Lower previsions induced by multi-valued mappings*

Enrique Miranda[†], Gert de Cooman [‡] and Inés Couso [§]

Abstract

We discuss how lower previsions induced by multi-valued mappings fit into the framework of the behavioural theory of imprecise probabilities, and show how the notions of coherence and natural extension from that theory can be used to prove and generalise existing results in an elegant and straightforward manner. This provides a clear example for their explanatory and unifying power.

Keywords: Coherence, conditioning, evidence theory, imprecise probabilities, lower inverse, lower prevision, multi-valued mapping, natural extension, random set.
AMS Classification: 68T30, 54C60, 68T37.

1 Introduction

The term 'imprecise probabilities' covers different mathematical models such as upper and lower probabilities induced by multi-valued mappings ([10], [24]), upper and lower expectations ([4], [11], [26]), sets of probability measures ([3], [5], [12], [19]), upper and lower previsions, sets of desirable gambles, and preference orderings [28]. These models arise as an alternative to, or as an extension of, the classical or precise probability theory ([7], [17]), which in a number of situations makes assumptions that are arguably too strict in order to model the available information. Several such uncertainty models give different interpretations to lower and upper probabilities. Two prominent types of interpretation are the evidential ([24]) and the behavioural ([28]). The former regards the imprecise probability of an event as a link between the event and the available evidence, while the latter interprets the probability in terms of behaviour.

*This is an expanded and updated version, with proofs, of a conference paper [22] presented at the Ninth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002, Annecy, France, July 1–5, 2002).

[†]Corresponding author: Department of Informatics, Statistics and Telematics, Rey Juan Carlos University, C-Tulipán, s/n. 28933 Móstoles (Spain). Tel: (+34) 914888132, fax: (+34) 914887338. e-mail: emiranda@escet.urjc.es

[‡]SYSTeMS Research Group, Ghent University, Belgium. e-mail:gert.decooman@ugent.be

[§]Department of Statistics and Operations Research, University of Oviedo, Spain. e-mail: couso@pinon.ccu.uniovi.es

Random sets, and multi-valued mappings in general, have been used successfully to model imprecision and uncertainty in the relation between the elements of two different spaces. This is for instance made evident in [18] and [20]. The upper and lower probabilities induced by a multi-valued mapping ([9], [10]) were given an evidential interpretation by Shafer ([24]). However, as far as we are aware, they have not been connected with the behavioural interpretation of imprecise probabilities in any thorough or detailed manner, nor have the mathematical consequences of such a connection been systematically explored. The problem that we are treating in this paper can be summarised through the following types of questions: (i) what does it mean, in terms of the behaviour of a subject, that his lower probability for an event, induced by a random set, is 0.4; and (ii) can we use the information represented by a random set to deduce not only lower and upper probabilities of events, but also lower and upper previsions for random variables?

To answer these questions, we intend to establish a link between random sets and the behavioural theory of imprecise probabilities ([28]). To our knowledge, the only previous work in this direction was done by Walley ([28, Section 4.3.5]), and De Cooman ([6]). Though we derive a basic formula (see Eq. (8) further on) that is essentially the same as Walley's, our course of reasoning is different, and our analysis is more detailed.

The paper is organised as follows: in Section 2, we give a brief review of the main ideas behind the behavioural interpretation of imprecise probabilities that we shall need in the rest of the paper. In Section 3, we recall the definition of lower and upper probabilities induced by a random set, and we give a first and immediate generalisation. Section 4 gives a fairly general treatment of how to use the information conveyed by a multi-valued mapping in the context of the behavioural theory of imprecise probabilities. This discussion allows us to give a satisfactory answer to the two types of questions mentioned above. It also allows us to prove generalisations of a number of classical results on random sets in the subsequent Section 5.

2 Basic notions from the behavioural theory of imprecise probabilities

For a proper understanding of the course of reasoning in this paper, more than a casual acquaintance is required with the basic ideas underlying the behavioural theory of imprecise probabilities. We refer to Walley's book ([28]) for extensive discussion and motivation, and for many of the results and formulae that we shall use below. Our aim in this section is to familiarise the reader with a number of these ideas, in the hope that this will make the main message of the paper understandable to a wider audience. We also introduce some basic notation.

2.1 Basic notation and behavioural interpretation

Consider a subject who is uncertain about something, say, the outcome of an experiment. Let Ω be the space of all possible outcomes, then a bounded real-valued function on Ω is called a *gamble* on Ω. The set of all gambles on Ω is denoted by $\mathscr{L}(\Omega)$. A gamble is an uncertain reward: if the outcome of the experiment turns out to be $\omega \in \Omega$, then the corresponding reward will be $X(\omega)$ (positive or negative), expressed in units of some (predetermined) linear utility.

The subject's *lower prevision* $\underline{P}(X)$ for a gamble X is defined as his supremum acceptable price for buying X, i.e., it is the highest price μ such that the subject will accept to buy X for all prices strictly smaller than μ (buying X for a price x is the same thing as accepting the uncertain reward $X - x$). Similarly, a subject's *upper prevision* $\overline{P}(X)$ for X is his infimum acceptable selling price for X. Clearly, $\overline{P}(X) = -\underline{P}(-X)$ since selling X for a price x is the same thing as buying $-X$ for the price $-x$. This *conjugacy relation* implies that we can limit our attention to lower previsions.

A subset A of Ω is called an *event*, and it can be identified with its indicator (function) I_A, which is a gamble on Ω. The *lower probability* of A is nothing but the lower prevision $\underline{P}(I_A)$ of its indicator. We shall often identify an event with its indicator, and write $\underline{P}(A)$ instead of $\underline{P}(I_A)$. We prefer to work with gambles, rather than to restrict ourselves to events: as Walley has pointed out ([28]), the language of gambles is much more powerful than that of events when working with imprecise probabilities.

2.2 Rationality requirements

Assume that the subject has given lower prevision assessments $\underline{P}(X)$ for all gambles X in some set of gambles $\mathscr{K} \subseteq \mathscr{L}(\Omega)$, which need not have any predefined structure. Since these assessments represent commitments of the subject to act in certain ways, they are subject to a number of rationality requirements. The strongest such requirement is that $\underline{P}$ should be *coherent*. Coherence means first of all that the subject's assessments *avoid sure loss*: for any n in the set of positive natural numbers $\mathbb{N}$ and for any $X_1, \dots, X_n$ in $\mathscr{K}$ we require that

$$\sup_{\omega\in\Omega} \left[\sum_{k=1}^{n} [X_k(\omega) - \underline{P}(X_k)] \right] \geq 0.$$

Otherwise, there would be some $\varepsilon > 0$ such that for all ω in Ω:

$$\sum_{k=1}^{n} [X_k(\omega) - \underline{P}(X_k) + \varepsilon] \leq -\varepsilon,$$

i.e., the net reward of buying the gambles X_k for the acceptable prices $\underline{P}(X_k) - \varepsilon$ is sure to lead to a loss of at least ε, whatever the outcome of the experiment!

But coherence also means that if we consider any $X \in \mathscr{K}$, we cannot force the subject to accept X for a price strictly higher than his specified supremum buying price $\underline{P}(X)$, by exploiting buying transactions implicit in his lower previsions $\underline{P}(X_k)$ for a finite number of gambles X_k in $\mathscr{K}$, which he is committed to accept. More explicitly, we require that for any n and m in $\mathbb{N}$, and $X_0, \dots, X_n$ in $\mathscr{K}$:

$$\sup_{\omega\in\Omega} \left[\sum_{k=1}^{n} [X_k(\omega) - \underline{P}(X_k)] - m[X_0(\omega) - \underline{P}(X_0)] \right] \geq 0.$$

Otherwise, there would be some $\varepsilon > 0$ such that $m[X_0 - [\underline{P}(X_0) + \varepsilon]]$ pointwise dominates the acceptable combination of buying transactions $\sum_{k=1}^{n}[X_k - \underline{P}(X_k) + \varepsilon]$, and is therefore acceptable as well. This would mean that by combining these acceptable transactions derived from his assessments, the subject can be effectively forced to buy X_0 at the price $\underline{P}(X_0) + \varepsilon$, which is strictly higher than the supremum acceptable buying price $\underline{P}(X_0)$ that he has specified for it. This is an inconsistency that is to be avoided.

2.3 Natural extension

We can always extend a coherent lower prevision $\underline{P}$ to a coherent lower prevision $\underline{E}$ on the set of all gambles $\mathscr{L}(\Omega)$, through a procedure called *natural extension*. The natural extension $\underline{E}$ of $\underline{P}$ is the smallest coherent lower prevision on $\mathscr{L}(\Omega)$ that coincides on $\mathscr{K}$ with $\underline{P}$. It is given for all $X \in \mathscr{L}(\Omega)$ by

$$\begin{aligned}\underline{E}(X) = \sup_{\substack{n\geq 0\\ X_1,\ldots,X_n\in\mathscr{K}\\ \mu_1,\ldots,\mu_n\geq 0}} \sup\left\{\alpha\colon X-\alpha \geq \sum_{k=1}^{n}\mu_k[X_k-\underline{P}(X_k)]\right\}\\ = \sup_{\substack{n\geq 0\\ X_1,\ldots,X_n\in\mathscr{K}\\ \mu_1,\ldots,\mu_n\geq 0}} \inf_{\omega\in\Omega}\left[X(\omega)-\sum_{k=1}^{n}\mu_k[X_k(\omega)-\underline{P}(X_k)]\right], \quad (1)\end{aligned}$$

where the $\mu_1, \ldots, \mu_n$ in the suprema are non-negative real numbers. The natural extension summarises the behavioural implications of $\underline{P}$: $\underline{E}(X)$ is the supremum buying price for X that can be derived from the lower prevision $\underline{P}$ by arguments of coherence alone: we see from its definition above that it is the supremum of all prices that the subject can be effectively forced to buy the gamble X for, by combining finite numbers of buying transactions implicit in his lower prevision assessments $\underline{P}$.

2.4 Relation to precise probability theory

When $\underline{P}(X) = \overline{P}(X)$, the subject's supremum buying price coincides with his infimum selling price, and this common value is a *(linear) prevision* or *fair price* for the gamble X, in the sense of de Finetti ([7]). For fair prices, the notion of coherence essentially coincides with de Finetti's definition. This means that the theory of imprecise probabilities includes de Finetti's approach (precise previsions and/or probabilities) as a special case. But it is also much more flexible than the latter, because it doesn't require a subject's supremum buying price and his infimum selling price to coincide. More specifically, it allows that there are prices x such that $\underline{P}(X) < x < \overline{P}(X)$, i.e., for which the subject refrains from committing himself to buy or sell the gamble X for price x: he is allowed to be undecided! Indecision seems a natural thing to allow, especially if the subject has very little information about the outcome ω.

A linear prevision P on the set $\mathscr{L}(\Omega)$ can also be characterised as a linear functional that is positive (if $X \geq 0$ then $P(X) \geq 0$) and has unit norm ($P(I_\Omega) = 1$). Its restriction to events is a finitely additive probability. Let us denote by $\mathbb{P}(\Omega)$ the set of all linear previsions on $\mathscr{L}(\Omega)$.

