CONDITIONAL MODELS: COHERENCE AND INFERENCE THROUGH SEQUENCES OF JOINT MASS FUNCTIONS

ENRIQUE MIRANDA AND MARCO ZAFFALON

Abstract. We call a *conditional model* any set of statements made of conditional probabilities or expectations. We take conditional models as primitive compared to unconditional probability, in the sense that conditional statements do not need to be derived from an unconditional probability. We focus on two problems: (*coherence*) giving conditions to guarantee that a conditional model is self-consistent; (*inference*) delivering methods to derive new probabilistic statements from a self-consistent conditional model. We address these problems in the case where the probabilistic statements can be specified imprecisely through sets of probabilities, while restricting the attention to finite spaces of possibilities. Using Walley's theory of *coherent lower previsions*, we fully characterise the question of coherence, and specialise it for the case of precisely specified probabilities, which is the most common case addressed in the literature. This shows that coherent conditional models are equivalent to sequences of (possibly sets of) unconditional mass functions. In turn, this implies that the inferences from a conditional model are the limits of the conditional inferences obtained by applying Bayes' rule, when possible, to the elements of the sequence. In doing so, we unveil the tight connection between conditional models and zero-probability events.

1. Introduction

Motivation. This paper deals with conditional probability models, where these are given without having to derive them from an unconditional probability. We investigate the rules to deal with this kind of models. Let us start with a simple example:

Example 1. Let X_1, X_2 be variables taking values in $\{1, 2\}$, about which you express the conditional probabilities $P(X_1 = 1|X_2 = 1) = 1 = P(X_1 = 2|X_2 = 2)$ and $P(X_2 = 1|X_1 = 2) = 1 = P(X_2 = 2|X_1 = 1)$. Unfortunately, this amounts to code the contradictory statements $X_1 = X_2$ and $X_1 \neq X_2$. ♦

This shows that when working with conditional probabilities we should make sure that they are not expressing contradictions, which may not always be as easy to spot as in this example.

One idea is to reject conditional probabilities for which there is not a 'compatible' joint mass function: i.e., in the example we should require the existence of probabilities $P(X_1 = x_1, X_2 = x_2)$, for all $x_1, x_2 \in \{1, 2\}$, that lead to the conditional ones through Bayes' rule. It can be checked that this allows us to detect the contradiction in the above conditional probabilities.

Key words and phrases. Sets of mass functions, coherent lower previsions, weak and strong coherence, natural extension, regular extension, desirable gambles.

This is an updated and expanded version of a paper [24] presented at ISIPTA '09: the *Sixth International Symposium on Imprecise Probability: Theories and Applications*.

Although this seems to be a useful necessary consistency condition, it still leaves room for inconsistencies; consider the following example adapted from [27, Section 7.3.5].

Example 2. Assume now that X_1, X_2 take values in $\{1, 2, 3\}$. You insist in coding the contradictory statements $X_1 = X_2$ and $X_1 \neq X_2$ on $\{1, 2\}$, while assuming in addition that $X_1 = 3$ if and only if $X_2 = 3$. You express this by the conditional probabilities $P(X_1 = 1|X_2 = 1) = 1 = P(X_1 = 2|X_2 = 2)$, $P(X_2 = 1|X_1 = 2) = 1 = P(X_2 = 2|X_1 = 1)$, and $P(X_1 = 3|X_2 = 3) = 1 = P(X_2 = 3|X_1 = 3)$. It can be checked that, contrary to what happened in Example 1, in this case the conditional probabilities are compatible with a joint mass function: the one determined by $P(X_1 = 3, X_2 = 3) = 1$. In fact, this is the *only* compatible joint. ♦

The key of this example is that the compatible joint assigns zero probability to the conditioning events that are involved in the inconsistency. Therefore Bayes' rule cannot be used just when it would be much needed to find out the contradiction.

One question at this point would be if it is worth dealing with these, let us call them, *conditional models*. After all, this type of models are not in the probability mainstream. The influential Kolmogorovian tradition is focused on unconditional probability as the primitive notion, and conditional probability as a derived one. This viewpoint bypasses the question of zero probabilities because it forbids to condition on events of probability zero. Moreover, since all the conditionals are obtained from a global model in the form of an unconditional probability, the question of originating contradictions is virtually not met.

