# A survey of the theory of coherent lower previsions

Enrique Miranda*

**Abstract**

This paper presents a summary of Peter Walley's theory of coherent lower previsions. We introduce three representations of coherent assessments: coherent lower and upper previsions, closed and convex sets of linear previsions, and sets of desirable gambles. We show also how the notion of coherence can be used to update our beliefs with new information, and a number of possibilities to model the notion of independence with coherent lower previsions. Next, we comment on the connection with other approaches in the literature: de Finetti's and Williams' earlier work, Kuznetsov's and Weischelberger's work on interval-valued probabilities, Dempster-Shafer theory of evidence and Shafer and Vovk's game-theoretic approach. Finally, we present a brief survey of some applications and summarize the main strengths and challenges of the theory.

**Keywords.** Subjective probability, imprecision, avoiding sure loss, coherence, desirability, conditional lower previsions, independence.

## 1 Introduction

This paper aims at presenting the main facts about the theory of coherent lower previsions. This theory falls within the subjective approach to probability, where the probability of an event represents our information about how likely is this event to happen. This interpretation of probability is mostly used in the framework of decision making, and is sometimes referred to as *epistemic* probability [30, 43].

Subjective probabilities can be given a number of different interpretations. One of them is the *behavioural* one we consider in this paper: the probability of an event is interpreted in terms of some behaviour that depends on the appearance of the event, for instance as betting rates on or against the event, or buying and selling prices on the event.

The main goal of the theory of coherent lower previsions is to provide a number of rationality criteria for reasoning with subjective probabilities. By *reasoning* we shall mean here the cognitive process aimed at solving problems, reaching conclusions or making decisions. The *rationality* of some reasoning can be characterised by means of

*Rey Juan Carlos University, Dept. of Statistics and Operations Research. C-Tulipán, s/n. 28933 Móstoles (Spain). e-mail: enrique.miranda@urjc.es

a number of principles and standards that determine the quality of the reasoning. In the case of probabilistic reasoning, one can consider two different types of rationality. The *internal* one, which studies to which extent our model is self-consistent, is modelled in Walley's theory using the notion of coherence. But there is also an *external* part of rationality, which studies whether our model is consistent with the available evidence. The allowance for imprecision in Walley's theory is related to this type of rationality.

Within the subjective approach to probability, there are two problems that may affect the estimation of the probability of an event: those of indeterminacy and incompleteness. Indeterminacy, also called *indeterminate uncertainty* in [34], happens when there exist events which are not equivalent for our subject, but for which he has no preference, meaning that he cannot decide if he would bet on one over the other. It can be due for instance to lack of information, conflicting beliefs, or conflicting information. On the other hand, incompleteness is due to difficulties in the elicitation of the model, meaning that our subject may not be capable of estimating the subjective probability of an event with an arbitrary degree of precision. This can be caused by a lack of introspection or a lack of assessment strategies, or to the limits of the computational abilities. Both indeterminacy and incompleteness are a source of imprecision in probability models.

One of the first to talk about the presence of imprecision when modelling uncertainty was Keynes in [33], although there were already some comments about it in earlier works by Bernoulli and Lambert [58]. Keynes considered an ordering between the probability of the different outcomes of an experiment which need only be partial. His ideas were later formalized by Koopman in [35, 36, 37]. Other work in this direction was made by Borel [4], Smith [61], Good [31], Kyburg [43] and Levi [45]. In 1975, Williams [85] made a first attempt to make a detailed study of imprecise subjective probability theory, based on the work that de Finetti had done on subjective probability [21, 23] and considering lower and upper previsions instead of precise previsions. This was developed in much more detail by Walley in [71], who established the arguably more mature theory that we shall survey here.

The terms *indeterminate* or *imprecise* probabilities are used in the literature to refer to any model using upper or lower probabilities on some domain, i.e., for a model where the assumption of the existence of a precise and additive probability model is dropped. In this sense, we can consider credal sets [45], 2- and $n$-monotone set functions [5], possibility measures [15, 27, 88], p-boxes [29], fuzzy measures [26, 32], etc. Our focus here is on what we shall call the *behavioural* theory of coherent lower previsions, as developed by Peter Walley in [71]. We are interested in this model mainly for two reasons: from a mathematical point of view, it subsumes most of the other models in the literature as particular cases, having therefore a unifying character. On the other hand, it also has a clear interpretation in terms of acceptable transactions. This interpretation lends itself naturally to decision making [71, Section 3.9]. Our aim in this paper is to give a gentle introduction to the theory for the reader who first approaches the field, and to serve him as a guide on his way through. However, this work by no means pretends to be an alternative to Walley's book, and we refer to [71] for a much more detailed account of the properties we shall present and for a thorough justification of the theory.

In order to ease the transition for the interested reader from this survey to Walley's book, let us give a short outline of the different chapters of the book. The book starts with an introduction to reasoning and behavior in Chapter 1. Chapter 2 introduces coherent lower and upper previsions, and studies their main properties. Chapter 3 shows how to coherently extend our assessments, through the notion of natural extension, and provides also the expression in terms of sets of linear previsions or almost-desirable gambles. Chapter 4 discusses the assessment and the elicitation of imprecise probabilities. Chapter 5 studies the different sources of imprecision in probabilities, investigates the adequacy of precise models for cases of complete ignorance and comments on other imprecise probability models. The study of conditional lower and upper previsions starts in Chapter 6, with the definition of separate coherence and the coherence of an unconditional and a conditional lower prevision. This is generalized in Chapter 7 to the case of a finite number of conditional lower previsions, focusing on a number of statistical models. Chapter 8 establishes a general theory of natural extension of several coherent conditional previsions. Finally, Chapter 9 is devoted to the modelling of the notion of independence.

In this paper, we shall summarize the main aspects of this book and the relationships between Walley's theory of coherent lower previsions and some other approaches to imprecise probabilities. The paper is structured as follows: in Section 2, we present the main features of unconditional coherent lower previsions. We give three representations of the available information: coherent lower and upper previsions, sets of desirable gambles, and sets of linear previsions, and show how to extend the assessments to larger domains. In Section 3, we outline how we can use the theory of coherent lower previsions to update the assessments with new information, and how to combine information from different sources. Thus, we make a study of conditional lower previsions. Section 4 is devoted to the notion of independence. In Section 5, we compare Walley's theory with other approaches to subjective probability: the seminal work of de Finetti, first generalized to the imprecise case by Williams, Kuznetsov's and Weischelberger's work on interval-valued probabilities, the Dempster-Shafer theory of evidence, and Shafer and Vovk's recent work on game-theoretic probability. In Section 6, we review a number of applications. We conclude the paper in Section 7 with an overview of some questions and remaining challenges in the field.

# 2 Coherent lower previsions and other equivalent representations

In this section, we present the main facts about coherent lower previsions, their behavioural interpretation and the notion of natural extension. We show that the information provided by a coherent lower prevision can also be expressed by means of a set of linear previsions or by a set of desirable gambles. Although this last approach is arguably better suited to understanding the ideas behind the behavioural interpretation, we have opted for starting with the notion of coherent lower previsions, because this will help to understand the differences with classical probability theory and moreover they will be the ones we use when talking about conditional lower previsions in Section 3. We refer to [20] for an alternative introduction.

## 2.1 Coherent lower previsions

Consider a non-empty space $\Omega$, representing the set of outcomes of an experiment. The behavioural theory of imprecise probabilities provides tools to model our information about the likelihood of the different outcomes in terms of our betting behavior on some gambles that depend on these outcomes. Specifically, a gamble $f$ on $\Omega$ is a bounded real-valued function on $\Omega$. It represents an uncertain reward, meaning that we obtain the price $f(\omega)$ if the outcome of the experiment is $\omega \in \Omega$. This reward is expressed in units of some linear utility scale, see [71, Section 2.2] for details. We shall denote by $\mathscr{L}(\Omega)$ the set of gambles on $\Omega$.[1] A particular case of gambles are the indicators of events, that is, the gambles $I_A$ defined by $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ otherwise for some $A \subseteq \Omega$.

*Example* 1. Jack has made it to the final of a TV contest. He has already won 50000 euros,[2] and he can add to this the amount he gets by playing with *The Magic Urn*. He must draw a ball from the urn, and he wins or loses money depending on its color: if he draws a green ball, he gets 10000 euros; if he draws a red ball, he gets 5000 euros; and if he draws a black ball, he gets nothing. Mathematically, the set of outcomes of the experiment Jack is going to make (drawing a ball from the urn) is $\Omega = \{green, red, black\}$, and the gamble $f_1$ which determines his prize is given by $f_1(green) = 10000, f_1(red) = 5000, f_1(black) = 0.$♦

Let $\mathscr{K}$ be a set of gambles on $\Omega$. A *lower prevision* on $\mathscr{K}$ is a functional $\underline{P} : \mathscr{K} \to \mathbb{R}$. For any gamble $f$ in $\mathscr{K}$, $\underline{P}(f)$ represents a subject's supremum acceptable buying price for $f$; this means that he is disposed to pay $\underline{P}(f) - \varepsilon$ for the uncertain reward determined by $f$ and the outcome of the experiment, or, in other words, that the transaction $f - \underline{P}(f) + \varepsilon$, understood as a point-wise operation, is acceptable to him for every $\varepsilon > 0$ (however, nothing is said about whether he would buy $f$ for the price $\underline{P}(f)$).[3]

Given a gamble $f$ we can also consider our subject's infimum acceptable *selling* price for $f$, which we shall denote by $\overline{P}(f)$. It means that the transaction $\overline{P}(f) + \varepsilon - f$ is acceptable to him for every $\varepsilon > 0$ (but nothing is said about the transaction $\overline{P}(f) - f$). We obtain in this way an *upper prevision* $\overline{P}(f)$ on some set of gambles $\mathscr{K}'$.

We shall see in Section 2.3 an equivalent formulation of lower and upper previsions in terms of sets of desirable gambles. For the time being, and in order to justify the rationality requirements we shall introduce, we shall only assume the following:

1. A transaction that makes our subject lose utiles, no matter the outcome of the experiment, is not acceptable for him.

[1] Although Walley's theory assumes the variables involved are bounded, the theory has also been generalized to unbounded random variables in [63, 62]. The related formulation of the theory from a game-theoretic point of view, as we will present it in Section 5.6, has also been made for arbitrary gambles. See also Section 5.3.

[2] Such an amount of money is only added in order to justify the linearity of the utility scale, which in the case of money only holds if the amounts at stake are small compared to the subject's capital.

[3] In this paper, we use Walley's notation of $\underline{P}$ for lower previsions and $\overline{P}$ for upper previsions; these can be seen as lower and upper expectations, and will only be interpreted as lower and upper probabilities when the gamble $f$ is the indicator of some event.

2. If he considers acceptable a transaction, he should also accept any other transaction that gives him a greater reward, no matter the outcome of the experiment.

3. A positive linear combination of acceptable transactions should also be acceptable.

*Example 1(cont.).* If Jack pays $x$ euros in order to play at *The Magic Urn*, then the increase in his wealth is given by $10000-x$ euros if he draws a green ball, $5000-x$ euros if he draws a red ball, and of $-x$ euros if he draws a black ball (i.e., he loses $x$ euros in that case). The supremum amount of money that he is disposed to pay will be his lower prevision for the gamble $f_1$. If for instance he is certain that there are no black balls in the urn, he should be disposed to pay as much as 5000 euros, because his wealth is not going to decrease, no matter which color is the ball he draws. And if he has no additional information about the composition of the urn and wants to be cautious, he will not pay more than 5000 euros, because it could happen that all the balls in the urn are red, and by paying more than 5000 euros he would end up always losing money.

On the other hand, and after Jack has drawn a ball from the urn, and before he sees the color, he can sell the unknown prize attached to it to the 2nd ranked player (Kate) for $y$ euros. If Jack does this, his increase in wealth will be $y-10000$ euros if the ball is green, $y-5000$ euros if the ball is red, and $y$ euros if the ball is black. The minimum amount of money that he requires in order to sell it will be his upper prevision for the gamble $f_1$. If he knows that there are only red and green balls in the urn, he should accept to sell it for more than 10000 euros, because he is not going to lose money by this, but not for less: it could be that all the balls are green and then by selling it for less that 10000 euros he would end up losing money, no matter the outcome. ♦

Since by selling a gamble $f$ for a price $\mu$ or alternatively by buying the gamble $-f$ for the price $-\mu$ our subject increases his wealth in $\mu - f$ in both cases, it may be argued that he should be disposed to accept these transactions under the same conditions. Hence, his infimum acceptable selling price for $f$ should agree with the opposite of his supremum acceptable buying price for $-f$. As a consequence, given a lower prevision on a set of gambles $\mathscr{K}$ we can define an upper prevision $\overline{P}$ on $-\mathscr{K} := \{-f : f \in \mathscr{K}\}$, by $\overline{P}(-f) = -\underline{P}(f)$, and vice versa. Taking this into account, all the developments for lower previsions can also be done for upper previsions, and vice versa. We shall concentrate in this survey on lower previsions.

Lower previsions are subject to a number of rationality criteria, which assure the consistency of the judgements they represent. First of all, a positive linear combination of a number of acceptable gambles should never result in a transaction that makes our subject lose utiles, no matter the outcome. This is modeled through the notion of *avoiding sure loss*: for every natural number $n \geq 0$ and $f_1, \ldots, f_n$ in $\mathscr{K}$, we should have

$$\sup_{\omega \in \Omega} \sum_{i=1}^{n} [f_i(\omega) - \underline{P}(f_i)] \geq 0.$$

Otherwise, there would exist some $\delta > 0$ such that $\sum_{i=1}^{n} [f_i(\omega) - (\underline{P}(f_i) - \delta)] \leq -\delta$, meaning that the sum of acceptable transactions $[f_i - (\underline{P}(f_i) - \delta)]$ results in a loss of at least $\delta$, no matter the outcome of the experiment.

*Example 1(cont.).* Jack decides to pay 5000 euros in order to play at *The Magic Urn*. The TV host offers him also another prize depending on the same outcome: he will win 9000 euros if he draws a black ball, 5000 if he draws a red ball and nothing if he draws a green ball. Let $f_2$ denote this other prize. Jack decides to pay additionally as much as 5500 euros to also get this other prize. But then he is paying 10500 euros, and total prize he is going to win is 10000 euros with a green or a red ball, and 9000 euros with a black ball. Hence, he is going to lose money, no matter the outcome of the experiment. Mathematically, this means that the assessments $\underline{P}(f_1) = 5000$ and $\underline{P}(f_2) = 5500$ incur a sure loss. ♦

But there is a stronger requirement, namely that of *coherence*. Coherence means that our subject's supremum acceptable buying price for a gamble $f$ should not be raised by considering a positive linear combination of a finite number of other acceptable gambles. Formally, this means that for every non-negative integers $n$ and $m$ and gambles $f_0, f_1, \dots, f_n$ in $\mathscr{K}$, we have

$$\sup_{\omega\in\Omega} \sum_{i=1}^{n} [f_i(\omega) - \underline{P}(f_i)] - m[f_0(\omega) - \underline{P}(f_0)] \geq 0. \tag{1}$$

A lower prevision satisfying this condition will in particular avoid sure loss, by considering the particular case of $m = 0$.

Let us suppose that Equation (1) does not hold for some non-negative integers $n, m$ and some $f_0, \dots, f_n$ in $\mathscr{K}$. If $m = 0$, this means that $\underline{P}$ incurs in a sure loss, which we have already argued is an inconsistency. Assume then that $m > 0$. Then there exists some $\delta > 0$ such that

$$\sum_{i=1}^{n} [f_i(\omega) - (\underline{P}(f_i) - \delta)] \leq m[f_0(\omega) - (\underline{P}(\omega) + \delta)]$$

for all $\omega \in \Omega$. The left-hand side is a positive linear combination of acceptable buying transactions, which should then be acceptable. The right-hand side, which gives a bigger reward than this acceptable transaction, should be acceptable as well. But this means that our subject should be disposed to buy the gamble $f_0$ at the price $\underline{P}(f_0) + \delta$, which is greater than the supremum buying price he established before. This is an inconsistency.