The notions of avoiding sure loss, coherence, and natural extension can be characterised in terms of sets of linear previsions. Consider a lower prevision $\underline{P}$ defined on a set of gambles $\mathscr{K}$. Its set of dominating linear previsions is given by

$$\mathscr{M}(\underline{P}) = \{P \in \mathbb{P}(\Omega) \colon (\forall X \in \mathscr{K})(P(X) \geq \underline{P}(X))\}.$$

Then $\underline{P}$ avoids sure loss iff $\mathscr{M}(\underline{P}) \neq \emptyset$, i.e., if it has a dominating linear prevision. $\underline{P}$ is coherent iff $\underline{P}(X) = \min\{P(X) \colon P \in \mathscr{M}(\underline{P})\}$ for all X in $\mathscr{K}$, i.e., if it is the *lower envelope* of $\mathscr{M}(\underline{P})$. And the natural extension $\underline{E}$ of $\underline{P}$ is given by $\underline{E}(X) = \min\{P(X) \colon P \in \mathscr{M}(\underline{P})\}$ for all X in $\mathscr{L}(\Omega)$. Moreover, the lower envelope of any set of linear previsions is always a coherent lower prevision.

2.5 Vacuous previsions

We now give an important example of a lower prevision that is in general not a linear prevision. Assume that we want to model the piece of information that the event $A \subseteq \Omega$ occurs, or in other words that the outcome of the experiment we consider assumes a value in A. This can be represented by the so-called *vacuous lower prevision relative to A*, which will be denoted by $\underline{P}_A$, and is given by

$$\underline{P}_A(X) = \inf_{\omega \in A} X(\omega),$$

for all gambles X on Ω. $\underline{P}_A$ is a coherent lower prevision on $\mathscr{L}(\Omega)$ and its conjugate upper prevision is given by

$$\overline{P}_A(X) = \sup_{\omega \in A} X(\omega).$$

There are several lines of reasoning to motivate that this lower prevision indeed is the appropriate model for the given information. First of all, if our subject knows that A occurs, *and nothing more*, he should be willing to buy a gamble X for any price s strictly lower than $\inf_{\omega \in A} X(\omega)$ because doing so results in a sure gain; but he should not be willing to pay a price t strictly higher than that, because then there is some $\omega \in A$ such that $t > X(\omega)$, and *for all our subject knows*, ω might be the actual outcome of the experiment!

A second justification for $\underline{P}_A$ is that it is the natural extension of the single precise probability assessment $P(A) = 1$, which is equivalent to $\underline{P}(A) = 1$. Using Eq. (1), we find indeed that the natural extension of this assessment is given by

$$\begin{aligned}
&\sup_{\lambda \geq 0} \inf_{\omega \in \Omega} [X(\omega) - \lambda [I_A(\omega) - \underline{P}(A)]] \\
&\qquad = \sup_{\lambda \geq 0} \min \left\{ \inf_{\omega \in A} X(\omega), \inf_{\omega \in A^c} [X(\omega) + \lambda] \right\} = \inf_{\omega \in A} X(\omega) = \underline{P}_A(X)
\end{aligned}$$

for all gambles X on Ω. This shows that the vacuous lower prevision relative to A follows *uniquely* from the subject's single assessment that the probability of event A is equal to 1, or equivalently, that the subject is practically certain that A occurs (since he is prepared to bet at all odds on the occurrence of A).

We find a third justification for $\underline{P}_A$ if we consider the set $\mathscr{M}(\underline{P}_A)$ of those linear previsions that dominate $\underline{P}_A$. It is easy to show that

$$\mathscr{M}(\underline{P}_A) = \{P \in \mathbb{P}(\Omega) \colon P(A) = 1\},$$

so $\underline{P}_A$ is again seen to be equivalent to the statement that $P(A) = 1$.

All of this tells us that $\underline{P}_A$ is the smallest, and therefore most conservative, coherent lower prevision $\underline{P}$ on $\mathscr{L}(\Omega)$ that satisfies $\underline{P}(A) = 1$ (and therefore $\overline{P}(A) = P(A) = 1$). Thus, in the context of the theory of lower previsions, $\underline{P}_A$ is the appropriate model for the piece of information that A occurs *and nothing more*: any other coherent lower prevision $\underline{P}$ that satisfies $\underline{P}(A) = 1$ dominates $\underline{P}_A$, and therefore represents stronger commitments than those required by coherence and this piece of information alone.

2.6 Conditioning

If $\mathscr{B}$ is a partition of Ω, then we can define for every $B \in \mathscr{B}$ our subject's *conditional lower prevision* $\underline{P}(X|B)$ *of* X*, given* B as the supremum price he would currently be willing to pay for X, if he came to know at some later time that the outcome of the experiment took a value in B (and nothing else). Alternatively, it could be defined as the subject's supremum buying price for the so-called *contingent* gamble $I_B X$.[1]. If we assume that the conditional lower previsions $\underline{P}(\cdot|B)$ are defined on the same domain $\mathscr{H}$ for all $B \in \mathscr{B}$,[2] then we can summarise all these conditional lower previsions through $\underline{P}(\cdot|\mathscr{B})$, where for all $X \in \mathscr{H}$

$$\underline{P}(X|\mathscr{B}) = \sum_{B \in \mathscr{B}} I_B \underline{P}(X|B).$$

We shall also call the object $\underline{P}(\cdot|\mathscr{B})$ a *conditional lower prevision*. It is a function of two things: gambles and elements of the partition. On the one hand, if we fix the gamble X, then its partial function $\underline{P}(X|\mathscr{B})$ is a gamble on Ω that takes the constant value $\underline{P}(X|B)$ on the element B of $\mathscr{B}$. On the other hand, if we fix the element B of the partition $\mathscr{B}$, then its partial function $\underline{P}(\cdot|B)$ is a lower prevision defined on the set of gambles $\mathscr{H}$.

Conditional lower previsions $\underline{P}(\cdot|\mathscr{B})$ are also subject to rationality criteria. First of all, there is the requirement of *separate coherence*: for each $B \in \mathscr{B}$, $\underline{P}(\cdot|B)$ should be a coherent lower prevision on $\mathscr{H}$ in the sense defined above, and moreover we should obviously also require that $\underline{P}(B|B) = 1$: a subject should be willing to bet at all odds on the occurrence of the event B after observing it. But, if besides $\underline{P}(\cdot|\mathscr{B})$ a coherent unconditional lower prevision $\underline{P}$ on $\mathscr{K}$ is also specified, we should require, besides the separate coherence, that the assessments in $\underline{P}(\cdot|\mathscr{B})$ should be consistent with those in

[1] Some authors (see for instance [13]) prefer to work explicitly with conditional objects, and would define $\underline{P}(X|B)$ as a supremum buying price for a conditional object $X|B$. To avoid possible confusion, we emphasise that this is *not* the approach we follow here: on our behavioural interpretation, $X|B$ does not have any meaning *per se*, and it is only the value $\underline{P}(X|B)$ that expresses our behavioural dispositions, in the manner described above.

[2] This is no essential restriction; for more details see [28, Section 6.2.4].

$\underline{P}$. This leads to the requirement of the *(joint) coherence* of $\underline{P}$ and $\underline{P}(\cdot|\mathscr{B})$, which is studied in much detail in [28, Chapter 6].[3]

Let us remark in passing that, when the conditional and unconditional lower previsions are actually fair prices, then joint coherence reduces, under some additional technical conditions on the domains $\mathscr{K}$ and $\mathscr{H}$, to Bayes' rule plus the so-called $\mathscr{B}$-conglomerability of $\underline{P}$ ([28, Section 6.8]).

If $\underline{P}$ and $\underline{P}(\cdot|\mathscr{B})$ are (jointly) coherent—we then also say that $\underline{P}$ is coherent with $\underline{P}(\cdot|\mathscr{B})$—a procedure of natural extension allows us to extend them to a pair $\underline{P}'$ and $\underline{P}'(\cdot|\mathscr{B})$ which is the smallest jointly coherent pair that extends the pair $\underline{P}$ and $\underline{P}(\cdot|\mathscr{B})$ to all gambles on Ω. Note that $\underline{P}'$ is not necessarily equal to the (unconditional) natural extension of $\underline{P}$ alone, as it also has to take into account the behavioural consequences of the assessments that are present in $\underline{P}(\cdot|\mathscr{B})$! As is the case for unconditional natural extension, the natural extensions $\underline{P}'$ and $\underline{P}'(\cdot|\mathscr{B})$ summarise the behavioural implications of $\underline{P}$ and $\underline{P}(\cdot|\mathscr{B})$, only taking into account the consequences of (separate and) joint coherence.

3 A first step toward generalising random sets

Let us also briefly recall the basic concepts in the theory of random sets. Consider a so-called multi-valued mapping Γ taking elements of an *initial space* Λ to subsets of a *final space* Ω, i.e., $\Gamma\colon \Lambda \to \wp(\Omega)$. We assume throughout that $\Gamma(\lambda) \neq \emptyset$ for all $\lambda \in \Lambda$ (but see Technical Remarks 1 and 2 below).

This kind of mapping has been given many different uses and interpretations. It has for instance been employed as a model for the available information about a random variable ([18]), when there is uncertainty due to missing data, measurement errors, or simply imprecision in the observations. More generally, it has been used to model relations between two spaces, in such diverse fields as economy ([8]) and stochastic geometry ([16, 20]). In particular, as Walley has pointed out in [28], the refinement or coarsening of a space can be modelled by means of a multi-valued mapping.

Dempster ([10], see also [15]) and, some years before him, Strassen ([25]) have argued that the multi-valued mapping Γ turns a probability measure on the initial space into a conjugate pair of lower and upper probabilities on the final space, in the following manner.

Consider a σ-field $\mathscr{A}_\Lambda$ on Λ, a σ-field $\mathscr{A}_\Omega$ on Ω, and a probability measure P defined on the measurable space $(\Lambda, \mathscr{A}_\Lambda)$. We say that the multi-valued mapping Γ is *strongly measurable*[4] when A_* (and therefore also A^*) belong to $\mathscr{A}_\Lambda$ for all $A \in \mathscr{A}_\Omega$, where

$$A_* = \{\lambda \in \Lambda\colon \Gamma(\lambda) \subseteq A\}$$

is the so-called *lower inverse* and

$$A^* = \{\lambda \in \Lambda\colon \Gamma(\lambda) \cap A \neq \emptyset\}$$

[3] A detailed discussion of this notion is beyond the scope of this paper. We refer to [28, Chapter 6] for an explicit formulation of the requirement of joint coherence, and for some of the more technical results that we shall need in a number of the proofs that follow.

