FORWARD IRRELEVANCE

GERT DE COOMAN AND ENRIQUE MIRANDA

ABSTRACT. We investigate how to combine marginal assessments about the values that random variables assume separately into a model for the values that they assume jointly, when (i) these marginal assessments are modelled by means of coherent lower previsions, and (ii) we have the additional assumption that the random variables are forward epistemically irrelevant to each other. We consider and provide arguments for two possible combinations, namely the forward irrelevant natural extension and the forward irrelevant product, and we study the relationships between them. Our treatment also uncovers an interesting connection between the behavioural theory of coherent lower previsions, and Shafer and Vovk's game-theoretic approach to probability theory.

1. INTRODUCTION

In probability and statistics, assessments of independence are often useful as they allow us to reduce the complexity of inference problems. To give an example, and to set the stage for the developments in this paper, we consider two random variables X_1 and X_2, taking values in the respective *finite* sets $\mathscr{X}_1$ and $\mathscr{X}_2$. Suppose that a subject is uncertain about the values of these variables, but that he has some model expressing his beliefs about them. Then we say that X_1 is *epistemically irrelevant* to X_2 for the subject when he assesses that learning the actual value of X_1 won't change his beliefs (or belief model) about the value of X_2. We say that X_1 and X_2 are *epistemically independent* when X_1 and X_2 are epistemically irrelevant to one another; the terminology is borrowed from Walley (1991, Chapter 9).

Let us first look at what these general definitions yield when the belief models our subject uses are precise probabilities. If the subject has a marginal probability mass function $p_1(x_1)$ for the first variable X_1, and a conditional mass function $q_2(x_2|x_1)$ for the second variable X_2 conditional on the first, then we can calculate his joint mass function $p(x_1,x_2)$ using Bayes's rule: $p(x_1,x_2) = p_1(x_1)q_2(x_2|x_1)$. Now consider any real-valued function f on $\mathscr{X}_1 \times \mathscr{X}_2$. We shall call such functions *gambles*, because they can be interpreted as uncertain rewards. We find for the prevision (or expectation, or fair price, we use de Finetti's (1974–1975) terminology and notation throughout this paper.) of such a gamble f that:

$$\begin{aligned} P(f) &= \sum_{(x_1,x_2)\in\mathscr{X}_1\times\mathscr{X}_2} f(x_1,x_2)p_1(x_1)q_2(x_2|x_1) \\ &= \sum_{x_1\in\mathscr{X}_1} p_1(x_1) \sum_{x_2\in\mathscr{X}_2} f(x_1,x_2)q_2(x_2|x_1) = \sum_{x_1\in\mathscr{X}_1} p_1(x_1)Q_2(f(x_1,\cdot)|x_1) \\ &= P_1(Q_2(f|X_1)), \end{aligned} \tag{1}$$

Date: November 9, 2007.

2000 *Mathematics Subject Classification.* 60A99,60G99,91A40.

Key words and phrases. Imprecise probabilities, coherent lower previsions, forward irrelevance, game-theoretic probability, epistemic irrelevance, epistemic independence, event trees.

Supported by research grant G.0139.01 of the Flemish Fund for Scientific Research (FWO), and by MCYT, projects MTM2004-01269, TSI2004-06801-C04-01.

where we let $Q_2(f|X_1)$ be the subject's conditional prevision of f given X_1, which is a gamble on $\mathscr{X}_1$ whose value in x_1,

$$Q_2(f|x_1) := Q_2(f(x_1,\cdot)|x_1) = \sum_{x_2\in\mathscr{X}_2} f(x_1,x_2)q_2(x_2|x_1),$$

is the subject's conditional prevision of f given that $X_1 = x_1$. We also let P_1 be the subject's marginal prevision (operator) for the first random variable, associated with the marginal mass function p_1: $P_1(g) := \sum_{x_1\in\mathscr{X}_1} g(x_1)p_1(x_1)$ for all gambles g on $\mathscr{X}_1$.

When the subject judges X_1 to be (epistemically) irrelevant to X_2, then we get for all $x_1 \in \mathscr{X}_1$ and $x_2 \in \mathscr{X}_2$ that

$$q_2(x_2|x_1) = p_2(x_2), \tag{2}$$

where p_2 is the subject's marginal mass function for the second variable X_2 that we can derive from the joint p using $p_2(x_2) := \sum_{x_1\in\mathscr{X}_1} p(x_1,x_2)$. The equality (2) expresses that learning that $X_1 = x_1$ doesn't change the subject's probability model for the value of the second variable. Condition (2) is equivalent to requiring that for all $x_1 \in \mathscr{X}$ and all gambles f on $\mathscr{X}_1 \times \mathscr{X}_2$,

$$Q_2(f(x_1,\cdot)|x_1) = P_2(f(x_1,\cdot)), \tag{3}$$

where now P_2 is the subject's marginal prevision (operator) for the second variable, associated with the marginal mass function p_2. We can then write for the joint prevision:

$$P(f) = P_1(P_2(f)), \tag{4}$$

where f is any gamble on $\mathscr{X}_1 \times \mathscr{X}_2$, and where we let $P_2(f)$ be the gamble on $\mathscr{X}_1$ that assumes the value $P_2(f(x_1,\cdot))$ in $x_1 \in \mathscr{X}_1$.

Similarly, when X_2 is epistemically irrelevant to X_1 for our subject, then

$$q_1(x_1|x_2) = p_1(x_1) \tag{5}$$

for all $x_1 \in \mathscr{X}_1$ and $x_2 \in \mathscr{X}_2$. Here $q_1(x_1|x_2)$ is the subject's mass function for the first variable X_1 conditional on the second. This leads to another expression for the joint prevision:

$$P(f) = P_2(P_1(f)). \tag{6}$$

Expressions (4) and (6) for the joint are equivalent, as generally $P_1(P_2(f)) = P_2(P_1(f))$. This is related to the fact that Conditions (2) and (5) are equivalent: if X_1 is epistemically irrelevant to X_2 then X_2 is epistemically irrelevant to X_1, and *vice versa*. In other words, *for precise probability models, epistemic irrelevance is equivalent to epistemic independence.*

Some caution is needed here: this equivalence is only guaranteed if the marginal mass functions are everywhere non-zero. If some events have zero probability, then it can still be guaranteed provided we slightly change the definition of epistemic irrelevance, and for instance impose $q_2(x_2|x_1) = p_2(x_2)$ only when $p_1(x_1) > 0$.

All of this will seem tritely obvious to anyone with a basic knowledge of probability theory, but the point we want to make, is that the situation changes dramatically when we use belief models that are more general (and arguably more realistic) than the precise (Bayesian) ones, such as Walley's (1991) imprecise probability models.

On Walley's view, a subject may not generally be disposed to specify a fair price $P(f)$ for any gamble f, but we can always ask for his *lower prevision* $\underline{P}(f)$, which is his supremum acceptable price for buying the uncertain reward f, and his *upper prevision* $\overline{P}(f)$, which is his infimum acceptable price for selling f. We give a fairly detailed introduction to Walley's theory in Section 2.

On this new approach, if X_1 is epistemically irrelevant to X_2 for our subject, then [compare with Condition (3)]

$$\underline{Q}_2(f(x_1,\cdot)|x_1) = \underline{P}_2(f(x_1,\cdot))$$

for all gambles f on $\mathscr{X}_1 \times \mathscr{X}_2$ and all $x_1 \in \mathscr{X}_1$. Here, similar to what we did before, $\underline{P}_2$ is the subject's marginal lower prevision (operator) for X_2, and $\underline{Q}_2(\cdot|X_1)$ is his lower prevision (operator) for X_2 *conditional* on X_1. We shall see in Section 3 that a reasonable joint model[1] for the value that (X_1, X_2) assumes in $\mathscr{X}_1 \times \mathscr{X}_2$ is then given by [compare with Eqs. (1) and (4)]

$$\underline{P}(f) = \underline{P}_1(\underline{Q}_2(f|X_1)) = \underline{P}_1(\underline{P}_2(f)) \tag{7}$$

for all gambles f on $\mathscr{X}_1 \times \mathscr{X}_2$, where $\underline{P}_1$ is the subject's marginal lower prevision (operator) for X_1, and where we also let $\underline{P}_2(f)$ be the gamble on $\mathscr{X}_1$ that assumes the value $\underline{P}_2(f(x_1,\cdot))$ in any $x_1 \in \mathscr{X}_1$.

On the other hand, when our subject judges X_2 to be epistemically irrelevant to X_1, we are eventually led to the joint model

$$\underline{P}'(f) = \underline{P}_2(\underline{P}_1(f)),$$

where now $\underline{P}_1(f)$ is the gamble on $\mathscr{X}_2$ that assumes the value $\underline{P}_1(f(\cdot,x_2))$ in any $x_2 \in \mathscr{X}_1$. Interestingly, it now is no longer guaranteed that $\underline{P} = \underline{P}'$, or in other words that $\underline{P}_1(\underline{P}_2(\cdot)) = \underline{P}_2(\underline{P}_1(\cdot))$. We give an example in Section 4, where we also argue that this asymmetry (i) isn't caused by 'pathological' consequences involving zero probabilities, and can't be remedied by simple tricks as the one mentioned above for (precise) probabilities; and (ii) is *endemic*, as it seems to be there for *all* imprecise models, apart from the extreme (linear or vacuous) ones. It results that, here, the notion of *epistemic irrelevance is fundamentally asymmetrical, and is no longer equivalent to the symmetrical notion of epistemic independence.* This was discussed in much more detail by Couso et al. (2000).

But then, as the two notions are no longer equivalent here, it becomes quite important to distinguish between them when we actually represent beliefs using imprecise probability models. There are a number of reasons why a subject shouldn't use the consequences of epistemic independence automatically, when he is only assessing epistemic irrelevance.

First of all, an assessment of epistemic independence is stronger, and leads to higher joint lower previsions. As lower previsions represent supremum buying prices, higher values represent stronger commitments, and these may be unwarranted when it is only epistemic irrelevance that our subject wants to model.

Secondly, joint lower previsions based on an epistemic irrelevance assessment are generally relatively straightforward to calculate, as Eq. (7) testifies. But calculating joint lower previsions from marginals based on an epistemic independence assessment is quite often a very complicated affair; see for instance the expressions in Section 9.3.2 of Walley (1991).

Moreover, there are special but nevertheless important situations where we want to argue that it may be natural to make an epistemic irrelevance assessment, but not one of independence. Suppose, for instance that we consider two random variables, X_1 and X_2, where our subject knows that the value of X_1 *will be revealed to him before* that of X_2.[2] Then assessing that X_1 and X_2 are epistemically independent amounts to assessing that

(i) X_1 is epistemically irrelevant to X_2: getting to know the value of X_1 doesn't change our subject's beliefs about X_2;

[1] This is the most conservative joint lower prevision that is coherent with $\underline{P}_1$ and $\underline{Q}_2(\cdot|X_1)$, see also (Walley, 1991, Section 6.7).

[2] The discussion that follows here, as well as the one in Appendix A, generalises naturally to random processes, where the values of a process are revealed at subsequent points in time.

(ii) X_2 is epistemically irrelevant to X_1: getting to know the value of X_2 doesn't change our subject's beliefs about X_1.

But since the subject knows that he will always know the value of X_1 before that of X_2, (ii) is effectively a counter-factual statement for him: "if I got to the value of X_2 first, then learning that value wouldn't affect my beliefs about X_1". It's not clear that making such an assessment has any real value, and we feel it is much more natural in such situations context to let go of (ii) and therefore to resort to epistemic irrelevance.

This line of reasoning can also be related to Shafer's (1985) idea that conditioning must always be associated with a *protocol*. A subject can then only *meaningfully* update (or condition) a probability model on events that he envisages may happen (according to the established protocol). In the specific situation described above, conditioning on the variable X_2 could only legitimately be done if it were possible to find out the value of X_2 without getting to know that of X_1, *quod non*. Therefore, it isn't legitimate to consider the conditional lower prevision $\underline{Q}_1(\cdot|X_2)$ expressing the beliefs about X_1 conditional on X_2, and we therefore can't meaningfully impose (ii), as it requires that $\underline{Q}_1(\cdot|X_2) = \underline{P}_1$. Shafer has developed and formalised his ideas about protocols and conditioning using the notion of an event tree, in an interesting book dealing with causal reasoning (Shafer, 1996). In Appendix A, we formulate a simple example, where X_1 and X_2 are the outcomes of successive coin tosses, in Shafer's event-tree language. We show that in this specific case, the general notion of *event-tree independence* that he develops in his book, is effectively equivalent to the requirement that X_1 should be *epistemically irrelevant* to X_2.

For all these reasons, we feel that a study of the joint lower previsions that result from epistemic irrelevance assessments is quite important, also from a practical point of view. We take the first steps towards such a study in this paper.

We shall consider a finite number of variables X_1, ..., X_N taking values in respective sets $\mathscr{X}_1, \dots, \mathscr{X}_N$. We are going to assume moreover that for $k = 2, \dots, N$ the variables X_1, ..., X_{k-1} are epistemically irrelevant to X_k: we shall call such an assessment *forward irrelevance*. It means that we aren't learning from the 'past',[3] and it will be in general weaker than an assessment of epistemic independence. We shall study which are the inferences that can be made, based on such assessments.

In order to model the information we have about the variables X_1, ..., X_N, we use the behavioural theory of imprecise probabilities, developed mainly by Walley (1991), with influences from earlier work by de Finetti (1974–1975) and Williams (1975), amongst others. This theory constitutes a generalisation of de Finetti's account of subjective probability, and uses (coherent) lower and upper previsions to represent a subject's behavioural dispositions. We give a brief introduction to the basic ideas behind coherent lower previsions in Section 2, and we explain how they can be identified with sets of (finitely additive) probability measures. This introductory section can be skipped by anyone with a reasonable working knowledge of coherent lower previsions.

It may appear at first sight that, because of this choice of model, the results we obtain have a limited interest for people working outside the field of imprecise probabilities. We think that this is not necessarily so, for two reasons: on the one hand, the mathematical theory of coherent lower and upper previsions subsumes a number of approaches to uncertainty modelling in the literature, like probability charges (Bhaskara Rao

[3] Of course, we can only speak of 'the past' in this context when the index k refers to the 'time' that the actual value of a variable X_k is revealed. This is the specific situation where we argue that the notion of epistemic irrelevance is more natural than epistemic independence. But we don't question the interest of epistemic independence in other contexts, of course.

and Bhaskara Rao, 1983), 2- and n-monotone set functions (Choquet, 1953–1954), possibility measures (De Cooman, 2001; De Cooman and Aeyels, 1999, 2000; Dubois and Prade, 1988), and p-boxes (Ferson et al., 2003). This means that the results we establish here will also be valid for any of these models.