*Example 1(cont.).* After thinking again, Jack decides to pay as much as 5000 euros for the first prize ($f_1$) and 4000 euros for the second ($f_2$). He also decides that he will sell this second prize to Kate for anything bigger than 6000 euros, but not for less. In the language of lower and upper previsions, this means that $\underline{P}(f_1) = 5000, \underline{P}(f_2) = 4000, \overline{P}(f_2) = 6000$, or, equivalently, $\underline{P}(-f_2) = -6000$. The rise in his wealth is given by the following table:

| | green | red | black |
|---|---|---|---|
| Buy $f_1$ for 5000 | 5000 | 0 | $-5000$ |
| Buy $f_2$ for 4000 | $-4000$ | 1000 | 5000 |
| Sell $f_2$ for 6000 | 6000 | 1000 | $-3000$ |

Table 1: Increase on Jack's wealth depending on the color of the ball he draws.

These assessments avoid sure loss. However, the first one already implies that he should sell $f_2$ for anything bigger than 5000 euros: if he does so, the increase in his wealth is 5000 euros with a green ball, 0 with a red ball, and $-4000$ with a black ball. This situation is better than the one produced by buying $f_1$ for 5000 euros, which he considered acceptable. This is an inconsistency with his assessment of 6000 euros as infimum acceptable selling price for $f_2$. ♦

When the domain $\mathscr{K}$ of a lower prevision is a *linear space*, i.e., closed under addition of gambles and multiplication of gambles by real numbers, coherence takes a simpler form. It can be checked that in that case $\underline{P}$ is coherent if and only if the following three axioms hold for all $f,g$ in $\mathscr{K}$ and all $\lambda > 0$:

(P1) $\underline{P}(f) \geq \inf f$ [accepting sure gains].

(P2) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [positive homogeneity].

(P3) $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ [superlinearity].

A consequence of these axioms is that a convex combination, a lower envelope and a point-wise limit of coherent lower previsions is again a coherent lower prevision.

A coherent lower prevision defined on indicators of events only is called a *coherent lower probability*. The lower probability of an event $A$ can also be seen as our subject's supremum betting rate *on* the event, where betting on $A$ means that he gets 1 if $A$ appears and 0 if it does not; similarly, the upper probability of an event $A$ can be interpreted as our subject infimum betting rate *against* the event $A$. Here, in contradistinction with the usual approaches to classical probability theory, we start from previsions (of gambles) and deduce the probabilities (of events), instead of going from probabilities to previsions using some expectation operator (but see also [83] for a similar approach in the case of precise probabilities).

As Walley himself argues in [71, Section 2.11], the notion of coherence can be considered too weak to fully characterize the rationality of probabilistic reasoning. Indeed, other additional requirements could be added in order to achieve this: we may for instance use symmetry or independent principles, or the principle of direct inference discussed in [71, Section 1.7.2]. In this sense, most of the notions of upper and lower probabilities considered in the literature (2- and $n$-monotone capacities, belief functions, necessity measures) are particular cases of coherent lower previsions. Nevertheless, the main point here is that coherence is a necessary condition for rationality, and that we should at least require our subject's assessments to satisfy it. Other possible rationality axioms, such as the notion of conglomerability, will be discussed later in this paper.

The assessments expressed by means of a lower prevision can also be made in terms of two alternative representations: sets of linear previsions and sets of almost-desirable gambles.

## 2.2 Linear previsions

When the supremum buying price and the infimum selling price for a gamble $f$ coincide, then the common value $P(f) := \underline{P}(f) = \overline{P}(f)$ is called a *fair price* or *prevision* for $f$. More generally, a real-valued function defined on a set of gambles $\mathscr{K}$ is called a *linear prevision* when for all natural numbers $m,n$ and gambles $f_1,\dots,f_m,g_1,\dots,g_n$ in the domain,

$$\sup_{\omega\in\Omega}[\sum_{i=1}^{m}[f_i(\omega)-P(f_i)]-\sum_{j=1}^{n}[g_j(\omega)-P(g_j)]]\geq 0. \tag{2}$$

Assume that this condition does not hold for certain gambles $f_1,\dots,f_m,g_1,\dots,g_n$ in $\mathscr{K}$. Then there exists some $\delta>0$ such that

$$\sum_{i=1}^{m}[f_i(\omega)-P(f_i)+\delta]+\sum_{j=1}^{n}[P(g_j)-g_j+\delta]<-\delta \tag{3}$$

for all $\omega\in\Omega$. Since $P(f_i)$ is our subject's supremum acceptable buying price for the gamble $f_i$, he is disposed to pay $P(f_i)-\delta$ for it, so the transaction $f_i-P(f_i)+\delta$ is acceptable for him; on the other hand, since $P(g_j)$ is his infimum acceptable selling price for $g_j$, he is disposed to sell it for the price $P(g_j)+\delta$, so the transaction $P(g_j)+\delta-g_j$ is acceptable. But Equation (3) tells us that the sum of these acceptable transactions produces a loss of at least $\delta$, no matter the outcome of the experiment!

A linear prevision $P$ is coherent, both when interpreted as a lower and as an upper prevision; the former means that $\underline{P}:=P$ is a coherent lower prevision on $\mathscr{K}$, the latter that the functional $\underline{P}_1$ on $-\mathscr{K}:=\{-f : f\in\mathscr{K}\}$ given by $\underline{P}_1(f)=-P(-f)$ is a coherent lower prevision on $-\mathscr{K}$. However, not every functional which is coherent both as a lower and as an upper prevision is a linear prevision. This is because coherence as a lower prevision only guarantees that Equation (2) holds for $n\leq 1$, and coherence as an upper prevision only guarantees that the same equation holds for the case where $m\leq 1$. An example showing that these two properties do not imply that Equation (2) holds for all non-negative natural numbers $n,m$ is the following:

*Example* 2. Let us consider a five element space $\Omega:=\{x_1,x_2,x_3,x_4,x_5\}$, and a functional $P$ defined on a subset $\{g_1,g_2,g_3,g_4\}$ of $\mathscr{L}(\Omega)$ such that the gambles $g_i$ are given by:

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|
| $g_1$ | 0 | 0 | 1 | 1 | -2 |
| $g_2$ | 0 | 0 | 0 | -2 | 1 |
| $g_3$ | 1 | 0 | 2 | 0 | 0 |
| $g_4$ | 0 | 1 | 0 | 0 | 0 |

Consider $P$ given by $P(g_i) = 0$ for $i = 1, \dots, 4$. Then it can be checked that $P$ is coherent both as a lower and as an upper prevision. However, the gamble $g_1 + g_2 - g_3 - g_4$ has the constant value $-1$, and this implies that Equation (2) does not hold. Therefore $P$ is not a linear prevision. ⧫

As it was the case with the coherence condition for lower previsions, when the domain $\mathscr{K}$ satisfies some additional conditions, then Equation (2) can be simplified. For instance, if the domain $\mathscr{K}$ is self-conjugate, meaning that $-\mathscr{K} = \mathscr{K}$, then $P$ is a linear prevision if and only if it avoids sure loss and satisfies $P(f) = -P(-f)$ for all $f \in \mathscr{K}$. On the other hand, when $\mathscr{K}$ is a linear space of gambles, then $P$ is a linear prevision if and only if

(P0) $P(f+g) = P(f) + P(g)$ for all $f, g \in \mathscr{K}$.

(P1) $P(f) \geq \inf f$ for all $f \in \mathscr{K}$.

It follows from these two conditions that $P$ also satisfies

(P2) $P(\lambda f) = \lambda P(f)$ for all $f \in \mathscr{K}, \lambda > 0$.

Hence, when the domain is a linear space of gambles linear previsions are coherent lower previsions which are additive instead of simply super-additive.

A linear prevision on a domain $\mathscr{K}$ is always the restriction of a linear prevision on all gambles, which is, by conditions (P0) and (P1), a linear functional on $\mathscr{L}(\Omega)$. We shall denote by $\mathbb{P}(\Omega)$ the set of linear previsions on $\Omega$.

Given a linear prevision $P$ on all gambles, i.e., an element of $\mathbb{P}(\Omega)$, we can consider its restriction $Q$ to the set of indicators of events. This restriction can be seen in particular as a functional on the class $\mathscr{P}(\Omega)$ of subsets of $\Omega$, using the identification $Q(A) = P(I_A)$. It can be checked that this functional is a finitely additive probability, and $P$ is simply the expectation with respect to $Q$. Hence, in case of linear previsions there is no difference in expressive power between representing the information in terms of events or in terms of gambles: the restriction to events (the probability) determines the value on gambles (the expectation) and vice versa. This is no longer true in the imprecise case, where there usually are infinitely many coherent extensions of a coherent lower probability [71, Section 2.7.3], and this is why the theory is formulated in general in terms of gambles. The fact that lower previsions of events do not determine uniquely the lower previsions of gambles is due to the fact that we are not dealing with additive functionals anymore.

A linear prevision $P$ whose domain is made up of the indicators of the events in some class $\mathscr{A}$ is called an *additive probability* on $\mathscr{A}$. If in particular $\mathscr{A}$ is a field of events, then $P$ is a finitely additive probability in the usual sense, and moreover condition (2) simplifies to the usual axioms of finite additivity:

(a) $P(A) \geq 0$ for all $A$ in $\mathscr{A}$.

(b) $P(\Omega) = 1$.

(c) $P(A \cup B) = P(A) + P(B)$ whenever $A \cap B = \emptyset$.

*Example 1(cont.).* Assume that Jack knows that there are only 10 balls in the urn, and that the drawing is fair, so that the probability of each color depends only on the proportion of balls of that color. If he knew the exact composition of the urn, for instance that there are 5 green balls, 4 red balls and 1 black ball, then his expected gain with the gamble $f_1$ would be $0.5 * 10000 + 0.4 * 5000 - 0.1 * 0 = 7000$ euros, and this should be his fair price for the gamble. Any linear prevision will be determined by its restriction to events via the expectation operator. This restriction to events corresponds to some particular composition of the urn: if he knows that there are 4 red balls out of 10 in the urn, then his betting rate on or against drawing a red ball (that is, his fair price for a gamble with reward 1 if he draws a red ball and 0 if he does not) should be 0.4. ⧫

We can characterize the coherence of a lower prevision $\underline{P}$ with domain $\mathscr{K}$ by means of its set of *dominating* linear previsions, which we shall denote as

$$\mathscr{M}(\underline{P}) := \{P \in \mathbb{P}(\Omega) : P(f) \geq \underline{P}(f) \text{ for all } f \text{ in } \mathscr{K}\}.$$

It can be checked that $\underline{P}$ is coherent if and only if it is the lower envelope of $\mathscr{M}(\underline{P})$, that is, if and only if

$$\underline{P}(f) = \min\{P(f) : P \in \mathscr{M}(\underline{P})\}$$

for all $f$ in $\mathscr{K}$. Moreover, the set of dominating linear previsions allows us to establish a one-to-one correspondence between coherent lower previsions $\underline{P}$ and closed (in the weak-* topology) and convex sets of linear previsions: given a coherent lower prevision $\underline{P}$ we can consider the (closed and convex) set of linear previsions $\mathscr{M}(\underline{P})$. Conversely, every closed and convex set $\mathscr{M}$ of linear previsions determines uniquely a coherent lower prevision $\underline{P}$ by taking lower envelopes. [4] Besides, it can be checked that these two operations (taking lower envelopes of closed convex sets of linear previsions and considering the linear previsions that dominate a given coherent lower prevision) commute. We should warn the reader, however, that this equivalence does not hold in general for the *conditional* lower previsions that we shall introduce in Section 3, although there exists an envelope result for the alternative approach developed by Williams that we shall present in Section 5.2.

The representation of coherent lower previsions in terms of sets of linear previsions allows us to give them a *Bayesian sensitivity analysis* representation: we may assume that there is a fair price for every gamble $f$ on $\Omega$, which results on a linear prevision $P \in \mathbb{P}(\Omega)$, and such that our subject's imperfect knowledge of $P$ only allows him to place it among a set of possible candidates, $\mathscr{M}$. Then the inferences he can make from $\mathscr{M}$ are equivalent to the ones he can make using the lower envelope $\underline{P}$ of this set. This lower envelope is a coherent lower prevision.

Hence, all the developments we make with (unconditional) coherent lower previsions can also be made with sets of finitely additive probabilities, or equivalently with the set of their associated expectation operators, which are linear previsions. In this sense, there is a strong link between this theory and robust Bayesian analysis [53]. We

[4] Remark nonetheless that convexity is not really an issue here, since a set of linear previsions represents the same behavioural dispositions (has the same lower and upper envelopes) as its convex hull, and in many cases, the set of linear previsions compatible with some beliefs will not be convex; see [10, 13] for more details, and [55] for a more critical approach to the assumption of convexity in the context of decision making.

shall see in Section 3 that, roughly speaking, this link only holds when we update our information if we deal with finite sets of categories.

*Example 1(cont.).* The lower previsions $\underline{P}(f_1) = 5000$ and $\underline{P}(f_2) = 4000$ that Jack established before for the gambles $f_1, f_2$ are coherent. The information $\underline{P}$ gives is equivalent to its set of dominating linear previsions, $\mathscr{M}(\underline{P})$. Any of these linear previsions is characterized by its restriction to events, which gives the probability of drawing a ball of some particular color. In this case, it can be checked that

$$\mathscr{M}(\underline{P}) := \{(p_1, p_2, p_3) : p_1 \geq p_3, p_2 \geq 4(p_1 - p_3) - p_3\},$$

where $p_1, p_2$ and $p_3$ denote the proportions of green, red and black balls in the urn, respectively. This means that his beliefs about the composition of the urn are that there are at least as many green as black balls, and that the number of red and black balls is at least four times the difference of green and black balls.

We see in this example that the actual set of possible compositions of the urns which are compatible with these beliefs determines the following set of compatible previsions (remember that we are assuming that there are 10 balls in the urn and that the drawing is fair):

| P( Green ) | P( Red ) | P( Black ) |
|---|---|---|
| 0 | 1 | 0 |
| 0.1 | 0.9 | 0 |
| 0.1 | 0.8 | 0.1 |
| 0.2 | 0.8 | 0 |
| 0.2 | 0.7 | 0.1 |
| 0.2 | 0.6 | 0.2 |
| 0.3 | 0.5 | 0.2 |
| 0.3 | 0.4 | 0.3 |
| 0.4 | 0.3 | 0.3 |
| 0.4 | 0.2 | 0.4 |
| 0.5 | 0.1 | 0.4 |
| 0.5 | 0 | 0.5 |

Table 2: The previsions compatible with Jack's beliefs about the composition of the urn and the extraction method.

This set is finite, and therefore non-convex; it provides nonetheless the same behavioural information as its convex hull $\mathscr{M}(\underline{P})$. It is also interesting that, as remarked by one of the referees, not all the probabilities in this table correspond to extreme points of $\mathscr{M}(\underline{P})$: for instance $(0.4, 0.2, 0.4) = 0.8 * (0.5, 0, 0.5) + 0.2 * (0, 1, 0)$. ♦

We will comment further on linear previsions when we discuss the connection with de Finetti's work, in Section 5.1.

## 2.3 Sets of desirable gambles

An alternative, more direct, representation of the behavioural assessments expressed by a lower prevision can be made by sets of *almost-desirable gambles*. A gamble $f$ is said to be almost-desirable when our subject is almost disposed to accept it, meaning that he considers the gamble $f+\varepsilon$ to be desirable for every $\varepsilon>0$ (although nothing is said about the desirability of $f$).

Following the behavioural interpretation we gave in Section 2.1 to lower previsions, we see that the gamble $f-\underline{P}(f)$ must be almost-desirable for our subject: if $\underline{P}(f)$ is his supremum acceptable buying price, then he must accept the gamble $f-\underline{P}(f)+\varepsilon$, which means paying $\underline{P}(f)-\varepsilon$ for the gamble $f$, for every $\varepsilon>0$. The lower prevision of $f$ is the maximum value of $\mu$ such that the gamble $f-\mu$ is almost-desirable to our subject.

We can also express the requirement of coherence in terms of sets of almost-desirable gambles. Assume that, out of the gambles in $\mathscr{L}(\Omega)$, our subject judges the gambles in $\mathscr{D}$ to be almost-desirable. His assessments are coherent, and we say that $\mathscr{D}$ is coherent, if and only if the following conditions hold:

(D0) If $f\in\mathscr{D}$, then $\sup f\geq 0$ [avoiding sure loss].