[4] Some authors impose different measurability conditions; for an overview see [15].

the *upper inverse* of A under Γ. In that case, we shall also call Γ a *random set*. The *lower probability* P_* and the *upper probability* P^* induced by Γ on $\mathscr{A}_\Omega$ are defined by

$$P_*(A) = P(A_*) \text{ and } P^*(A) = P(A^*), \quad A \in \mathscr{A}_\Omega.$$

We shall concentrate below on the lower inverse, which is used to induce the lower probability P_*. It should be clear that working with the upper inverse leads to an analogous, and completely equivalent, course of reasoning.

Now let us consider the slightly more general case, where P is replaced by a lower probability $\underline{P}$ on the σ-field $\mathscr{A}_\Lambda$ of subsets of Λ. We shall give detailed reasons for wanting to do so in Section 4; see also related work by Augustin ([1, 2]) for earlier explorations of this idea. Γ will now induce a lower probability $\underline{P}_*$ on the σ-field of subsets $\mathscr{A}_\Omega$ of Ω through the formula

$$\underline{P}_*(A) = \underline{P}(A_*), \quad A \in \mathscr{A}_\Omega, \tag{2}$$

provided that Γ is again strongly measurable. There is an interesting connection between potential properties of $\underline{P}$ and $\underline{P}_*$. Consider a set function μ defined on a field $\mathscr{A}$ of sets such that $\mu(\emptyset) = 0$ and $\mu(A) \geq 0$ for all $A \in \mathscr{A}$. Recall that μ is called *monotone* if for all A_1 and A_2 in $\mathscr{A}$:

$$A_1 \subseteq A_2 \Rightarrow \mu(A_1) \leq \mu(A_2);$$

μ is called *k-monotone* (where $k \geq 2$) if for all $A_1, \ldots, A_k$ in $\mathscr{A}$:

$$\mu(A_1 \cup \cdots \cup A_k) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \mu(\cap_{i\in I} A_i),$$

where $|I|$ denotes the cardinality of the set I. Observe that k-monotonicity implies monotonicity. Finally, μ is called a *necessity measure* if for any family $(A_i)_{i\in I}$ of elements of $\mathscr{A}$ such that $\cap_{i\in I} A_i \in \mathscr{A}$:

$$\mu(\cap_{i\in I} A_i) = \inf_{i\in I} \mu(A_i).$$

Theorem 1. *Let the lower probability $\underline{P}$ be defined on a σ-field $\mathscr{A}_\Lambda$ of subsets of Λ, consider a σ-field $\mathscr{A}_\Omega$ of subsets of Ω and assume that $\Gamma\colon \Lambda \to \wp(\Omega)$ is strongly measurable with respect to $\mathscr{A}_\Lambda$ and $\mathscr{A}_\Omega$.*[5] *If $\underline{P}_*$ is the lower probability induced by Γ through Eq.* (2)*, then: (i) if $\underline{P}$ avoids sure loss, so does $\underline{P}_*$; (ii) if $\underline{P}$ is coherent, so is $\underline{P}_*$; (iii) if $\underline{P}$ is monotone, so is $\underline{P}_*$; (iv) if $\underline{P}$ is k-monotone, so is $\underline{P}_*$ (where $k \geq 2$); and (v) if $\underline{P}$ is a necessity measure, so is $\underline{P}_*$.*[6]

[5] Notice that nothing essential is changed if we consider fields rather than σ-fields as domains.

[6] An alternative proof of this fifth point could be given using the fact that for any necessity measure $\underline{P}$, there is a so-called antitone random set $\Gamma_{\underline{P}}$ and a probability P inducing $\underline{P}$ as a lower probability ([6], [21]); indeed, any antitone random set and probability induce a necessity measure. Then, the lower probability induced by Γ would coincide with the one induced by the composition of Γ and $\Gamma_{\underline{P}}$. This composition is also antitone and thus still induces a necessity measure.

Proof. To prove (i), assume that $\underline{P}$ avoids sure loss. Consider n in $\mathbb{N}$ and $A_1, \dots, A_n$ in $\mathscr{A}_\Omega$. We first show that for all $\lambda \in \Lambda$:

$$\sup_{\omega\in\Omega}\left[\sum_{i=1}^{n} I_{A_i}(\omega)\right] \geq \sum_{i=1}^{n} I_{A_{i*}}(\lambda).$$

Fix λ in Λ and consider ω in $\Gamma(\lambda)$. If $\lambda \notin A_{i*}$, then clearly $I_{A_i}(\omega) \geq 0 = I_{A_{i*}}(\lambda)$. On the other hand, if $\lambda \in A_{i*}$, then $\Gamma(\lambda) \subseteq A_i$, whence $\omega \in A_i$. This implies that $\sum_{i=1}^{n} I_{A_i}(\omega) \geq \sum_{i=1}^{n} I_{A_{i*}}(\lambda)$ for all ω in $\Gamma(\lambda)$, which leads immediately to the desired inequality. Using this result, we find that

$$\begin{aligned}\sup_{\omega\in\Omega}\left[\sum_{i=1}^{n}[I_{A_i}(\omega) - \underline{P}_*(A_i)]\right] &= \sup_{\omega\in\Omega}\left[\sum_{i=1}^{n} I_{A_i}(\omega)\right] - \sum_{i=1}^{n}\underline{P}_*(A_i)\\ &\geq \sup_{\lambda\in\Lambda}\left[\sum_{i=1}^{n} I_{A_{i*}}(\lambda)\right] - \sum_{i=1}^{n}\underline{P}_*(A_i) = \sup_{\lambda\in\Lambda}\left[\sum_{i=1}^{n}[I_{A_{i*}}(\lambda) - \underline{P}(A_{i*})]\right] \geq 0,\end{aligned}$$

where the last inequality holds because $\underline{P}$ avoids sure loss. We conclude that $\underline{P}_*$ avoids sure loss as well.

To prove (ii), assume that $\underline{P}$ is coherent. Then since $\underline{P}$ in particular avoids sure loss, we deduce from (i) that $\underline{P}_*$ avoids sure loss as well. Moreover, consider n and m in $\mathbb{N}$ and $A_0, A_1, \dots, A_n$ in $\mathscr{A}_\Omega$. Let us first show that for all λ in Λ:

$$\sup_{\omega\in\Omega}\left[\sum_{i=1}^{n} I_{A_i}(\omega) - mI_{A_0}(\omega)\right] \geq \sum_{i=1}^{n} I_{A_{i*}}(\lambda) - mI_{A_{0*}}(\lambda).$$

Fix λ in Λ. There are two possibilities. If $\lambda \notin A_{0*}$, then there is some ω_1 in $\Gamma(\lambda)$ such that $\omega_1 \notin A_0$. For this ω_1 we find using a similar course of reasoning as before that

$$\sum_{i=1}^{n} I_{A_i}(\omega_1) - mI_{A_0}(\omega_1) = \sum_{i=1}^{n} I_{A_i}(\omega_1) \geq \sum_{i=1}^{n} I_{A_{i*}}(\lambda) = \sum_{i=1}^{n} I_{A_{i*}}(\lambda) - mI_{A_{0*}}(\lambda).$$

If on the other hand $\lambda \in A_{0*}$, then for any ω in $\Gamma(\lambda)$

$$\sum_{i=1}^{n} I_{A_i}(\omega) - mI_{A_0}(\omega) = \sum_{i=1}^{n} I_{A_i}(\omega) - m \geq \sum_{i=1}^{n} I_{A_{i*}}(\lambda) - m = \sum_{i=1}^{n} I_{A_{i*}}(\lambda) - mI_{A_{0*}}(\lambda).$$

This proves the desired inequality. Using this result, we find that

$$\begin{aligned}
&\sup_{\omega\in\Omega}\left[\sum_{i=1}^{n}[I_{A_i}(\omega)-\underline{P}_*(A_i)]-m[I_{A_0}(\omega)-\underline{P}_*(A_0)]\right]\\
&\qquad=\sup_{\omega\in\Omega}\left[\sum_{i=1}^{n}I_{A_i}(\omega)-mI_{A_0}(\omega)\right]-\sum_{i=1}^{n}\underline{P}_*(A_i)+m\underline{P}_*(A_0)\\
&\qquad\geq\sup_{\lambda\in\Lambda}\left[\sum_{i=1}^{n}I_{A_{i*}}(\lambda)-mI_{A_{0*}}(\lambda)\right]-\sum_{i=1}^{n}\underline{P}_*(A_i)+m\underline{P}_*(A_0)\\
&\qquad=\sup_{\lambda\in\Lambda}\left[\sum_{i=1}^{n}[I_{A_{i*}}(\lambda)-\underline{P}(A_{i*})]-m[I_{A_{0*}}(\lambda)-\underline{P}(A_{0*})]\right]\\
&\qquad\geq 0,
\end{aligned}$$

where the last inequality holds because $\underline{P}$ is coherent. We conclude that $\underline{P}_*$ is coherent as well.

The proof of (iii) is immediate taking into account that if $A_1\subseteq A_2$ then $A_{1*}\subseteq A_{2*}$ as well. To prove the last two statements, consider an arbitrary family $(A_i)_{i\in I}$ of elements of $\mathscr{A}_\Omega$. If $\cap_{i\in I}A_i\in\mathscr{A}_\Omega$ then

$$(\cap_{i\in I}A_i)_*=\{\lambda\in\Lambda\colon\Gamma(\lambda)\subseteq\cap_{i\in I}A_i\}=\cap_{i\in I}\{\lambda\in\Lambda\colon\Gamma(\lambda)\subseteq A_i\}=\cap_{i\in I}A_{i*}\tag{3}$$

and if $\cup_{i\in I}A_i\in\mathscr{A}_\Omega$ then

$$(\cup_{i\in I}A_i)_*=\{\lambda\in\Lambda\colon\Gamma(\lambda)\subseteq\cup_{i\in I}A_i\}\supseteq\cup_{i\in I}\{\lambda\in\Lambda\colon\Gamma(\lambda)\subseteq A_i\}=\cup_{i\in I}A_{i*}.\tag{4}$$

Assume that $\underline{P}$ is k-monotone, and consider $A_1,\dots,A_k$ in $\mathscr{A}_\Omega$. It then follows from the expressions (3)–(4) and the monotonicity of $\underline{P}$ that

$$\begin{aligned}
&\underline{P}_*(A_1\cup\dots\cup A_k)-\sum_{\emptyset\neq I\subseteq\{1,\dots,k\}}(-1)^{|I|+1}\underline{P}_*(\cap_{i\in I}A_i)\\
&\qquad\geq\underline{P}(A_{1*}\cup\dots\cup A_{k*})-\sum_{\emptyset\neq I\subseteq\{1,\dots,k\}}(-1)^{|I|+1}\underline{P}(\cap_{i\in I}A_{i*})\geq 0,
\end{aligned}$$

where the last inequality follows from the k-monotonicity of $\underline{P}$. It follows that $\underline{P}_*$ is k-monotone.