Moreover, the behavioural theory of imprecise probabilities can also be given a Bayesian sensitivity analysis interpretation: we may assume the existence of a precise but unknown probability model for the random variables $X_1, \dots, X_N$, and model our information about it by means of a set of possible models. As we shall see further on in Theorem 5, some of the results we shall find also make sense on such a sensitivity analysis interpretation.

In Section 3 we explain how a subject's assessment that he doesn't learn from the past can be used to combine a number of marginal lower previsions into a joint lower prevision, called their *forward irrelevant natural extension*. We study the properties of this combination, and show later that it can be related to specific types of *coherent probability protocols* introduced by Shafer and Vovk (2001). We also discuss another interesting way of combining marginal lower previsions into a joint, leading to their *forward irrelevant product*. This product has an interesting Bayesian sensitivity analysis interpretation. We also discuss its properties, and its relationship with the forward irrelevant natural extension. We show in particular that the forward irrelevant product generally dominates—is less conservative or more committal than—the forward irrelevant natural extension, and that these two coincide when the variables X_k we consider, can assume only a finite number of values.

As indicated above, our results also allow us to uncover a perhaps surprising relationship between Walley's (1991) behavioural theory of coherent lower previsions, and Shafer and Vovk's (2001) game-theoretic approach to probability theory. This is done in Section 5. In that same section, we also give an interesting financial interpretation for the forward irrelevant natural extension in terms of an investment game involving futures. We have gathered the proofs of the main results in Appendix B.

2. Coherent lower and upper previsions

Here, we present a succinct overview of the relevant main ideas underlying the behavioural theory of imprecise probabilities, in order to make it easier for the reader to understand the course of reasoning that we shall develop. We refer to Walley (1991) for extensive discussion and motivation.

2.1. **Basic notation and rationality requirements.** Consider a subject who is uncertain about something, say, the value that a random variable X assumes in a set of possible values[4] $\mathscr{X}$. Then, a bounded real-valued function on $\mathscr{X}$ is called a *gamble* on $\mathscr{X}$ (or on X), and the set of all gambles on $\mathscr{X}$ is denoted by $\mathscr{L}(\mathscr{X})$. Given a real number μ, we also use μ to denote the gamble that takes the constant value μ. A *lower prevision* $\underline{P}$ is a real-valued map (a functional) defined on some subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X})$, called its *domain*. For any gamble f in $\mathscr{K}$, $\underline{P}(f)$ is called the lower prevision of f.

A subset A of $\mathscr{X}$ is called an *event*, and it can be identified with its *indicator* I_A, which is the gamble on $\mathscr{X}$ that assumes the value one on A and zero elsewhere. The *lower probability* $\underline{P}(A)$ of A is defined as the lower prevision $\underline{P}(I_A)$ of its indicator I_A. On the other hand, given a lower prevision $\underline{P}$, its *conjugate upper prevision* $\overline{P}$ is defined on the set of gambles $-\mathscr{K} := \{-f : f \in \mathscr{K}\}$ by $\overline{P}(f) := -\underline{P}(-f)$ for every $-f$ in the domain of $\underline{P}$. This conjugacy relationship shows that we can restrict our attention to lower previsions only. If the domain of $\overline{P}$ contains only indicators, then we also call $\overline{P}$ an *upper probability*.

[4]We don't require $\mathscr{X}$ to be a subset of the reals, nor that X satisfies any kind of measurability condition.

A lower prevision $\underline{P}$ with domain $\mathscr{K}$ is called *coherent* when for any natural numbers $n \geq 0$ and $m \geq 0$, and $f_0, \dots, f_n$ in $\mathscr{K}$:

$$\sup_{x\in\mathscr{X}} \left[\sum_{k=1}^{n} [f_k(x) - \underline{P}(f_k)] - m[f_0(x) - \underline{P}(f_0)] \right] \geq 0. \tag{8}$$

Coherent lower previsions share a number of basic properties. For instance, given gambles f and g in $\mathscr{K}$, real numbers μ and non-negative real numbers λ, coherence implies that the following properties hold, whenever the gambles are in the domain $\mathscr{K}$ of $\underline{P}$:

(C1) $\underline{P}(f) \geq \inf_{x\in\mathscr{X}} f(x)$;
(C2) $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g)$ [super-additivity];
(C3) $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ [non-negative homogeneity];
(C4) $\underline{P}(f+\mu) = \underline{P}(f) + \mu$ [constant additivity].

Other properties can be found in Walley (1991, Section 2.6). It is important to mention here that when $\mathscr{K}$ is a linear space, coherence is equivalent to (C1)–(C3). More generally, a lower prevision on a general domain is coherent if and only if it can be extended to a coherent lower prevision on some linear space, i.e., a real functional satisfying (C1)–(C3).

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

$$\underline{E}(f) = \sup_{\substack{f_1,\dots,f_n\in\mathscr{K}\\ \mu_1,\dots,\mu_n\geq 0, n\geq 0}} \inf_{x\in\mathscr{X}} \left[f(x) - \sum_{k=1}^{n} \mu_k [f_k(x) - \underline{P}(f_k)] \right], \tag{9}$$

where the $\mu_1, \dots, \mu_n$ in the supremum are non-negative real numbers.

2.3. **Relation to precise probabilities.** A *linear prevision* P is a real-valued functional defined on a set of gambles $\mathscr{K}$, that satisfies

$$\sup \left[\sum_{i=1}^{n} f_i - \sum_{j=1}^{m} g_j \right] \geq \sum_{i=1}^{n} P(f_i) - \sum_{j=1}^{m} P(g_j) \tag{10}$$

for all natural numbers n and m, and all gambles $f_1, \dots, f_n, g_1, \dots, g_m$ in $\mathscr{K}$.

In particular, a linear prevision P on the set $\mathscr{L}(\mathscr{X})$ is a real linear functional that is positive (if $f \geq 0$ then $P(f) \geq 0$) and has unit norm ($P(I_{\mathscr{X}}) = 1$). Its restriction to events is a finitely additive probability. Moreover, any finitely additive probability defined on the set $\wp(\mathscr{X})$ of all events can be uniquely extended to a linear prevision on $\mathscr{L}(\mathscr{X})$. For this reason, we shall identify linear previsions on $\mathscr{L}(\mathscr{X})$ with finitely additive probabilities on $\wp(\mathscr{X})$. We denote by $\mathbb{P}(\mathscr{X})$ the set of all linear previsions on $\mathscr{L}(\mathscr{X})$.

Linear previsions are the *precise* probability models: they coincide with de Finetti's (1974–1975) notion of a coherent prevision or fair price. We call coherent lower and upper previsions *imprecise* probability models. That linear previsions are only required to be *finitely* additive, and not σ-additive, derives from the finitary character of the coherence requirement in Eq. (10).

Consider a lower prevision $\underline{P}$ defined on a set of gambles $\mathscr{K}$. Its *set of dominating linear previsions* $\mathscr{M}(\underline{P})$ is given by

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

Then $\underline{P}$ is coherent if and only if $\mathscr{M}(\underline{P}) \neq \emptyset$ and $\underline{P}(f) = \min\{P(f) \colon P \in \mathscr{M}(\underline{P})\}$ for all f in $\mathscr{K}$, i.e., if $\underline{P}$ is the *lower envelope* of $\mathscr{M}(\underline{P})$. And the natural extension $\underline{E}$ of a coherent $\underline{P}$ satisfies $\underline{E}(f) = \min\{P(f) \colon P \in \mathscr{M}(\underline{P})\}$ for all f in $\mathscr{L}(\mathscr{X})$. Moreover, the lower envelope of any set of linear previsions is always a coherent lower prevision.

These relationships allow us to provide coherent lower previsions with a *Bayesian sensitivity analysis* interpretation, which is different from the behavioural interpretation discussed below in Section 2.5: we might assume the existence of an ideal (but unknown) precise probability model P_T on $\mathscr{L}(\mathscr{X})$, and represent our imperfect knowledge about P_T by means of a (convex closed) set $\mathscr{M}$ of possible candidates for P_T. The information given by this set is equivalent to the one provided by its *lower envelope* $\underline{P}$, which is given by $\underline{P}(f) = \min_{P\in\mathscr{M}} P(f)$ for all f in $\mathscr{L}(\mathscr{X})$. This lower envelope $\underline{P}$ is a coherent lower prevision; and indeed, $P_T \in \mathscr{M}$ is equivalent to $P_T \geq \underline{P}$.

2.4. **Joint and marginal lower previsions.** Now consider two random variables X_1 and X_2 that may assume values in the respective sets $\mathscr{X}_1$ and $\mathscr{X}_2$. We assume that these variables are *logically independent*: the joint random variable (X_1, X_2) may assume all values in the product set $\mathscr{X}_1 \times \mathscr{X}_2$. A subject's coherent lower prevision $\underline{P}$ on a subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$ is a model for his uncertainty about the value that the joint random variable (X_1, X_2) assumes in $\mathscr{X}_1 \times \mathscr{X}_2$, and we call it a *joint lower prevision*.

We can associate with $\underline{P}$ its $\mathscr{X}_1$-*marginal* (lower prevision) $\underline{P}_1$, defined as follows:

$$\underline{P}_1(g) = \underline{P}(g')$$

for all gambles g on $\mathscr{X}_1$, such that the corresponding gamble g' on $\mathscr{X}_1 \times \mathscr{X}_2$, defined by $g'(x_1, x_2) := g(x_1)$ for all (x_1, x_2) in $\mathscr{X}_1 \times \mathscr{X}_2$, belongs to $\mathscr{K}$. The gamble g' is constant on the sets $\{x_1\} \times \mathscr{X}_2$, and we call it $\mathscr{X}_1$-*measurable*. In what follows, we identify g and g', and simply write $\underline{P}(g)$ rather than $\underline{P}(g')$. The marginal $\underline{P}_1$ is the corresponding model for the subject's uncertainty about the value that X_1 assumes in $\mathscr{X}_1$, irrespective of what value X_2 assumes in $\mathscr{X}_2$. The $\mathscr{X}_2$-marginal $\underline{P}_2$ is defined similarly. The coherence of the joint lower prevision $\underline{P}$ clearly implies the coherence of its marginals. If $\underline{P}$ is in particular a linear prevision on $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$, its marginals are linear previsions too.

Conversely, assume we start with two coherent marginal lower previsions $\underline{P}_1$ and $\underline{P}_2$, defined on the respective domains $\mathscr{K}_1 \subseteq \mathscr{L}(\mathscr{X}_1)$ and $\mathscr{K}_2 \subseteq \mathscr{L}(\mathscr{X}_2)$. We can interpret $\mathscr{K}_1$ as a set of gambles on $\mathscr{X}_1 \times \mathscr{X}_2$ that are $\mathscr{X}_1$-measurable, and similarly for $\mathscr{K}_2$. Any coherent joint lower prevision defined on a set $\mathscr{K}$ of gambles on $\mathscr{X}_1 \times \mathscr{X}_2$ that includes $\mathscr{K}_1$ and $\mathscr{K}_2$, and that coincides with $\underline{P}_1$ and $\underline{P}_2$ on their respective domains, i.e., has marginals $\underline{P}_1$ and $\underline{P}_2$, will be called a *product* of $\underline{P}_1$ and $\underline{P}_2$. We shall investigate various ways of defining such products further on in the paper.[5]

2.5. **The behavioural interpretation.** The mathematical theory presented above can be better understood if we consider the following *behavioural* interpretation.

We interpret a gamble as an uncertain reward: if the value of the random variable X turns out to be $x \in \mathscr{X}$, then the corresponding reward will be $f(x)$ (positive or negative), expressed in units of some (predetermined) linear utility. A subject's *lower prevision* $\underline{P}(f)$ for a gamble f is defined as his supremum acceptable price for buying f: it is the highest price μ such that the subject will accept to buy f for all prices $\alpha < \mu$ (buying f for a price

[5] It should be noted here that, in contradistinction with Walley (1991, Section 9.3.1), we don't intend the mere term 'product' to imply that the variables X_1 and X_2 are assumed to be independent in any way. On our approach, there may be many types of products, some of which may be associated with certain types of interdependence between the random variables X_1 and X_2. In other words, a product will be simply a joint distribution which is compatible with the given marginals.

α is the same thing as accepting the uncertain reward $f-\alpha$). A subject's *upper prevision* $\overline{P}(f)$ for f is his infimum acceptable selling price for f. Then $\overline{P}(f) = -\underline{P}(-f)$, since selling f for a price α is the same thing as buying $-f$ for the price $-\alpha$.

A lower prevision $\underline{P}$ with domain $\mathscr{K}$ is then coherent when a finite combination of acceptable buying transactions can't lead to a sure loss, and when moreover for any f in $\mathscr{K}$, we can't force the subject to accept f for a price strictly higher than his specified supremum buying price $\underline{P}(f)$, by exploiting buying transactions implicit in his lower previsions $\underline{P}(f_k)$ for a finite number of gambles f_k in $\mathscr{K}$, which he is committed to accept. This is the essence of the mathematical requirement (8).

The natural extension of a coherent lower prevision $\underline{P}$ is the smallest coherent extension to all gambles, and as such it summarises the behavioural implications of $\underline{P}$: $\underline{E}(f)$ is the supremum buying price for f that can be derived from the lower prevision $\underline{P}$ by arguments of coherence alone. We can see from its definition (9) that it is the supremum of all prices that the subject can be effectively forced to buy the gamble f for, by combining finite numbers of buying transactions implicit in his lower prevision assessments $\underline{P}$. In general $\underline{E}$ won't be the unique coherent extension of $\underline{P}$ to $\mathscr{L}(\mathscr{X})$; but any other coherent extension will point-wise dominate $\underline{E}$ and will therefore represent behavioural dispositions not implied by the assessments $\underline{P}$ and coherence alone.

Finally, a linear prevision P with a negation invariant domain $\mathscr{K} = -\mathscr{K}$ corresponds to the case where $\underline{P}(f) = \overline{P}(f)$, i.e., when the subject's supremum buying price coincides with his infimum selling price, and this common value is a *prevision* or *fair price* for the gamble f, in the sense of de Finetti (1974–1975). This means that our subject is disposed to buy the gamble f for any price $\mu < P(f)$, and to sell it for any price $\mu' > P(f)$ (but nothing is said about his behaviour for $\mu = P(f)$).

2.6. **Conditional lower previsions and separate coherence.** Consider a gamble h on $\mathscr{X}_1 \times \mathscr{X}_2$ and any value $x_1 \in \mathscr{X}_1$. Our subject's *conditional lower prevision* $\underline{P}(h|X_1 = x_1)$, also denoted as $\underline{P}(h|x_1)$, is the largest real number p for which the subject would buy the gamble h for any price strictly lower than p, if he knew in addition that the variable X_1 assumes the value x_1 (and nothing more!).

