural extension of $\mathcal{D}_0'$.

Example 10 Football. *In the football example, suppose that we observe a crowd of unhappy spectators leaving the game. If we interpret this as evidence that the home team did not win, and model it through likelihoods $(\frac{1}{2}, 1, 1)$ for (W, D, L), then we would simply update the initial set of judgements $\mathcal{D}_0 = \{D + L - W, W - D, D - L\}$ to the set $\mathcal{D}_0' = \{X/L : X \in \mathcal{D}_0\} = \{D + L - 2W, 2W - D, D - L\}$. We can obtain a new model $\mathcal{D}'$, and any inferences that are required, directly from $\mathcal{D}_0'$.*

Another argument in favour of partial preference orderings is that they are needed in a general theory of decision which allows imprecision in both probabilities and utilities, as in [7, 17, 21]. In a general theory of decision, the primary mathematical model will be some kind of partial preference ordering, of either the possible actions or more general objects such as randomized actions, Savage acts or horse lotteries. Such orderings might be constructed from separate assessments of imprecise probabilities and imprecise utilities, but it is important to recognize that not all the reasonable partial preference orderings can be constructed in this way. That is illustrated by the following example. See [17] for a similar conclusion.

Example 11 Intersection of Complete Preference Orders [21]. *Consider the simplest possible non-trivial decision problem, where there are two possible states of the world, labeled as ω and ω', and two possible consequences, c_1 and c_2. Denote the four possible acts by a_{ij}, where $a_{ij}(\omega) = c_i$ and $a_{ij}(\omega') = c_j$ for $i, j = 1, 2$.*

Suppose that a subject evaluates the acts by assessing both a probability $P(\omega) > \frac{1}{2}$ and utility values $U(c_1) > U(c_2)$. By ordering acts according to their expected utility, he obtains the complete preference ordering

$$a_{11} \succ a_{12} \succ a_{21} \succ a_{22}. \tag{9}$$

A second subject assesses $P(\omega) < \frac{1}{2}$ and $U(c_1) <$

$U(c_2)$ and he obtains the complete preference ordering

$$a_{22} \succ a_{12} \succ a_{21} \succ a_{11}. \tag{10}$$

The intersection of these two complete orderings is the partial ordering in which $a_{12} \succ a_{21}$ but all other pairs of acts are incomparable. This partial ordering models the 'consensus preferences' that the two rational subjects have in common. It would also be the appropriate model for an individual who produced the complete orderings (9) and (10) by analyzing the decision problem in two different ways, but who was undecided about which analysis to accept; such a person may have determined only that $a_{12} \succ a_{21}$. The partial ordering is therefore a reasonable model for preferences.

But this partial ordering cannot be obtained from a set of probability measures $\mathcal{M}$ and a set of utility functions $\mathcal{U}$, by taking $a \succ b$ if and only if a has greater expected utility than b under all combinations of a probability measure in $\mathcal{M}$ with a utility function in $\mathcal{U}$. To prove that, suppose that a partial ordering is obtained in this way. Then the comparison $a_{12} \succ a_{21}$ implies that $[1-2P(\omega)][U(c_2) - U(c_1)] > 0$ for all $P \in \mathcal{M}$ and $U \in \mathcal{U}$. This implies that either $P(\omega) > \frac{1}{2}$ and $U(c_1) > U(c_2)$ for all $P \in \mathcal{M}$ and $U \in \mathcal{U}$, giving the complete ordering (9), or $P(\omega) < \frac{1}{2}$ and $U(c_1) < U(c_2)$ for all $P \in \mathcal{M}$ and $U \in \mathcal{U}$, giving the complete ordering (10). So a partial preference ordering which is obtained from some sets $\mathcal{M}$ and $\mathcal{U}$, and which includes the preference $a_{12} \succ a_{21}$, must be a complete ordering.

Compare this example with Savage's result [16], that every reasonable complete preference ordering of acts can be constructed from separate assessments of a precise probability measure and a precise utility function. Partial preference orderings are more general than combinations of imprecise probability and imprecise utility. This is important because it shows that preferences need to be constructed in other ways, not just by assessing imprecise probabilities and imprecise utilities. A very general method of constructing a coherent partial preference ordering from simple judgements was outlined in [21].

7 Conclusions

Until now, most studies of imprecise probability have been concerned with special types of upper and lower probability or with comparative probability orderings. I have argued that these models are not sufficiently general to represent some common types of uncertainty. In advocating a more general model, I am not suggesting that we should stop studying coherent upper and lower probabilities, Choquet capacities, belief functions, possibility measures and other special kinds of model. As I said in the introduction, each of these models is useful in special kinds of application, and each has special mathematical properties which make it interesting from a theoretical point of view. However, I suggest that much more effort should be devoted to studying the more general models which are needed in many applications.

Lower previsions are much more general and informative than lower probabilities, and they seem to be adequate models in the great majority of applications that are concerned with uncertainty but not with utility, and those applications in which utilities are precisely known. They also have an advantage of familiarity over the more general models: they are closer to well established concepts of probability and expectation, and especially to de Finetti's concept of prevision [6].