Let us finally assume that $\underline{P}$ is a necessity measure. Consider any family $(A_i)_{i\in I}$ of elements of $\mathscr{A}_\Omega$ such that $\cap_{i\in I}A_i\in\mathscr{A}_\Omega$ as well. Then it follows from Eq. (3) that

$$\underline{P}_*(\cap_{i\in I}A_i)=\underline{P}((\cap_{i\in I}A_i)_*)=\underline{P}(\cap_{i\in I}A_{i*})=\inf_{i\in I}\underline{P}(A_{i*})=\inf_{i\in I}\underline{P}_*(A_i)$$

so $\underline{P}_*$ is a necessity measure as well. □

The fourth point of this theorem generalises a result from [23], where it is shown that the lower probability induced by a probability measure and a strongly measurable multi-valued mapping is always completely monotone, i.e., k-monotone for all natural $k\geq 2$. Theorem 1 tells us in particular that this result still holds if instead of a probability we have a completely monotone lower probability on the initial space.

4 A more general approach

Although the generalisation of random sets studied in the previous section is fairly intuitive and straightforward, it has a number of shortcomings. First of all, there is the emphasis on lower probabilities and events, rather than lower previsions and gambles. Although the languages of gambles and events are equally powerful when working with precise probabilities, Walley ([28]) has shown that gambles are much more expressive than events in the context of imprecise probabilities. Secondly, the domains $\mathscr{A}_\Lambda$ and $\mathscr{A}_\Omega$ of the respective lower probabilities $\underline{P}$ and $\underline{P}_*$ were assumed to be σ-fields, and Γ is assumed to be strongly measurable with respect to these σ-fields. This is something the theory of random sets has inherited from the Kolmogorov approach to precise probability theory, where probability measures can only be defined on σ-fields. We have already mentioned above (see Section 2) that in Walley's theory of lower previsions, and in de Finetti's approach to precise probability theory ([7]), coherent lower previsions as well as linear previsions can be defined on *arbitrary* sets of gambles that need not have any predefined structure at all, and they can be extended by natural extension to lower previsions defined on all gambles. We would therefore like to be able to find a method to use an *arbitrary* multi-valued mapping $\Gamma\colon \Lambda \to \wp(\Omega)$ to turn an *arbitrary* coherent lower prevision $\underline{P}$ on an *arbitrary* set $\mathscr{K}$ of gambles on Λ into a coherent lower prevision *defined on all gambles* on Ω, or even on $\Lambda \times \Omega$. Moreover, we would like this method to have a clear behavioural interpretation and motivation.

4.1 A behavioural interpretation of the basic model

First of all, let us shed some light on the meaning of the multi-valued mapping Γ. We consider two random variables L and O taking values in Λ and Ω respectively. These random variables are linked in the following way: *if L assumes the value $\lambda \in \Lambda$, then we know that O assumes a value in $\Gamma(\lambda) \subseteq \Omega$, and nothing else.*[7]

The problem we are faced with can then be analysed as follows. We consider a subject who has two sources of information. First, there is information about which value the random variable L will assume in Λ, and he models this through a *coherent lower prevision $\underline{P}$* on some set $\mathscr{K}$ of gambles on Λ.[8]

The second source of information is the given interpretation of the multi-valued mapping Γ: if the subject knows that the random variable L assumes the value λ in Λ, then he also knows that the random variable O assumes a value in $\Gamma(\lambda)$, and nothing else. We shall now argue that this type of information can be modelled by a special type of *conditional lower prevision*.

Consider the partition $\mathscr{B} = \{\{\lambda\} \times \Omega\colon \lambda \in \Lambda\}$ of the set $\Lambda \times \Omega$. For any λ in Λ, the occurrence of the event $\{\lambda\} \times \Omega$ corresponds to our subject knowing that the random variable L assumes the value λ, and the conditional conditional lower prevision $\underline{P}(\cdot|\{\lambda\} \times \Omega)$ models the available information about the value of (L,O) when he knows that $L = \lambda$. Taking into account the interpretation for Γ outlined above, this means that $\underline{P}(\cdot|\{\lambda\} \times \Omega)$ should reflect that the subject knows that (L,O) assumes a

[7] This interpretation of the multi-valued map Γ essentially goes back to Strassen [25].

[8] In Section 5, we shall pay attention to the particular case where the set of gambles $\mathscr{K}$ is a σ-field of events, and relate the results from the present section with those of Section 3.

value in the set $\{\lambda\}\times\Gamma(\lambda)$, and nothing more! We have seen in Section 2.5 that this information should be modelled by the vacuous lower prevision relative to $\{\lambda\}\times\Gamma(\lambda)$: for any gamble X on $\Lambda\times\Omega$

$$\underline{P}(X|\{\lambda\}\times\Omega)=\underline{P}_{\{\lambda\}\times\Gamma(\lambda)}(X)=\inf_{\omega\in\Gamma(\lambda)}X(\lambda,\omega). \tag{5}$$

We shall also use the short notation $\underline{P}(X|\lambda)$ for $\underline{P}(X|\{\lambda\}\times\Omega)$. $\underline{P}(X|\lambda)$ is the supremum price that our subject should be prepared to pay for the gamble X if he knew that L assumed the value λ and nothing else.

Eq. (5) gives us the value of the conditional lower prevision $\underline{P}(X|\lambda)$ for all $\lambda\in\Lambda$. As we explained in Section 2.6, we can summarise all this information through

$$\underline{P}(X|\mathscr{B})=\sum_{B\in\mathscr{B}}I_B\underline{P}(X|B)=\sum_{\lambda\in\Lambda}I_{\{\lambda\}\times\Omega}\underline{P}(X|\lambda),$$

which we also denote as $\underline{P}(X|\Lambda)$. Observe that $\underline{P}(X|\Lambda)$ can also be interpreted as a gamble on $\Lambda\times\Omega$, whose value in (λ,ω) is given by $\underline{P}(X|\lambda)$.

In summary, the information in the multi-valued map Γ can be represented by the conditional lower prevision $\underline{P}(\cdot|\Lambda)$. This conditional lower prevision is a function of two things: the gambles X and the elements λ of Λ.

Important Remark 1. We shall frequently identify the space $\mathscr{L}(\Lambda)$ with the space of those gambles on $\Lambda\times\Omega$ that are *constant* on Ω, that is, those gambles satisfying $Z(\lambda,\omega)=Z(\lambda,\omega')$ for all $\lambda\in\Lambda$ and all ω, ω' in Ω. Similarly, it is clear that we can identify the space $\mathscr{L}(\Omega)$ with the gambles on $\Lambda\times\Omega$ that are constant on Λ. Let us give two examples of this. First of all, given a gamble X on $\Lambda\times\Omega$ and $\lambda\in\Lambda$, we can define the gamble X^λ on $\Lambda\times\Omega$ by $X^\lambda(\nu,\omega)=X(\lambda,\omega)$ for all $(\nu,\omega)\in\Lambda\times\Omega$. Observe that X^λ is constant on Λ, so we can identify it with a gamble on Ω. Moreover,[9]

$$\underline{P}(X|\lambda)=\inf_{\omega\in\Gamma(\lambda)}X(\lambda,\omega)=\inf_{\omega\in\Gamma(\lambda)}X^\lambda(\lambda,\omega)=\underline{P}(X^\lambda|\lambda).$$

Secondly, we have just introduced the notation

$$\underline{P}(X|\Lambda)=\sum_{\lambda\in\Lambda}I_{\{\lambda\}\times\Omega}\underline{P}(X|\lambda)$$

for the gamble on $\Lambda\times\Omega$ that assumes the value $\underline{P}(X|\lambda)$ in the element (λ,ω) of $\Lambda\times\Omega$. Note that this gamble is constant on Ω, so we can identify it with a gamble on Λ. ⧫

If A is a subset of $\Lambda\times\Omega$, and we let $X=I_A$ in Eq. (5), then we find that for all λ in Λ:

$$\underline{P}(A|\lambda)=\inf_{\omega\in\Gamma(\lambda)}I_A(\lambda,\omega)=I_{A_\circ}(\lambda),$$

[9]This property does not only hold for the special choice of the conditional lower prevision we made in Eq. (5). One can show (see [28, Lemma 6.2.4]) that it is a necessary consequence of the *separate coherence* of any conditional lower prevision $\underline{P}(\cdot|\Lambda)$. Hence, on our approach, it is this consequence of separate coherence that *a posteriori* allows us to treat a conditional lower prevision $\underline{P}(Y|B)$ as if it were a lower prevision of some conditional object $Y|B$; see also footnote 1.

where $A_\circ$ is the subset of Λ defined by

$$A_\circ = \{\lambda \in \Lambda \colon \{\lambda\} \times \Gamma(\lambda) \subseteq A\}. \tag{6}$$

If in particular A is a subset of Ω, then it follows immediately that $(\Lambda \times A)_\circ = A_*$, i.e., $\underline{P}(\Lambda \times A|\Lambda)$ is the indicator function of the lower inverse $A_* = \{\lambda \in \Lambda \colon \Gamma(\lambda) \subseteq A\}$ of A, defined in the previous section. This means that the conditional lower prevision $\underline{P}(\cdot|\Lambda)$ extends the notion of a lower inverse.

Definition 1. Given a multi-valued mapping $\Gamma \colon \Lambda \to \wp(\Omega)$, we can associate with any gamble X on $\Lambda \times \Omega$ its *lower inverse* $X_\circ = \underline{P}(X|\Lambda)$, which is a gamble on Λ defined by $X_\circ(\lambda) = \underline{P}(X|\lambda) = \inf_{\omega \in \Gamma(\lambda)} X(\lambda, \omega)$ for all λ in Λ.

Since for any λ in Λ, $\underline{P}(\cdot|\lambda)$ has been defined as the vacuous lower prevision on $\mathscr{L}(\Lambda \times \Omega)$ relative to the set $\{\lambda\} \times \Gamma(\lambda)$, it is a coherent lower prevision. Since moreover $\underline{P}(\{\lambda\} \times \Omega|\lambda) = 1$ since $\Gamma(\lambda) \neq \emptyset$, we may conclude that the conditional lower prevision $\underline{P}(\cdot|\Lambda)$ is *separately coherent*.

Technical Remark 1. We assume in this paper that the set $\Lambda_\emptyset = \{\lambda \in \Lambda \colon \Gamma(\lambda) = \emptyset\}$ is empty. But, for technical mathematical reasons (see for instance [6]) it is sometimes necessary to consider multi-valued mappings $\Gamma \colon \Lambda \to \wp(\Omega)$ for which there are λ such that $\Gamma(\lambda) = \emptyset$, i.e., for which $\Lambda_\emptyset$ is non-empty. This would only be compatible with the given information that 'O assumes a value in $\Gamma(\lambda)$ if $L = \lambda$', and could thus only then be incorporated in the present model if our subject were absolutely certain that L could never assume any value in $\Lambda_\emptyset$, or in other words if

$$\Lambda_\emptyset^c \in \mathscr{K} \text{ and } \underline{P}(\Lambda_\emptyset^c) = 1, \tag{7}$$

where $\Lambda_\emptyset^c = \{\lambda \in \Lambda \colon \Gamma(\lambda) \neq \emptyset\}$ is the set-theoretic complement of the set $\Lambda_\emptyset$. This means that the subject is prepared to bet at all odds against the event that L assumes a value in $\Lambda_\emptyset$. For $\lambda \in \Lambda_\emptyset$ and $X \in \mathscr{L}(\Lambda \times \Omega)$ we could then let $\underline{P}(X|\lambda) = \underline{P}_\lambda(X^\lambda)$, where $\underline{P}_\lambda$ is *any* coherent lower prevision defined on $\mathscr{L}(\Omega)$ and X^λ is the gamble derived from X in Important Remark 1. It turns out that, under this proviso, the course of reasoning followed below would still be valid, and that the final result would not be influenced by the choice of the $\underline{P}_\lambda$ (see Technical Remark 2 further on for more details). ⧫

In summary, the available information has been represented by means of the coherent lower prevision $\underline{P}$ on $\mathscr{K}$ and the separately coherent conditional lower prevision $\underline{P}(\cdot|\Lambda)$ on $\mathscr{L}(\Lambda \times \Omega)$. These represent our subject's commitments to buy certain gambles, based on this information.