We shall denote by $\underline{P}(h|X_1)$ the *gamble* on $\mathscr{X}_1$ that assumes the value $\underline{P}(h|X_1 = x_1) = \underline{P}(h|x_1)$ in any x_1 in $\mathscr{X}_1$. We can assume that $\underline{P}(h|X_1)$ is defined for all gambles h in some subset $\mathscr{H}$ of $\mathscr{X}_1 \times \mathscr{X}_2$, and we call $\underline{P}(\cdot|X_1)$ a *conditional lower prevision* on $\mathscr{H}$. It is important to recognise that $\underline{P}(\cdot|X_1)$ maps any gamble h on $\mathscr{X}_1 \times \mathscr{X}_2$ to the gamble $\underline{P}(h|X_1)$ on $\mathscr{X}_1$. We also use the notations

$$G(h|x_1) := I_{\{x_1\}\times\mathscr{X}_2}[h - \underline{P}(h|x_1)], \quad G(h|X_1) = h - \underline{P}(h|X_1) := \sum_{x_1 \in \mathscr{X}_1} G(h|x_1);$$

$G(h|X_1)$ is a gamble on $\mathscr{X}_1$ as well.

That the domain of $\underline{P}(\cdot|x_1)$ is the same set $\mathscr{H}$ for all $x_1 \in \mathscr{X}_1$ is a consequence of the notion of separate coherence that we shall introduce next (Walley, 1991, Section 6.2.4), and that we shall assume for all the conditional lower previsions in this paper. We say that $\underline{P}(\cdot|X_1)$ is *separately coherent* if (i) for all x_1 in $\mathscr{X}_1$, $\underline{P}(\cdot|x_1)$ is a coherent lower prevision on its domain, and if moreover (ii) $\underline{P}(\{x_1\} \times \mathscr{X}_2|x_1) = 1$. It is a very important consequence of this definition that for all x_1 in $\mathscr{X}_1$ and all gambles h on the domain of $\underline{P}(\cdot|x_1)$,

$$\underline{P}(h|x_1) = \underline{P}(h(x_1,\cdot)|x_1).$$

This implies that a separately coherent $\underline{P}(\cdot|X_1)$ is completely determined by the values $\underline{P}(f|X_1)$ that it assumes in the gambles f on $\mathscr{X}_2$ alone. We shall use this very useful property repeatedly throughout the paper.

2.7. **Joint coherence and marginal extension.** If besides the (separately coherent) conditional lower prevision $\underline{P}(\cdot|X_1)$ on some subset $\mathscr{H}$ of $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$, the subject has also specified a coherent joint (unconditional) lower prevision $\underline{P}$ on some subset $\mathscr{K}$ of $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$, then $\underline{P}$ and $\underline{P}(\cdot|X_1)$ should in addition satisfy the consistency criterion of *joint coherence*, which is discussed and motivated at great length in Walley (1991, Chapter 6), and to a lesser extent in Appendix B (Section B.1).

Now, suppose our subject has specified a coherent marginal lower prevision $\underline{P}_1$ on some subset $\mathscr{K}_1$ of $\mathscr{L}(\mathscr{X}_1)$, modelling the available information about the value that X_1 assumes in $\mathscr{X}_1$. And, when modelling the available information about X_2, he specifies a separately coherent conditional lower prevision $\underline{P}(\cdot|X_1)$ on some subset $\mathscr{H}$ of $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$. We can then always extend $\underline{P}_1$ and $\underline{P}(\cdot|X_1)$ to a pair $\underline{M}$ and $\underline{M}(\cdot|X_1)$ defined on all of $\mathscr{L}(\mathscr{X}_1 \times \mathscr{X}_2)$, which is the point-wise smallest jointly coherent pair that coincides with $\underline{P}_1$ and $\underline{P}(\cdot|X_1)$ on their respective domains $\mathscr{K}_1$ and $\mathscr{H}$. $\underline{M}$ and $\underline{M}(\cdot|X_1)$ are called the *marginal extensions* of $\underline{P}_1$ and $\underline{P}(\cdot|X_1)$, and they are given, for all gambles f on $\mathscr{X}_1 \times \mathscr{X}_2$, by

$$\underline{M}(f|x_1) = \underline{E}(f(x_1,\cdot)|x_1) \text{ and } \underline{M}(f) = \underline{E}_1(\underline{E}(f|X_1)),$$

where for each x_1 in $\mathscr{X}_1$, $\underline{E}(\cdot|x_1)$ is the (unconditional) natural extension of the coherent lower prevision $\underline{P}(\cdot|x_1)$ to all gambles on $\mathscr{X}_2$, and $\underline{E}_1$ is the (unconditional) natural extension of $\underline{P}_1$ to all gambles on $\mathscr{X}_1$. This result is called the Marginal Extension Theorem in Walley (1991, Theorem 6.7.2). Note that $\underline{M}$ coincides with $\underline{E}_1$ on $\mathscr{X}_1$-measurable gambles, but also that $\underline{M}$ is not necessarily equal to the (unconditional) natural extension of $\underline{P}_1$ to all gambles on $\mathscr{X}_1 \times \mathscr{X}_2$, as it also has to take into account the behavioural consequences of the assessments that are present in $\underline{P}(\cdot|X_1)$. As is the case for unconditional natural extension, the marginal extensions $\underline{M}$ and $\underline{M}(\cdot|X_1)$ summarise the behavioural implications of $\underline{P}_1$ and $\underline{P}(\cdot|X_1)$, only taking into account the consequences of (separate and) joint coherence.

Further on, we consider a more general situation, where we work with N random variables X_1 ,..., X_N taking values in the respective sets $\mathscr{X}_1, \dots, \mathscr{X}_N$, and we apply a generalisation of the Marginal Extension Theorem proved in Miranda and De Cooman (2007).

3. FORWARD IRRELEVANT NATURAL EXTENSION AND FORWARD IRRELEVANT PRODUCT

We are ready to begin our detailed discussion of how to combine marginal lower previsions into a joint, in such a way as to take into account epistemic irrelevance assessments.

3.1. **Marginal information.** Consider N random variables $X_1, \dots, X_N$ taking values in the respective non-empty sets $\mathscr{X}_1, \dots, \mathscr{X}_N$. *We do not assume* that these random variables are real-valued, i.e., that the $\mathscr{X}_k$ are subsets of the set of real numbers $\mathbb{R}$.

For each variable X_k, a subject has beliefs about the values that it assumes in $\mathscr{X}_k$, expressed in the form of a coherent *marginal lower prevision* $\underline{P}_k$ defined on some set of gambles $\mathscr{K}_k \subseteq \mathscr{L}(\mathscr{X}_k)$.

Now, if we know $\underline{P}_k(f)$ for some f, then coherence implies that $\underline{P}_k(\lambda f+\mu) = \lambda \underline{P}_k(f) + \mu$ for all $\lambda \geq 0$ and $\mu \in \mathbb{R}$, so $\underline{P}_k$ can be uniquely extended to a coherent lower prevision on all $\lambda f + \mu$. We may therefore assume, without loss of generality, that $\underline{P}_k$ is actually defined on the set of gambles $\mathscr{K}_k^*$ (a cone containing all constant gambles), given by:

$$\mathscr{K}_k^* := \{\lambda f + \mu \colon \lambda \geq 0, \mu \in \mathbb{R} \text{ and } f \in \mathscr{K}_k\}. \tag{11}$$

We can extend the marginal lower previsions $\underline{P}_k$ defined on $\mathscr{K}_k$ (or on $\mathscr{K}_k^*$) to marginal lower previsions $\underline{E}_k$ defined on all of $\mathscr{L}(\mathscr{X}_k)$, for $1 \leq k \leq N$, through the procedure of natural extension, as explained in Section 2.2. Applying (9), this yields:

$$\underline{E}_k(h) = \sup_{\substack{g_{ki_k}\in\mathscr{K}_k^*\\ i_k=1,\dots,n_k, n_k\geq 0}} \inf_{x_k\in\mathscr{X}_k} \left[h(x_k) - \sum_{i_k=1}^{n_k} [g_{ki_k}(x_k) - \underline{P}_k(g_{ki_k})]\right] \tag{12}$$

for all gambles h on $\mathscr{X}_k$, also taking into account that $\mathscr{K}_k^*$ is a cone. Recall that $\underline{E}_k$ is the point-wise smallest (least-committal) coherent extension of $\underline{P}_k$: it is the extension that takes into account only the consequences of coherence.

For any $1 \leq k \leq N$, we define the set

$$\mathscr{X}^k := \times_{i=1}^k \mathscr{X}_i = \{(x_1,\dots,x_k) \colon x_i \in \mathscr{X}_i, i = 1,\dots,k\}$$

and the random variable $X^k := (X_1,\dots,X_k)$ taking values in the set $\mathscr{X}^k$. Our subject judges the random variables X_1, ..., X_N to be *logically independent*, which means that according to him, the X^k can assume all values in the corresponding Cartesian product sets $\mathscr{X}^k$.

3.2. **Expressing forward irrelevance.** We now express the following *forward irrelevance* assessment: for each $2 \leq k \leq N$, our subject assesses that his beliefs about the value that the variable X_k assumes in $\mathscr{X}_k$ will not be influenced by any additional information about the value that the 'previous' variables $X^{k-1} = (X_1,\dots,X_{k-1})$ assume in $\mathscr{X}^{k-1} = \times_{i=1}^{k-1}\mathscr{X}_i$. To use Walley's (1991) terminology, the variables X_1, ..., X_{k-1} are *epistemically irrelevant* to the variable X_k, for $2 \leq k \leq N$.

To make this forward irrelevance condition more explicit, we define the sets of gambles $\mathscr{K}^k$ on the product sets $\mathscr{X}^k$: let $\mathscr{K}^1 := \mathscr{K}_1^*$ and for $2 \leq k \leq N$, let $\mathscr{K}^k$ be the set of all gambles f on $\mathscr{X}^k$ such that all partial maps $f(x,\cdot)$ are in $\mathscr{K}_k^*$ for $x \in \mathscr{X}^{k-1}$, i.e., $\mathscr{K}^k := \{f \in \mathscr{L}(\mathscr{X}^k) \colon (\forall x \in \mathscr{X}^{k-1})(f(x,\cdot) \in \mathscr{K}_k^*)\}$. It follows from Eq. (11) that $\mathscr{K}^k$ is a cone as well. In fact, we have something stronger: that $\lambda f + \mu \in \mathscr{K}^k$ for all f in $\mathscr{K}^k$, and all *gambles* $\lambda \geq 0$ and μ on $\mathscr{X}^{k-1}$. The forward irrelevance assessment can now be used to define conditional lower previsions $\underline{P}(\cdot|X^{k-1})$: let $\underline{P}(f|x_1,\dots,x_{k-1}) := \underline{P}_k(f)$ for all f in $\mathscr{K}_k$. Invoking separate coherence (Section 2.6), they can actually be defined on all g in $\mathscr{K}^k$ by

$$\underline{P}(g|x_1,\dots,x_{k-1}) := \underline{P}_k(g(x_1,\dots,x_{k-1},\cdot)) \tag{13}$$

for all $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$, where $2 \leq k \leq N$.

In summary, we have the following assessments: an marginal lower prevision $\underline{P}_1$ defined on $\mathscr{K}^1$, and conditional lower previsions $\underline{P}(\cdot|X^{k-1})$ defined on $\mathscr{K}^k$, which are derived from the marginals $\underline{P}_k$ and the forward irrelevance assessment (13), for $2 \leq k \leq N$.

3.3. **The forward irrelevant natural extension.** We now investigate what are the minimal behavioural consequences of these (conditional) lower prevision assessments. In particular, if a subject has specified the marginal lower previsions $\underline{P}_k$ summarising his dispositions to buy gambles f_k in $\mathscr{K}_k$ for prices up to $\underline{P}_k(f_k)$, and if he makes the forward irrelevance assessment expressed through Eq. (13), then what is the smallest (most conservative, or least-committal) price that these assessments and coherence imply he should be willing to pay for a gamble f on the product space $\mathscr{X}^N$?

The lower prevision that represents these least-committal supremum acceptable buying prices is identified in the following theorem, which is proved in Appendix B, Section B.2.

Theorem 1. *The so-called* natural extension $\underline{E}^N$ *of* $\underline{P}_1$ *and the* $\underline{P}(\cdot|X^{k-1})$ *to a lower prevision on* $\mathscr{L}(\mathscr{X}^N)$*, defined by*

$$\underline{E}^N(f) = \sup_{\substack{g_{ki_k}\in\mathscr{K}^k, i_k=1,\dots,n_k\\ n_k\geq 0, k=1,\dots,N}} \inf_{x\in\mathscr{X}^N} \left[f(x) - \sum_{k=1}^{N}\sum_{i_k=1}^{n_k} G(g_{ki_k}|X^{k-1})(x) \right] \tag{14}$$

is the point-wise smallest coherent extension of the marginal lower prevision $\underline{P}_1$ *to* $\mathscr{L}(\mathscr{X}^N)$ *that is jointly coherent with the conditional lower previsions* $\underline{P}(\cdot|X_1)$*, ...,* $\underline{P}(\cdot|X^{N-1})$ *obtained from the marginal lower previsions* $\underline{P}_2$*, ...,* $\underline{P}_N$ *through the forward irrelevance assessments* (13)*.*

The expression generalises Walley's definition of natural extension (Walley, 1991, Section 8.1) from linear spaces to more general domains (cones).

The joint lower prevision $\underline{E}^N$ is actually a *product* of these marginal lower previsions: we shall see in Proposition 7 that its marginals coincide with the lower previsions $\underline{P}_1$, ..., $\underline{P}_N$, on their respective domains $\mathscr{K}_1$, ..., $\mathscr{K}_N$. In other words, $\underline{E}^N$ provides a way to combine the marginal lower previsions $\underline{P}_k$ into a joint lower prevision, taking into account the assessment of forward irrelevance. We call $\underline{E}^N$ the *forward irrelevant natural extension* of the given marginals. Above, we have used the following notations, for all $x\in\mathscr{X}^N$:

$$G(g|X^0)(x) = g(x_1) - \underline{P}_1(g)$$

for any g in $\mathscr{K}^1 = \mathscr{K}_1^*$, and

$$G(g|X^{k-1})(x) = \sum_{y\in\mathscr{X}^{k-1}} I_{\{y\}}(x_1,\dots,x_{k-1})[g(x_1,\dots,x_k) - \underline{P}(g|y)]$$

for all $2\leq k\leq N$ and $g\in\mathscr{K}^k$, which can be simplified to

$$G(g|X^{k-1})(x) = g(x_1,\dots,x_k) - \underline{P}_k(g(x_1,\dots,x_{k-1},\cdot)),$$

taking into account the forward irrelevance condition (13). This means that we can further simplify the given expression for the forward irrelevant natural extension $\underline{E}^N$, also taking into account that each $\mathscr{K}^k$ is a cone (and with some obvious abuse of notation for $k=1$):

$$\underline{E}^N(f) = \sup_{\substack{g_{ki_k}\in\mathscr{K}^k, i_k=1,\dots,n_k\\ n_k\geq 0, k=1,\dots,N}} \inf_{x\in\mathscr{X}^N} \Bigg[f(x) - \sum_{k=1}^{N}\sum_{i_k=1}^{n_k} [g_{ki_k}(x_1,\dots,x_k) - \underline{P}_k(g_{ki_k}(x_1,\dots,x_{k-1},\cdot))] \Bigg]. \tag{15}$$

3.4. **The forward irrelevant product.** The forward irrelevant natural extension $\underline{E}^N$ is the smallest joint lower prevision on $\mathscr{L}(\mathscr{X}^N)$ that is coherent with the given assessments. In some situations, we might be interested not only in coherently extending the given assessments to a *joint* lower prevision, but we also might want to coherently extend the *conditional* lower previsions $\underline{P}(\cdot|X^k)$, defined on $\mathscr{K}^{k+1}$ ($k=1,\dots,N-1$), to all of $\mathscr{L}(\mathscr{X}^{k+1})$, or to $\mathscr{L}(\mathscr{X}^N)$ for that matter.