(D1) If $\inf f>0$, then $f\in\mathscr{D}$ [accepting sure gains].

(D2) If $f\in\mathscr{D},\lambda>0$, then $\lambda f\in\mathscr{D}$ [positive homogeneity].

(D3) If $f,g\in\mathscr{D}$, then $f+g\in\mathscr{D}$ [addition].

(D4) If $f+\delta\in\mathscr{D}$ for all $\delta>0$, then $f\in\mathscr{D}$ [closure].

Let us give an interpretation of these axioms. D0 states that a gamble $f$ which makes him lose utiles, no matter the outcome, should not be acceptable for our subject. D1 on the other hand tells that he should accept a gamble which will always increase his wealth. D2 means that the almost-desirability of a gamble $f$ should not depend on the scale of utility we are considering. D3 states that if two gambles $f$ and $g$ are almost-desirable, our subject should be disposed to accept their combined transaction. Axioms D0-D3 characterize the so-called *desirable* gambles (see [71, Section 2.2.4]). A class of desirable gambles is always a convex cone of gambles. Axiom D4 is a closure property, and allows us to give an interpretation in terms of *almost*-desirability.

It is a consequence of axioms (D1) and (D4) that any non-negative gamble $f$ is almost-desirable. From this and (D3), we can deduce that if a gamble $f$ dominates an almost-desirable gamble $g$, then $f$ (which is the sum of the almost-desirable gambles $g$ and $f-g$) is also almost-desirable. Finally, it follows from (D2) and (D3) that a positive linear combination of almost-desirable gambles is also almost-desirable. These are the rationality requirements we have used in Section 2.1 to justify the notion of coherence.

A coherent set of almost-desirable gambles is a closed convex cone containing all non-negative gambles and not containing any uniformly negative gambles. There is, moreover, a one-to-one correspondence between coherent lower previsions on $\mathscr{L}(\Omega)$ and coherent sets of almost-desirable gambles: given a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\Omega)$, the class

$$\mathscr{D}_{\underline{P}}:=\{f\in\mathscr{L}(\Omega):\underline{P}(f)\geq 0\} \tag{4}$$

is a coherent set of almost-desirable gambles. Conversely, if $\mathscr{D}$ is a coherent set of almost-desirable gambles, the lower prevision $\underline{P}_{\mathscr{D}}$ given by

$$\underline{P}_{\mathscr{D}}(f):=\max\{\mu:f-\mu\in\mathscr{D}\} \tag{5}$$

is coherent. Moreover, it can be checked that the operations given in Equations (4) and (5) commute, meaning that if we consider a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\Omega)$ and the set $\mathscr{D}_{\underline{P}}$ of almost-desirable gambles that we can obtain from it via Equation (4), then the lower prevision $\underline{P}_{\mathscr{D}_{\underline{P}}}$ that we define on $\mathscr{L}(\Omega)$ by Equation (5) coincides with $\underline{P}$; and similarly if we start with a set of almost-desirable gambles.

Since the assessments expressed by means of a coherent lower prevision $\underline{P}$ can also be expressed by means of its set of dominating linear previsions $\mathscr{M}(\underline{P})$, we see that this set is also equivalent to the set of almost-desirable gambles $\mathscr{D}$ we have just derived. Indeed, given a coherent set of almost-desirable gambles $\mathscr{D}$, we can consider the set of linear previsions

$$\mathscr{M}_{\mathscr{D}} := \{P \in \mathbb{P}(\Omega) : P(f) \geq 0 \text{ for all } f \text{ in } \mathscr{D}\}.$$

$\mathscr{M}_{\mathscr{D}}$ is a closed and convex set of linear previsions, and its lower envelope $\underline{P}$ coincides with the coherent lower prevision induced by $\mathscr{D}$ through Equation (5).

Conversely, given a closed and convex set of linear previsions $\mathscr{M}$, we can consider the set of gambles

$$\mathscr{D}_{\mathscr{M}} = \{f \in \mathscr{L}(\Omega) : P(f) \geq 0 \text{ for all } P \text{ in } \mathscr{M}\}.$$

$\mathscr{D}_{\mathscr{M}}$ is a coherent set of almost-desirable gambles, and the lower prevision $\underline{P}_{\mathscr{D}_{\mathscr{M}}}$ it induces is equal to the lower envelope of $\mathscr{M}$.

*Example 1(cont.).* From the set of linear previsions $\mathscr{M}(\underline{P})$ compatible with Jack's coherent assessments, we obtain the class of almost-desirable gambles

$$\begin{aligned}\mathscr{D} = \{&(a_1,a_2,a_3) : \\ &a_1+a_3 \geq 0,\ a_2 \geq 0,\ a_1 \geq \max\{-4a_2, -3(a_1+a_2+a_3), -a_2-4(a_1+a_3)\}\},\end{aligned}$$

where $a_1 = f(green), a_2 = f(red), a_3 = f(black)$. ♦

Hence, we have three equivalent representations of coherent assessments: coherent lower previsions, closed and convex sets of linear previsions, and coherent sets of almost-desirable gambles. The use of one or another of these representations will depend on the context: for instance, a representation in terms of sets of linear previsions may be more useful if we want to give our model a Bayesian sensitivity analysis interpretation, while the use of sets of almost-desirable gambles may be more interesting in connection when decision making.

Note however that these representations do not tell us anything about our subject's buying behavior for the gambles $f$ at the price $\underline{P}(f)$: he may accept it, as he does for the price $\underline{P}(f) - \varepsilon$ for all $\varepsilon > 0$, but then he also might not. If we want to give information about the behavior for $\underline{P}(f)$, we have to consider a more informative model: sets of *really desirable gambles*. These sets allow to distinguish between desirability and almost desirability, and they solve moreover some of the difficulties we shall see in Section 3 when talking about conditioning on sets of probability zero [77]. We refer to [71, Appendix F], [77] and [20] for a more detailed account of this more general model.

## 2.4 Natural extension

Assume now that our subject has established his acceptable buying prices rates for all gambles on some domain $\mathscr{K}$. He may then wish to check which are the consequences of these assessments for other gambles. If for instance he is disposed to pay a price $\mu_1$ for a gamble $f_1$ and a price $\mu_2$ for a gamble $f_2$, he should be disposed to pay at least the price $\mu_1 + \mu_2$ for their sum $f_1 + f_2$. In general, given a gamble $f$ which is not in the domain, he would like to know which is the supremum buying price that he should find acceptable for $f$, taking into account his previous assessments ($\underline{P}$), and using *only* the condition of coherence.

Assume that for a given price $\mu$ there exist gambles $g_1, \dots, g_n$ in $\mathscr{K}$ and non-negative real numbers $\lambda_1, \dots, \lambda_n$, such that

$$f(\omega) - \mu \geq \sum_{i=1}^{n} \lambda_i (g_i(\omega) - \underline{P}(g_i)).$$

Since all the transactions in the sum of the right-hand side are acceptable to him, so should be the left-hand side, which dominates their sum. Hence, he should be disposed to pay the price $\mu$ for the gamble $f$, and therefore his supremum acceptable buying price should be greater than or equal to $\mu$. To use the language of the previous section, the right-hand side of the inequality is an almost-desirable gamble, and as a consequence so must be the gamble on the left-hand side. And the correspondence between almost-desirable gambles and coherent lower previsions implies then that the lower prevision of $f$ should be greater than, or equal to, $\mu$.

The lower prevision that provides these supremum acceptable buying prices is called the *natural extension* of $\underline{P}$. It is given, for $f \in \mathscr{L}(\Omega)$, by

$$\underline{E}(f) = \sup_{\substack{g_i \in \mathscr{K}, \lambda_i \geq 0, \\ i=1,\dots,n_i, n_i \in \mathbb{N}}} \inf_{\omega \in \Omega} \left\{ f(\omega) - \sum_{i=1}^{n} \lambda_i [g_i(\omega) - \underline{P}(g_i)] \right\}. \tag{6}$$

The reasoning above tells us that our subject should be disposed to pay the price $\underline{E}(f) - \varepsilon$ for the gamble $f$, and this for every $\varepsilon > 0$. Hence, his supremum acceptable buying price should dominate $\underline{E}(f)$. But we can check also that this value is sufficient to achieve coherence ([71, Theorem 3.1.2]).

Therefore, $\underline{E}(f)$ is the smallest, or more conservative, value we can give to the buying price of $f$ in order to achieve coherence with the assessments in $\underline{P}$. There may be other coherent extensions, which may be interesting in some situations; however, any other of these less conservative, coherent extensions will represent stronger assessments than the ones that can be derived purely from $\underline{P}$ and the notion of coherence. This is why we usually adopt $\underline{E}$ as our inferred model.

Coherent lower previsions constitute a very general model for uncertainty: they include for instance as particular cases 2- and $n$-monotone capacities, belief functions, or probability measures. In this style, the procedure of natural extension includes as particular cases many of the extension procedures present in the literature: Choquet integration of 2- and $n$-monotone capacities, Lebesgue integration of probability measures, or Bayes's rule for updating probability measures. It provides the consequences for other gambles of our previous assessments and the notion of coherence. If for instance we consider a probability measure on some $\sigma$-field of events $\mathscr{A}$, the natural extension to all events will determine the set of finitely additive extensions of $P$ to $\mathscr{P}(\Omega)$, and it will be equal to the lower envelope of this set. It coincides moreover with the inner measure of $P$.

*Example 1(cont.).* Jack is offered next a new gamble $f_3$, whose reward is 2000 euros if he draws a green ball, 3000 if he draws a red ball, and 4000 euros if he draws a black ball. Taking into account his previous assessments, $\underline{P}(f_1) = 5000$, $\underline{P}(f_2) = 4000$, he should at least pay

$$\begin{aligned}&\underline{E}(f_3)\\ &= \sup_{\lambda_1,\lambda_2\geq 0} \min\{2000-5000\lambda_1+4000\lambda_2, 3000-1000\lambda_2, 4000+5000\lambda_1-5000\lambda_2\}.\end{aligned}$$

It can be checked that this supremum is achieved for $\lambda_1 = 0, \lambda_2 = 0.2$. We obtain then $\underline{E}(f_3) = 2800$. This is the supremum acceptable buying price for $f_3$ that Jack can derive from his previous assessments and the notion of coherence. ♦

If the lower prevision $\underline{P}$ does not avoid sure loss, then Equation (6) yields $\underline{E}(f) = \infty$ for all $f \in \mathscr{L}(\Omega)$. The idea is that if our subject's initial assessments are such that he can end up losing utiles no matter the outcome of the experiment, he will also lose utiles with any other gamble that they offer to him. Because of this, the first thing we have to verify is whether the initial assessments avoid sure loss, and only then we can consider their consequences on other gambles.

When $\underline{P}$ avoids sure loss, $\underline{E}$ is the smallest coherent lower prevision on all gambles that dominates $\underline{P}$ on $\mathscr{K}$, in the sense that $\underline{E}$ is coherent and any other coherent lower prevision $\underline{E}'$ on $\mathscr{L}(\Omega)$ such that $\underline{E}'(f) \geq \underline{P}(f)$ for all $f \in \mathscr{K}$ will satisfy $\underline{E}'(f) \geq \underline{E}(f)$ for all $f$ in $\mathscr{L}(\Omega)$. $\underline{E}$ is not in general an extension of $\underline{P}$; it will only be so when $\underline{P}$ is coherent itself. Otherwise, the natural extension will correct the assessments present in $\underline{P}$ into the smallest possible coherent lower prevision. Hence, the notion of natural extension can be used to modify the initial assessments into other assessments that satisfy the notion of coherence, and it does so in the least-committal way, i.e., it provides the smallest coherent lower prevision with the same property.

*Example 1(cont.).* Let us consider again the assessments $\underline{P}(f_1) = 5000, \underline{P}(f_2) = 4000$ and $\overline{P}(f_2) = 6000$. These imply the acceptable buying transactions in Table 1, which, as we showed, avoid sure loss but are incoherent. If we apply Equation (6) to them we obtain that their natural extension is $\underline{E}(f_1) = 5000, \underline{E}(f_2) = 4000, \overline{E}(f_2) = 5000$. Hence, it is a consequence of coherence that Jack should be disposed to sell the gamble $f_2$ for anything bigger than 5000 euros. ♦

The natural extension of the assessments given by a coherent lower prevision $\underline{P}$ can also be calculated in terms of the equivalent representations we have given in Sections 2.2 and 2.3.

Consider a coherent lower prevision $\underline{P}$ with domain $\mathscr{K}$, and let $\mathscr{D}_{\underline{P}}$ be the set of almost-desirable gambles associated with the lower prevision $\underline{P}$ by Equation (4):

$$\mathscr{D}_{\underline{P}} := \{f \in \mathscr{K} : \underline{P}(f) \geq 0\}.$$

The natural extension $\mathscr{E}_{\mathscr{D}_{\underline{P}}}$ of $\mathscr{D}_{\underline{P}}$ provides the smallest set of almost-desirable gambles that contains $\mathscr{D}_{\underline{P}}$ and is coherent. It is the closure (in the supremum norm topology) of

$$\{f : \exists f_j \in \mathscr{D}_{\underline{P}}, \lambda_j > 0 \text{ such that } f \geq \sum_{j=1}^{n} \lambda_j f_j\}, \tag{7}$$

which is the smallest convex cone that contains $\mathscr{D}_{\underline{P}}$ and all non-negative gambles. Then the natural extension of $\underline{P}$ to all gambles is given by

$$\underline{E}(f) = \sup\{\mu : f - \mu \in \mathscr{E}_{\mathscr{D}_{\underline{P}}}\}.$$

If we consider the set $\mathscr{M}(\underline{P})$ of linear previsions that dominate $\underline{P}$ on $\mathscr{K}$, then

$$\underline{E}(f) = \min\{P(f) : P \in \mathscr{M}(\underline{P})\}. \tag{8}$$

This last expression also makes sense if we consider the Bayesian sensitivity analysis interpretation we have given to coherent lower previsions in Section 2.2: there is a linear prevision modelling our subject's information, but his imperfect knowledge of it makes him consider a set of linear previsions $\mathscr{M}(\underline{P})$, whose lower envelope is $\underline{P}$. If he wants to extend $\underline{P}$ to a bigger domain, he should consider all the linear previsions in $\mathscr{M}(\underline{P})$ as possible models (he has no additional information allowing to disregard any of them), or equivalently their lower envelope. He obtains then that $\mathscr{M}(\underline{P}) = \mathscr{M}(\underline{E})$.

The procedure of natural extension preserves the equivalence between the different representations of our assessments: if we consider for instance a coherent lower prevision $\underline{P}$ with $\mathscr{K}$ and the set of almost-desirable gambles $\mathscr{D}_{\underline{P}}$ we derive from Equation (4), then we can consider the natural extension of $\mathscr{D}_{\underline{P}}$ via Equation (7). The coherent lower prevision we can derive from this set of acceptable gambles using Equation (5) coincides with the natural extension $\underline{E}$ of $\underline{P}$. This is because the notion of natural extension, both for lower or linear previsions or for sets of desirable gambles determines the least-committal extension of our initial model that satisfies the notion of coherence.

On the other hand, we can also consider the natural extension of a lower prevision $\underline{P}$ from a domain $\mathscr{K}$ to a bigger domain $\mathscr{K}_1$ (not necessarily equal to $\mathscr{L}(\Omega)$). It can be checked then that the procedure of natural extension is *transitive*, in the following sense: if $\underline{E}_1$ denotes the natural extension of $\underline{P}$ to $\mathscr{K}_1$ and we later consider the natural extension $\underline{E}_2$ of $\underline{E}_1$ to some bigger domain $\mathscr{K}_2 \supset \mathscr{K}_1$, then $\underline{E}_2$ agrees with the natural extension of $\underline{P}$ from $\mathscr{K}$ to $\mathscr{K}_2$: in both cases we are only considering the behavioural consequences of the assessments on $\mathscr{K}$ and the condition of coherence. This is easiest to see using Equation (8): we have $\mathscr{M}(\underline{E}_1) = \mathscr{M}(\underline{E}_2) = \mathscr{M}(\underline{P})$.