4.2 Making inferences based on the basic model

We are, however, interested in what can be inferred from these commitments, and in particular, what they imply about whether or not the subject should buy a gamble $X \in \mathscr{L}(\Lambda \times \Omega)$ for a given price. As we have seen in Section 2, there is in the behavioural theory of lower previsions a general reasoning technique based on the notion of coherence, called *natural extension*, which can be used to make inferences from assessments of lower previsions. In the special case considered here, it follows from Theorem 2 below that there is a smallest (and therefore most conservative or least committal) unconditional lower prevision $\underline{P}_\circ$, defined on $\mathscr{L}(\Lambda\times\Omega)$ by Eq. (8), that extends the unconditional lower prevision $\underline{P}$ from $\mathscr{K}$ to $\mathscr{L}(\Lambda\times\Omega)$ and that is coherent with the conditional lower prevision $\underline{P}(\cdot|\Lambda)$. It is called the *natural extension* of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ to an unconditional lower prevision on $\mathscr{L}(\Lambda\times\Omega)$. Its behavioural meaning is clear: for any gamble X on $\Lambda\times\Omega$, $\underline{P}_\circ(X)$ is the supremum price for which the subject can be forced to buy the gamble X by constructing a finite combination of buying transactions that are implicit in his $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ and that he is therefore committed to accept. In this sense, $\underline{P}_\circ(X)$ is the supremum buying price for X that we can infer from the assessments $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$.

Theorem 2. *Let $\underline{P}$ be a coherent unconditional lower prevision defined on $\mathscr{K}\subseteq\mathscr{L}(\Lambda)$, and let $\underline{P}(\cdot|\Lambda)$ be the separately coherent conditional lower prevision defined on $\mathscr{L}(\Lambda\times\Omega)$ by Eq.* (5)*. Then $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ are (jointly) coherent. The smallest coherent lower prevision on $\mathscr{L}(\Lambda\times\Omega)$ that extends $\underline{P}$ from $\mathscr{K}$ to $\mathscr{L}(\Lambda\times\Omega)$ and that is coherent with $\underline{P}(\cdot|\Lambda)$ is given by*

$$\underline{P}_\circ(X)=\underline{E}(\underline{P}(X|\Lambda)) \tag{8}$$

for all $X\in\mathscr{L}(\Lambda\times\Omega)$, where $\underline{E}$ is the (unconditional) natural extension of the lower prevision $\underline{P}$ from $\mathscr{K}$ to $\mathscr{L}(\Lambda)$, i.e., for all gambles Y on Λ,

$$\underline{E}(Y)=\sup_{\substack{n\geq 0\\ Z_1,\dots,Z_n\in\mathscr{K}\\ \mu_1,\dots,\mu_n\geq 0}}\inf_{\lambda\in\Lambda}\left[Y(\lambda)-\sum_{k=1}^{n}\mu_k[Z_k(\lambda)-\underline{P}(Z_k)]\right].$$

This theorem is an immediate consequence of Walley's Marginal Extension Theorem ([28, Theorem 6.7.2]). Note also that for any gamble Z on Λ (or any gamble Z constant on Ω) we have indeed (by separate coherence) that $Z=\underline{P}(Z|\Lambda)$ whence $\underline{P}_\circ(Z)=\underline{E}(Z)$. In particular, this tells us that $\underline{P}_\circ(Z)=\underline{P}(Z)$ for all $Z\in\mathscr{K}$, or in other words that $\underline{P}_\circ$ is a coherent extension of $\underline{P}$, i.e. a coherent lower prevision that coincides with $\underline{P}$ on $\mathscr{K}$.

If we recall Definition 1, we see that $\underline{P}_\circ(X)=\underline{E}(X_\circ)$ for all X in $\mathscr{L}(\Lambda\times\Omega)$, and therefore the concept of natural extension allows us to (i) give a behavioural motivation for the notion of a lower probability induced by a multi-valued mapping, defined in Section 3; and (ii) extend it from lower probabilities to lower previsions (or from events to gambles). We have thus proven in a general context what Walley showed in a different way and in a finitary setting in [28, Section 4.3]. We shall discuss these issues in more detail in Section 5.

Technical Remark 2. If we allow that Γ may assume empty values, i.e., $\Lambda_\emptyset\neq\emptyset$, then Theorem 2 will still hold, provided that the lower previsions $\underline{P}_\lambda$, defined on $\mathscr{L}(\Omega)$ for all $\lambda\in\Lambda_\emptyset$ and mentioned in Technical Remark 1, are coherent, so that the conditional lower prevision $\underline{P}(\cdot|\Lambda)$ is still separately coherent. For any gamble X on $\Lambda\times\Omega$, its lower inverse $X_\circ$ will now assume the value $\underline{P}_\lambda(X^\lambda)$ in any element λ of Λ such that $\Gamma(\lambda)=\emptyset$, so $X_\circ$ will depend on the (arbitrary) choice of the $\underline{P}_\lambda$. But if we make the assumptions summarised in Eq. (7), which are necessary for our behavioural interpretation of Γ to make any sense, then it turns out that *the natural extension $\underline{P}_\circ$ is not influenced in any way by the choice of the $\underline{P}_\lambda$!* Indeed, it follows from the coherence of the lower prevision $\underline{P}_\circ$ and $\overline{P}_\circ(\Lambda_\emptyset \times \Omega) = 1 - \underline{P}_\circ(\Lambda_\emptyset^c \times \Omega) = 1 - \underline{P}(\Lambda_\emptyset^c) = 0$, that $\underline{P}_\circ(X) = \underline{P}_\circ(XI_{\Lambda_\emptyset^c \times \Omega})$ and $\overline{P}_\circ(X) = \overline{P}_\circ(XI_{\Lambda_\emptyset^c \times \Omega})$. ♦

The natural extension $\underline{P}_\circ$ can also be seen as a lower envelope of induced lower previsions. This is an immediate consequence of a more general result on conditioning, proven by Walley ([28, Section 6.7 and in particular 6.7.5]).

Theorem 3. *Let $\underline{P}$ be a coherent lower prevision defined on a subset $\mathscr{K}$ of $\mathscr{L}(\Lambda)$, and let $\underline{P}(\cdot|\Lambda)$ be the separately coherent conditional lower prevision defined on $\mathscr{L}(\Lambda\times\Omega)$ by Eq.* (5). *Let $\mathscr{M}(\underline{P})$ be the set of linear previsions on $\mathscr{L}(\Lambda)$ that dominate $\underline{P}$:*

$$\mathscr{M}(\underline{P}) = \{P \in \mathbb{P}(\Lambda) \colon (\forall X \in \mathscr{K})(P(X) \geq \underline{P}(X))\}.$$

Then the natural extension $\underline{P}_\circ$ of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ is the lower envelope of the natural extensions $P_\circ$ of P and $\underline{P}(\cdot|\Lambda)$ where $P \in \mathscr{M}(\underline{P})$: for all X in $\mathscr{L}(\Lambda\times\Omega)$

$$\underline{P}_\circ(X) = \min_{P\in\mathscr{M}(\underline{P})} P_\circ(X) = \min_{P\in\mathscr{M}(\underline{P})} P(X_\circ).$$

4.3 What is the essential information?

We see in Theorem 2 that $\underline{P}_\circ$ is calculated using the natural extension $\underline{E}$ of the lower prevision $\underline{P}$. One reason for this is of course that, although we have seen that the lower inverse $X_\circ = \underline{P}(X|\Lambda)$ of a gamble $X \in \mathscr{L}(\Lambda\times\Omega)$ can be considered as a gamble on Λ, we have no guarantee that $X_\circ$ should belong to the domain $\mathscr{K}$ of $\underline{P}$. We should therefore extend $\underline{P}$ to the coherent lower prevision $\underline{E}$ on the larger domain $\mathscr{L}(\Lambda)$.

But this brings us to an interesting problem. If we restrict the gambles X that we consider to the set

$${}_\circ\mathscr{K} = \{X \in \mathscr{L}(\Lambda\times\Omega) \colon X_\circ \in \mathscr{K}\}$$

of those gambles whose lower inverse belongs to $\mathscr{K}$, we can calculate $\underline{P}_\circ(X)$ using $\underline{P}$ rather than its natural extension $\underline{E}$, as $\underline{P}_\circ(X) = \underline{E}(X_\circ) = \underline{P}(X_\circ)$! In other words, using $\underline{P}$ alone, we can define the lower prevision $\underline{P}_1$ on the set of gambles ${}_\circ\mathscr{K} \subseteq \mathscr{L}(\Lambda\times\Omega)$ by

$$\underline{P}_1(X) = \underline{P}(X_\circ) = \underline{P}(\underline{P}(X|\Lambda)), \quad X \in {}_\circ\mathscr{K}. \tag{9}$$

Note that if we identify gambles on Λ with gambles on $\Lambda\times\Omega$ that are constant on Ω, we may write that $\mathscr{K} \subseteq {}_\circ\mathscr{K}$, or in other words, $\underline{P}_1$ actually extends $\underline{P}$.

Of course, since $\underline{P}$ is assumed to be coherent and therefore coincides with its natural extension $\underline{E}$ on its domain $\mathscr{K}$, $\underline{P}_1$ is nothing but the restriction of the coherent lower prevision $\underline{P}_\circ$ to ${}_\circ\mathscr{K}$, and is therefore coherent as well.