We have shown in Miranda and De Cooman (2007) that it is always possible to coherently extend any (separately) coherent (conditional) lower previsions $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$ in a least-committal way: there always are (separately and) jointly coherent (conditional) lower previsions $\underline{M}^N$, $\underline{M}^N(\cdot|X^1)$, ..., $\underline{M}^N(\cdot|X^{N-1})$, defined on $\mathscr{L}(\mathscr{X}^N)$, that coincide with the respective original (conditional) lower previsions $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$ on their respective domains $\mathscr{K}^1$, $\mathscr{K}^2$, ..., $\mathscr{K}^N$, and that are the same time point-wise dominated by (i.e., more conservative than) all the other (separately and) jointly coherent extensions of the original (conditional) lower previsions. This is the essence of our Marginal Extension Theorem (Miranda and De Cooman, 2007, Theorem 4 and Section 7). We refer to Miranda and De Cooman (2007) for a detailed introduction (with proofs) to the concept of marginal extension and its properties. If we apply our general Marginal Extension Theorem to the special case considered here, where the $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, $\dots$, $\underline{P}(\cdot|X^{N-1})$ are constructed from the marginals $\underline{P}_1,\dots,\underline{P}_N$ using the forward irrelevance assessment (13), we come to the conclusions summarised in Theorem 2 below.

Consider, for $1\leq\ell\leq N$, the (unconditional) natural extensions $\underline{E}_\ell$ to $\mathscr{L}(\mathscr{X}_\ell)$ of the marginals $\underline{P}_\ell$ on $\mathscr{K}_\ell$, and define the lower prevision $\underline{M}^\ell$ on $\mathscr{L}(\mathscr{X}^\ell)$ by

$$\underline{M}^\ell(h):=\underline{E}_1(\underline{E}_2(\dots(\underline{E}_\ell(h))\dots)),$$

for any gamble h on $\mathscr{X}^\ell$. We use the general convention that for any gamble g on $\mathscr{X}^k$, $\underline{E}_k(g)$ denotes the gamble on $\mathscr{X}^{k-1}$, whose value in an element $(x_1,\dots,x_{k-1})$ of $\mathscr{X}^{k-1}$ is given by $\underline{E}_k(g(x_1,\dots,x_{k-1},\cdot))$. In general, we have the following recursion formula

$$\underline{M}^k(f)=\underline{M}^{k-1}(\underline{E}_k(f))$$

for $k=2,\dots,N$ and f in $\mathscr{L}(\mathscr{X}^k)$. Observe that $\underline{M}^1=\underline{E}_1$.

We can now use the lower prevision $\underline{M}^N$ to define the so-called *marginal extensions* $\underline{M}^N,\underline{M}^N(\cdot|X^1),\dots,\underline{M}^N(\cdot|X^{N-1})$ to $\mathscr{L}(\mathscr{X}^N)$ of the original (conditional) lower previsions $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, $\dots$, $\underline{P}(\cdot|X^{N-1})$. Consider any gamble f on $\mathscr{X}^N$. Then obviously[6]

$$\underline{M}^N(f)=\underline{E}_1(\underline{E}_2(\dots(\underline{E}_N(f))\dots)), \tag{16}$$

and similarly, for any $x=(x_1,\dots,x_N)$ in $\mathscr{X}^N$,

$$\begin{aligned}\underline{M}^N(f|x_1)&=\underline{M}^N(f(x_1,\cdot))=\underline{E}_2(\underline{E}_3(\dots(\underline{E}_N(f(x_1,\cdot)))\dots))\\ \underline{M}^N(f|x_1,x_2)&=\underline{M}^N(f(x_1,x_2,\cdot))=\underline{E}_3(\underline{E}_4(\dots(\underline{E}_N(f(x_1,x_2,\cdot)))\dots))\\ &\dots\\ \underline{M}^N(f|x_1,\dots,x_{N-1})&=\underline{M}^N(f(x_1,\dots,x_{N-1},\cdot))=\underline{E}_N(f(x_1,\dots,x_{N-1},\cdot)).\end{aligned} \tag{17}$$

Theorem 2. *The* marginal extensions $\underline{M}^N$, $\underline{M}^N(\cdot|X^1)$, $\dots$, $\underline{M}^N(\cdot|X^{N-1})$ *defined above in Eqs.* (16) *and* (17) *are the point-wise smallest jointly coherent extensions to* $\mathscr{L}(\mathscr{X}^N)$ *of the (conditional) lower previsions* $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, $\dots$, $\underline{P}(\cdot|X^{N-1})$*, obtained from the marginal lower previsions* $\underline{P}_1$, $\dots$, $\underline{P}_N$ *through the forward irrelevance assessments* (13) .

Since we shall see in Proposition 7 that the joint lower prevision $\underline{M}^N$ coincides with $\underline{P}_1$, $\dots$, $\underline{P}_N$ on their respective domains $\mathscr{K}_1$, $\dots$, $\mathscr{K}_N$, $\underline{M}^N$ is a product of these marginal lower previsions, and we shall call it their *forward irrelevant product*. It too provides a way of combining the marginal lower previsions $\underline{P}_k$ into a joint lower prevision, taking into account the assessment of forward irrelevance.

The procedure of *marginal extension preserves forward irrelevance*: the equalities (17) extend the equalities (13) to all gambles on $\mathscr{X}^N$, and not just the ones in the domains $\mathscr{K}^k$.

[6]To see that this expression makes sense, note that for any gamble f on $\mathscr{X}^N$, $\underline{E}_N(f)$ is a gamble on $\mathscr{X}^{N-1}$, and as such we can apply $\underline{E}_{N-1}$ to it; then $\underline{E}_{N-1}(\underline{E}_N(f))$ is a gamble on $\mathscr{X}^{N-2}$, to which we can apply $\underline{E}_{N-2}$; and, finally, $\underline{E}_2(\dots(\underline{E}_N(f)))$ is a gamble on $\mathscr{X}_1$, to which we can apply $\underline{E}_1$ to obtain the real value of $\underline{M}^N(f)$.

3.5. **The relationship between the forward irrelevant natural extension and the forward irrelevant product.** Perhaps surprisingly, the forward irrelevant product and the forward irrelevant natural extension don't always coincide, unless the variables X_k may assume only a finite number of values, i.e., unless the sets $\mathscr{X}_k$ are finite. That they coincide in this case follows from Walley (1991, Theorem 8.1.9) or from Miranda and De Cooman (2007, Section 6). For an example showing that they don't necessarily coincide when the spaces $\mathscr{X}_k$ are infinite, check Example 1 and Section 7 in Miranda and De Cooman (2007). We summarise this as follows.

Theorem 3. *The forward irrelevant product dominates the forward natural extension:* $\underline{M}^N(f) \geq \underline{E}^N(f)$ *for all gambles* f *on* $\mathscr{X}^N$*. But* $\underline{E}^N$ *and* $\underline{M}^N$ *coincide if the sets* $\mathscr{X}_k$*,* $k = 1,\dots,N$ *are finite.*

The forward irrelevant natural extension and the forward irrelevant product also coincide when the initial domains $\mathscr{K}_k$ are actually equal to $\mathscr{L}(\mathscr{X}_k)$, for $k = 1,\dots,N$. This is stated in the following theorem, which is proved in Appendix B (Section B.3). When $\mathscr{K}_k = \mathscr{L}(\mathscr{X}_k)$ for $k = 1,\dots,N$, we obtain

$$\underline{M}^N(f) = \underline{P}_1(\underline{P}_2(\dots(\underline{P}_N(f))\dots)),$$

and, for any $x = (x_1,\dots,x_N)$ in $\mathscr{X}^N$,

$$\begin{aligned}\underline{M}^N(f|x_1) &= \underline{P}_2(\underline{P}_3(\dots(\underline{P}_N(f(x_1,\cdot)))\dots))\\ &\dots\\ \underline{M}^N(f|x_1,\dots,x_{N-1}) &= \underline{P}_N(f(x_1,\dots,x_{N-1},\cdot)).\end{aligned}$$

Theorem 4. *If the domain* $\mathscr{K}_k$ *of the marginal lower prevision* $\underline{P}_k$ *is* $\mathscr{L}(\mathscr{X}_k)$ *for* $k = 1,\dots,N$*, then* $\underline{E}^N$ *and* $\underline{M}^N$ *coincide.*

When $\underline{E}^N$ doesn't coincide with (i.e., is strictly dominated by) $\underline{M}^N$, it can't, of course, be jointly coherent with the conditional lower previsions $\underline{M}^N(\cdot|X^1)$, …, $\underline{M}^N(\cdot|X^{N-1})$, although it is, by Theorem 1, still coherent with their restrictions $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^{N-1})$ to the respective domains $\mathscr{K}^2$, …, $\mathscr{K}^N$. So it is all right to use the forward irrelevant natural extension if we are interested in coherently extending the given assessments to a *joint* lower prevision on $\mathscr{L}(\mathscr{X}^N)$ only. If, however, we also want to coherently extend the given assessments to *conditional* lower previsions, we need the forward irrelevant product. In this sense, the forward irrelevant product seems to be the better extension.

3.6. **Properties of the forward irrelevant natural extension and the forward irrelevant product.** Let us now devote some attention to the properties of the forward irrelevant natural extension $\underline{E}^N$ and product $\underline{M}^N$. First of all, $\underline{M}^N$ has an interesting Bayesian sensitivity analysis interpretation. We can construct $\underline{M}^N$ as a lower envelope of joint linear previsions, each of which can be obtained by combining the marginal linear previsions in the $\mathscr{M}(\underline{P}_k)$ in a special way. There seems to be no analogous construction for $\underline{E}^N$. The following theorem is an immediate special case of the more general Lower Envelope Theorem we have proved in (Miranda and De Cooman, 2007, Theorem 3).

Theorem 5 (Lower Envelope Theorem)**.** *Let* P_1 *be any element of* $\mathscr{M}(\underline{P}_1)$*, and for* $2 \leq k \leq N$ *and any* $(x_1,\dots,x_{k-1})$ *in* $\mathscr{X}^{k-1}$*, let* $P_k(\cdot|x_1,\dots,x_{k-1})$ *be any element of* $\mathscr{M}(\underline{P}_k)$*. Define, for any gamble* f_k *on* $\mathscr{X}^k$*,* $P_k(f_k|X^{k-1})$ *as the gamble on* $\mathscr{X}^{k-1}$ *that assumes the value* $P_k(f_k(x_1,\dots,x_{k-1},\cdot)|x_1,\dots,x_{k-1})$ *in the element* $(x_1,\dots,x_{k-1})$ *of* $\mathscr{X}^{k-1}$*, for* $2 \leq k \leq N$*. Finally let, for any gamble* f *on* $\mathscr{X}^N$*,* $P^N(f) = P_1(P_2(\dots(P_N(f|X^{N-1}))\dots|X^1))$*, i.e., apply Bayes' rule (or marginal extension) to combine the linear prevision* P_1 *and the conditional* *linear previsions* $P_2(\cdot|X^1)$, $\dots$, $P_N(\cdot|X^{N-1})$ *[see footnote 6]. Then the* P^N *constructed in this way is a linear prevision on* $\mathscr{L}(\mathscr{X}^N)$*. Moreover,* $\underline{M}^N$ *is the lower envelope of all such linear previsions, and for any gamble* f *on* $\mathscr{X}^N$ *there is such a linear prevision that coincides on* f *with* $\underline{M}^N$*.*

This theorem allows us to relate the forward irrelevant product to other products of marginal lower previsions, extant in the literature. First of all, the so-called *type-1 product*, or *strong product* of the marginals $\underline{P}_1$, $\dots$, $\underline{P}_N$ is obtained by choosing the *same* $P_k(\cdot|x_1,\dots,x_{k-1})$ in $\mathscr{M}(\underline{P}_k)$ for all $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$ in the above procedure, and then taking the lower envelope; see for instance Walley (1991, Section 9.3.5) and Couso et al. (2000). It therefore dominates the forward irrelevant product.

In case we have $\mathscr{X}_1=\dots=\mathscr{X}_N$, and $\underline{P}_1=\dots=\underline{P}_N=\underline{P}$, the so-called *type-2 product* of the marginals $\underline{P}_1$, $\dots$, $\underline{P}_N$ is obtained by choosing the same P_1 and $P_k(\cdot|x_1,\dots,x_{k-1})$ in $\mathscr{M}(\underline{P})$ for all $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$ *and* for all $2\leq k\leq N$ in the above procedure, and then taking the lower envelope; again see Walley (1991, Section 9.3.5) and Couso et al. (2000). It therefore dominates both the type-1 product and the forward irrelevant product.

Next, we show that the forward irrelevant natural extension and the forward irrelevant product have a number of properties that are similar to (but sometimes weaker than) the usual product of linear previsions.

Proposition 6 (External linearity)**.** *Let* f_k *be any gamble on* $\mathscr{X}_k$ *for* $1\leq k\leq N$*. Then*

$$\underline{E}^N\left(\sum_{k=1}^N f_k\right)=\underline{M}^N\left(\sum_{k=1}^N f_k\right)=\sum_{k=1}^N \underline{E}_k(f_k).$$

The forward irrelevant natural extension and product are indeed *products*: $\underline{E}^N$ and $\underline{M}^N$ are extensions of the marginals $\underline{P}_k$.

Proposition 7. *Let* f_k *be any gamble on* $\mathscr{X}_k$*. Then* $\underline{E}^N(f_k)=\underline{M}^N(f_k)=\underline{E}_k(f_k)$*. If in particular* f_k *belongs to* $\mathscr{K}_k$*, then* $\underline{E}^N(f_k)=\underline{M}^N(f_k)=\underline{P}_k(f_k)$*, for all* $1\leq k\leq N$*.*

The forward irrelevant natural extension and product also satisfy a (restricted) product rule.