There is a duality relationship between coherent lower previsions and sets of probability measures. Some aspects of the mathematical theory can be handled most conveniently with one model and some with the other. It is therefore important to be able to use both models and to exploit the duality. Coherent lower previsions have the advantage of being more closely related to preferences and behaviour than are sets of probability measures. Many authors, particularly those studying robust Bayesian inference, have not yet recognized that many of the things they are doing with sets of probability measures can be done more easily with coherent lower previsions. For example, simpler methods can be found for checking coherence and making inferences from precise or imprecise probability assessments [22, 24]. As another example, the main result of [28] was proved much more simply in [20] using only elementary properties of coherent lower previsions; see also [25].

Sets of desirable gambles and partial preference orderings are the most informative of the mathematical models I have discussed, and they seem to be able to model all the common types of uncertainty. They uniquely determine upper and lower previsions and conditional previsions, and they contain all the information about preferences that is relevant in making decisions. In many ways they are the simplest and most natural mathematical models. The coherence axioms and rules of inference (natural extension) for sets of desirable gambles are especially simple. In this paper I have advocated these models on the grounds of mathematical generality, but it is also arguable that they are the simplest and most natural models from the point of view of interpretation [22]. I conclude that sets of desirable gambles and partial preference orderings may be the best mathematical models for a general theory of imprecise probability.

Acknowledgements

I want to thank FAPESP and the Escola Politécnica, Universidade de São Paulo, for supporting this research.

References

[1] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5:131–295, 1953-54.

[2] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.

[3] D. Denneberg. *Non-Additive Measure and Integral*. Kluwer, Dordrecht, 1994.

[4] D. Dubois and H. Prade. *Possibility Theory*. Plenum Press, New York, 1988.

[5] T. L. Fine. An argument for comparative probability. In *Basic Problems in Methodology and Linguistics*, eds. R. E. Butts and J. Hintikka, 105–119. Reidel, Dordrecht, 1977.

[6] B. de Finetti. *Theory of Probability*, Volumes 1 and 2. Wiley, London, 1974-75.

[7] P. C. Fishburn. A theory of subjective expected utility with vague preferences. *Theory and Decision*, 6:287–310, 1975.

[8] F. J. Giron and S. Ríos. Quasi-Bayesian behaviour: a more realistic approach to decision making? (with discussion) In *Bayesian Statistics 1*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, 17–38 and 49–66. Valencia U. Press, Valencia, 1980.

[9] I. J. Good. Subjective probability as the measure of a non-measurable set. In *Logic, Methodology and Philosophy of Science*, eds. E. Nagel, P. Suppes and A. Tarski, 319–329. Stanford U. Press, California, 1962.

[10] J.-Y. Jaffray. Bayesian updating and belief functions. *IEEE Transactions on Systems, Man, and Cybernetics*, 22:1144–1152, 1992.

[11] J. M. Keynes. *A Treatise on Probability*. Macmillan, London, 1921.

[12] B. O. Koopman. The bases of probability. *Bull. Amer. Math. Soc.*, 46:763–774, 1940.

[13] H. E. Kyburg, Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, Connecticut, 1961.

[14] I. Levi. On indeterminate probabilities. *Journal of Philosophy*, 71:391–418, 1974.

[15] I. Levi. *The Enterprise of Knowledge*. MIT Press, London, 1980.

[16] L. J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.

[17] T. Seidenfeld, M. J. Schervish and J. B. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23:2168–2217, 1995.

[18] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[19] C. A. B. Smith. Consistency in statistical inference and decision (with discussion). *Journal of the Royal Statistical Society*, Series B, 23:1–37, 1961.

[20] P. Walley. Coherent lower (and upper) probabilities. *Research Report* 22. Department of Statistics, University of Warwick, Coventry, 1981.

[21] P. Walley. Rationality and Vagueness. Unpublished manuscript, 1984-85.

[22] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[23] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.

[24] P. Walley. Coherent upper and lower previsions. http://ensmain.rug.ac.be/∼ipp/documentation/ *The Imprecise Probabilities Project*, 1997.

[25] P. Walley. Statistical inferences based on a second-order possibility distribution. *International Journal of General Systems*, 26:337–383, 1997.

[26] P. Walley and G. de Cooman. A behavioural model for linguistic uncertainty. To appear in *Computing with Words*, ed. P. Wang. Wiley, New York, 1999.

[27] Z. Wang and G. J. Klir. *Fuzzy Measure Theory*. Plenum Press, New York, 1992.

[28] L. A. Wasserman and J. B. Kadane. Bayes' theorem for Choquet capacities. *Annals of Statistics*, 18:1328–1339, 1990.

[29] P. Whittle. *Probability*. Penguin, Baltimore, 1970.

[30] P. M. Williams. Coherence, strict coherence and zero probabilities. In *Fifth Int. Congress of Logic, Methodology and Philosophy of Science*, VI-29:30, 1975.

[31] P. M. Williams. Indeterminate probabilities. In *Formal Methods in the Methodology of Empirical Sciences*, eds. M. Przelecki, K. Szaniawski and R. Wojcicki, 229–246. Reidel, Dordrecht, 1976.

[32] N. Wilson and S. Moral. A logical view of probability. In *Proceedings of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, ed. A. Cohn, 386–390. Wiley, London, 1994.

[33] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.