Proposition 4. *If the lower prevision $\underline{P}$ on $\mathscr{K}$ is coherent, then $\underline{P}_1$ is a coherent lower prevision on ${}_\circ\mathscr{K}$.*

The question that now arises, is the following: does $\underline{P}_1$ essentially contain all the information in $\underline{P}_\circ$, or in other words, is $\underline{P}_\circ$ equal to the (unconditional) natural extension $\underline{E}_1$ of $\underline{P}_1$ to a lower prevision on $\mathscr{L}(\Lambda\times\Omega)$, given by

$$\underline{E}_1(X)=\sup_{\substack{n\geq 0\\ Z_1,\ldots,Z_n\in{}_\circ\mathscr{K}\\ \mu_1,\ldots,\mu_n\geq 0}}\ \inf_{(\lambda,\omega)\in\Lambda\times\Omega}\left[X(\lambda,\omega)-\sum_{k=1}^{n}\mu_k[Z_k(\lambda,\omega)-\underline{P}_1(Z_k)]\right] \tag{10}$$

for all $X\in\mathscr{L}(\Lambda\times\Omega)$? Or, to formulate it in yet another manner, does the following diagram commute?

$$\begin{array}{ccc}
\underline{P} & \xrightarrow{\text{composition with } \underline{P}(\cdot|\Lambda)} & \underline{P}_1\\
{\scriptstyle\text{nat. ext.}}\big\downarrow & & \big\downarrow{\scriptstyle\text{nat. ext.}}\\
\underline{E} & \xrightarrow{\text{composition with } \underline{P}(\cdot|\Lambda)} & \underline{P}_\circ
\end{array}$$

We now show that this is indeed the case. The following lemma allows us to put the question of whether $\underline{E}_1$ and $\underline{P}_\circ$ coincide in a different perspective.

Lemma 5. *The natural extension $\underline{E}_1$ of $\underline{P}_1$ to $\mathscr{L}(\Lambda\times\Omega)$ coincides with the natural extension $\underline{P}_\circ$ of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ if and only if the lower prevision $\underline{E}_1$ is coherent with the conditional lower prevision $\underline{P}(\cdot|\Lambda)$.*

Proof. It is clear that if $\underline{E}_1=\underline{P}_\circ$ then $\underline{E}_1$ is coherent with $\underline{P}(\cdot|\Lambda)$, because we know from Theorem 2 that $\underline{P}_\circ$ is. Conversely, if $\underline{E}_1$ is coherent with $\underline{P}(\cdot|\Lambda)$, then $\underline{E}_1$ will dominate on its domain $\mathscr{L}(\Lambda\times\Omega)$ the smallest coherent extension $\underline{P}_\circ$ of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ to an unconditional lower prevision on $\mathscr{L}(\Lambda\times\Omega)$. But, since $\underline{P}_\circ$ is a coherent extension of $\underline{P}_1$, it always dominates the smallest coherent extension $\underline{E}_1$ of $\underline{P}_1$ on $\mathscr{L}(\Lambda\times\Omega)$. This proves that $\underline{E}_1$ and $\underline{P}_\circ$ are equal. □

We now proceed to prove that $\underline{E}_1$ and $\underline{P}(\cdot|\Lambda)$ are indeed guaranteed to be coherent.

Lemma 6. *Consider two gambles X_1 and X_2 on $\Lambda\times\Omega$. If $\underline{P}(X_1|\Lambda)=\underline{P}(X_2|\Lambda)$, then $\underline{E}_1(X_1)=\underline{E}_1(X_2)$. Consequently, if $\underline{P}(X_1|\Lambda)=0$ then $\underline{E}_1(X_1)=0$.*

Proof. First of all, we show that for any gamble X in ${}_\circ\mathscr{K}$, $\underline{E}_1(X)$ is the supremum S of all $\alpha\in\mathbb{R}$ such that there are integer $n\geq 0$, real $\mu_1,\ldots,\mu_n\geq 0$ and $Z_1,\ldots,Z_n$ in ${}_\circ\mathscr{K}$ such that for all $\lambda\in\Lambda$ and $\omega\in\Gamma(\lambda)$,

$$X(\lambda,\omega)-\alpha\geq\sum_{k=1}^{n}\mu_k[Z_k(\lambda,\omega)-\underline{P}(\underline{P}(Z_k|\Lambda))].$$

It is clear from Eq. (10) that $S\geq\underline{E}_1(X)$. Conversely, consider $n\geq 0$, real $\mu_1,\ldots,\mu_n\geq 0$ and $Z_1,\ldots,Z_n$ in ${}_\circ\mathscr{K}$ satisfying the above inequality. We can modify the gambles $Z_1,\ldots,Z_n$ on the sets $\{(\lambda,\omega)\colon\omega\notin\Gamma(\lambda)\}$, $\lambda\in\Lambda$, in such a way that the above inequality is extended to all elements of $\Lambda\times\Omega$. It is important to note that this change will not affect the $\underline{P}(Z_k|\Lambda)$, so the modified gambles will still belong to ${}_\circ\mathscr{K}$. So indeed $S\leq\underline{E}_1(X)$.

Next, assume that $\underline{P}(X_1|\Lambda) = \underline{P}(X_2|\Lambda)$ and let $\alpha < \underline{E}_1(X_1)$. Then we have just proven that there are integer $n \geq 0$, real $\mu_1, \dots \mu_n \geq 0$ and $Z_1, \dots, Z_n$ in ${}_\circ\mathscr{K}$ such that

$$X_1(\lambda,\omega) - \alpha \geq \sum_{k=1}^{n} \mu_k[Z_k(\lambda,\omega) - \underline{P}(\underline{P}(Z_k|\Lambda))],$$

for all $\lambda \in \Lambda$ and $\omega \in \Gamma(\lambda)$. Define the gambles Y_k on $\Lambda \times \Omega$ by $Y_k(\lambda,\omega) = \underline{P}(Z_k|\lambda)$ for all $(\lambda,\omega) \in \Lambda \times \Omega$. These Y_k are constant on Ω and $\underline{P}(Y_k|\Lambda) = \underline{P}(Z_k|\Lambda)$, so $Y_k \in {}_\circ\mathscr{K}$, and $Y_k(\lambda) = Y_k(\lambda,\omega) = \inf_{\nu \in \Gamma(\lambda)} Z_k(\lambda,\nu) \leq Z_k(\lambda,\omega)$ for all $\lambda \in \Lambda$ and $\omega \in \Gamma(\lambda)$. Consequently,

$$X_1(\lambda,\omega) - \alpha \geq \sum_{k=1}^{n} \mu_k[Y_k(\lambda) - \underline{P}(\underline{P}(Y_k|\Lambda))],$$

for all $\lambda \in \Lambda$ and $\omega \in \Gamma(\lambda)$. This is equivalent to

$$\underline{P}(X_1|\lambda) - \alpha = \inf_{\omega \in \Gamma(\lambda)} X_1(\lambda,\omega) - \alpha \geq \sum_{k=1}^{n} \mu_k[Y_k(\lambda) - \underline{P}(\underline{P}(Y_k|\Lambda))],$$

for all $\lambda \in \Lambda$, and since we can replace $\underline{P}(X_1|\lambda)$ with $\underline{P}(X_2|\lambda)$ in this inequality, this eventually leads back to

$$X_2(\lambda,\omega) - \alpha \geq \sum_{k=1}^{n} \mu_k[Y_k(\lambda) - \underline{P}(\underline{P}(Y_k|\Lambda))],$$

for all $\lambda \in \Lambda$ and $\omega \in \Gamma(\lambda)$. This implies that $\underline{E}_1(X_2) \geq \alpha$, whence $\underline{E}_1(X_2) \geq \underline{E}_1(X_1)$. The same argument with X_1 and X_2 interchanged leads to $\underline{E}_1(X_2) = \underline{E}_1(X_1)$. To prove the last statement, assume that $\underline{P}(X_1|\Lambda) = 0$ and let $X_2 = 0$. Then clearly $\underline{P}(X_2|\Lambda) = 0 = \underline{P}(X_1|\Lambda)$, whence $\underline{E}_1(X_1) = \underline{E}_1(X_2) = \underline{E}_1(0) = 0$, where the last equality follows from the coherence of $\underline{E}_1$. □

Theorem 7. *Let $\underline{P}$ be a coherent lower prevision defined on a subset $\mathscr{K}$ of $\mathscr{L}(\Lambda)$, and let $\underline{P}(\cdot|\Lambda)$ be the separately coherent conditional lower prevision defined on $\mathscr{L}(\Lambda \times \Omega)$ by Eq.* (5)*. Let $\underline{P}_1$ be the coherent lower prevision defined on ${}_\circ\mathscr{K}$ by Eq.* (9)*, and let $\underline{E}_1$ be its (unconditional) natural extension to $\mathscr{L}(\Lambda \times \Omega)$. Then $\underline{E}_1$ coincides with the natural extension $\underline{P}_\circ$ of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ to an unconditional lower prevision on $\mathscr{L}(\Lambda \times \Omega)$.*

Proof. Since both $\underline{E}_1$ and $\underline{P}(\cdot|\Lambda)$ are defined on the linear space $\mathscr{L}(\Lambda \times \Omega)$, we can deduce from Walley's Theorem 6.5.3 in [28] that $\underline{E}_1$ is coherent with $\underline{P}(\cdot|\Lambda)$, and therefore by Lemma 5 coincides with $\underline{P}_\circ$, if and only if the following conditions are verified for all X in $\mathscr{L}(\Lambda \times \Omega)$ and λ in Λ:

C1. $\underline{E}_1(X - \underline{P}(X|\Lambda)) \geq 0$;

C2. $\underline{E}_1\left(I_{\{\lambda\}\times\Omega}[X - \underline{P}(X|\lambda)]\right) = 0$.

We now show that this is indeed the case. First, we check that C1 holds. For any X in $\mathscr{L}(\Lambda \times \Omega)$,

$$X - \underline{P}(X|\Lambda) = \sum_{\nu \in \Lambda} I_{\{\nu\}\times\Omega}[X - \underline{P}(X|\nu)],$$

and so we find for all λ in Λ that

$$\begin{aligned}[X-\underline{P}(X|\Lambda)]_\circ(\lambda) &= \underline{P}(X-\underline{P}(X|\Lambda)|\lambda) \\ &= \inf_{\omega\in\Gamma(\lambda)} \sum_{\nu\in\Lambda} I_{\{\nu\}\times\Omega}(\lambda,\omega)[X(\lambda,\omega)-\underline{P}(X|\nu)] \\ &= \inf_{\omega\in\Gamma(\lambda)} [X(\lambda,\omega)-\underline{P}(X|\lambda)] = 0.\end{aligned}$$

¿From Lemma 6, we deduce that $\underline{E}_1(X-\underline{P}(X|\Lambda)) = 0$, so C1 holds. To check that C2 holds, observe that for X in $\mathscr{L}(\Lambda\times\Omega)$ and λ and ν in Λ:

$$\begin{aligned}\underline{P}\left(I_{\{\lambda\}\times\Omega}[X-\underline{P}(X|\lambda)]|\nu\right) &= \inf_{\omega\in\Gamma(\nu)} I_{\{\lambda\}\times\Omega}(\nu,\omega)[X(\nu,\omega)-\underline{P}(X|\lambda)] \\ &= I_{\{\lambda\}}(\nu) \inf_{\omega\in\Gamma(\nu)} [X(\nu,\omega)-\underline{P}(X|\nu)] = 0.\end{aligned}$$

This tells us that $\left[I_{\{\lambda\}\times\Omega}[X-\underline{P}(X|\lambda)]\right]_\circ = 0$, and from Lemma 6 we may then infer that also $\underline{E}_1\left(I_{\{\lambda\}\times\Omega}[X-\underline{P}(X|\lambda)]\right) = 0$, so C2 holds and therefore $\underline{E}_1$ and $\underline{P}(\cdot|\Lambda)$ are coherent. □

5 Application to random sets

Let us now apply what we have learnt in the previous section to the problem we studied in Section 3. We again consider a coherent lower probability $\underline{P}$ defined on a σ-field $\mathscr{A}_\Lambda$ of subsets of Λ, as well as a σ-field $\mathscr{A}_\Omega$ of subsets of Ω, and a multi-valued mapping $\Gamma\colon \Lambda \to \wp(\Omega)$ that is strongly measurable with respect to these σ-fields.