Proposition 8 (Product rule)**.** *Let* f_k *be a non-negative gamble on* $\mathscr{X}_k$ *for* $1\leq k\leq N$*. Then*

$$\underline{E}^N(f_1f_2\dots f_N)=\underline{M}^N(f_1f_2\dots f_N)=\underline{E}_1(f_1)\underline{E}_2(f_2)\dots\underline{E}_N(f_N)$$
$$\overline{E}^N(f_1f_2\dots f_N)=\overline{M}^N(f_1f_2\dots f_N)=\overline{E}_1(f_1)\overline{E}_2(f_2)\dots\overline{E}_N(f_N).$$

In particular, let A_k *be any subset of* $\mathscr{X}_k$ *for* $1\leq k\leq N$*. Then*

$$\underline{E}^N(A_1\times A_2\times\dots\times A_N)=\underline{M}^N(A_1\times A_2\times\dots\times A_N)=\underline{E}_1(A_1)\underline{E}_2(A_2)\dots\underline{E}_N(A_N)$$
$$\overline{E}^N(A_1\times A_2\times\dots\times A_N)=\overline{M}^N(A_1\times A_2\times\dots\times A_N)=\overline{E}_1(A_1)\overline{E}_2(A_2)\dots\overline{E}_N(A_N).$$

Finally, both the forward irrelevant product and the forward irrelevant natural extension satisfy the so-called *forward factorising property*. This property allows us to establish laws of large numbers for these lower previsions (De Cooman and Miranda, 2006).

Proposition 9 (Forward factorisation)**.** *Let* f_k *be a non-negative gamble on* $\mathscr{X}_k$ *for* $1\leq k\leq N-1$*, and let* f_N *be a gamble on* $\mathscr{X}_N$*. Then*

$$\underline{E}^N(f_1f_2\dots f_{N-1}[f_N-\underline{E}_N(f_N)])=\underline{M}^N(f_1f_2\dots f_{N-1}[f_N-\underline{E}_N(f_N)])=0.$$

The following proposition gives an equivalent formulation. We prove in De Cooman and Miranda (2006) that in the precise case the forward factorisation property is equivalent to $E^N(gf_N)=E^N(g)E^N(f_N)$ for all $g\in\mathscr{X}^{N-1}$ and all $f_N\in\mathscr{X}_N$, hence its name.

Proposition 10. *Let g be a non-negative gamble on $\mathscr{X}^{N-1}$, and let f_N be a gamble on $\mathscr{X}_N$. Then $\underline{E}^N(g[f_N-\underline{E}_N(f_N)])=\underline{M}^N(g[f_N-\underline{E}_N(f_N)])=0$.*

We refer to Miranda and De Cooman (2007, Section 7) for the proofs of Propositions 6–8. Propositions 9 and 10 are proved in Appendix B (Sections B.4 and B.5).

3.7. Examples of forward irrelevant natural extension and product.

3.7.1. *Linear marginals.* Suppose our subject's marginal lower previsions are precise in the following sense: for each $k=1,\dots,N$, $\underline{P}_k=P_k$ is a linear prevision on $\mathscr{K}_k=\mathscr{L}(\mathscr{X}_k)$. Then it follows from Eq. (16) that the forward irrelevant product $\underline{M}^N$ satisfies:

$$\underline{M}^N(f)=\overline{M}^N(f)=P_1(P_2(\dots(P_{N-1}(P_N(f)))\dots))$$

for all gambles f on $\mathscr{X}^N$, or in other words, $\underline{M}^N=\overline{M}^N=M^N$ is a linear prevision on $\mathscr{L}(\mathscr{X}^N)$ that is the usual product[7] of the marginals $P_1,\dots,P_N$.

Since here the domain of P_k is $\mathscr{L}(\mathscr{X}_k)$, for $k=1,\dots,n$, we deduce from Theorem 4 that the forward irrelevant natural extension coincides with the forward irrelevant product.

3.7.2. *Vacuous marginals.* Suppose that our subject has the following information: the variable X_k assumes a value in a non-empty subset A_k of $\mathscr{X}_k$. It has been argued elsewhere (for instance, by Walley (1991) and De Cooman and Troffaes (2004)) that he can model this using the so-called *vacuous lower prevision $\underline{P}_{A_k}$ relative to A_k*, where for any gamble f on $\mathscr{X}_k$: $\underline{P}_{A_k}(f)=\inf_{x_k\in A_k}f(x_k)$. Both forward irrelevant natural extension and product of these marginal lower previsions $\underline{P}_k=\underline{P}_{A_k}$ coincide with the vacuous lower prevision relative to the Cartesian product $A_1\times\dots\times A_N$:

$$\underline{M}^N(f)=\underline{E}^N(f)=\underline{P}_{A_1\times\dots\times A_N}(f)=\inf_{(x_1,\dots,x_N)\in A_1\times\dots\times A_N}f(x_1,\dots,x_N)$$

for all gambles f on $\mathscr{X}^N$. For $\underline{M}^N$, this is an immediate consequence of Eq. (16). That this result also holds for $\underline{E}^N$ follows from Theorem 4.

4. Epistemic irrelevance versus epistemic independence

An assessment of forward irrelevance is actually weaker than one of epistemic independence. Indeed, an assessment of *epistemic independence* would mean that the subject doesn't change his beliefs about any variable X_k after observing the values of a collection of variables $\{X_\ell\colon \ell\in T\}$, where T is *any* non-empty subset of $\{1,\dots,N\}$ not containing k; see for instance Walley (1991, Section 9.3) and Couso et al. (2000).

Epistemic independence clearly implies both forward and backward irrelevance, where of course, *backward irrelevance* simply means that the subject doesn't change his beliefs about any X_k after observing the variables $X_{k+1},\dots,X_N$. We now show neither forward nor backward irrelevance generally imply epistemic independence.

Consider the special case of two random variables X_1 and X_2 taking values on the same finite space $\mathscr{X}$, with identical marginal lower previsions $\underline{P}_1=\underline{P}_2=\underline{P}$. Clearly, they are

---

[7] Some care is needed here, however, since for general linear previsions, contrary to the σ-additive case (Fubini's Theorem), this product is not necessarily 'commutative', meaning that 'the order of integration' cannot generally be 'permuted', unless the spaces $\mathscr{X}_k$ are finite. This means that if we consider a permutation π on $\{1,\dots,n\}$ and construct the forward irrelevant product $P_{\pi(1)}(P_{\pi(2)}(\dots(P_{\pi(n)}(\cdot))))$, it will not coincide in general with M^N. In Walley's (1991) words, we then say that the linear marginals $P_1,\dots,P_N$ are *incompatible*, meaning that there is no (jointly) coherent lower prevision with marginals $P_1,\dots,P_N$ that furthermore expresses the epistemic independence of the random variables $X_1,\dots,X_N$; see also Section 4 and Walley (1991, Section 9.3) for more details about epistemic independence.

epistemically independent if and only if there is both forward and backward irrelevance. Suppose that, in general, forward irrelevance implied epistemic independence, then the forward irrelevant product (or the forward irrelevant natural extension, since they coincide on finite spaces) $\underline{P}_1(\underline{P}_2(\cdot))$ of $\underline{P}_1$ and $\underline{P}_2$ would generally coincide with the epistemic independent product and with the backward irrelevant product $\underline{P}_2(\underline{P}_1(\cdot))$. The following simple counterexample shows that this is not the case.

Let $\mathscr{X} = \{a,b\}$ and let the linear prevision P on $\mathscr{X}$ be determined by $P(\{a\}) = \alpha$ and $P(\{b\}) = 1-\alpha$, where $0 \leq \alpha \leq 1$. Let the coherent marginal lower previsions $\underline{P}_1 = \underline{P}_2 = \underline{P}$ be the so-called *linear-vacuous mixture*, or *ε-contamination* of P, given by

$$\underline{P}(f) = (1-\varepsilon)P(f) + \varepsilon \underline{P}_{\mathscr{X}}(f) = (1-\varepsilon)[\alpha f(a) + (1-\alpha)f(b)] + \varepsilon \min\{f(a), f(b)\}$$

for all gambles f on $\mathscr{X}$, where $0 \leq \varepsilon \leq 1$. It is, by the way, easy to see that all coherent lower previsions on a two-element space are such linear-vacuous mixtures, which implies that by varying α and ε we generate all possible coherent lower previsions on $\mathscr{X}$.

Consider the gambles $h_1 = I_{\{a\}} + 2I_{\{b\}}$ and $h_2 = I_{\{a\}} - 2I_{\{b\}}$ on $\mathscr{X}$, and the gamble h on $\mathscr{X} \times \mathscr{X}$ given by $h(x,y) = h_1(x)h_2(y)$, i.e.,

$$h = I_{\{(a,a)\}} - 2I_{\{(a,b)\}} + 2I_{\{(b,a)\}} - 4I_{\{(b,b)\}}$$

In Fig. 1 we have plotted the difference $\underline{P}_1(\underline{P}_2(\cdot)) - \underline{P}_2(\underline{P}_1(\cdot))$ between the forward and the backward irrelevant natural extensions/products of the marginal lower previsions $\underline{P}_1$ and $\underline{P}_2$, as a function of α and ε, for the gambles h and $-h$. These plots clearly show that, unless the marginals are precise ($\varepsilon = 0$) or completely vacuous ($\varepsilon = 1$), the forward and backward irrelevant natural extensions/products are always different in some gamble. This example also shows that the product rule of Proposition 8 cannot in general be extended to gambles that don't have constant sign (but see also Propositions 9 and 10).

We get this inequality even though the two marginals are the same. Moreover, since

$$\underline{P}_1(\{a\}) = \underline{P}_2(\{a\}) = \alpha(1-\varepsilon) \quad \text{and} \quad \underline{P}_1(\{b\}) = \underline{P}_2(\{b\}) = (1-\alpha)(1-\varepsilon),$$

the inequality is in no way caused by any 'pathological' consequences of conditioning on sets with probability zero.

We learn from this example that, generally speaking, in the context of the theory of coherent lower previsions, the fact that one variable is irrelevant to another doesn't imply the converse! Epistemic irrelevance is an asymmetrical notion, but when we restrict ourselves to precise models (linear previsions), the asymmetry usually collapses into symmetry; see also Section 1 and footnote 7. For more detailed discussion, we refer to Couso et al. (2000).

The so-called *independent natural extension* of $\underline{P}_1, \ldots, \underline{P}_N$ is defined as the point-wise smallest (most conservative) product for which there is epistemic independence (Walley, 1991, Section 9.3). Since epistemic independence is generally a stronger requirement than forward irrelevance, the independent natural extension will generally dominate the forward irrelevant natural extension.

5. Further developments of the model

In the particular case that all the domains $\mathscr{K}_k$ are finite, or in other words, when the marginal lower previsions $\underline{P}_k$ are based on a finite number of assessments, we can simplify the formula for the forward irrelevant natural extension $\underline{E}^N$:

$$\underline{E}^N(f) = \sup_{\substack{g_{kh_k} \in \mathscr{L}^+(\mathscr{X}^{k-1}) \\ h_k \in \mathscr{K}_k, k=1,\ldots,N}} \inf_{x \in \mathscr{X}^N} \left[f(x) - \sum_{k=1}^{N} \sum_{h_k \in \mathscr{K}_k} g_{kh_k}(x_1,\ldots,x_{k-1})[h_k(x_k) - \underline{P}_k(h_k)] \right] \quad (18)$$

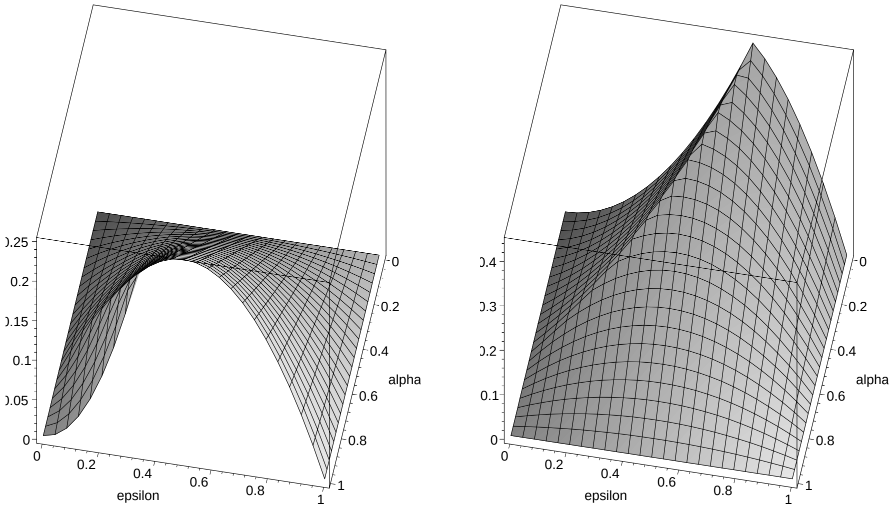

FIGURE 1. The difference $\underline{P}_1(\underline{P}_2(\cdot)) - \underline{P}_2(\underline{P}_1(\cdot))$ between the forward and the backward irrelevant natural extensions/products of the marginal lower previsions $\underline{P}_1$ and $\underline{P}_2$, as a function of α and ε, for the gambles h (left) and $-h$ (right).

where $\mathscr{L}^+(\mathscr{X}^k)$ denotes the set of the non-negative gambles on $\mathscr{X}^k$ for $1 \leq k \leq N-1$, and with some abuse of notation, $\mathscr{L}^+(\mathscr{X}^0)$ denotes the set of non-negative real numbers; see Appendix B, Section B.6 for a proof.

So $\underline{E}^N(f)$ is the solution of a special linear programming problem (with a possibly infinite number of linear constraints and variables). Interestingly, the dual form of this problem consists in minimising the linear expression $P(f)$ where the linear prevision P is subject to the linear constraints implicit in the condition $P \in \mathscr{M}(\underline{E}^N)$.

It is instructive to compare this expression with the one for the natural extension of the marginals $\underline{P}_k$ without making any assumption about the interdependence of the variables $X_1, \ldots, X_N$. This is simply the natural extension $\underline{F}^N$ to $\mathscr{L}(\mathscr{X}^N)$ of the lower prevision $\underline{Q}^N$ defined on the gambles h_k in $\mathscr{K}_k$ by $\underline{Q}^N(h_k) = \underline{P}_k(h_k)$ for $k = 1, \ldots, N$, or in other words, the point-wise smallest coherent lower prevision on $\mathscr{L}(\mathscr{X}^N)$ that coincides with the $\underline{P}_k$ on $\mathscr{K}_k$. Using the expression (9) for natural extension given in Section 2.2, we easily find that

$$\underline{F}^N(f) = \sup_{\substack{\lambda_{kh_k} \in \mathbb{R}^+ \\ h_k \in \mathscr{K}_k, k=1,\ldots,N}} \inf_{x \in \mathscr{X}^N} \left[f(x) - \sum_{k=1}^{N} \sum_{h_k \in \mathscr{K}_k} \lambda_{kh_k} [h_k(x_k) - \underline{P}_k(h_k)] \right]$$

for any gamble f on $\mathscr{X}^N$. We see that these formulae are quite similar, apart from the fact that the non-negative *constants* λ_{kh_k} in the expression for $\underline{F}^N(f)$ are replaced in the expression for $\underline{E}^N(f)$ by non-negative *gambles* g_{kh_k} that may depend on the first $k-1$ variables $x_1, \ldots, x_{k-1}$. This makes sure that the forward irrelevant natural extension $\underline{E}^N$ dominates the natural extension $\underline{F}^N$: making an extra assessment of forward irrelevance makes the resultant model more precise.