# 3 Updating and combining information

So far, we have assumed that the only information that our subject possesses about the outcome of the experiment is that it belongs to the set $\Omega$. But it may happen that, after he has made his assessments, he comes to have some additional information about this outcome, for instance that it belongs to some element of a partition $\mathscr{B}$ of $\Omega$. He then has to update his assessments, and the way to do this is by means of what we shall call *conditional lower previsions*.

## 3.1 Conditional lower previsions

Let $B$ be a subset of the sampling space $\Omega$, and consider a gamble $f$ on $\Omega$. Walley's theory of coherent lower previsions gives two different interpretations of $\underline{P}(f|B)$: the *updated* and the *contingent* one. The most natural in our view is the contingent interpretation, under which $\underline{P}(f|B)$ is our subject's current supremum buying price for the gamble $f$ contingent on $B$, that is, the supremum value of $\mu$ such that the gamble $I_B(f-\mu)$ is desirable for our subject.

In order to relate our subject's current dispositions on a gamble $f$ contingent on $B$ with his dispositions towards this gamble if he later shall come to know whether the outcome of the experiment belongs to $B$, Walley introduces the so-called *updating principle*. We say a gamble $f$ is $B$-desirable for our subject when he is currently disposed to accept $f$ provided he later observes that the outcome belongs to $B$. Then the updating principle requires that a gamble is $B$-desirable if and only if $I_B f$ is desirable. In this way, we can relate the current and future dispositions of our subject.

Under the updated interpretation of conditional lower previsions, $\underline{P}(f|B)$ is the subject's supremum acceptable buying price he would pay for the gamble $f$ now if he came to know later that the outcome belongs to the set $B$, and nothing more. It coincides with the value determined by the contingent interpretation of $\underline{P}(f|B)$ because of the updating principle.

Let $\mathscr{B}$ be a partition of our sampling space, $\Omega$, and consider an element $B$ of this partition. This partition could be for instance a class of categories of the set of outcomes. Assume that our subject has given conditional assessments $\underline{P}(f|B)$ for all gambles $f$ on some domain $\mathscr{H}_B$. As it was the case for (unconditional) lower previsions, we should require that these assessments are consistent with each other. We say that the conditional lower prevision $\underline{P}(\cdot|B)$ is *separately coherent* when the following two conditions are satisfied:

(SC1) It is coherent as an unconditional prevision, i.e.,

$$\sup_{\omega\in\Omega}\sum_{i=1}^{n}[f_i(\omega)-\underline{P}(f_i|B)]-m[f_0(\omega)-\underline{P}(f_0|B)]\geq 0 \tag{9}$$

for all non-negative integers $n,m$ and all gambles $f_0,\dots,f_n$ in $\mathscr{H}_B$;

(SC2) the indicator function of $B$ belongs to $\mathscr{H}_B$ and $\underline{P}$ satisfies $\underline{P}(B|B)=1$.

The coherence requirement (9) can be given a behavioural interpretation in the same way as with (unconditional) coherence in Equation (1): if it does not hold for some non-negative integers $n,m$ and gambles $f_0,\dots,f_n$ in $\mathscr{H}_B$, then it can be checked that either: (i) the almost-desirable gamble $\sum_{i=1}^{n}[f_i-\underline{P}(f_i|B)]$ incurs in a sure loss (if $m=0$) or (ii) we can raise $\underline{P}(f_0|B)$ in some positive quantity $\delta$, contradicting our interpretation of it as his supremum acceptable buying price (if $m>0$).

On the other hand, Equation (9) already implies that $\underline{P}(B|B)$ should be smaller than, or equal to, 1. That we require it to be equal to one means just that our subject should bet at all odds on the occurrence of the event $B$ after having observed it.

In this way, we can obtain separately coherent conditional lower previsions $\underline{P}(\cdot|B)$ with domains $\mathscr{H}_B$ for all events $B$ in the partition $\mathscr{B}$. It is a consequence of separate coherence that the conditional lower prevision $\underline{P}(f|B)$ does only depend on the values that $f$ takes on $B$, i.e, for every two gambles $f$ and $g$ such that $f(\omega) = g(\omega)$ for all $\omega \in B$, we should have $\underline{P}(f|B) = \underline{P}(g|B)$. This property implies that all the domains $\mathscr{H}_B$ can be extended to the common domain $\mathscr{H} := \{f = \sum_{B\in\mathscr{B}} f_B : f_B \in \mathscr{H}_B\ \forall B\}$, and we can define on $\mathscr{H}$ a *conditional lower prevision* $\underline{P}(\cdot|\mathscr{B})$ by

$$\underline{P}(f|\mathscr{B}) := \sum_{B\in\mathscr{B}} I_B \underline{P}(f|B),$$

i.e., the gamble on $\Omega$ that assumes the value $\underline{P}(f|B)$ on all elements of $B$. This conditional lower prevision is then called *separately coherent* when $\underline{P}(\cdot|B)$ is separately coherent for all $B \in \mathscr{B}$. It provides the updated supremum buying price after learning that the outcome of the experiment belongs to some particular element of $\mathscr{B}$. We shall later use the notation

$$G(f|B) := I_B(f - \underline{P}(f|B)),\ G(f|\mathscr{B}) := \sum_{B\in\mathscr{B}} G(f|B) = f - \underline{P}(f|\mathscr{B}). \tag{10}$$

When the domain $\mathscr{H}$ of $\underline{P}(\cdot|\mathscr{B})$ is a linear space containing all constant gambles, then separate coherence is equivalent to:

(C1) $\underline{P}(f|B) \geq \inf_{\omega\in B} f(\omega)$.

(C2) $\underline{P}(\lambda f|\mathscr{B}) = \lambda \underline{P}(f|\mathscr{B})$.

(C3) $\underline{P}(f+g|\mathscr{B}) \geq \underline{P}(f|\mathscr{B}) + \underline{P}(g|\mathscr{B})$.

for all positive real $\lambda$, $B \in \mathscr{B}$ and gambles $f, g$ in $\mathscr{H}$. The first requirement shows that the conditional lower prevision on $B$ should only depend on the behavior of $f$ on this set; conditions (C2) and (C3) are the counterparts of the requirements (P2) and (P3) we made for unconditional lower previsions, respectively.

*Example 1(cont.).* For the gambles $f_1$ and $f_2$ whose reward in terms of the color of the ball drawn is given by the following table

| | green | red | black |
|---|---|---|---|
| $f_1$ | 10000 | 5000 | 0 |
| $f_2$ | 0 | 5000 | 9000 |

Table 3: Increase on Jack's wealth depending on the color of the ball he draws.

Jack had established the coherent assessments $\underline{P}(f_1) = 5000$ and $\underline{P}(f_2) = 4000$. But he may also establish now his supremum acceptable buying prices for these gambles depending on some future information on the color of the ball drawn. If for instance he is informed that the ball drawn is not green, Jack should update his lower prevision for the gamble $f_2$, because he is sure that in that case he would get at least a prize of 5000 euros out of it. On the other hand, if he keeps the supremum buying prize of 5000 euros for $f_1$ he is implying that he is sure that the ball that has been drawn is red once he comes to know that it is not green.

If for instance he considers as possible models the ones in Table 2 and updates them using Bayes's rule, then the updated supremum buying prices he should give by taking lower envelopes would be $\underline{P}(f_1|$ not green $) = 0$ and $\underline{P}(f_2|$ not green $) = 5000$.[5] ♦

## 3.2 Coherence of a finite number of conditional lower previsions

In practice, it is not uncommon to have lower previsions conditional on different partitions $\mathscr{B}_1,\dots,\mathscr{B}_n$ of $\Omega$. We can think for instance of different sets of categories, or of information provided in a sequential way. We end up then with a finite number of conditional lower previsions $\underline{P}(\cdot|\mathscr{B}_1),\dots,\underline{P}(\cdot|\mathscr{B}_n)$ with respective domains $\mathscr{H}_1,\dots,\mathscr{H}_n \subseteq \mathscr{L}(\Omega)$, which we shall assume are separately coherent.

As it was the case with unconditional lower previsions, before making any inference based on these assessments we have to verify that they are consistent with each other. And again, by 'consistent' we shall mean that a combination of acceptable buying prices should neither lead to a sure loss, nor to an increase of the supposedly supremum acceptable buying price for a gamble $f$.

To see which form coherence takes now, we need to introduce the second pillar of Walley's theory of conditional previsions (the other is the updating principle): the *conglomerative principle*. This rationality principle requires that if a gamble is $B$-desirable for every set $B$ in a partition $\mathscr{B}$ of $\Omega$, then $f$ is also desirable. Taking into account the updating principle, this means that if $I_B f$ is desirable for every $B$ in $\mathscr{B}$, then $f$ should be desirable.

It follows from this principle that for every gamble $f$ in the domain of $\underline{P}(\cdot|\mathscr{B})$, the gamble $G(f|\mathscr{B})$ given by Equation (10) should be almost-desirable. This is the basis of the following definition of coherence. For simplicity, we shall assume that the domains $\mathscr{H}_1,\dots,\mathscr{H}_n$ are linear spaces of gambles. A possible generalization to non-linear domains can be found in [46]. Given $f_i \in \mathscr{H}_i$, we shall denote by $S_i(f_i) := \{B \in \mathscr{B}_i : I_B f_i \neq 0\}$ the *$\mathscr{B}_i$-support* of $f_i$. It is the set of elements of $\mathscr{B}_i$ where $f_i$ is not identically zero. It follows from the separate coherence of $\underline{P}(\cdot|\mathscr{B}_i)$ that $\underline{P}(0|\mathscr{B}_i) = 0$ for all $i = 1,\dots,n$, and as a consequence the gamble $\underline{P}(f|\mathscr{B}_i)$ (or $G(f|\mathscr{B}_i)$, for that matter) is identically zero outside $S_i(f_i)$.

We say that $\underline{P}(\cdot|\mathscr{B}_1),\dots,\underline{P}(\cdot|\mathscr{B}_n)$ are (jointly) *coherent* when for all $f_i \in \mathscr{H}_i, i = 1,\dots,n$ and all $f_0 \in \mathscr{H}_j, B_0 \in \mathscr{B}_j$ for some $j \in \{1,\dots,n\}$,

$$\sup_{\omega \in B} \left[ \sum_{i=1}^{n} G(f_i|\mathscr{B}_i) - G(f_0|B_0) \right](\omega) \geq 0 \tag{11}$$

for some $B \in \{B_0\} \cup \bigcup_{i=1}^{n} S_i(f_i)$.

Assume that Equation (11) does not hold. Then, there is some $\delta > 0$ such that $G(f_0|B_0) + I_{B_0}\delta$ dominates the almost-desirable gamble $\sum_{i=1}^{n} G(f_i|\mathscr{B}_i)$ on every $B \in \{B_0\} \cup \bigcup_{i=1}^{n} S_i(f_i)$. As a consequence, the gamble $G(f_0|B_0) + I_{B_0}\delta$ should also be almost-desirable, and this means that $\underline{P}(f|B_0) + \delta$ should be an acceptable buying price for $f$, contingent on $B_0$. This is an inconsistency.

[5] This is an instance of a procedure called *regular extension*, that can sometimes be used to coherently update beliefs; see [71, Appendix J] for more details.

**Remark 1.** *The sum* $\sum_{i=1}^{n} G(f_i|\mathscr{B}_i) - G(f_0|B_0)]$ *in the left-hand side in Equation* (11) *is identically zero outside the union of the sets in the family* $\{B_0\} \cup \bigcup_{i=1}^{n} S_i(f_i)$*. Hence, if in Equation* (11) *we consider the supremum over all* $\Omega$ *(a condition called* weak coherence *by Walley [71, Section 7.1.4]) instead of over the sets in the family* $\{B_0\} \cup \bigcup_{i=1}^{m} S_i(f_i)$*, the condition will be automatically satisfied whenever the union of these sets is not equal to* $\Omega$*, no matter how inconsistent these assessments are with each other. This is one of the reasons to consider this stronger version as our definition of coherence. See [71, Example 7.3.5] for other undesirable properties of weak coherence.*

### 3.2.1 Natural extension of conditional lower previsions

Assume then that our subject has provided a finite number of (separately and jointly) coherent lower previsions $\underline{P}(\cdot|\mathscr{B}_1), \dots, \underline{P}(\cdot|\mathscr{B}_n)$ defined on respective linear subsets $\mathscr{H}_1, \dots, \mathscr{H}_n$ of $\mathscr{L}(\Omega)$. Then he may wish to see which are the behavioural implications of these assessments on gambles which are not in the domain. The way to do this is through the notion of *natural extension*. Given $f \in \mathscr{L}(\Omega)$ and $B_0 \in \mathscr{B}_i$, $\underline{E}(f|B_0)$ is defined as the supremum value of $\mu$ for which there are $f_i \in \mathscr{H}_i$ such that

$$\sup_{\omega \in B} \left[ \sum_{i=1}^{n} G(f_i|\mathscr{B}_i) - I_{B_0}(f - \mu) \right](\omega) < 0 \tag{12}$$

for all $B$ in the class $\{B_0\} \cup \cup_{i=1}^{n} S_i(f_i)$.

In the particular case where we only have an unconditional lower prevision, i.e., when $n = 1$ and $\mathscr{B}_1 = \{\Omega\}$, this notion coincides with the unconditional natural extension we introduced in Section 2.4. If we have a number of conditional lower previsions $\underline{P}(\cdot|\mathscr{B}_1), \dots, \underline{P}(\cdot|\mathscr{B}_n)$, we can calculate their natural extensions $\underline{E}(\cdot|\mathscr{B}_1), \dots, \underline{E}(\cdot|\mathscr{B}_n)$ to all gambles using Equation (12). If the partitions $\mathscr{B}_1, \dots, \mathscr{B}_n$ are finite, then these natural extensions share some of the properties of the unconditional natural extension:

1. They coincide with $\underline{P}(\cdot|\mathscr{B}_1), \dots, \underline{P}(\cdot|\mathscr{B}_n)$ if and only if these conditional lower previsions are coherent.
2. They are the smallest coherent extensions of $\underline{P}(\cdot|\mathscr{B}_1), \dots, \underline{P}(\cdot|\mathscr{B}_n)$ to all gambles.
3. They are the lower envelope of a family of coherent conditional *linear* previsions, $\{P_\gamma(\cdot|\mathscr{B}_1), \dots, P_\gamma(\cdot|\mathscr{B}_n) : \gamma \in \Gamma\}$.

Hence, when the partitions are finite, the notion of natural extension of a number of conditional lower prevision also provides us with the consequences of the assessments present on these previsions and the notion of (joint) coherence, to all gambles in the domain, and it can also be given a Bayesian sensitivity analysis interpretation. This is interesting for many applications, where we must deal with finite spaces only; however, there are also interesting situations, such as parametric inference, where we must deal with infinite spaces and where we end up with partitions that have an infinite number of different elements. In that case, it is easy to see that in order to achieve coherence, $\underline{P}(f|B_0)$ must be at least as large as the supremum $\mu$ that satisfies Equation (12) for some $f_i \in \mathscr{H}_i$, $i = 1, \dots, n$. However, and unlike the case of finite partitions, we cannot guarantee that these values provide coherent extensions of $\underline{P}(\cdot|\mathscr{B}_1), \dots, \underline{P}(\cdot|\mathscr{B}_n)$ to all gambles: in general these will only be lower bounds of all the coherent extensions. Indeed, when the partitions are infinite, we can have a number of problems:

1. There may be no coherent extensions, and as a consequence the natural extensions may not be coherent [71, Sections 6.6.6 and 6.6.7].
2. Even if the smallest coherent extensions exist, they may differ from the natural extensions, which are not coherent [71, Section 8.1.3].
3. The minimal coherent extensions, and as a consequence also the natural extensions, may not be lower envelopes of coherent linear collections [71, Sections 6.6.9 and 6.6.10].