Then we can extend $\underline{P}$ to a lower prevision $\underline{P}_\circ = \underline{E}(\underline{P}(\cdot|\Lambda))$ on $\mathscr{L}(\Lambda\times\Omega)$ that is the natural extension of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$. Here $\underline{E}$ is the (unconditional) natural extension of the lower probability $\underline{P}$ to a lower prevision on $\mathscr{L}(\Lambda)$. In particular, we conclude from Eq. (6) that for any event A in $\mathscr{A}_\Omega$, $\underline{P}(\Lambda\times A|\Lambda)$ is the indicator function of the set $A_* = \{\lambda\in\Lambda\colon \Gamma(\lambda)\subseteq A\}$, which belongs to $\mathscr{A}_\Lambda$ since Γ is strongly measurable. Consequently,

$$\underline{P}_\circ(\Lambda\times A) = \underline{P}(\{\lambda\in\Lambda\colon \Gamma(\lambda)\subseteq A\}) = \underline{P}_*(A).$$

In other words, the natural extension of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ to the set of events $\mathscr{A}_\Omega$ is nothing but the induced lower probability $\underline{P}_*$ defined in Section 3! This gives a behavioural interpretation to this induced lower probability, but at the same time, it provides us with a natural way to extend it to a lower prevision on $\mathscr{L}(\Omega)$, i.e., to associate a kind of 'integral' with the induced lower probability, by restricting $\underline{P}_\circ$ to gambles on Ω (or constant on Λ): for all X in $\mathscr{L}(\Omega)$, $\underline{P}_\circ(X) = \underline{E}(\inf_{\omega\in\Gamma(\cdot)} X(\omega))$.

To give an example, if $\underline{P}$ is a finitely additive probability P on $\mathscr{A}_\Lambda$, and therefore in particular completely monotone, the induced lower probability P_* on $\mathscr{A}_\Omega$ is completely monotone as well by Theorem 1 (it is a so-called belief function ([24]) if Λ is finite). This lower probability was given an evidential interpretation in [24], and it can now be given a clear behavioural interpretation: $P_*(A)$ is the supremum acceptable buying price for the gamble I_A—or equivalently, the supremum rate for betting on the event

A—taking into account the behavioural dispositions expressed both by P and by the conditional lower prevision $\underline{P}(\cdot|\Lambda)$ associated with the multi-valued mapping Γ.

This leads us to another interesting question. As before, let us denote by $\underline{P}_*$ the restriction of $\underline{P}_\circ$ to $\mathscr{A}_\Omega$, i.e., $\underline{P}_*$ is the induced lower probability mentioned in Section 3, and let $\underline{E}_*$ be its (unconditional) natural extension to all gambles on Ω. Then we can ask ourselves if $\underline{P}_*$ contains enough information in order to calculate $\underline{P}_\circ$ on $\mathscr{A}_\Omega$-measurable gambles. In other words, do $\underline{E}_*$ and $\underline{P}_\circ$ coincide on $\mathscr{A}_\Omega$-measurable gambles?

To illustrate this question, let us again consider the case that $\underline{P}$ is a finitely additive probability P, so $P_* = \underline{P}_*$ is completely monotone. Since the natural extension of a 2-monotone lower probability is given by Choquet integration ([27], see also [11]), we find that $\underline{E}(X_\circ) = (C)\int_\Lambda X_\circ \, \mathrm{d}P$. We would like to know if this coincides with the natural extension $\underline{E}_*$ of P_* to $\mathscr{L}(\Omega)$, which is given by $\underline{E}_*(X) = (C)\int X \, \mathrm{d}P_*$. In this particular case, the equality for $\mathscr{A}_\Omega$-measurable X indeed follows from a result by Wasserman (which we can find in [14]). But we would like to find out whether such an equality can be proven for more general types of lower probabilities $\underline{P}$.

We shall see below that this is only guaranteed to succeed if the lower probability $\underline{P}$ is 2-monotone. But, as a first step, let us motivate by means of a counter-example why we only seek to prove the equality for $\mathscr{A}_\Omega$-measurable gambles.

Example 1. Consider an arbitrary non-empty set Λ that contains at least two elements, and let $\Omega = \Lambda$. Let $\mathscr{A}_\Lambda = \wp(\Lambda)$, $\mathscr{A}_\Omega = \{\emptyset, \Omega\}$ and let $\Gamma\colon \Lambda \to \wp(\Omega)$ be defined by $\Gamma(\lambda) = \{\lambda\}$ for all $\lambda \in \Lambda$. This Γ is strongly measurable with respect to $\mathscr{A}_\Lambda$ and $\mathscr{A}_\Omega$. Consider a coherent lower probability $\underline{P}$ defined on $\mathscr{A}_\Lambda$, and such that there is some proper subset A of Λ with $0 < \underline{P}(A) < 1$. Clearly, $A \notin \mathscr{A}_\Omega$. We know that $(\Lambda \times A)_\circ = \{\lambda \in \Lambda\colon \{\lambda\} \times \Gamma(\lambda) \subseteq \Lambda \times A\} = A$, whence $\underline{P}_\circ(\Lambda \times A) = \underline{P}(A) > 0$. On the other hand, the lower probability $\underline{P}_*$ induced on $\mathscr{A}_\Omega$ is completely specified by $\underline{P}_*(\emptyset) = 0$ and $\underline{P}_*(\Omega) = 1$. It is easy to show using the definition (1) of natural extension that $\underline{E}_*(A) = 0$ since $A \neq \Omega$.[10] This shows that generally speaking $\underline{P}_\circ$ and $\underline{E}_*$ need not coincide on events (and gambles) that are not $\mathscr{A}_\Omega$-measurable. ⧫

Let us now first assume that the lower probability $\underline{P}$ on $\mathscr{A}_\Lambda$ is 2-monotone. We have seen in Theorem 1 that the induced lower probability $\underline{P}_*$ on $\mathscr{A}_\Omega$ is 2-monotone as well. We then have the following theorem.

Theorem 8. *Let $\underline{P}$ be a 2-monotone lower probability defined on the σ-field $\mathscr{A}_\Lambda$ on Λ, and let $\underline{P}_*$ be the 2-monotone lower probability defined on the σ-field $\mathscr{A}_\Omega$ by*

$$\underline{P}_*(A) = \underline{P}(\{\lambda \in \Lambda\colon \Gamma(\lambda) \subseteq A\}), \quad A \in \mathscr{A}_\Omega,$$

where $\Gamma\colon \Lambda \to \wp(\Omega)$ is strongly measurable with respect to $\mathscr{A}_\Lambda$ and $\mathscr{A}_\Omega$. Then the natural extension $\underline{E}_*$ of $\underline{P}_*$ coincides on $\mathscr{A}_\Omega$-measurable gambles with the natural extension $\underline{P}_\circ$ of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$. In particular, we have for any $\mathscr{A}_\Omega$-measurable gamble X on Ω that

$$\underline{P}_*(X) = (C)\int_\Omega X \, \mathrm{d}\underline{P}_* = (C)\int_\Lambda X_\circ \, \mathrm{d}\underline{P} = \underline{P}_\circ(X),$$

where the integrals are Choquet integrals [11].

[10] Alternatively, observe that $\underline{P}_*$ is vacuous, so its natural extension is vacuous as well.

Proof. Let X be a gamble on Ω that is $\mathscr{A}_\Omega$-measurable. Then there is a sequence (X_n) of $\mathscr{A}_\Omega$-measurable simple gambles converging uniformly to X. Since, by Walley's Theorem 2.6.1(l) in [28], $\underline{Q}(X_n) \to \underline{Q}(X)$ for any coherent lower prevision $\underline{Q}$ defined on the $\mathscr{A}_\Omega$-measurable gambles, we see that it suffices to prove the equality of the coherent $\underline{E}_*$ and $\underline{P}_\circ$ for simple $\mathscr{A}_\Omega$-measurable gambles. Assume therefore that X is simple. Then we can write $X = \sum_{k=1}^n x_k I_{F_k}$, where $F_k \in \mathscr{A}_\Lambda$, $F_k \subseteq F_{k+1}$ for $k = 1, \dots, n-1$, $F_n = \Omega$, and $x_k \geq 0$ for $k = 1, \dots, n-1$. Since the natural extension of a 2-monotone lower prevision is given by Choquet integration, we find that

$$\underline{E}_*(X) = (C)\int_\Omega X \, \mathrm{d}\underline{P}_* = \sum_{k=1}^n x_k \underline{P}_*(F_k) = \sum_{k=1}^n x_k \underline{P}(F_{k*}).$$

It is not difficult to prove that $X_\circ$ is a simple gamble on Λ with representation $X_\circ = \sum_{k=1}^n x_k I_{F_{k*}}$, where also $F_{k*} \in \mathscr{A}_\Lambda$, $F_{k*} \subseteq F_{k+1*}$, for $k = 1, \dots, n-1$ and $F_{n*} = \Lambda$; hence, we see that

$$\sum_{k=1}^n x_k \underline{P}(F_{k*}) = (C)\int_\Lambda X_\circ \, \mathrm{d}\underline{P} = \underline{E}(X_\circ) = \underline{P}_\circ(X),$$

where the one but last equality follows from the fact that the natural extension of the 2-monotone lower probability $\underline{P}$ is also given by Choquet integration. □

As stated before, this theorem generalises results given by Wasserman; it also proves what was essentially hinted at in [6, Remark 2]. It seems easy to generalise it to not necessarily bounded gambles, using limit arguments. It could also fairly easily be generalised to lower probabilities $\underline{P}$ defined on a field, rather than a σ-field of events.

The following counterexample shows that $\underline{P}_*$ and $\underline{P}_\circ$ need not coincide for all $\mathscr{A}_\Omega$-measurable gambles when $\underline{P}$ is not 2-monotone. It is based on an idea expressed in Theorem 6.2 in [27], and it uses the central result (Theorem 7) of the previous section.

Example 2. Consider $\Lambda = \{a, b, c, d\}$ and let $\Omega = \Lambda$. Let $\mathscr{A}_\Lambda = \wp(\Lambda)$, and let $\mathscr{A}_\Omega$ be the (σ-)field generated by the partition $\{\{a\}, \{b\}, \{c, d\}\}$. Let $\Gamma \colon \Lambda \to \wp(\Omega)$ be defined by $\Gamma(\lambda) = \{\lambda\}$ for all $\lambda \in \Lambda$. Then Γ is strongly measurable with respect to $\mathscr{A}_\Lambda$ and $\mathscr{A}_\Omega$.