For the conjugate upper prevision $\overline{E}^N$ of $\underline{E}_N$, we get

$$\overline{E}^N(f) = \inf_{\substack{g_{kh_k}\in\mathscr{L}^+(\mathscr{X}^{k-1})\\ h_k\in\mathscr{K}_k,k=1,\dots,N}} \sup_{x\in\mathscr{X}^N}\left[f(x)+\sum_{k=1}^{N}\sum_{h_k\in\mathscr{K}_k} g_{kh_k}(x_1,\dots,x_{k-1})[h_k(x_k)-\underline{P}_k(h_k)]\right] \quad (19)$$

But Eq. (19) also allows us to establish an intriguing relationship between the forward irrelevant natural extension in the theory of coherent lower previsions and Shafer and Vovk's (2001) game-theoretic approach to probability theory. To see how this comes about, let us show how we can interpret $\overline{E}^N$ in terms of a special investment game.

Assume that the random variable X_k represents some meaningful part of the state of the market at times $k = 1,\dots,N$. A gamble h_k is a real-valued function of this variable: for each possible state x_k of the market at time k, it yields a (possibly negative) amount of utility $h_k(x_k)$. As an example, h_k could be the unknown market value of a share at time k, and this value is determined by the (unknown) state of the market X_k at that time.

Now let us consider two players: *Bank* and *Investor*. These players are called House and Gambler in Shafer et al. (2003), and Forecaster and Skeptic in Shafer and Vovk (2001), respectively. Before time $k = 1$ (for instance, at time 0), Bank models his uncertainty about the values of the random variable X_k in terms of lower prevision assessments $\underline{P}_k$ on finite sets of gambles $\mathscr{K}_k \subseteq \mathscr{L}(\mathscr{X}_k)$, and this for all $k = 1,\dots,N$.

Recall that the lower prevision $\underline{P}_k$ on $\mathscr{K}_k$ represents Bank's commitments to buying the gambles h_k in $\mathscr{K}_k$ for the price $\underline{P}_k(h_k)$.[8] In fact, Bank is also committed to buying λh_k for the price $\lambda\underline{P}_k(h_k)$ for all non-negative λ. As an example, if h_k is the uncertain market value of a share at time k, then $\underline{P}_k(h_k)$ is the supremum price, announced by the Bank at time 0, that he shall pay at time k for buying the share (at that time).

We assume that the actual value of the random variable X_k becomes known to Bank and Investor at time k. Investor will now try to exploit Bank's commitments in order to achieve some goal. At each time $k-\frac{1}{2}$, i.e., some time after the value of X_{k-1} and before that of X_k becomes known, she chooses for each of the gambles h_k in $\mathscr{K}_k$ (for each of the shares Bank has promised to buy at time k), a non-negative multiplier $g_{kh_k}(x_1,\dots,x_{k-1})$, which represents the number of shares she will sell to Bank at time k, for the price $\underline{P}_k(h_k)$ that Bank has promised to pay for them. The number of shares $g_{kh_k}(x_1,\dots,x_{k-1})$ that Investor sells at time k, can depend on the previously observed states of the market $X_1, \dots, X_{k-1}$. If the state of the market X_k at time k turns out to be x_k, then the utility Investor derives from the transactions at time k is given by: $-\sum_{h_k\in\mathscr{K}_k} g_{kh_k}(x_1,\dots,x_{k-1})[h_k(x_k)-\underline{P}_k(h_k)]$.

Investor could also specify the *functions* $g_{kh_k}\colon \mathscr{X}^k \to \mathbb{R}^+$ *in advance*, i.e., at time 0, and in that case these functions constitute her *strategy* for exploiting Bank's commitments. If Investor starts at time 0 with an initial capital β, then by following this strategy her capital at time N will be $\beta-\sum_{k=1}^{N}\sum_{h_k\in\mathscr{K}_k} g_{kh_k}(x_1,\dots,x_{k-1})[h_k(x_k)-\underline{P}_k(h_k)]$.

Now consider a gamble f on $\mathscr{X}^N$, that we shall interpret as an *investment* for Investor at time N. The price $f(x_1,\dots,x_N)$ that Investor has to pay for it generally depends on the history of $(x_1,\dots,x_N)$ of the market. Clearly, Investor can hedge the investment f with the chosen strategy and an initial capital β if and only if for all $x = (x_1,\dots,x_N)$ in $\mathscr{X}^N$:

$$f(x) \leq \beta - \sum_{k=1}^{N}\sum_{h_k\in\mathscr{K}_k} g_{kh_k}(x_1,\dots,x_{k-1})[h_k(x_k)-\underline{P}_k(h_k)].$$

---

[8] Actually, Bank is only committed to buying h_k for all prices $\underline{P}_k(h_k)-\varepsilon$, $\varepsilon > 0$. But it is simpler to assume that this holds for $\varepsilon = 0$ as well, and this won't affect the conclusions we reach.

And the infimum capital for which there is some strategy that allows her to hedge the investment f is called the *upper price* for f by Shafer and Vovk (2001), and given by

$$\overline{\mathbb{E}}(f) = \inf_{\substack{g_{kh_k} \in \mathscr{L}^+(\mathscr{X}^{k-1}) \\ h_k \in \mathscr{K}_k, k=1,\dots,N}} \sup_{x \in \mathscr{X}^N} \left[f(x) + \sum_{k=1}^{N} \sum_{h_k \in \mathscr{K}_k} g_{kh_k}(x_1,\dots,x_{k-1})[h_k(x_k) - \underline{P}_k(h_k)] \right].$$

This is precisely equal to the upper prevision $\overline{E}^N(f)$ for f associated with the forward irrelevant natural extension of Bank's marginal lower previsions!

Theorem 11. *The infimum capital for which Investor has some strategy that allows her to hedge an investment f is equal to Bank's infimum selling price for f, based on his marginal lower previsions $\underline{P}_k$ and on his assessment that he can't learn about the current state of the market by observing the previous states: $\overline{\mathbb{E}}(f) = \overline{E}^N(f)$.*

By itself, this is a surprising, non-trivial and interesting result. But the equality of upper prices and the (conjugate) forward irrelevant natural extension also allows us to bring together two approaches to probability theory that until now were taken to be quite different: Walley's (1991) behavioural theory of coherent lower previsions, and Shafer and Vovk's (2001) game-theoretic approach to probability theory. On the one hand, our Theorem 11 allows us to incorporate many of the results in Shafer and Vovk's work into the theory of coherent lower previsions. And on the other hand, it shows that all the properties we have proved for the forward irrelevant natural extension in Section 3 are also valid for Shafer and Vovk's upper prices in this particular type of investment game.

6. Discussion

Why do we believe that the results presented here merit attention?

First of all, the material about forward irrelevance in Section 3 constitutes a significant contribution to the theory of coherent lower previsions. We have shown that the forward irrelevant product, given by the *concatenation formula* (16), is the appropriate way to combine marginal lower previsions into a joint lower prevision, based on an assessment of forward irrelevance; see the comments near the end of Section 3.5. Such a concatenation is sometimes used to combine marginal lower previsions; see for instance Denneberg (1994, Chapter 12). The fact that, even on finite spaces, changing the order of the marginals in the concatenation generally produces different joints, or in other words that a Fubini-like result doesn't generally hold for lower previsions (unless they are linear or vacuous, see Section 4), is related to the important observation that epistemic irrelevance (as opposed to the more involved notion of epistemic independence) is an asymmetrical notion.

Besides the forward irrelevant product, we have also introduced the forward irrelevant natural extension as a way to combine marginal into joint lower previsions, based on an assessment of forward irrelevance. This forward irrelevant natural extension turns out to coincide with the upper prices for specific types of coherent probability protocols in Shafer and Vovk's approach. This is our main reason for deeming the forward irrelevant natural extension useful. It allows us to embed many of Shafer and Vovk's ideas into the theory of coherent lower previsions. We feel, however, that our approach to interpreting the forward irrelevant product has certain benefit. One of them is that it has a more direct behavioural interpretation. For one thing, in contradistinction with the general approach described by Shafer and Vovk (2001) and Shafer et al. (2003), it has no need of *Cournot's bridge* to link upper and lower previsions and probabilities to behaviour (see Shafer and Vovk (2001, Section 2.2) for more details).

Both the forward irrelevant product and the forward irrelevant marginal extension can be regarded as joint lower previsions compatible with a number of given marginals under forward irrelevance. The forward irrelevant natural extension of $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^{N-1})$ to $\mathscr{L}(\mathscr{X}^N)$ is the smallest coherent lower prevision that extends $\underline{P}_1$ and is coherent with $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^{N-1})$. But being coherent with $\underline{P}(\cdot|X^k)$, $k = 1,\dots,N-1$ means in particular extending the marginals $\underline{P}_2$, …, $\underline{P}_N$. Hence, the forward irrelevant natural extension is the smallest coherent joint lower prevision on $\mathscr{X}^N$ that is coherent with the given marginals, *also taking into account the assessment of forward irrelevance*.

If we are only interested in coherently constructing a joint lower prevision from a number of given marginals under forward irrelevance, the forward irrelevant natural extension seems to be an appropriate choice. If on the other hand, we wish to coherently extend the marginals to joint and conditional lower previsions on all gambles that satisfy the conditions for forward irrelevance everywhere (see Section 3.4), then the procedure of marginal extension, leading to the forward irrelevant product, seems preferable.

Finally, we have also related the forward irrelevant natural extension and product to a number of other ways of combining marginals into joints: they are more conservative than independent natural extension, type-1 and type-2 products.

APPENDIX A. EVENT-TREE INDEPENDENCE AND EPISTEMIC IRRELEVANCE

In this Appendix, we shed more light on why we believe that in certain contexts, and especially in random processes, where the values of random variables become known one after the other, the notion of epistemic irrelevance (rather than epistemic independence) is a natural one to consider. In particular, we show that for a random process with only two random variables, Shafer's (1996) notion of (event-tree) independence reduces to forward epistemic irrelevance. This can be generalised to random processes with more than two observations (or time instants), but the derivation is cumbersome, and the essential ideas remain the same as in the special case we consider here.

Consider an experiment, where two coins are flipped one after the other. The random variable X_1 is the outcome of the first coin flip, and X_2 the outcome of the second. Both variables assume values in the set $\{h,t\}$. The event tree for this experiment is depicted in Fig. 2. The labels for the *situations* (or nodes in the tree) should be intuitively clear: e.g., in

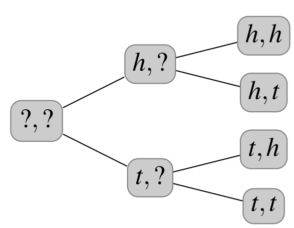

FIGURE 2. The event tree associated with two successive coin tosses.

the *initial* situation '?, ?' none of the coins have been flipped, in the *non-terminal* situation 'h, ?' the first coin has landed 'heads' and the second coin hasn't been flipped yet, and in the *terminal* situation 't,t' both coins have been flipped and have landed 'tails'.

For each variable X_k there is, in each situation s, a lower prevision (operator) $\underline{P}_s^{X_k}$ representing some subject's beliefs about the value of X_k conditional on the experiment getting to situation s. Table 1 identifies these models in the language used in the Introduction.

$\underline{P}^{X_1}_{?,?}(f)$	$\underline{P}^{X_1}_{h,?}(f)$	$\underline{P}^{X_1}_{t,?}(f)$	$\underline{P}^{X_1}_{h,h}(f)$	$\underline{P}^{X_1}_{h,t}(f)$	$\underline{P}^{X_1}_{t,h}(f)$	$\underline{P}^{X_1}_{t,t}(f)$
$\underline{P}_1(f)$	$f(h)$	$f(t)$	$f(h)$	$f(h)$	$f(t)$	$f(t)$

$\underline{P}^{X_2}_{?,?}(g)$	$\underline{P}^{X_2}_{h,?}(g)$	$\underline{P}^{X_2}_{t,?}(g)$	$\underline{P}^{X_2}_{h,h}(g)$	$\underline{P}^{X_2}_{h,t}(g)$	$\underline{P}^{X_2}_{t,h}(g)$	$\underline{P}^{X_2}_{t,t}(g)$
$\underline{P}_2(g)$	$\underline{Q}_2(g\vert h)$	$\underline{Q}_2(g\vert t)$	$g(h)$	$g(t)$	$g(h)$	$g(t)$

TABLE 1. The conditional lower previsions $\underline{P}^{X_1}_s$ and $\underline{P}^{X_2}_s$. f is any gamble on the outcome of the first variable X_1, and g any gamble on the outcome of the second variable X_2.

In Shafer's (1996) language, (the experiment in a) *situation s influences a variable X* if s has at least one child situation c such that the subject's model is changed if the experiment moves from s to c: for our imprecise probability models this means that $\underline{P}^X_s \neq \underline{P}^X_c$. Two variables X and Y are then *event-tree independent* if there are no situations that influence them both. Shafer proves in Section 8.1 of his book that for event trees with (precise) probabilistic models, event-tree independence implies the usual notion of independence as discussed in the Introduction.

Let us now investigate under what conditions the variables X_1 and X_2 will be event-tree independent. It is obvious that the only situations that can influence these variables, are the non-terminal situations, as these are the only ones with children. In Table 2, we use the identifications of Table 1 to indicate if and when each of these situations influences X_1 and/or X_2. It is clear from this table that X_1 and X_2 are event-tree independent if and only if $\underline{Q}_2(\cdot|X_1) = \underline{P}_2$, or in other words if X_1 is epistemically irrelevant to X_2.

	$?,?$	$h,?$	$t,?$
X_1	always	never	never
X_2	if $\underline{Q}_2(\cdot\vert h) \neq \underline{P}_2$ or $\underline{Q}_2(\cdot\vert t) \neq \underline{P}_2$	always	always

TABLE 2. When do the non-terminal situations influence the variables X_1 and X_2?