The natural extensions are the minimal coherent extensions of the lower previsions $\underline{P}(\cdot|\mathscr{B}_1)$,...,$\underline{P}(\cdot|\mathscr{B}_n)$ if and only if they are jointly coherent themselves. But we need some additional conditions to guarantee the joint coherence of $\underline{E}(\cdot|\mathscr{B}_1)$,...,$\underline{E}(\cdot|\mathscr{B}_n)$. One of these conditions is that all the partitions $\mathscr{B}_i$ are finite. But even when the partitions are infinite it may happen that we are able to characterize the minimal coherent extensions, but they differ from the natural extensions. One of the reasons for this defective behaviour of the natural extension in the conditional case is the notion of conglomerability, that we shall treat in detail in the following section, and that becomes trivial in the case where the partitions are finite.

*Example* 3. An example where the natural extension fails to provide the minimal coherent extensions is given in [48, Example 1]. Let us consider the possibility space $\Omega = \mathscr{X}_1 \times \mathscr{X}_2$, where $\mathscr{X}_1 = \mathscr{X}_2 = [0,1]$, and the partition $\mathscr{B} = \{B_{x_1} : x_1 \in \mathscr{X}_1\}$, where $B_{x_1} := \{x_1\} \times \mathscr{X}_2$. Let $\mathscr{K} = \{\lambda\pi_1 : \lambda \in \mathbb{R}\}$ and $\mathscr{H} = \{g\pi_2 : g \in \mathscr{L}(\mathscr{X}_1)\}$, where the gamble $\lambda\pi_1$ is defined by $\lambda\pi_1(x_1,x_2) = \lambda x_1$, and the gamble $g\pi_2$ by $g\pi_2(x_1,x_2) = g(x_1)x_2$. Let us define the linear (and therefore coherent lower) prevision $P$ on $\mathscr{K}$ by

$$P(\lambda\pi_1) = \lambda,$$

and the conditional linear prevision $P(\cdot|\mathscr{B})$ on $\mathscr{H}$ by

$$P(g\pi_2|B_{x_1}) = g(x_1)$$

for all $x_1$ in $\mathscr{X}_1$. Then it can be checked that $P$ and $P(\cdot|\mathscr{B})$ are coherent. Given the gamble $f$ on $\Omega$, given by

$$f(x_1,x_2) = \begin{cases} 0 & \text{if } (x_1,x_2) = (1-\frac{1}{n}, 1-\frac{1}{n}) \text{ for some } n > 0 \\ 1 & \text{otherwise,} \end{cases}$$

it can be checked that the natural extension $\underline{E}$ of $P$, $P(\cdot|\mathscr{B})$ provides $\underline{E}(f) = 0$. However, in this case the smallest coherent extensions $\underline{M}, \underline{M}(\cdot|\mathscr{B})$ of $P, P(\cdot|\mathscr{B})$ can be calculated, and we obtain $\underline{M}(f) = 1$. The reason for this discrepancy is that $\underline{E}$ only gives an extension of $P$ which is coherent with $P(\cdot|\mathscr{B})$; if we also want to extend $P(\cdot|\mathscr{B})$ to all gambles the natural extension may not guarantee coherence. ♦

This example provides an instance of the *marginal extension* of a number of conditional and unconditional lower previsions. When these previsions are conditioning on a sequence of increasingly finer partitions, the marginal extension can be used to determine the smallest coherent extensions to all gambles. See [71, Section 6.7.2] and [48] for more information.

## 3.3 Coherence of an unconditional and a conditional lower prevision

Let us consider in more detail the case where we have a conditional lower prevision $\underline{P}(\cdot|\mathscr{B})$ on some domain $\mathscr{H}$ and an unconditional lower prevision $\underline{P}$ on some set of gambles $\mathscr{K}$. Assume that they satisfy the following conditions:

(a) $\mathscr{K}, \mathscr{H}$ are linear subspaces of $\mathscr{L}(\Omega)$.

(b) Given $f \in \mathscr{H}$, the gambles $\underline{P}(f|\mathscr{B})$ and $I_B f$ belong to $\mathscr{H}$ for all $B \in \mathscr{B}$.

(c) $\underline{P}(\cdot|\mathscr{B})$ is separately coherent on $\mathscr{H}$ and $\underline{P}$ is coherent on $\mathscr{K}$.

As Walley points out, the first two assumptions are made for mathematical convenience only; (b) can be assumed without loss of generality [71, Section 6.3.1], and the results can be extended to situations where (a) does not hold [46]. Note that the unconditional lower prevision can be seen as a conditional lower prevision by simply considering the partition $\{\Omega\}$, and then Equation (10) becomes $G(f) = f - \underline{P}(f)$.

The lower previsions $\underline{P}$ and $\underline{P}(\cdot|\mathscr{B})$ are *jointly coherent*, i.e., they satisfy Equation (11), if and only if

(JC1) $\sup_{\omega\in\Omega}[G(f_1) + G(g_1|\mathscr{B}) - G(f_2)] \geq 0$.

(JC2) $\sup_{\omega\in\Omega}[G(f_1) + G(g_1|\mathscr{B}) - G(g_2|B_0)] \geq 0$.

for all $f_1, f_2 \in \mathscr{K}, g_1, g_2 \in \mathscr{H}$ and $B_0 \in \mathscr{B}$. These conditions can be simplified under some additional assumptions on the domains (see [71, Section 6.5] for details).

Again, it can be checked that if any of these conditions fails, the assessments of our subject produce inconsistencies. Assume first that (JC1) does not hold. If $f_2 = 0$, then we have a sum of acceptable transactions that produces a sure loss. If $f_2 \neq 0$, then there is some $\delta > 0$ such that the gamble $G(f_2) - \delta$ dominates the desirable gamble $G(f_1) + G(g_1|\mathscr{B}) + \delta$. This means that our subject is willing to increase his supremum acceptable buying price for $f_2$ in $\delta$, a contradiction.

Similarly, if (JC2) does not hold and $g_2 = 0$ we have a sum of acceptable transactions that produces a sure loss; and if $g_2 \neq 0$ there is some $\delta > 0$ such that $G(g_2|B_0) - \delta$ dominates $G(f_1) + G(g_1|\mathscr{B}) + \delta$ and is therefore desirable. Hence, our subject should be willing to pay $\underline{P}(g_2|B_0) + \delta$ for $g_2$ contingent on $B_0$, a contradiction.

It is a consequence of the joint coherence of $\underline{P}$, $\underline{P}(\cdot|\mathscr{B})$ that, given $f \in \mathscr{H}$ and $B \in \mathscr{B}$, $\underline{P}(G(f|B)) = \underline{P}(I_B(f - \underline{P}(f|B))) = 0$, where $G(f|B)$ is defined in Equation (10). When $\underline{P}(B) > 0$, there is a unique value $\mu$ such that $\underline{P}(I_B(f - \mu)) = 0$, and therefore this $\mu$ must be the conditional lower prevision $\underline{P}(f|B)$. This is called the *Generalised Bayes Rule* (GBR). This rule has a number of interesting properties:

1. It is a generalization of Bayes's rule in classical probability theory.
2. If $\underline{P}(B) > 0$ and we define $\underline{P}(f|B)$ via the Generalized Bayes Rule, then it is the lower envelope of the conditional linear previsions $P(f|B)$ that we can define using Bayes's rule on the elements of $\mathscr{M}(\underline{P})$.
3. When the partition $\mathscr{B}$ is finite and $\underline{P}(B) > 0$ for all $B \in \mathscr{B}$, then the GBR uniquely determines the conditional lower prevision $\underline{P}(\cdot|\mathscr{B})$.

*Example* 4. Three horses (a,b and c) take part in a race. Our a priori lower probability for each horse being the winner is $\underline{P}(\{a\}) = 0.1, \underline{P}(\{b\}) = 0.25, \underline{P}(\{c\}) = 0.3, \underline{P}(\{a,b\}) = 0.4, \underline{P}(\{a,c\}) = 0.6, \underline{P}(\{b,c\}) = 0.7$. Since there are rumors that c is not going to take part in the race due to some injury, we may provide our updated lower probabilities for that case using the Generalised Bayes Rule. Taking into account that we are dealing with finite spaces and that the conditioning event has positive lower probability, applying the Generalised Bayes Rule is equivalent to taking the lower envelope of the linear conditional previsions that we obtain applying Bayes's rule on the elements of $\mathscr{M}(\underline{P})$. Thus, we obtain:

$$\underline{P}(\{a\}|\{a,b\}) = \inf\left\{\frac{P(\{a\})}{P(\{a,b\})} : P \in \mathscr{M}(\underline{P})\right\} = 0.1/0.5 = 0.2,$$

$$\underline{P}(\{b\}|\{a,b\}) = \inf\left\{\frac{P(\{b\})}{P(\{a,b\})} : P \in \mathscr{M}(\underline{P})\right\} = 0.25/0.55 = 0.45. \blacklozenge$$

We saw in the previous section that a number of conditional lower previsions may be coherent and still have some undesirable properties, when the partitions are infinite. Something similar applies to the case where we have only a conditional and an unconditional lower prevision. For instance, a coherent pair $\underline{P}, \underline{P}(\cdot|\mathscr{B})$ is not necessarily the lower envelope of coherent pairs of linear unconditional and conditional previsions, $P, P(\cdot|\mathscr{B})$ (these even may not exist). On the other hand, there are linear previsions $P$ for which there is no conditional linear prevision $P(\cdot|\mathscr{B})$ such that $P, P(\cdot|\mathscr{B})$ are coherent in the sense of Equation (11), i.e., linear previsions that cannot be updated in a coherent way to a linear conditional prevision $P(\cdot|\mathscr{B})$, but which can be updated to a conditional lower prevision.

Taking this into account, given an unconditional prevision $\underline{P}$ representing our subject's beliefs and a partition $\mathscr{B}$ of $\Omega$, he may be interested in considering the conditional lower previsions $\underline{P}(\cdot|\mathscr{B})$ which are coherent with $\underline{P}$, i.e., those for which conditions (JC1) and (JC2) are satisfied. A necessary and sufficient condition for the existence of such $\underline{P}(\cdot|\mathscr{B})$ is that $\underline{P}$ is *$\mathscr{B}$-conglomerable*: this is the case when given distinct sets $B_1, B_2, \dots$ in $\mathscr{B}$ such that $\underline{P}(B_n) > 0$ for all $n$ and a gamble $f$ such that $\underline{P}(I_{B_n} f) \geq 0$ for all $n$, it holds that $\underline{P}(I_{\cup_n B_n} f) \geq 0$.

The condition of $\mathscr{B}$-conglomerability holds trivially when the partition $\mathscr{B}$ is finite, or when $\underline{P}(B) = 0$ for every set $B$ in the partition. It only becomes non-trivial when we consider a partition $\mathscr{B}$ for which there are infinitely many elements $B$ satisfying $\underline{P}(B) > 0$. It makes sense as a rationality axiom once we accept the updating and conglomerability principles: to see this, consider that if $\{B_n\}_n$ is a partition of $\Omega$ with $\underline{P}(B_n) > 0$ and $\underline{P}(I_{B_n} f) \geq 0$, then for every $\delta > 0$ the gamble $I_{B_n}(f + \delta)$ is desirable.

The updating and conglomerative principles imply then that $f+\delta$ is desirable, whence $\underline{P}(f+\delta) \geq 0$. Since this holds for all $\delta > 0$, we deduce that $\underline{P}(f) \geq 0$.

More generally, Walley says that a coherent lower prevision $\underline{P}$ is *fully conglomerable* when it is $\mathscr{B}$-conglomerable for every partition $\mathscr{B}$. A full conglomerable coherent lower prevision can be coherently updated to a conditional lower prevision $\underline{P}(\cdot|\mathscr{B})$ for any partition $\mathscr{B}$ of $\Omega$. Again, full conglomerability can be accepted as an axiom of rationality provided we accept the updating and conglomerability principles, and also provided that when we define our coherent lower prevision we want to be able to updated for all possible partitions of our set of values.

There is an important connection between full conglomerability and countable additivity: given a linear prevision $P$ on $\mathscr{L}(\Omega)$ taking infinitely many values, it is fully conglomerable if and only for every countable partition $\{B_n\}_n$ of $\Omega$ it satisfies $\sum_n P(B_n) = 1$.

Full conglomerability is one of the points of disagreement between Walley's and de Finetti's work, that we shall present in more detail in Section 5.1. De Fintetti rejects the assumption of countable additivity on probabilities, and taking into account the above relationship also the property of full conglomerability. One key observation here is that de Finetti does not assume the conglomerative principle as a rationality axiom, and full conglomerability can be seen as a consequence of it.

When $\underline{P}$ is $\mathscr{B}$-conglomerable and $B \in \mathscr{B}$, the conditional lower prevision $\underline{P}(f|B)$ is uniquely determined by the Generalised Bayes Rule if $\underline{P}(B) > 0$. If $\underline{P}(B) = 0$, however, there is not a unique value for $\underline{P}(f|B)$ for which we achieve coherence. The smallest conditional lower prevision $\underline{P}(\cdot|B)$ which is coherent with $\underline{P}$ is the vacuous conditional prevision, given by $\underline{P}(f|B) = \inf_{\omega \in B} f(\omega)$, and if we want to have more informative assessments we may need some additional assumptions. Indeed, the approach to conditioning on sets of probability zero is one of the differences in the approach to conditioning by Walley (and also by de Finetti and Williams) and that by Kolmogorov. In Kolmogorov's approach, conditioning is made on a $\sigma$-field $\mathscr{A}$, and a conditional prevision $P(f|\mathscr{A})$ is any $\mathscr{A}$-measurable gamble which satisfies $P(gP(f|\mathscr{A})) = P(gf)$ for every $\mathscr{A}$-measurable gamble $g$. In particular, if we consider an event $B$ of probability zero, Kolmogorov allows the prevision $P(\cdot|B)$ to be completely arbitrary. Walley's coherence condition is more general because it can be applied on previsions conditional on partitions, and is more restrictive when dealing with sets of probability zero than Kolmogorov's (although it may be argued that it also makes more sense). In particular, for a given linear prevision $P$ there may not exist linear conditional previsions which are coherent with $P$.

One interesting approach to conditioning on sets of probability zero in a coherent setting is the use of *zero-layers* by Coletti and Scozzafava [8, Chapter 12], which also appears in some earlier work by Krauss [38]. Their approach to conditioning is nevertheless slightly different from Walley, since they consider conditional previsions as previsions whose domain is a class of conditional events. See [6, 7, 8] for further information on Coletti and Scozzafava's work, and [71, Section 6.10] and [78, Section 1.4] for further details on Walley's approach to conditioning on events of lower probability zero.

# 4 Independence

Next, we see how we can define the concept of independence in the context of coherent lower previsions. Let us consider two random variables $X_1, X_2$ taking values in respective sets $\mathscr{X}_1, \mathscr{X}_2$. In the classical setting, we call the two variables (stochastically) independent when, given the probability measure $P$ that models the value that $(X_1, X_2)$ assume jointly, any of the following conditions holds[6]

(a) $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$ for all $x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2$ [decomposition].

(b) $P(X_1 = x_1 | X_2 = x_2) = P(X_1 = x_1)$ for all $x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2$ [marginalization].

**Remark 2.** *These two conditions are equivalent provided the marginal distributions are everywhere non-zero, that is, provided we are not conditioning on sets of probability zero. But if $P(X_2 = x_2) = 0$ for instance, then condition (a) holds trivially, while there are many values of $P(X_1 = x_1 | X_2 = x_2)$ for which condition (b) may not hold, and this even under the more restrictive treatment of conditioning of sets of probability zero that we presented in the previous section. To simplify this section, we shall assume throughout that the conditioning events have all positive lower probability.* ⧫

Provided we are conditioning on sets of positive probability, independence is a symmetrical notion: if (b) holds, then we also have

$$P(X_2 = x_2 | X_1 = x_1) = P(X_2 = x_2)$$

for all $x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2$.

When our knowledge about the value that $(X_1, X_2)$ assume jointly is represented by means of a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$, there is no unique way of extending the notion of independence. The properties of decomposition and marginalization are no longer equivalent, and moreover symmetry is not immediate anymore, meaning that we must distinguish between irrelevance (an asymmetrical notion) and independence (its symmetric counterpart). On the other hand, all our definitions must be made in terms of variables and not of events, since events do not keep all the information about the coherent lower prevision $\underline{P}$.