Consider the lower probability $\underline{P}$ on $\mathscr{A}_\Lambda$ given by the lower envelope of the probability measures with mass functions $(0.25, 0.25, 0.25, 0.25)$, $(0.5, 0.5, 0, 0)$. It is clear that $\underline{P}$ is coherent, since any lower envelope is. However, $\underline{P}$ is not 2-monotone:

$$\underline{P}(\{a, c\}) + \underline{P}(\{a, d\}) = 1 > \frac{3}{4} = \underline{P}(\{a, c, d\}) + \underline{P}(\{a\}).$$

On the other hand, the lower probability $\underline{P}_*$ induced by Γ on $\mathscr{A}_\Omega$ is nothing but the restriction of $\underline{P}$ to $\mathscr{A}_\Omega$, which is 2-monotone, since $\mathscr{A}_\Omega$ is generated by only three atoms, and it is easy to show that any coherent lower probability defined a field generated by at most three atoms is necessarily 2-monotone. Consider the simple $\mathscr{A}_\Omega$-measurable gamble

$$Z = I_{\{a,c\}} + I_{\{a,d\}} = I_{\{a\}} + I_{\{a,c,d\}}.$$

Then, since the natural extension $\underline{E}_*$ of the 2-monotone $\underline{P}_*$ is given by Choquet-integration for $\mathscr{A}_\Omega$-measurable gambles, we see that

$$\underline{E}_*(Z) = (C)\int_\Omega Z \,\mathrm{d}\underline{P}_* = \underline{P}_*(\{a\}) + \underline{P}_*(\{a,c,d\}) = \underline{P}(\{a\}) + \underline{P}(\{a,c,d\}) = \frac{3}{4}.$$

On the other hand, if we use the notation established in the previous section, the set

$$_\circ\mathscr{A}_\Lambda = \{X \in \mathscr{L}(\Lambda \times \Omega) \colon X_\circ \in \mathscr{A}_\Lambda\}$$

contains the set $\{\Lambda \times A \colon A \subseteq \Omega\}$, or, with some abuse of notation, contains $\wp(\Omega)$. Note that the restriction of $\underline{P}_1$ to $\wp(\Omega)$ is nothing but the original lower probability $\underline{P}$, so the natural extension $\underline{E}$ of $\underline{P}$ will be dominated on its domain $\mathscr{L}(\Omega)$ – recall that $\Omega = \Lambda$ – by the natural extension $\underline{P}_\circ$ of $\underline{P}_1$ (see Theorem 7 for more details). In particular, this means that, since $\underline{E}$ is coherent and therefore superadditive, and since it coincides with $\underline{P}$ on events (because $\underline{P}$ is coherent):

$$\begin{aligned} \underline{P}_\circ(Z) \geq \underline{E}(Z) = \underline{E}(I_{\{a,c\}} + I_{\{a,d\}}) &\geq \underline{E}(I_{\{a,c\}}) + \underline{E}(I_{\{a,d\}}) \\ &= \underline{P}(\{a,c\}) + \underline{P}(\{a,d\}) = 1. \end{aligned}$$

This means that $\underline{P}_\circ(Z) > \underline{E}_*(Z)$, and therefore the natural extension $\underline{E}_*$ of $\underline{P}_*$ does not coincide with the natural extension $\underline{P}_\circ$ of $\underline{P}$ and $\underline{P}(\cdot|\Lambda)$ for all $\mathscr{A}_\Omega$-measurable gambles. ♦

6 Conclusions

In this paper, we have started a study of lower probabilities and lower previsions induced by multi-valued mappings, from the behavioural point of view, following the suggestions in [28] and [6]. We have given a behavioural interpretation and justification for the definitions of induced lower (and upper) probabilities that are commonly used in the literature. This justification is based on the notions of coherence and natural extension, which play a central part in the behavioural theory of imprecise probabilities. It leads in a very natural way to the generalisation of the notions of induced lower (and upper) probabilities that we have studied above, and which allows us to associate a kind of integral with them. We have seen that (only) under some conditions this integral is a Choquet integral.

It is moreover clear from the results we proved in the previous section that our approach also allows us to study existing problems in the theory of random sets from a new, different point of view. It turns out that it is once again the notions of coherence and natural extension that lead to alternative proofs for, and generalisations of, existing theorems. This provides evidence for their unifying and explanatory power.

There is a simple idea underlying the arguments of this paper, namely that *a multi-valued map represents conditional information, and that this information can be represented by a (specific) conditional lower prevision.* To put this idea in its proper perspective, let us first consider the simpler case of a *single*-valued map γ between the initial space Λ and the final space Ω. Given a precise probability P on the initial space Λ, such a map induces a precise probability P_γ on the final space Ω by

$$P_\gamma(A) = P(\gamma^{-1}(A)) = P(\{\lambda \in \Lambda \colon \gamma(\lambda) \in A\})$$

for subsets A of Ω.[11] For gambles X on Ω, we have the well-known 'change of variables result' for previsions or expectations (i.e., for the Lebesgue integrals associated with the probabilities):

$$E_{P_\gamma}(X) = \int_\Omega X \,\mathrm{d}P_\gamma = \int_\Lambda X \circ \gamma \mathrm{d}P = E_P(X \circ \gamma). \tag{11}$$

The interpretation of E_{P_γ} becomes immediate if we interpret the map γ as conditional information: *if the random variable L assumes the value λ in Λ, then we know that O assumes the value $\gamma(\lambda)$ in Ω.* This information can be represented by the following conditional linear prevision $P(\cdot|\Lambda)$: for all gambles X on Ω and all $\lambda \in \Lambda$,

$$P(X|\lambda) = X(\gamma(\lambda)) = (X \circ \gamma)(\lambda),$$

i.e., $P(\cdot|\lambda)$ is the precise probability on Ω all of whose probability mass lies in the point $\gamma(\lambda)$. We can now find the prevision (or expectation) of the gamble X by combining the marginal prevision E_P on the space Λ, and the conditional prevision $P(\cdot|\Lambda)$: using an appropriate version of the law of total probability, or equivalently, the precise version of Walley's Marginal Extension Theorem ([28, Theorem 6.7.2]), or in other words, natural extension, we find for the prevision $P(X)$ of the gamble X that

$$P(X) = E_P(P(X|\Lambda)) = E_P(X \circ \gamma) = E_{P_\gamma}(X), \tag{12}$$

where the last equality follows from Eq. (11). This tells us that P_γ (or equivalently, E_{P_γ}) is the probabilistic information on Ω that can be deduced form the marginal model P on Λ and the conditional model $P(\cdot|\Lambda)$ that represents the information present in the map γ.

When we want to generalise this course of reasoning from single-valued maps γ to multi-valued maps Γ, we face the following problem: the information present in the multi-valued mapping Γ can no longer be represented by a conditional *linear* prevision. In other words, if we want to remain within the framework of precise probability theory, we must abandon the simple and powerful device of interpreting Γ as conditional information. But if we work with the theory of imprecise probabilities, it is still perfectly possible to interpret Γ as conditional information that can be represented by a special conditional *lower* prevision (see Eq. (5)). And so, the whole argument outlined above can be extended from single-to multi-valued mappings, which is essentially what we have done in this paper. Observe, in this respect, that Theorem 2 generalises Eq. (12), and that Eq. (11) is a special case of Theorem 8.

As a topic for future research, we intend to investigate whether these notions also allow us to shed new light on other existing problems in the theory of multi-valued mappings, such as how to define independence, and how to provide a behavioural interpretation to existing definitions of independence.

[11] Let us dispense with technical aspects of measurability here.

Acknowledgements

This paper was written during a visit of Enrique Miranda to the SYSTeMS group at Ghent University, and has been partially funded by FEDER-MCYT, grant number BFM2001-3515, and URJC, PPR-2003-41, as well as by Research Project G.0139.01 of the Fund for Scientific Research, Flanders, Belgium. The scientific responsability rests with the authors. They would also like to acknowledge a number of insightful comments and suggestions for improvement made by two anonymous referees.

References

[1] Th. Augustin. Modelling weak information with generalised probability assignments. In H. H. Bock and W. Polasek, editors, *Data Analysis and Information Systems – Statistical and Conceptual Approaches*, pages 101–113. Springer Verlag, Heidelberg, 1996.

[2] Th. Augustin. Generalized basic probability assignaments. *International Journal of General Systems*, 2004. Conditionally accepted for publication.

[3] J. O. Berger. The robust Bayesian viewpoint. In J. B. Kadane, editor, *Robustness of Bayesian Analyses*, pages 63-144. Elsevier Science, Amsterdam, 1984.

[4] G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953–1954.

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

[6] G. de Cooman and D. Aeyels. A random set description of a possibility measure and its natural extension. *IEEE Transactions on Systems, Man and Cybernetics—Part A: Systems and Humans*, 30(2):124–130, 2000.

[7] B. de Finetti. *Theory of Probability*. John Wiley & Sons, Chichester, 1974–1975. English Translation of *Teoria delle Probabilità*.

[8] G. Debreu. Integration of correspondences. In *Proceedings of the Fifth Berkeley Symposium of Mathematical Statistics and Probability*, pages 351–372, Berkeley, 1965.

[9] A. P. Dempster. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics*, 39:957–966, 1967.

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

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

[12] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics*, pages 17–38. Valencia University Press, Valencia, 1980.

[13] I. R. Goodman, H. T. Nguyen and E. A. Walker. *Conditional inference and logic for intelligent systems*. North-Holland, Amsterdam, 1991.

[14] M. Grabisch, H. T. Nguyen, and E. A. Walker. *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Kluwer Academic Publishers, Dordrecht, 1995.

[15] C. J. Himmelberg. Measurable relations. *Fundamenta Mathematicae*, 87:53–72, 1975.

[16] D. G. Kendall. Foundations of a theory of random sets. In E. F. Harding and D. G. Kendall, editors, *Stochastic Geometry*, pages 322–376. Wiley, New York, 1974.

[17] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.

[18] R. Kruse and K. D. Meyer. *Statistics with Vague Data*. D. Reidel Publishing Company, Dordrecht, 1987.

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

[20] G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, New York, 1975.

[21] E. Miranda, I. Couso, and P. Gil. A random set characterization of possibility measures. *Information Sciences*, 2004. Accepted for publication.

[22] E. Miranda, G. de Cooman, and I. Couso. Imprecise probabilities induced by multi-valued mappings. In *Proceedings of the Ninth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002, Annecy, France, July 1–5, 2002)*, pages 1061–1068. Gutenberg, 2002.

[23] H. T. Nguyen. On random sets and belief functions. *Journal of Mathematical Analysis and Applications*, 65(3):531–542, 1978.

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

[25] V. Strassen. Meßfehler und Information. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:273–305, 1964.

[26] M. Sugeno. *The Theory of Fuzzy Integrals and Its Applications*. PhD thesis, Tokyo Institute of Technology, Tokyo, 1974.

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

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