APPENDIX B. PROOFS OF THEOREMS

B.1. **Preliminaries.** Let us begin by making a few preliminary remarks, and fixing a number of notations. For $1 \leq j \leq N$, let $\mathscr{X}^N_j = \times_{\ell=j}^N \mathscr{X}_\ell$. For $1 \leq j \leq N-1$, we call a gamble f on $\mathscr{X}^N$ $\mathscr{X}^j$*-measurable* if it is constant on the elements of $\{(x_1,\dots,x_j)\} \times \mathscr{X}^N_{j+1}$, for all $(x_1,\dots,x_j)$ in $\mathscr{X}^j$.

We define the $\mathscr{X}^{j-1}$-*support* $S_j(f)$ of a gamble f on $\mathscr{X}^N$, as the set of those subsets $\{(x_1,\dots,x_{j-1})\}\times\mathscr{X}_j^N$ of $\mathscr{X}^N$ such that the gamble f is not identically zero on this set. In particular, $S_1(f)=\mathscr{X}^N$ if f is not identically 0. If f happens to be $\mathscr{X}^j$-measurable, then

$$S_j(f) := \{\{(x_1,\dots,x_{j-1})\}\times\mathscr{X}_j^N\colon f(x_1,\dots,x_{j-1},x_j)\neq 0 \text{ for some } x_j\in\mathscr{X}_j\}.$$

In what follows, we need to prove joint coherence of conditional lower previsions whose domains are not necessarily linear spaces. Since Walley (1991, Section 8.1) has only given definitions of joint coherence for conditional lower previsions defined on linear spaces, we must check the following, fairly straightforward, generalisation to non-linear domains of Walley's coherence condition. See Miranda and De Cooman (2005) for more details.

Consider a coherent lower prevision $\underline{Q}$ defined on a subset $\mathscr{H}^1$ of $\mathscr{L}(\mathscr{X}^N)$, and separately coherent conditional lower previsions $\underline{Q}(\cdot|X^1)$, …, $\underline{Q}(\cdot|X^{N-1})$ defined on the respective subsets $\mathscr{H}^2$, …, $\mathscr{H}^N$ of $\mathscr{L}(\mathscr{X}^N)$. We may assume without loss of generality that these domains are *cones* (see the discussion in Section 3.2). Then these (conditional) lower previsions are called *jointly coherent* if for any g_ℓ^j in $\mathscr{H}^\ell$, $j=1,\dots,n_\ell$, $n_\ell\geq 0$ and $\ell=1,\dots,N$ and any $k\in\{1,\dots,N\}$ and $g_0\in\mathscr{H}^k$, the following conditions are satisfied:

(JC1) If $k=1$, we must have that

$$\sup_{x\in\mathscr{X}^N}\left[\sum_{j=1}^{n_1}G(g_1^j)(x)+\sum_{\ell=2}^{N}\sum_{j=1}^{n_\ell}G(g_\ell^j|X^{\ell-1})(x)-G(g_0)(x)\right]\geq 0. \tag{20}$$

(JC2) If $k>1$, we must have that for any $(y_1,\dots,y_{k-1})\in\mathscr{X}^{k-1}$, there is some B in

$$\{\{(y_1,\dots,y_{k-1})\}\times\mathscr{X}_k^N\}\cup\bigcup_{\ell=1}^{N}\bigcup_{j=1}^{n_\ell}S_\ell(g_\ell^j)$$

such that

$$\sup_{x\in B}\left[\sum_{j=1}^{n_1}G(g_1^j)(x)+\sum_{\ell=2}^{N}\sum_{j=1}^{n_\ell}G(g_\ell^j|X^{\ell-1})(x)-G(g_0|y_1,\dots,y_{k-1})(x)\right]\geq 0. \tag{21}$$

Here, for any $x\in\mathscr{X}^N$, and $g_\ell^j\in\mathscr{H}^\ell$,

$$G(g_\ell^j)(x)=g_\ell^j(x)-\underline{Q}(g_\ell^j) \tag{22}$$

when $\ell=1$, and

$$G(g_\ell^j|X^{\ell-1})(x)=g_\ell^j(x_1,\dots,x_\ell)-\underline{Q}(g_\ell^j|(x_1,\dots,x_{\ell-1})) \tag{23}$$

when $\ell>1$. The idea behind this condition is, as for (unconditional) coherence (see Section 2), that we shouldn't be able to raise the supremum acceptable buying price we have given for a gamble f by considering the acceptable transactions implicit in other gambles in the domains. For instance, if condition (20) fails, then there is some $\delta>0$ such that the gamble $G(f_0)-\delta=f_0-(\underline{Q}(f_0)+\delta)$ dominates the acceptable transaction $\sum_{j=1}^{n_1}G(g_1^j)+\sum_{\ell=2}^{N}\sum_{j=1}^{n_\ell}G(g_\ell^j|X^{\ell-1})+\delta$, meaning that $\underline{Q}(f_0)+\delta$ must be an acceptable buying price for f_0.

Using Eqs. (22) and (23), it is easy to see that conditions (20) and (21) are equivalent, respectively, to

$$\sup_{x\in\mathscr{X}^N}\left[\sum_{j=1}^{n_1}[g_1^j(x)-\underline{Q}(g_1^j)]+\sum_{\ell=2}^{N}\sum_{j=1}^{n_\ell}[g_\ell^j(x)-\underline{Q}(g_\ell^j|x_1,\dots,x_{\ell-1})]-[g_0(x)-\underline{Q}(g_0)]\right]\geq 0,$$

and

$$\sup_{x\in B}\Bigg[\sum_{j=1}^{n_1}[g_1^j(x)-\underline{Q}(g_1^j)]+\sum_{\ell=2}^{N}\sum_{j=1}^{n_\ell}[g_\ell^j(x)-\underline{Q}(g_\ell^j|x_1,\dots,x_{\ell-1})] - I_{\{(y_1,\dots,y_{k-1})\}}(x_1,\dots,x_{k-1})[g_0(x)-\underline{Q}(g_0|y_1,\dots,y_{k-1})]\Bigg]\geq 0.$$

B.2. **Proof of Theorem 1.** Theorem 1 is a consequence of the following three lemmas.

Lemma 12. *The lower previsions $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^N)$ are jointly coherent.*

Proof. We can prove *independently* (using the Marginal Extension Theorem, see Theorem 2) that $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^N)$ can be extended to jointly coherent (conditional) lower previsions on all of $\mathscr{L}(\mathscr{X}^N)$. This means that the restrictions $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^N)$ to the respective domains $\mathscr{K}^1$, …, $\mathscr{K}^N$ of these jointly coherent (conditional) lower previsions, are of course jointly coherent as well. □

Lemma 13. *The lower prevision $\underline{E}^N$ is jointly coherent with the conditional lower previsions $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^{N-1})$.*

Proof. We have to check that the immediate generalisation (see Section B.1) to non-linear domains of Walley's joint coherence requirements is satisfied. We use the fact that $\underline{E}^N$ is a coherent lower prevision on the linear space $\mathscr{L}(\mathscr{X}^N)$, and that the domains $\mathscr{K}^i$ are cones. Consider $g_1\in\mathscr{L}(\mathscr{X}^N)$ and $g_i^j\in\mathscr{K}^i$ for $i=2,\dots,N,\ j=1,\dots,n_i$. Let $k\in\{1,\dots,N\}$, $g_0\in\mathscr{K}^k$ and $(y_1,\dots,y_{k-1})\in\mathscr{X}^{k-1}$. Assume first that $k>1$, i.e., let us prove (JC2). If $g_1=0$, then the result follows from the joint coherence of $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^{N-1})$.

Assume then that $g_1\neq 0$. Then (JC2) is equivalent to

$$\sup_{x\in\mathscr{X}^N}\left[[g_1(x)-\underline{E}^N(g_1)]+\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-G(g_0|y_1,\dots,y_{k-1})(x)\right]\geq 0. \tag{24}$$

From the definition of $\underline{E}^N$ [see Eq. (14)] and the fact that all relevant domains are cones, we see that for any $\varepsilon>0$, there are $m_i\geq 0$ and $h_i^j\in\mathscr{K}^i$ for $i=1,\dots,N$ and $j=1,\dots,m_i$, such that

$$g_1(x)-\sum_{i=1}^{N}\sum_{j=1}^{m_i}G(h_i^j|X^{i-1})(x)\geq\underline{E}^N(g_1)-\varepsilon \tag{25}$$

for all $x\in\mathscr{X}^N$. Hence

$$\begin{aligned}&\sup_{x\in\mathscr{X}^N}\left[[g_1(x)-\underline{E}^N(g_1)]+\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-G(g_0|y_1,\dots,y_{k-1})(x)\right]\\ &\geq\sup_{x\in\mathscr{X}^N}\left[\sum_{i=1}^{N}\sum_{j=1}^{m_i}G(h_i^j|X^{i-1})(x)+\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-G(g_0|y_1,\dots,y_{k-1})(x)\right]-\varepsilon\\ &\geq-\varepsilon,\end{aligned}$$

where the last inequality follows from the joint coherence of $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, …, $\underline{P}(\cdot|X^{N-1})$, by Lemma 12 [also see Eq. (21)]. Since this holds for any $\varepsilon>0$, we deduce that Eq. (24) holds.

Next, we consider (JC1). Assume then that $k=1$, i.e., $g_0\in\mathscr{L}(\mathscr{X}^N)$. We show that

$$\sup_{x\in\mathscr{X}^N}\left[[g_1(x)-\underline{E}^N(g_1)]+\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-[g_0(x)-\underline{E}^N(g_0)]\right]\geq 0. \tag{26}$$

Assume *ex absurdo* that there is some $\delta > 0$ such that this supremum is smaller than $-\delta$. Then, using the approximation of $\underline{E}^N(g_1)$ given by Eq. (25) for $\varepsilon = \frac{\delta}{2}$, we deduce that

$$\begin{aligned}-\delta &> \sup_{x\in\mathscr{X}^N}\left[[g_1(x)-\underline{E}^N(g_1)]+\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-[g_0-\underline{E}^N(g_0)]\right]\\ &\geq -\frac{\delta}{2}+\sup_{x\in\mathscr{X}^N}\left[\sum_{i=1}^{N}\sum_{j=1}^{m_i}G(h_i^j|X^{i-1})(x)+\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-[g_0(x)-\underline{E}^N(g_0)]\right],\end{aligned}$$

whence

$$\inf_{x\in\mathscr{X}^N}\left[g_0(x)-\sum_{i=1}^{N}\sum_{j=1}^{m_i}G(h_i^j|X^{i-1})(x)-\sum_{i=2}^{N}\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)\right] > \underline{E}^N(g_0)+\frac{\delta}{2},$$

and this contradicts the definition of $\underline{E}^N(g_0)$. Hence, Eq. (26) holds and we conclude that the (conditional) lower previsions $\underline{E}^N$, $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$ are jointly coherent. □

Lemma 14. *$\underline{E}^N$ is an extension of the lower prevision $\underline{P}_1$ to $\mathscr{L}(\mathscr{X}^N)$, i.e., $\underline{E}^N(f)=\underline{P}_1(f)$ for all f in $\mathscr{K}^1$.*

Proof. Consider $f\in\mathscr{K}^1$. From (15) we deduce that $\underline{E}^N(f)\geq\inf_{x\in\mathscr{X}^N}[f(x)-G_1(f)(x)]=\underline{P}_1(f)$. Assume *ex absurdo* that $\underline{E}^N(f)>\underline{P}_1(f)$. Then there are gambles $g_i^j\in\mathscr{K}_i$ for $i=1,\dots,N$ and $j=1,\dots,n_i$, such that $\inf_{x\in\mathscr{X}^N}[f(x)-\sum_{i=1}^N\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)]>\underline{P}_1(f)$. But this means that $\sup_{x\in\mathscr{X}^N}[\sum_{i=1}^N\sum_{j=1}^{n_i}G(g_i^j|X^{i-1})(x)-[f(x)-\underline{P}_1(f)]]<0$, and this contradicts the joint coherence of $\underline{P}_1$, $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$, which we have proved in Lemma 12 [also see Eqs. (20) and (21)]. Consequently, $\underline{E}^N(f)=\underline{P}_1(f)$. □

Proof of Theorem 1. Taking into account the previous lemmas, it only remains to prove that $\underline{E}^N$ is the point-wise smallest coherent extension of $\underline{P}_1$ to $\mathscr{L}(\mathscr{X}^N)$ that is coherent with $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$. Let therefore $\underline{F}$ be another coherent extension of $\underline{P}_1$ to $\mathscr{L}(\mathscr{X}^N)$ that is moreover jointly coherent with the conditional lower previsions $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$, and consider any $f\in\mathscr{L}(\mathscr{X}^N)$. Then, for any $\varepsilon>0$, the expression (14) for $\underline{E}^N(f)$ implies that there are $h_i^j\in\mathscr{K}^i$ for $i=1,\dots,N$ and $j=1,\dots,n_i$ such that

$$\sup_{x\in\mathscr{X}^N}\left[\sum_{i=1}^{N}\sum_{j=1}^{n_i}G(h_i^j|X^{i-1})(x)-f(x)\right]\leq-\underline{E}^N(f)+\varepsilon.$$

On the other hand, the coherence of $\underline{F}$, $\underline{P}(\cdot|X^1)$, ..., $\underline{P}(\cdot|X^{N-1})$ implies that

$$\sup_{x\in\mathscr{X}^N}\left[\sum_{i=1}^{N}\sum_{j=1}^{n_i}G(h_i^j|X^{i-1})(x)-[f(x)-\underline{F}(f)]\right]\geq 0,$$

whence

$$-\underline{F}(f)\leq\sup_{x\in\mathscr{X}^N}\left[\sum_{i=1}^{N}\sum_{j=1}^{n_i}G(h_i^j|X^{i-1})(x)-f(x)\right]\leq-\underline{E}^N(f)+\varepsilon.$$

Therefore, $\underline{F}(f)\geq\underline{E}^N(f)-\varepsilon$ for any $\varepsilon>0$, so $\underline{F}$ dominates $\underline{E}^N$ on $\mathscr{L}(\mathscr{X}^N)$. □

B.3. **Proof of Theorem 4.** From Theorem 3 we infer that $\underline{E}^N \leq \underline{M}^N$, so we only need to prove the converse inequality. Note first of all that if $\mathscr{K}_k = \mathscr{L}(\mathscr{X}_k)$, then the domain $\mathscr{K}^k$ of $\underline{P}(\cdot|X^{k-1})$ is $\mathscr{L}(\mathscr{X}^k)$, for $k = 1, \dots, N$.