In this section, we shall present some of the generalizations proposed in the literature and the relationships between them. To fix things, consider a coherent lower prevision $\underline{P}$ on $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$ representing our knowledge about the value that $X_1, X_2$ assume jointly. [7] We shall assume throughout that these two variables are *logically independent*, meaning that the joint variable $(X_1, X_2)$ can assume any value in the product space $\mathscr{X}_1 \times \mathscr{X}_2$. Our information about the random variable $X_1$ is given by the *marginal* lower prevision $\underline{P}_1$, where

$$\underline{P}_1(f) = \underline{P}(\hat{f})$$

[6] In this definition we assume that the sets $\mathscr{X}_1, \mathscr{X}_2$ are finite in order to simplify the notation; in the infinite case we would simply consider density functions instead of probability mass functions. We also use $X_1 = x_1$ for denoting the event $X_1^{-1}(\{x_1\})$ to simplify the notation.

[7] Although here we are assuming for simplicity that the domain of $\underline{P}$ is $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$, all the developments we shall make in this section can be generalized to the case where the domain of $\underline{P}$ is some subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$.

for all $f \in \mathscr{L}(\mathscr{X}_1)$, where $\hat{f}$ is the gamble given by $\hat{f}(x_1,x_2) = f(x_1)$ for all $(x_1,x_2)$ in $\mathscr{X}_1 \times \mathscr{X}_2$ (such a gamble is called $\mathscr{X}_1$-measurable). Similarly, we can express our information about the outcome of $X_2$ by means of a coherent lower prevision $\underline{P}_2$ on $\mathscr{L}(\mathscr{X}_2)$, given by

$$\underline{P}_2(g) = \underline{P}(\hat{g})$$

for all $g \in \mathscr{L}(\mathscr{X}_2)$, where $\hat{g}$ is the gamble given by $\hat{g}(x_1,x_2) = g(x_2)$ for all $(x_1,x_2)$ in $\mathscr{X}_1 \times \mathscr{X}_2$. We shall also assume that we have conditional lower previsions $\underline{P}(\cdot|X_1)$, $\underline{P}(\cdot|X_2)$ on $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$ such that $\underline{P}, \underline{P}(\cdot|X_1)$ and $\underline{P}(\cdot|X_2)$ are jointly coherent (i.e., they satisfy Equation (11)). These conditional lower previsions represent our beliefs about the outcome of one of the experiments provided we observe the outcome of the other. $\underline{P}(\cdot|X_1)$ is a lower prevision conditional on the partition $\{X_1 = x_1 : x_1 \in \mathscr{X}_1\}$ of $\Omega$, and similarly $\underline{P}(\cdot|X_2)$ is a lower prevision conditional on the partition $\{X_2 = x_2 : x_2 \in \mathscr{X}_2\}$ of $\Omega$. To simplify the notation we shall sometimes use $\underline{P}(\cdot|x_1)$ to denote $\underline{P}(\cdot|X_1 = x_1)$ and $\underline{P}(\cdot|x_2)$ to denote $\underline{P}(\cdot|X_2 = x_2)$ for any $x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2$.

In the examples we shall consider in this section we shall deal with finite spaces only, and the conditioning events shall always have positive lower probability; this will simplify the calculations of the conditional lower probabilities, because (i) they will be uniquely determined by the Generalised Bayes Rule and (ii) they will also be the lower envelope of the conditional *precise* probabilities that can be obtained by applying Bayes's rule to the set of compatible (precise) models.

## 4.1 Epistemic irrelevance

The first generalization of the concept of independence to imprecise probabilities is based on the marginalization property, and is called epistemic irrelevance. We say that the experiment $X_1$ is *epistemically irrelevant* for $X_2$ when our beliefs about the value that $X_2$ takes do not change after we learn the value that $X_1$ has taken. Formally, this holds if and only if

$$\underline{P}(f|X_1 = x_1) = \underline{P}_2(f)$$

for all $\mathscr{X}_2$-measurable gambles $f$, and all $x_1 \in \mathscr{X}_1$.

The notion of epistemic irrelevance can also be defined in terms of sets of linear previsions [10, 11]: we say that $X_1$ is epistemically irrelevant for $X_2$ when

$$\{P_2(\cdot|x_1) : P_2 \in \mathscr{M}(\underline{P})\} = \mathscr{M}(\underline{P}_2),$$

for all $x_1$ in $\mathscr{X}_1$, i.e., when learning the outcome about the first experiment does not change our uncertainty (the set of possible precise models) about the second experiment. Similarly, this notion can also be expressed in terms of sets of desirable gambles: epistemic irrelevance means that the set of acceptable gambles for $X_2$ should not change after learning the outcome of $X_1$ [49].

*Example* 5. We consider three urns with green and red balls. The first urn has one red ball and one green ball. In the second urn, we have two green balls, one red ball and two other balls of unknown color (they may be red or green). In the third urn, we have one green ball, two red balls and two other balls of unknown color.

We select a ball from the first urn. If it is green, then we select a ball from the second urn; if it is red, we select a ball from the third urn. Let $X_1$ be the color of the first ball selected and $X_2$ the color of the second ball. We have

$$\begin{aligned}\underline{P}(\text{ the second ball is green} \mid \text{the first ball is green}) &= 2/5\\ \underline{P}(\text{ the second ball is green} \mid \text{the first ball is red}) &= 1/5,\end{aligned}$$

and thus the first experiment is not epistemically irrelevant to the second. ♦

## 4.2 Epistemic independence

The notion of epistemic irrelevance is an asymmetric notion, meaning that $X_1$ can be epistemic irrelevant for $X_2$ while $X_2$ is not epistemic irrelevant for $X_1$. When $X_1$ and $X_2$ are epistemically irrelevant to each other, we say that the two experiments are *epistemically independent*. This holds if and only if

$$\underline{P}(f|X_1 = x_1) = \underline{P}_2(f) \text{ and } \underline{P}(g|X_2 = x_2) = \underline{P}_1(g)$$

for all $\mathscr{X}_2$-measurable gambles $f$, $\mathscr{X}_1$-measurable gambles $g$, $x_1 \in \mathscr{X}_1$, and $x_2 \in \mathscr{X}_2$.

In terms of sets of linear previsions, the notion of epistemic independence means that, for every $x_1 \in \mathscr{X}_1$ and $x_2 \in \mathscr{X}_2$,

$$\{P_2(\cdot|x_1) : P_2 \in \mathscr{M}(\underline{P})\} = \mathscr{M}(\underline{P}_2) \text{ and } \{P_1(\cdot|x_2) : P_1 \in \mathscr{M}(\underline{P})\} = \mathscr{M}(\underline{P}_1).$$

The behavioural interpretation of these definitions is that our sets of desirable gambles for either experiment do not change after we learn the outcome of the other experiment.

*Example 5(cont.).* Assume now that in the third urn we also have 2 green balls, 1 red ball and 2 balls of unknown color. We have

$$\begin{aligned}\underline{P}(\text{ the second ball is green} \mid \text{the first ball is green}) &= 2/5 \text{ and}\\ \underline{P}(\text{ the second ball is green} \mid \text{the first ball is red}) &= 2/5,\end{aligned}$$

and similarly

$$\begin{aligned}\underline{P}(\text{ the second ball is red} \mid \text{the first ball is green}) &= 1/5 \text{ and}\\ \underline{P}(\text{ the second ball is red} \mid \text{the first ball is red}) &= 1/5.\end{aligned}$$

Hence, the first variable is epistemically irrelevant for the second. However,

$$\underline{P}(\text{ the first ball is green} \mid \text{the second ball is green}) = 1/3;$$

(use the composition where we have 2 green balls in the second urn and 4 in the third one) while $P(\text{ the first ball is green }) = 0.5$. We conclude that the second variable is not epistemically irrelevant to the first, and therefore they are not epistemically independent. ♦

This example provides an instance of the phenomena of *dilation*, where by updating our beliefs we make them more imprecise. We indeed have

$$\overline{P}(\text{ the first ball is green} \mid \text{the second ball is green}) = 2/3 \text{ and}$$
$$\underline{P}(\text{ the first ball is green} \mid \text{the second ball is green}) = 1/3,$$

so knowing the color of the second ball makes our beliefs about the probability of drawing a red ball in the first urn change from being precise in 0.5 to belong to the interval $[\frac{1}{3}, \frac{2}{3}]$. See [54] for more information about dilation.

## 4.3 Independent envelopes

As we said before, in the classical case, if all the marginals are everywhere non-zero, then independence can be expressed equivalently through the decomposition and the marginalization notions. These two properties, however, are no longer equivalent in the imprecise case, and we can have two epistemically independent experiments for which $\underline{P}(x_1, x_2) \neq \underline{P}_1(x_1) \cdot \underline{P}_2(x_2)$ for some $x_1 \in \mathscr{X}_1, x_2 \in \mathscr{X}_2$.

The first two notions we have considered are based on extending the marginalization property to the imprecise case. Next, we introduce two conditions of independence which are based on the decomposition property. These conditions are only applied to the extreme points of the set $\mathscr{M}(\underline{P})$, because these keep all the information about $\underline{P}$ and moreover it can be checked that a decomposition property cannot hold for all the elements of the convex set $\mathscr{M}(\underline{P})$.

We say that $\underline{P}$ is an *independent envelope* when every extreme point $P$ of $\mathscr{M}(\underline{P})$ factorizes as $P = P_1 \times P_2$, where $P_1$ is the marginal distribution of $P$ on $\mathscr{X}_1$ and $P_2$ is the marginal distribution of $P$ on $\mathscr{X}_2$. This concept is called *independence in the selection* in [10] and *type-2 independence* in [14].

It is easier to understand the ideas behind this notion if we consider the Bayesian sensitivity analysis interpretation of our beliefs. Assume then that we have two experiments $X_1$, $X_2$, and that the probability modelling the first experiment belongs to some set $\mathscr{M}(\underline{P}_1)$, while the probability modelling the second experiment belongs to another set $\mathscr{M}(\underline{P}_2)$. Let $\mathscr{M}(\underline{P})$ be the set of possible models for the behavior of the two experiments, taken together. This set of models can be characterized by its set of extreme points. Independence in the selection means that each of these extreme points satisfies the classical notion of independence, in the sense that it can be written as a product of its marginals.

*Example 5(cont.).* Assume now that in the first urn we have one red ball, one green ball and one ball of unknown color, and that in the second and third urns we have two green balls, one red ball and two balls of unknown color. Then the variables $X_1$ and $X_2$ are epistemically independent: reasoning as in the example in Section 4.2 we deduce that the color of the first ball drawn is irrelevant for the second; conversely, we have that $\underline{P}(\text{ the first ball is green }) = \underline{P}(\text{the first ball is red }) = \frac{1}{3}$, and we deduce from the same example that the color of the second ball is epistemically irrelevant for the first.

Let $P$ be the probability distribution associated with the following composition or the urns:

| Urn 1 | Urn 2 | Urn 3 |
|---|---|---|
| 2 red, 1 green | 2 green, 3 red | 4 green, 1 red |

It is easy to check that this is an extreme point of the set of all the probability distributions compatible with the available information (i.e., those associated with the possible compositions of the urns). However, it does not factorize: we have for instance

$$P(\text{ the 2nd ball is green and the 1st is red })$$
$$= P(\text{ the 2nd ball is green}|\text{ the 1st is red }) \cdot P(\text{ the 1st ball is red }) = \frac{4}{5} \cdot \frac{2}{3} = \frac{8}{15},$$

while

$$P(\text{ the 2nd ball is green}) \cdot P(\text{ the 1st ball is red }) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}.$$

Hence, the set of possible models is not an independent envelope. ♦

It is important to remark that this condition does not prevent the existence of some dependence between the experiments: we could think for instance of two urns with one ball of the same color, but such that we do not know if it is red or green. Then knowing the color of the ball in the first urn completely determines the color of the second, so there is no independence; however, the set of possible models is an independent envelope. Another example, where we do not condition on events with lower probability zero, is given in next section.

## 4.4 Strong independence

Independence in the selection implies that $\mathcal{M}(\underline{P})$ is included in the convex hull of the product set

$$\{P_1 \times P_2 : P_1 \in \mathcal{M}(\underline{P}_1), P_2 \in \mathcal{M}(\underline{P}_2)\},$$

but the two sets are not necessarily equal. When we do have the equality, we say that the two experiments are *strongly independent*. This is called *type-3 independence* in [14]. This equality means that if we consider a compatible model with our beliefs for each of the experiments, then we can construct their independent product and we obtain a compatible model with the information we have about the behavior of the two experiments, taken together. Hence, we do not have any information about the experiments that allows us to rule out any of the combinations of the marginal distributions.

Note nevertheless that this information may be misleading: if we have no information at all about the two experiments, we should consider the set of all previsions on $\mathcal{L}(\mathcal{X}_1 \times \mathcal{X}_2)$, and this set satisfies strong independence. Hence, strong independence implies that we have no information pointing towards some dependence between the two experiments.

*Example* 6. Consider that we have two urns with one red ball, one green ball and one ball of unknown color, either red or green, but the same in both cases. The set of possible compositions is given by the following table:

Let us select a ball at random from each of the urns, and let $X_i$ be the random variable denoting the color of the ball selected in the urn $i$, for $i = 1, 2$. Let $\underline{P}$ be the

| | Urn 1 | Urn 2 |
|---|---|---|
| Composition 1 | 2 red, 1 green | 2 red, 1 green |
| Composition 2 | 1 red, 2 green | 1 red, 2 green |

lower prevision associated with $(X_1, X_2)$, and let $\underline{P}_i$ be its marginals, representing our information about the outcome of $X_i$, for $i = 1, 2$. The associated lower prevision $\underline{P}$ is the lower envelope of the previsions $P, P'$ that are determined by the compositions in the table above. Both these previsions factorise, and from this we can deduce that $\underline{P}$ is an independent envelope. However, we do not have strong independence: consider the previsions $P_1$, $P_2$ given by $P_1(\text{red}) = 1/3$, $P_2(\text{red}) = 2/3$. Then $P_1 \geq \underline{P}_1$ and $P_2 \geq \underline{P}_2$. However, their product satisfies $(P_1 \times P_2)(\text{ first red, second green }) = 1/9 < 2/9 = \underline{P}(\text{ first red, second green })$. ♦

To summarize, and assuming that we are not conditioning on sets of probability zero, we have that Strong independence ⇒ Independence in the selection ⇒ Epistemic independence ⇒ Epistemic irrelevance, and all these conditions are equivalent in the precise case.

There are other conditions of independence that can be considered for convex sets of probabilities, or coherent lower previsions [10, 14]. In particular, the important notion of conditional independence is studied in [14]. On the other hand, it has also been studied how to extend the assessment given by marginal distributions under some assumption of independence. This is the basis for the notions of independent natural extension, and type-1 or type-2 products [71, Chapter 9]), that we do not present here.

# 5 Relationships with other uncertainty models

In this section, we survey briefly a number of mathematical models which are closely related to the Walley's behavioural theory of coherent lower previsions. We refer also to [71, Chapter 5] for some additional discussion.

## 5.1 The work of de Finetti

As we mentioned in the introduction, the behavioural theory of coherent lower previsions is based on the behavioural approach to subjective probability, championed by Bruno de Finetti [21, 23]. In this approach, the probabilistic information is represented by means of a *prevision* on a set of gambles. This prevision can be interpreted as an expectation, but bearing in mind that it is defined directly, without using integration. From a mathematical point of view this distinction is not important in the precise case, because the restriction of the prevision to events determines the prevision on all gambles using integration; however, lower previsions are in general more informative than their restrictions to events, as we have said.

De Finetti defined the prevision of a gamble $f$ for a subject as the unique value $\mu$ such that he accepts the transaction $c(f - \mu)$ for every real number $c$: this means that he is disposed to buy $f$ for the price $\mu$ and also to sell it for the same price, so $\mu$ is a *fair price* for the gamble $f$. Then a prevision is coherent when there is no combination of acceptable transactions which leads to a sure loss. Mathematically, this is expressed by

$$\sup \sum_{i=1}^{n} \lambda_i (f_i - P(f_i)) \geq 0$$

for every natural number $n$, real $\lambda_1, \ldots, \lambda_n$ and gambles $f_1, \ldots, f_n$ in $\mathscr{K}$. It is easy to see that this condition is equivalent to the coherence condition we have given in Eq. (2) for linear previsions.