However, for many researchers in probability, like de Finetti, and others in artificial intelligence (AI), it has made sense to take conditional probability as a primitive concept. Quoting [26, pp. 360–361]:

> "It is important to understand that even in the special case of conditional probabilities de Finetti was not merely interested in the Kolmogorovian set-up where conditional probability is *defined* from unconditional (σ–additive) probability. De Finetti focused on functions $P(\cdot|\cdot)$ such that $P(\cdot|A)$ is a (finitely additive) probability measure for any event A, $P(A|A) = 1$ for any A, and $P(A \cap B) = P(A|B)P(B)$ for any A and B. Such measures had been discussed already by Keynes [15] and appeared (later than de Finetti's first proposals) in various works in different research areas [6, 10]. Many AI techniques connected to ordinal uncertainty, default reasoning and counterfactual reasoning can be linked to such measures [1, 5, 16]. We should note that these 'full' conditional measures do allow one to define conditional probability $P(A|B)$ even when $P(B)$ is equal to zero."

In the statistical literature, these models, which go under the name of *conditionally specified distributions*, have also been the subject of considerable attention (e.g., see [2, 3, 4, 13, 30] and the references therein). The main reasons for this attention are that conditional models arise somewhat naturally in the context of subjective probability or within the AI community: it often makes sense to start with *any* set of probabilistic premises, whether conditional or not, in order to derive probabilistic conclusions. The Kolmogorovian setup may be seen as too narrow within these frameworks.

Nevertheless, the use of conditional models comes with the (possibly) contradictory statements that may follow from conditioning on events of zero probability. We think that this is a problem that should be properly dealt with, for several reasons. One is that zero probabilities might be hidden within a problem formulation:

one can write down a certain number of probabilistic statements without becoming aware that they imply some other probabilities to be equal to zero. Example 2 is a simple instance of this situation. A standpoint with more of a philosophical flavor is that the probability of an event being equal to zero does not imply that such an event is not going to occur.[1] And thirdly, conditioning on zero-probability events is often unavoidable when we consider conditional models which are *imprecisely specified*, which means that we consider thus *sets* of mass functions. In that case, it may happen that some of them give zero probability to a conditioning event while others do not; therefore it does make sense, and it is very often necessary, to consider probabilities conditional on such an event, because it is consistent with an assignment of positive probability, although it could actually also be given probability zero.

When all the assessments represented in some conditional models are consistent, it becomes also interesting to investigate which are their implications. This is another area where the presence of zero probabilities may complicate matters, because most of the approaches are focused on the notion of compatibility, even when the models are imprecisely specified, as in [3]. We shall find proper tools within the literature of subjective probability: specifically, in de Finetti's theory, and, more recently, in the theories of imprecise probability proposed by Williams [31], and later Walley [27].

Goal. Our aim in this paper is to provide results that tell us how we should work with conditional models, and how to relate those theories to the question of compatibility, which is central to the more common treatments of conditional models. We are particularly interested in two questions:

- *When is a conditional model self-consistent*? This question concerns the issue of avoiding contradictions mentioned initially and more generally of inconsistencies other than contradictions that can arise with these models.
- *How should we make inferences with a conditional model*? This question concerns the logical derivation of new probabilities from an initial set of conditional and unconditional probabilities.

We shall provide answers to these questions in the following, quite general, setup. We consider variables $X_1, \dots, X_n$ taking values from respective sets $\mathcal{X}_1, \dots, \mathcal{X}_n$. We assume these sets to be finite throughout the paper. The input probabilistic information is provided through a collection of *lower expectation functionals* $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$. For the generic term $\underline{P}_j(X_{O_j}|X_{I_j})$, O_j, I_j are disjoint subsets of $\{1, \dots, n\}$, with $O_j \neq \emptyset$; the symbols X_{O_j}, X_{I_j} denote vectors defined by $X_{O_j} := (X_k)_{k \in O_j}, X_{I_j} := (X_k)_{k \in I_j}$,[2] and taking values from the respective sets $\mathcal{X}_{O_j} := \times_{k \in O_j} \mathcal{X}_k, \mathcal{X}_{I_j} := \times_{k \in I_j} \mathcal{X}_k$. For some random variable f of X_{O_j}, and $z \in \mathcal{X}_{I_j}$, $\underline{P}_j(f|X_{I_j} = z)$ defines the lower expectation of f conditional on $X_{I_j} = z$. We shall show in Section 2.4 that the generic functional $\underline{P}_j(f|X_{I_j} = z)$ is in one-to-one correspondence with a closed and convex set of joint mass functions for X_{O_j}, conditional on $X_{I_j} = z$, in such a way that $\underline{P}_j(f|X_{I_j} = z)$ is actually the lower envelope of the expectations obtained from such a set. This is to say

[1] This is for instance the situation in continuous sample spaces.