Consider any gamble f on $\mathscr{X}^N$, and recursively define $f_N := f$, $f_{N-1} := \underline{P}_N(f_N)$, ..., $f_k := \underline{P}_{k+1}(f_{k+1})$, ..., $f_1 := \underline{P}_2(f_2)$. Then clearly f_k belongs to $\mathscr{L}(\mathscr{X}^k)$, and $\underline{P}_1(f_1) = \underline{M}^N(f)$, $\underline{P}(f_2|X^1) = \underline{P}_2(f_2) = f_1$, ..., $\underline{P}(f_k|X^{k-1}) = \underline{P}_k(f_k) = f_{k-1}$, ..., $\underline{P}(f_N|X^{N-1}) = \underline{P}_N(f_N) = f_{N-1}$. Hence

$$f - \sum_{k=1}^{N} G(f_k|X^{k-1}) = f - [f_1 - \underline{M}^N(f)] - \sum_{k=2}^{N-1} [f_k - f_{k-1}] - [f - f_{N-1}] = \underline{M}^N(f).$$

We therefore deduce from Eq. (14) that indeed $\underline{E}^N(f) \geq \underline{M}^N(f)$.

B.4. **Proof of Proposition 9.** First of all, we can assume without loss of generality that f_N is non-negative: otherwise, it suffices to take $g_N = f_N - \inf f_N$, which is non-negative and satisfies $g_N - \underline{E}_N(g_N) = f_N - \underline{E}_N(f_N)$.

From Proposition 8, we infer that

$$\begin{aligned}\underline{M}^N(f_1 \dots f_N) = \underline{E}^N(f_1 \dots f_N) &= \underline{E}_1(f_1) \dots \underline{E}_N(f_N) \\ &= \underline{M}^N(f_1 \dots f_{N-1}\underline{E}_N(f_N)) = \underline{E}^N(f_1 \dots f_{N-1}\underline{E}_N(f_N)),\end{aligned}$$

taking into account that $\underline{E}_N(f_N) \geq 0$ because we are assuming f_N to be non-negative, and the non-negative homogeneity of a coherent lower prevision. Using the super-additivity [due to coherence] of $\underline{M}^N$ and $\underline{E}^N$, we deduce that

$$\begin{aligned}\underline{M}^N(f_1 \dots f_{N-1}[f_N - \underline{E}_N(f_N)]) &\leq \underline{M}^N(f_1 \dots f_N) - \underline{M}^N(f_1 \dots f_{N-1}\underline{E}_N(f_N)) = 0 \\ \underline{E}^N(f_1 \dots f_{N-1}[f_N - \underline{E}_N(f_N)]) &\leq \underline{E}^N(f_1 \dots f_N) - \underline{E}^N(f_1 \dots f_{N-1}\underline{E}_N(f_N)) = 0.\end{aligned}$$

We are now going to show that, conversely, $\underline{E}^N(f_1 \dots f_{N-1}[f_N - \underline{E}_N(f_N)]) \geq 0$. This will complete the proof, since we know that $\underline{M}^N$ dominates $\underline{E}^N$, and it will therefore also follow that $\underline{M}^N(f_1 \dots f_{N-1}[f_N - \underline{E}_N(f_N)]) \geq 0$. So, fix $\varepsilon > 0$, then it follows from Eq. (12) that there are $n_N \geq 0$ and gambles g_N^j in $\mathscr{K}_N^*$ for $j = 1, \dots, n_N$ such that

$$\inf_{x_N \in \mathscr{X}_N} \left[f_N(x_N) - \sum_{j=1}^{n_N} [g_N^j(x_N) - \underline{P}_N(g_N^j)] \right] \geq \underline{E}_N(f_N) - \varepsilon. \tag{27}$$

Define the gambles h_N^j on $\mathscr{X}^N$ by $h_N^j := f_1 \dots f_{N-1} g_N^j$. All these gambles clearly belong to $\mathscr{K}^N$, so, using Eq. (15), we get that $\underline{E}^N(f_1 \dots f_{N-1}[f_N - \underline{E}_N(f_N)])$ is greater than or equal to

$$\begin{aligned}\sup_{\substack{g_{ki_k} \in \mathscr{K}^k, i_k = 1, \dots, n_k \\ n_k \geq 0, k = 1, \dots, N-1}} \inf_{x \in \mathscr{X}^N} \Bigg[& f_1 \dots f_{N-1}[f_N - \underline{E}_N(f_N)](x) \\ & - \sum_{k=1}^{N-1} \sum_{i_k=1}^{n_k} [g_{ki_k}(x_1, \dots, x_k) - \underline{P}_k(g_{ki_k}(x_1, \dots, x_{k-1}, \cdot))] \\ & - \sum_{j=1}^{n_N} [h_N^j(x_1, \dots, x_N) - \underline{P}_N(h_N^j(x_1, \dots, x_{N-1}, \cdot))] \Bigg]\end{aligned}$$

and after some manipulations, using the coherence of $\underline{P}_N$ and the fact that all the f_k are non-negative, this can be rewritten as

$$\sup_{\substack{g_{ki_k}\in\mathscr{K}^k, i_k=1,\dots,n_k\\ n_k\geq 0, k=1,\dots,N-1}} \inf_{x\in\mathscr{X}^{N-1}} \Bigg[-\sum_{k=1}^{N-1}\sum_{i_k=1}^{n_k}[g_{ki_k}(x_1,\dots,x_k)-\underline{P}_k(g_{ki_k}(x_1,\dots,x_{k-1},\cdot))] \\ + f_1(x_1)\dots f_{N-1}(x_{N-1})\inf_{x_N\in\mathscr{X}_N}\Bigg[f_N(x_N)-\underline{E}_N(f_N)-\sum_{j=1}^{n_N}[g_N^j(x_N)-\underline{P}_N(g_N j)]\Bigg]\Bigg]$$

Now if we use the inequality (27), we see that this is greater than or equal to

$$\underline{E}^N(f_1\dots f_{N-1}[-\varepsilon]) = -\varepsilon\overline{E}^N(f_1\dots f_{N-1}) \geq -\varepsilon\sup f_1\dots\sup f_{N-1},$$

where the last inequality is a consequence of coherence and the non-negativity of the f_k. Since this holds for any $\varepsilon>0$, we deduce that indeed $\underline{E}^N(f_1\dots f_{N-1}[f_N-\underline{E}_N(f_N)])\geq 0$.

B.5. **Proof of Proposition 10.** The proof of the inequality $\underline{E}^N(g[f_N-\underline{E}_N(f_N)])\geq 0$ is completely analogous to that in Proposition 9, with g taking the role of $f_1\dots f_{N-1}$. We deduce from this that $\underline{M}^N(g[f_N-\underline{E}_N(f_N)])\geq \underline{E}^N(g[f_N-\underline{E}_N(f_N)])\geq 0$.

Conversely, we get from Theorem 5 that $\underline{M}^N(gf_N)=\underline{M}^N(g)\underline{M}^N(f_N)=\underline{M}^{N-1}(g)\underline{E}_N(f_N)$ for any non-negative gamble g on $\mathscr{X}^{N-1}$ and any $f_N\in\mathscr{L}(\mathscr{X}_N)$. By the super-additivity [coherence] of $\underline{M}^N$ we get $\underline{M}^N(g[f_N-\underline{E}_N(f_N)])\leq \underline{M}^N(gf_N)-\underline{M}^N(g\underline{E}_N(f_N))\leq 0$, and as a consequence $\underline{M}^N(g(f_N-\underline{E}_N(f_N)))=\underline{E}^N(g(f_N-\underline{E}_N(f_N)))=0$.

B.6. **Proof of Eq.** (18)**.** Assume that for each X_k, we have a coherent lower prevision $\underline{P}_k$ defined on a *finite* subset $\mathscr{K}_k=\{f_k^1,\dots,f_k^{m_k}\}$ of $\mathscr{L}(\mathscr{X}_k)$. We may assume without loss of generality that the $\mathscr{K}_k$ contain no constant gambles. Then, for any $g_{kh_k}\in\mathscr{K}^k$ and $(x_1,\dots,x_{k-1})\in\mathscr{X}^{k-1}$, the gamble $g_{kh_k}(x_1,\dots,x_{k-1},\cdot)$ belongs by definition to $\mathscr{K}_k^*$, and consequently there are non-negative $\lambda_{(x_1,\dots,x_{k-1})kh_k}$, real $\mu_{(x_1,\dots,x_{k-1})kh_k}$ and gambles $g_{(x_1,\dots,x_{k-1})kh_k}$ in $\mathscr{K}_k$ such that

$$g_{kh_k}(x_1,\dots,x_{k-1},\cdot)=\lambda_{(x_1,\dots,x_{k-1})kh_k}g_{(x_1,\dots,x_{k-1})kh_k}+\mu_{(x_1,\dots,x_{k-1})kh_k}.$$

Hence

$$\begin{aligned} g_{kh_k}(x_1,\dots,x_{k-1},\cdot)-\underline{P}_k(g_{kh_k}|x_1,\dots,x_{k-1}) &= \lambda_{(x_1,\dots,x_{k-1})kh_k}g_{(x_1,\dots,x_{k-1})kh_k}+\mu_{(x_1,\dots,x_{k-1})kh_k} \\ &\quad -[\lambda_{(x_1,\dots,x_{k-1})kh_k}\underline{P}_k(g_{(x_1,\dots,x_{k-1})kh_k})+\mu_{(x_1,\dots,x_{k-1})kh_k}] \\ &= \lambda_{(x_1,\dots,x_{k-1})kh_k}[g_{(x_1,\dots,x_{k-1})kh_k}-\underline{P}_k(g_{(x_1,\dots,x_{k-1})kh_k})]. \end{aligned}$$

Define the gamble $\lambda_{kh_k}^j$ on $\mathscr{X}^{k-1}$ by

$$\lambda_{kh_k}^j(x_1,\dots,x_{k-1})=\begin{cases}\lambda_{(x_1,\dots,x_{k-1})kh_k} & \text{if } g_{(x_1,\dots,x_{k-1})kh_k}=f_k^j\\ 0 & \text{otherwise,}\end{cases}$$

for all $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$, $j=1,\dots,m_k$. Let us show that $\lambda_{kh_k}^j$ is indeed a gamble, i.e., that it is bounded. Since f_k^j is not a constant gamble by assumption, there are x_k^1 and x_k^2 in $\mathscr{X}_k$ such that $f_k^j(x_k^1)-f_k^j(x_k^2)>0$. Now, if $\lambda_{kh_k}^j(x_1,\dots,x_{k-1})$ were not bounded, there would exist, for any natural number M, some $(x_1,\dots,x_{k-1})$ in $\mathscr{X}^{k-1}$ such that $\lambda^j_{kh_k}(x_1,\dots,x_{k-1}) > M$. But then

$$g_{kh_k}(x_1,\dots,x_{k-1},x_k^1) - g_{kh_k}(x_1,\dots,x_{k-1},x_k^2) \\ = \lambda^j_{kh_k}(x_1,\dots,x_{k-1})(f_k^j(x_k^1) - f_k^j(x_k^2)) > M(f_k^j(x_k^1) - f_k^j(x_k^2)),$$

and this would imply that g_{kh_k} isn't bounded, a contradiction. Now,

$$g_{kh_k}(x_1,\dots,x_k) - \underline{P}_k(g_{kh_k}|(x_1,\dots,x_{k-1})) = \sum_{j=1}^{m_k} \lambda^j_{kh_k}(x_1,\dots,x_{k-1})(f_k^j(x_k) - \underline{P}_k(f_k^j)),$$

whence

$$\sum_{h_k=1}^{n_k} [g_{kh_k}(x_1,\dots,x_k) - \underline{P}_k(g_{kh_k}|(x_1,\dots,x_{k-1}))] \\ = \sum_{j=1}^{m_k} [f_k^j(x_k) - \underline{P}_k(f_k^j)] \sum_{h_k=1}^{n_k} \lambda^j_{kh_k}(x_1,\dots,x_{k-1})$$

and consequently

$$\underline{E}^N(f) = \sup_{\substack{g_{kh_k}\in\mathscr{K}^k \\ h_k=1,\dots,n_k,k=1,\dots,N}} \inf_{x\in\mathscr{X}^N} \left[f(x) - \sum_{k=1}^N \sum_{h_k=1}^{n_k} G(g_{kh_k}|X^{k-1}) \right] \\ \leq \sup_{\substack{\lambda_k^{j_k}\in\mathscr{L}^+(\mathscr{X}^{k-1}) \\ j_k=1,\dots,m_k,k=1,\dots,N}} \inf_{x\in\mathscr{X}^N} \left[f(x) - \sum_{k=1}^N \sum_{j=1}^{m_k} \lambda_k^{j_k}(x_1,\dots,x_{k-1})[f_k^j(x_k) - \underline{P}_k(f_k^j)] \right].$$

The converse inequality follows taking into account that for any $\lambda_k^{j_k}$ in $\mathscr{L}^+(\mathscr{X}^{k-1})$, the gamble g_{kj_k} in $\mathscr{X}^k$ given by $g_{kj_k} = \lambda_k^{j_k} f_k^j$ belongs to $\mathscr{K}^k$. This shows that we can indeed calculate the forward irrelevant natural extension $\underline{E}^N$ using Eq. (18).

ACKNOWLEDGEMENTS

We would like to thank the reviewers for their many comments, which were quite helpful in making this paper clearer, shorter and more readable.

REFERENCES

K. P. S. Bhaskara Rao and M. Bhaskara Rao. *Theory of Charges*. Academic Press, London, 1983.

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

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

G. de Cooman. Integration and conditioning in numerical possibility theory. *Annals of Mathematics and Artificial Intelligence*, 32:87–123, 2001.

G. de Cooman and D. Aeyels. Supremum preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.

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*, 30:124–130, 2000.

G. de Cooman and E. Miranda. Laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 2006. Submitted for publication.

G. de Cooman and M. C. M. Troffaes. Coherent lower previsions in systems modelling: products and aggregation rules. *Reliability Engineering and System Safety*, 85:113–134, 2004.

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

B. de Finetti. *Theory of Probability*. John Wiley & Sons, Chichester, 1974–1975. English translation of de Finetti (1970), two volumes.

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

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

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

E. Miranda and G. de Cooman. Coherence and independence in non-linear spaces. Technical report, Rey Juan Carlos University, TR05/10, May 2005. http://bayes.escet.urjc.es/publicaciones/wp05-10.pdf.

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.

G. Shafer. Conditional probability. *International Statistical Review*, 53:261–277, 1985.

G. Shafer. *The Art of Causal Conjecture*. The MIT Press, Cambridge, MA, 1996.

G. Shafer and V. Vovk. *Probability and Finance: It's Only a Game*. Wiley, New York, 2001.

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

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

P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975.

GHENT UNIVERSITY, SYSTEMS RESEARCH GROUP, TECHNOLOGIEPARK – ZWIJNAARDE 914, 9052 ZWIJNAARDE, BELGIUM

E-mail address: gert.decooman@UGent.be

REY JUAN CARLOS UNIVERSITY, DEPT. OF STATISTICS AND OPERATIONS RESEARCH – C-TULIPÁN, S/N, 28933 MÓSTOLES, SPAIN

E-mail address: enrique.miranda@urjc.es