Even if most of de Finetti's work on coherent previsions can be embedded in Walley's theory of coherent lower and upper previsions, Walley's work is not a generalization of de Finetti's work, because there are substantial differences between the two:

- The most fundamental one is that Walley does not require our subject to determine a fair price for all gambles $f$ in the domain. See [71, Section 5.7] for Walley's arguments against the assumption of precision.

- Although he did not provide a detailed theory of conditional previsions, his interpretation for them is slightly different from Walley's: he regards them as previsions defined on a family of conditional events $E|H$, where $E|H$ is a logical entity which is true if both $E$ and $H$ are true, false if $H$ is true and $E$ is false, and void if $H$ is false.

- Another difference between the two approaches is their regard towards the condition of conglomerability, which is rejected by de Finetti in [23, Section 4.19]: he gives two examples where conglomerability is not satisfied. One argument here is that full conglomerability rules out linear previsions which are not countably additive (see Section 3.3 for a more precise formulation). The key of the matter, as Walley says in [71, p. 327] is that

  > "Those who insist on additivity and linearity must either reject conglomerability or the possibility of extensions to larger domains."

## 5.2 The work of Williams

A first approach to generalizing de Finetti's work taking into account the presence of imprecision was considered by Peter Williams in [84, 85, 86] (see also [67] for a recent review of Williams' work). It was also motivated by the treatment of conditional previsions in de Finetti's work. Williams defines a coherent conditional lower prevision by means of a set of acceptable gambles, in the sense that $\underline{P}(f|B)$ is defined as the supremum value of $\mu$ such that $I_B(f-\mu)$ is an acceptable transaction for our subject. In particular, if we consider $B=\Omega$, $\underline{P}(f|\Omega)$ agrees with Walley's behavioural interpretation of a lower prevision (see [86] for a more detailed account of the unconditional case). Similarly, $\overline{P}(f|B)$ is defined as the infimum value of $\mu$ such that the gamble $I_B(\mu - f)$ is acceptable for our subject, which in the case of unconditional lower previsions ($B=\Omega$) becomes the infimum selling price for $f$.

In order to define coherence, he requires the set of acceptable gambles to satisfy conditions which are slightly weaker than (D0)–(D3).[8] He proves then that a necessary condition for coherence is that

$$\sup_{x \in \cup_{i=0}^{n} B_i} [\sum_{i=1}^{n} \lambda_i I_{B_i}(f_i - \underline{P}(f_i|B_i)) - \lambda_0 I_{B_0}(f_0 - \underline{P}(f_0|B_0))](x) \geq 0, \tag{13}$$

for every natural number $n$, $B_i$ included in $\Omega$, non-negative real $\lambda_i$ and $f_i$ in the domain of $\underline{P}(\cdot|B_i)$ for $i = 0, \ldots, n$. This condition is sufficient if we require the set of acceptable gambles to satisfy (D0)–(D3).

Let us compare this Condition, that we shall call $W$-coherence, with Walley's notion of coherence for conditional lower previsions (11). We see that, under Walley's terminology (see Section 3.2), Condition (13) corresponds to the particular case of Equation (11) where the sets $S_i(f_i)$ are finite for all $i$. Hence, if we consider a number of conditional lower previsions $\underline{P}(\cdot|\mathscr{B}_1), \ldots, \underline{P}(\cdot|\mathscr{B}_n)$ where $\mathscr{B}_1, \ldots, \mathscr{B}_n$ are finite partitions of $\Omega$, then they are coherent in the sense of Walley if and only if they are W-coherent. However, for arbitrary (not necessarily finite) partitions $\mathscr{B}_1, \ldots, \mathscr{B}_n$, Walley's notion of coherence is stronger than that of Williams.

The use of W-coherence for conditional lower previsions has a number of technical advantages over Walley's condition:

- W-coherent conditional lower previsions are always lower envelopes of sets of W-coherent conditional linear previsions ([85, Theorem 2], see also [71, Section 8.1]), and as a consequence they can be given a Bayesian sensitivity analysis interpretation.

- We can always construct the smallest W-coherent extension of a W-coherent conditional lower prevision [85, Theorem 1]. Moreover, this result can be generalised towards unbounded gambles [63].

There are also, however, a number of drawbacks:

- If we express the condition using Walley's terminology, it assumes the finiteness of the sets $S_i(f_i)$ involved, whereas for many statistical applications it is important to consider non-finite spaces and partitions.

- It does not satisfy in general the conglomerability condition, as a consequence when we consider infinite partitions we may end up with acceptable transactions (W-coherent) which incur in a sure loss [71, Example 6.6.9].

As a summary, Williams' definition of coherence is a valid and interesting alternative to Walley, that avoids some of the drawbacks of Walley's approach to conditional lower previsions. Moreover, it seems to be better suited for extending the theory to unbounded gambles. The choice between the two is based in whether we accept full conglomerability as a rationality requirement, which means essentially that we assume countable additivity of the involved previsions (but without measurability requirements): see Section 3.3 and [71, Section 6.9] for more details.

[8] The difference is that he does not require a gamble $f \neq 0$ with $\inf_{x \in supp(f)} f(x) = 0$ to be accepted, nor a gamble $f \neq 0$ with $\sup_{x \in supp(f)}(x) = 0$ to be rejected, where $supp(f) = \{\omega \in \Omega : f(\omega) \neq 0\}$ is the *support* of $f$. This is related also to Walley's discussion of *really desirable* gambles [71, Appendix F].

## 5.3 Kuznetsov's work on interval-valued probabilities

In a relatively unknown book on interval valued probabilities [40], the Russian mathematician V. Kuznetsov established, in parallel to Walley, a theory of interval-valued probabilities and previsions which has many things in common with the behavioural theory we have presented. Starting from some axioms which are equivalent to Walley, he deduces many of the properties that can also be found in Walley's book. He also obtains some interesting limit results for coherent lower and upper previsions. The main differences between both theories are:

- Kuznetsov does not consider the behavioural interpretation for coherent lower and upper previsions in terms of buying and selling prices.
- He makes some assumptions in the domains, and therefore we cannot consider previsions with arbitrary domains as is the case with Walley's theory.
- His theory is valid for unbounded gambles, and the domain of his upper previsions contains at least those gambles which are bounded above (and similarly, the domain of his lower previsions includes at least the gambles which are bounded below).
- When conditioning on a set $B$ of lower probability zero and positive upper probability, he suggests taking the limit on conditional dominating previsions, which give positive probability to $B$. This is related to the notion of *regular extension* in Walley [71, Appendix K].

See [41, 42] for two short papers in English presenting Kuznetsov's work, and [12, 66] for some works derived from his theory.

## 5.4 Weischelberger's $F$- and $R$-probabilities

Recently, Kurt Weischleberger and some of his colleagues have also established a theory of interval-valued probabilities [79, 80, 81] that is more statistically oriented than Walley's. This theory considers two non-additive measures $L \leq U$ determining the lower and upper bounds for our probabilities. The set of (finitely additive) probabilities $\mathscr{M}$ that lie between $L$ and $U$ are called the *structure* determined by $L, U$.

$L$ and $U$ are said to determine an *R-probability* with structure $\mathscr{M}$ when $\mathscr{M} \neq \emptyset$, and an *F-probability* when $L, U$ are the infimum and the supremum of the probabilities in $\mathscr{M}$, respectively. Hence, these concepts are equivalent to the notions of avoiding sure loss and coherence for imprecise probabilities.

Weischleberger's theory is more related to the classical works on probability theory, in the sense that it focuses on probabilities instead of previsions and that it also makes some measurability assumptions that are not present in Walley's theory. Another difference is that Weischelberger's work does not use a subjective interpretation of lower and upper probabilities, and it allows to connect the notions of avoiding sure loss and coherence with a frequentist interpretation of probability (see also [69] for another paper in this direction). The treatment of conditional lower and upper probabilities and of the notion of independence is also different [82].

See [2, 82] for some recent work based on this theory.

## 5.5 The Dempster-Shafer theory of evidence

Next, we outline the main features of the Dempster-Shafer theory of evidence [57]. The origin of this theory is in Dempster's work on random sets, or measurable multi-valued mappings [24, 25].

Let $\mathscr{A}$ be a field of subsets of $\Omega$, and consider a lower probability $\underline{P} : \mathscr{A} \to [0,1]$ that satisfies $\underline{P}(\emptyset) = 0, \underline{P}(\Omega) = 1$. $\underline{P}$ is said to be *completely monotone* if it is monotone and for any events $A_1, \ldots, A_n$ in $\mathscr{A}$ and $n \geq 2$,

$$\underline{P}(\cup_{i=1}^{n} A_i) \geq \sum_{I \subseteq \{1,\ldots,n\}} (-1)^{|I|+1} \underline{P}(\cap_{i \in I} A_i), \tag{14}$$

where $|I|$ denotes the cardinality of the set $I$. A lower probability that satisfies Equation (14) for all $k \leq p$ is called $p$-monotone. A complete monotone lower probability is then one that is $n$-monotone for every natural number $n$. When $\Omega$ is a finite space, and the field $\mathscr{A}$ is the class of all subsets of $\Omega$, a completely monotone lower probability is called a *belief function*. A belief function is determined by its so-called *basic probability assignment m*, which is given by

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \underline{P}(B).$$

Then $\underline{P}$ is given, for all $A \subseteq X$, by $\underline{P}(A) = \sum_{B \subseteq A} m(B)$. The upper probability that we can generate from a completely monotone lower probability using conjugacy is said to be *completely alternating*, and, in the particular case where $\Omega$ is finite, a *plausibility function*. The relationship between a plausibility function $\overline{P}$ and the basic probability assignment $m$ of its conjugate belief function $\underline{P}$ is $\overline{P}(A) = \sum_{B \cap A \neq \emptyset} m(B)$. The basic probability assignment of a set $A$ can be interpreted as the measure of belief that is committed exactly to $A$. Then $\underline{P}(A)$ would be our subject's belief that the result of the experiment belongs to the set $A$, while $\overline{P}(A)$ is the measure of his belief that the outcome of the experiment may belong to $A$ (or, perhaps more clearly, the measure of belief that does not rule out that the outcome is in $A$). This is the so-called *evidential interpretation*.

From the behavioural point of view, multi-valued mappings have been studied in [71, Section 4.3.5] and in [47]. On the other hand, 2- (and in particular completely-) monotone functions were studied by Walley in [68]. Some of his results generalize to the case where the domain is a lattice of events instead of a field [18]. It can be checked [68] that a 2-monotone lower probability is in particular coherent, and therefore so are $n$- and completely monotone capacities. More generally, we can consider $n$-monotone lower *previsions*, which appeared in [5, Chapter 3] as functionals between an Abelian semigroup and an Abelian group, and were studied from a behavioural point of view in [17, 18]. Hence, from a static point of view (without considering updating or combination of information) Dempster-Shafer theory can be embedded into Walley's theory of coherent lower previsions.

One of the main features of the evidential theory, and Walley's main point of criticism (see [71, Section 5.13] and [74]) is *Dempster's Rule of Combination* [24, 57]. It is used to combine several belief functions which are based on different sources of information into one belief function. Let $\underline{P}_1, \underline{P}_2$ be two belief functions with basic probability assignments $m_1, m_2$, respectively. The combined belief function $\underline{P}$ is the one related to the basic probability assignment

$$m(C) = \sum_{A \cap B = C} m_1(A) m_2(B), \tag{15}$$

where, if necessary, we normalize it by multiplying by

$$\left[1 - \sum_{A \cap B = \emptyset} m_1(A) m_2(B)\right]^{-1},$$

which is a measure of the extent of the conflict between $\underline{P}_1$ and $\underline{P}_2$.

This rule of combination is only reasonable for Walley provided we make some additional assumptions on the relationship between the experiments that underlie $\underline{P}_1, \underline{P}_2$ ([71, Section 5.13.7]). In particular, the conditioning rule that we may derive from it may fail to produce conditional lower previsions which are coherent with the unconditional ones. We may have to make some extra assumptions of conditional independence if we want it to satisfy coherence, and this requirement may be unreasonable in some cases. See also [44, 50, 87] for further criticism of this rule.

## 5.6 The game-theoretic approach to imprecise probabilities

We would like to conclude this section by making a brief summary of recent work by Shafer and Vovk [59], where a connection is established between probability and finance through a game-theoretic approach to coherent lower previsions. The ideas of this approach are roughly similar to the ones behind the behavioural theory we have outlined: we also have here acceptable transactions and study whether they can lead to a sure loss. There are, however, a number of important differences between the two: Shafer and Vovk consider a game between two players, whereas Walley's theory is focused on one subject who establishes buying and selling prices;[9] there is no assumption of boundedness of the variables involved; and they focus on the consequences and not so much on probability.

Let us make a short summary of the main ideas in this work. We refer to [60, 59] for more details. Consider a game between two players, called World and Skeptic, who play according to a certain protocol. The first player (World) can make a number of moves, where the next move may depend on his previous moves, but not on the ones made by Skeptic. These moves can be represented in a tree, and we call a *situation* some connected path in the tree that starts at the root of the tree.

In each situation $t$, Skeptic has a number of possible moves, and a gain function associated with the move he makes and World's next move, representing the change in Skeptic capital associated with these two moves. A strategy is then a function determining Skeptic next move for each non-final situation in the tree representing World's moves, and the associated capital if he starts the game with zero capital is his capital process.

[9] A two player model is already present in [61].

Under this interpretation, Shafer and Vovk define upper and lower prices for real variables. The *upper price* for a variable $f$ in a non-final situation $t$ is the lowest price at which Skeptic can buy the variable $f$ at the end of the game, no matter what moves World makes, using some strategy. The *lower price* for a variable $f$ in a non-final situation $t$ can be given an analogous interpretation using conjugacy.

A gambling protocol is called a *probability protocol* when the following two conditions are satisfied:

(PP1) The set of possible moves for Skeptic in any situation $t$ is a convex cone.

(PP2) For each non-final situation, Skeptic's gain function is linear with respect to Skeptic's moves, for any move of World fixed.

It is called *coherent* when for each non-final situation and each possible move $s$ of Skeptic there is another move $w$ of World such that the gain associated with $(s,w)$ is non-positive.

For many of their developments Shafer and Vovk decompose World into multiple players. They consider for instance *Reality* and *Forecaster*, where Forecaster determines for each non-final situation the set of possible moves and the gain function for Skeptic, while Reality determines the move of World. Shafer and Vovk establish a number of limit results, such as weak and strong laws of large numbers, for coherent probability protocols.

Under the terminology we have considered throughout this survey, our subject would take the role of Forecaster, in the sense that he establishes the supremum buying prices for all gambles in the domain. Then, we check the coherence of these prices using the roles of Skeptic (selecting some gambles in the domain) and Reality (determining the outcome of the experiment, that is, the value $\omega$ in $\Omega$). Walley's notion of coherence means that the buying prices established by Forecaster should not be exploitable by Skeptic in order to assure himself a gain, no matter how World moves.

We see then that, although some of the concepts used by Shafer and Vovk are similar to the ones in Walley's theory of coherent lower previsions (the behavioural interpretation, the notion of coherence, etc), there are also differences between the two approaches, like the use of two players and the extension to unbounded gambles. It is therefore interesting that Shafer and Vovk's work on coherent probability protocols has been recently connected to Walley's theory of coherent lower previsions in [19]. It has been proven that for every coherent probability protocol there exists an immediate prediction model such that both lead to identical buying prices, and vice versa. See [19] for more information. In this sense, in spite of some of their differences, the results from one theory can be embedded into the other.[10]

# 6 Applications

Let us give some comments now on the use of the theory of coherent lower previsions in other contexts. Since the origins of subjective probability are in decision making, it

[10]The (possible) unboundedness of the variables in the models is overcome in the case of coherent lower previsions by working with sets of (really) desirable gambles, instead of almost-desirable ones ([19, Sections 3 and 4]).

is not really surprising that much of the work made on practical uses of coherent lower previsions, which stem from subjective probability theory, is related to decision theory. Indeed, imprecise probabilities can be seen as a more realistic model for practical decision problems, because they allow us to treat some of the most common problems that appear in practice: lack of available information, conflicting sources of beliefs, and difficulties in eliciting expert's opinions.

The theory of coherent lower previsions, in different forms, has been employed in a number of applications, in different frameworks such as the environmental [70] or the bio-medical [64, 72, 90]. One of the consequences of the use of sets of probabilities will be that sometimes we cannot decide which of two alternatives must be preferred, something which is to be expected on the other hand in the case of little information. Another feature is that there are many possible ways of generalizing the principle of maximum expected utility to coherent lower previsions, such as E-admissibility, interval dominance, $\Gamma$-maximinity and maximality, etc. We refer to [65] for an overview of the different criteria, and to [56] for a comparison of E-admissibility and $\Gamma$-maximinity.

There have also been several studies connecting the behavioural theory of coherent lower previsions with different aspects of statistics [73, 75]. One of the most important is the work in [73], where Walley defined the so-called Imprecise Dirichlet Model. This is a model for inference which relates past and future observations about some variables. See also [3, 76] for additional information about this model. The Imprecise Dirichlet Model has been applied in different contexts, such as classification [1, 91, 90], reliability analysis [9] or game theory [52]. See also [16, 89] for further theoretical developments on this subject.

Finally, coherent lower previsions have also been used for dealing with uncertainty scenarios in climate change [39] and forestry [28].

# 7 Strengths and challenges

The results mentioned in this paper constitute an attempt to overview the main features of the behavioural theory of coherent lower previsions. As a summary, the main virtues of this theory are:

1. It is better suited for situations where the available information does not justify the use of a precise probability distribution.

2. It encompasses most of the other generalizations considered in the literature (belief functions, 2-monotone capacities, coherent lower probabilities). This is done mainly in two ways:

   - By focusing on the behavior towards variables (gambles), and therefore on the notion of expectation, without going through the behavior on events first.
   - By moving away from the measurability requirements on classical probability theory and considering previsions for arbitrary sets of gambles.

3. It provides a behavioural interpretation that leads naturally to decision making, and it can at the same time be given a sensitivity analysis representation in some cases.

Nevertheless, the theory of coherent lower previsions has a number of challenges that need to be studied in more detail:

1. The concept of coherence, although a reasonable rationality requirement, may be too weak to produce informative inferences.

2. The theory is established for bounded random variables, or gambles, and the generalization towards unbounded random variables is not straightforward.

3. It assumes the linearity of the utility scale, something which may not be reasonable in practice.[11]

4. Walley's notion of natural extension of conditional lower previsions has some undesirable properties, which do not seem to be resolvable without giving up on conglomerability.

5. Computing with coherent lower previsions in practice means often solving non-linear problems, whence for big problems the computational costs may be high.

6. There have been until now relatively few applications, and more are needed in order to effectively compare the applicability of this theory with classical tools.

7. More generally, we need to develop tools to be able to compare the effectiveness of precise and imprecise approaches in a number of situations.

8. There are several notions of independence as well as different optimality criteria, and it is not always clear which of them is better suited for each situation.

9. From the theoretical point of view there are a number of notions and results from classical probability theory which have not been generalized, such as ergodic theorems.

These problems constitute the main avenues for future research in the theory.

## Acknowledgements

I would like also to express my gratitude to Gert de Cooman, Marco Zaffalon, Roman Effremov and the anonymous referees for some very helpful comments that improved substantially earlier versions of this paper. I also acknowledge the financial support of the projects MTM2004-01269 and TSI2004-06801-C04-01.

[11] But see [51] for an interesting solution to this problem.

## References


[1] A. Antonucci, A. Salvetti, and M. Zaffalon. Credal networks for hazard assessment of debris flows. In J. Kropp and J. Scheffran, editors, *Advanced Methods for Decision Making and Risk Management in Sustainability Science*, pages 237–256. Nova Science, 2007.

[2] T. Augustin and F. Coolen. Non parametric predictive inference and interval probability. *Journal of Statitical Planning and Inference*, 124:251-272, 2004.

[3] J. M. Bernard. An introduction to the imprecise Dirichlet model for inference with multinomial data. *International Journal of Approximate Reasoning*, 39(2–3):123–150, 2005.

[4] E. Borel. *Probabilities and life*. Dover, New York, 1962.

[5] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier. Université de Grenoble*, 5:131–295, 1953.

[6] G. Coletti. Coherent numerical and ordinal probability assessments. *IEEE Transactions on Systems, Man and Cybernetics*, 24:1747–1754, 1994.

[7] G. Coletti and R. Scozzafava. Zero probabilities in stochastic independence. In R.R. Yager, B. Bouchon-Meunier and L.A. Zadeh, editors, *Information, Uncertainty, Fusion*, pages 185–196. Kluwer, 2000.

[8] G. Coletti and R. Scozzafava. *Probabilistic logic in a coherent setting*. Kluwer, New York, 2002.

[9] F. Coolen. An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety*, 56:61–68, 1997.

[10] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5(2):165–181, 2000.

[11] F. Cozman. Irrelevance and independence relations in Quasi-Bayesian networks. In G.F. Coopen and S. Moral, editors, *Proceedings of the 14th conference on Uncertainty in Artificial Intelligence*, pages 89–96, 1998.

[12] F. Cozman. Constructing sets of probability measures through Kuznetsov's independence condition. In *Proceedings of the 2nd International Symposium on Imprecise Probabilities and Their Applications*, pages 104–111, 2001.

[13] F. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45(1–2):173–195, 2005.

[14] L. M. de Campos and S. Moral. Independence concepts for convex sets of probabilities. In *Proceedings of 11th Conference on Uncertainty in Artificial Intelligence*, pages 108–115, 1995.

[15] G. de Cooman. Possibility theory I-III. *International Journal of General Systems*, 25(4):291–371, 1997.

[16] G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, 159(1–2):75–125, 2004.

[17] G. de Cooman, M. C. M. Troffaes, and E. Miranda. $n$-Monotone lower previsions and lower integrals. In *Proceedings of the 4th International Symposium on Imprecise Probabilities and Their Applications*, pages 145–154. 2005.

[18] G. de Cooman, M. C. M. Troffaes, and E. Miranda. $n$-Monotone lower previsions. *Journal of Intelligent and Fuzzy Systems*, 16(4):253–263, 2005.

[19] G. de Cooman and F. Hermans. On coherent immediate prediction: connecting two theories of imprecise probability. In *Proceedings of the 5th International Symposium on Imprecise Probabilities and Their Applications*, pages 97–106, 2007.

[20] G. de Cooman and E. Miranda. Symmetry of models versus models of symmetry. In W. L. Harper and G. R. Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, pages 67–149. King's College Publications, 2007.

[21] B. de Finetti. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937.

[22] B. de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[23] B. de Finetti. *Theory of probability*. John Wiley and Sons, Chichester, 1974–1975. English translation of [22].

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

[25] A. P. Dempster. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39:957–966, 1968.

[26] D. Denneberg. *Non-additive measure and integral*. Kluwer Academic Publishers, Dordrecht, 1994.

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

[28] M. Ducey. Imprecision in a timber asset sale model: motivation, specification, and behavioural implications. In *Proceedings of 2nd International Symposium on Imprecise Probabilities and Their Applications*, pages 162–170, 2001.

[29] S. Ferson, V. Kreinovich, L. Ginzburg, D. Myers, and Kari Sentz. Constructing probability boxes and Dempster-Shafer structures. Technical Report SAND2002–4015, Sandia National Laboratories, January 2003.

[30] T. Fine. *Theories of Probability*. Academic Press, New York, 1973.

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

[32] M. Grabisch, H. T. Nguyen, and E. A. Walker. *Fundamentals of uncertainty calculi with applications to fuzzy inference*. Kluwer Academic Publishers, Dordrecht, 1995.

[33] J. M. Keynes. *A Treatise on Probability*. McMillan, London, 1921.

[34] F. H. Knight. *Risk, Uncertainty and Profit*. University of Chicago Press, 1971. First published in 1921.

[35] B. Koopman. The axioms and algebra of intuitive probability. *Annals of Mathematics*, 41:269–292, 1940.

[36] B. Koopman. The bases of probability. *Bulletin of the American Mathematical Society*, 46:763–774, 1940.

[37] B. Koopman. Intuitive probabilities and sequences. *Annals of Mathematics*, 42:169–187, 1941.

[38] P. H. Krauss. Representation of conditional probability measures on Boolean algebras. *Acta Mathematica Academiae Scientiarum Hungaricae*, 19:229–241, 1968.

[39] E. Kriegler. Utilizing belief functions for the estimation of future climate change. *International Journal of Approximate Reasoning*, 39(2–3):185–209, 2005.

[40] V. Kuznetsov. *Interval statistical models*. Radio and communication, 1991. In Russian.

[41] V. Kuznetsov. Interval methods for processing statistical characteristics. *Reliable Computing*, supplement on Extended Abstracts of APIC'95, pages 116–122, 1995.

[42] V. Kuznetsov. Auxiliary problems of statistical data processing: interval approach. *Reliable Computing*, supplement on Extended Abstracts of APIC'95, pages 123–129, 1995.

[43] H. E. Kyburg and H.E. Smokler. *Studies in subjective probability*. Wiley, New York, 1964.

[44] H. E. Kyburg. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence Journal*, 31:271–294, 1987.

[45] I. Levi. *The enterprise of knowledge*. MIT Press, Cambridge, 1980.

[46] E. Miranda and G. de Cooman. Coherence and independence on non-linear spaces. Technical Reports on Statistics and Decision Sciences, TR05/10, May 2005. http://www.deio.urjc.es/publicaciones.php

[47] E. Miranda, G. de Cooman, and I. Couso. Lower previsions induced by multi-valued mappings. *Journal of Statistical Planning and Inference*, 133:173–197, 2005.

[48] E. Miranda and G. de Cooman. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 46(1): 188-225, 2007.

[49] S. Moral. Epistemic irrelevance on sets of desirable gambles. *Annals of Mathematics in Artificial Intelligence*, 45(1-2):197–214, 2005.

[50] J. Pearl. Reasoning with belief functions: an analysis of compatibility. *International Journal of Approximate Reasoning*, 4:363–390, 1990.

[51] R. Pelessoni and P. Vicig. Uncertainty modelling and conditioning with convex imprecise previsions. *International Journal of Approximate Reasoning*, 39 (2–3):297–319, 2005.

[52] E. Quaeghebeur and G. de Cooman. Game theoretic learning using the imprecise Dirichlet model. In *Proceedings of the 3rd International Symposium on Imprecise Probabilities and Their Applications*, pages 452–466, 2003.

[53] D. Ríos Insua and F. Ruggieri. *Robust Bayesian Analysis*. Springer, 2000.

[54] T. Seidenfeld and L. Wasserman. Dilation for sets of probabilities. *Annals of Statistics*, 21(3):1139–1154, 1993.

[55] T. Seidenfeld, M. Schervisch and J. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23(6):2168–2217, 1995.

[56] T. Seidenfeld. A contrast between two decision rules for use with (convex) sets of probabilities: Gamma-maximin versus E-admisibility. *Synthese*, 140(1–2):69–88, 2004.

[57] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, New Jersey, 1976.

[58] G. Shafer. Non-additive probabilities in the work of Bernoulli and Lambert. *Archive for History of Exact Sciences*, 19(4):309–370, 1978.

[59] G. Shafer and V. Vovk. *Probability and finance: it's only a game!* John Wiley and Sons, New York, 2001.

[60] G. Shafer, P. Gillett, and R Scherl. A new understanding of subjective probability and its generalization to lower and upper prevision. *International Journal of Approximate Reasoning*, 33:1–49, 2003.

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

[62] M. Troffaes and G. de Cooman. Extensions of coherent lower previsions to unbounded random variables. In B. Bouchon-Meunier, L. Foulloy, and R. Yager, editors, *Intelligent systems for information processing: from representation to applications*, pages 277–288. North-Holland, Amsterdam, 2003.

[63] M. Troffaes. Conditional lower previsions for unbounded random quantities. In J. Lawry, E. Miranda, A. Bugarin, S. Li, M.A. Gil, P. Grzegiorzewski, and O. Hryniewicz, editors, *Soft Methods in Integrated Uncertainty Modelling*, pages 201–210. Springer, 2006.

[64] M. Troffaes. Efficient and robust global amino acid sequence alignment with uncertain evolutionary distances. In B. Bouchon-Meunier, G. Coletti, and R. Yager, editors, *Modern Information Processing: from Theory to Applications*, pages 371–381. Elsevier, 2006.

[65] M. Troffaes. Decision making under uncertainty using imprecise probabilities. *International Journal of Approximate Reasoning*, 45(1):17–29, 2007.

[66] L. Utkin and I. Kozine. Different faces of the natural extension. In *Proceedings of the 2nd International Symposium on Imprecise Probabilities and Their Applications*, pages 316–323, 2001.

[67] P. Vicig, M. Zaffalon, and F. Cozman. Notes on 'Notes on conditional previsions'. *International Journal of Approximate Reasoning*, 44(3):358–365, 2007.

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

[69] P. Walley and T. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10:741–761, 1982.

[70] P. Walley and F. Campello de Souza. Uncertainty and indeterminacy in assessing the economic viability of energy options: a case study of solar heating systems in brazil. *Energy, Systems and Policy*, 14:281–304, 1990.

[71] P. Walley. *Statistical reasoning with imprecise probabilities*. Chapman and Hall, London, 1991.

[72] P. Walley, L. Gurrin, and P. Burton. Analysis of clinical data using imprecise prior probabilities. *The Statistician*, 45:457–485, 1996.

[73] P. Walley. Inferences from multinomial data: learning about a bag of marbels. *Journal of the Royal Statistical Society, B*, 58:3–57, 1996.

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

[75] P. Walley. A bounded derivative model for prior ignorance about a real-valued parameter. *Scandinavian Journal of Statistics*, 24:463–483, 1997.

[76] P. Walley and J. M. Bernard. Imprecise probabilistic prediction for categorical data. Technical Report CAF9901, Université Paris 8, 1999.

[77] P. Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[78] P. Walley, R. Pelessoni, and P. Vicig. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, 126:119–151, 2004.

[79] K. Weischelberger. Axiomatic foundations of the theory of interval probability. In *Proceedings of the 2nd Gauss Symposium*, pages 47–64, 1995.

[80] K. Weischelberger. The theory of interval probability as a unifying model for uncertainty. *International Journal of Approximate Reasoning*, 24:149–170, 2000.

[81] K. Weichselberger. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I: Intervallwahrscheinlichkeit als umfassendes Konzept*, in cooperation with T. Augustin and A. Wallner. Physica, Heidelberg, 2001.

[82] K. Weischleberger. The logical concept of probability: foundation and interpretation. In *Proceedings of the 5th International Symposium on Imprecise Probabilities and Their Applications*, pages 455–464, 2007.

[83] P. Whittle. *Probability via expectation*. Springer, 2000. 4th edition.

[84] P. Williams. Coherence, strict coherence and zero probabilities. In *Proceedings of the Fifth International Congress on Logic, Methodology and Philosohy of Science*, pages 29–33, 1975.

[85] P. Williams. Notes on conditional previsions. Technical report, University of Sussex, 1975. Reprinted in a revised form in the International Journal of Approximate Reasoning, 44(3):366-383, 2007.

[86] P. Williams. Indeterminate probabilities. In M. Przelecki, K. Szaniawski, and R. Wojcicki, editors, *Formal methods in the methodology of empirical sciences*, pages 229–246. D. Reidel Publishing Company, 1976.

[87] P. Williams. On a new theory of epistemic probability. *British Journal for the Philosphy of Science*, 29(4):375–378, 1978.

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

[89] M. Zaffalon. Exact credal treatment of missing data. *Journal of Statistical Planning and Inference*, 105(1):105–122, 2002.

[90] M. Zaffalon, K. Wesnes, and O. Petrini. Reliable diagnoses of dementia by the naive credal classifier from incomplete cognitive data. *Artificial Intelligence in Medicine*, 29(1–2):61–79, 2003.

[91] M. Zaffalon. Credible classification for environmental problems. *Environmental modelling and software*, 20(8):1003–1012, 2005.