[2] We use the symbol ':=' to denote a definition.

that, although we find convenient to formulate the problems using lower expectation functionals, our formulation is equivalent to one made of a collection of sets of conditional mass functions.

We use sets of mass functions instead of single mass functions because we want our results to be valid also in the case of imprecisely specified conditional models, such as those in [3]. Of course, also the case of precise probability is encompassed by our formulation by taking a single mass function in each set.

Using lower expectation functionals allows us in particular to make a bridge between (imprecisely specified) conditional models and Walley's theory of coherent lower previsions [27]. In fact, a *coherent lower prevision* is nothing else but a lower expectation functional. Walley's theory can be regarded as a generalisation of the Bayesian theory of probability (and statistics) to manage sets of probabilities,[3] which are also referred to as *imprecise probabilities.* We give an introductory overview of Walley's theory in Section 2. What is particularly interesting for our work here is that Walley's theory is founded on a notion of self-consistency, and moreover that it defines inferential tools that can be readily applied to conditional models.

Results. The central notion of self-consistency within the theory of coherent lower previsions is called *joint* or *strong coherence.* It implies a notion called *weak coherence.* What we show in Section 4 is that weak coherence precisely characterises the notion of compatibility, even when this is extended to deal with imprecisely specified probabilities. But we know from Example 2 that compatibility is not sufficient to rule out all inconsistencies. We shall see that the stronger concept of coherence is what allows one to get rid also of the inconsistencies that arise on top of zero probabilities.

Moreover, we note that in the case of precise probabilities, weak coherence is equivalent to a simpler notion of consistency called *avoiding uniform sure loss*, and coherence to another called *avoiding partial loss.* This allows us to give in Section 3 (Proposition 4) a specific characterisation of self-consistency for the precise case.

Regarding the second question related to probabilistic inference, we initially focus in Section 4 on conditional models that admit a compatible joint $\underline{P}(X_1, \dots, X_n)$. When all the mass functions in the corresponding set assign positive probability to the chosen conditioning event, it is possible to condition each of them in order to obtain an updated set of mass functions. We shall show that this procedure matches a procedure in Walley's theory that relies on weak coherence and that we call the *weak natural extension*; however, such a procedure is too weak as it can lead to inferences that are excessively conservative.

Indeed, Walley's theory focuses on a different inferential procedure that is based on strong coherence and is called the natural extension. The *natural extension* is shown by Walley to deliver the strongest least-committal inferences that logically follow from a conditional model.[4] The final part of this paper, in Section 5, is concerned with our most important result, that is, relating the natural extension to an improved form of compatibility that we define in this paper. This is important because the intuition behind Walley's procedure of the natural extension does not

[3]The theory is developed also for infinite spaces of possibilities, not only for the finite case. In that case the sets are made of finitely additive probabilities.

[4]This is not necessarily the case if the spaces of possibilities are infinite, although this situation lies outside the scope of this paper.

easily carry over to more traditional approaches to probability; relating the natural extension to a kind of compatibility makes it instead easier to access the rationale and the implications of the natural extension.

To illustrate our result, we need first to introduce one more notion. Consider a closed and convex set of joint mass functions for $X_1, \ldots, X_n$ such that for any event there is at least one mass function in the set that assigns positive probability to it (in other words, the set assigns positive *upper* probability to every event). We summarise the set by the lower expectation functional $\underline{P}(X_1, \ldots, X_n)$. Then the conditional models derived from $\underline{P}(X_1, \ldots, X_n)$ by *regular extension* are those obtained by applying Bayes' rule whenever possible (that is, whenever the conditioning event has positive probability) to the mass functions in the set.

We shall prove that a conditional model $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$ is strongly coherent if and only if there is a sequence of lower expectation functionals $\underline{P}_\epsilon(X_1, \ldots, X_n)$, $\epsilon \in \mathbb{R}^+$, that assign positive upper probability to each event, and with the following property: $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$ are the limits, when ϵ goes to zero, of $\underline{P}_1^\epsilon(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m^\epsilon(X_{O_m}|X_{I_m})$, the conditional model obtain from $\underline{P}_\epsilon(X_1, \ldots, X_n)$ by regular extension. This means that any coherent conditional model can be regarded as the limit of the conditionals obtained from a sequence of sets of joint mass functions. It can be shown that the need of the limit depends on the presence of zero probabilities.

Let us now focus on the natural extension. For any choice of O_{m+1}, I_{m+1}, disjoint subsets of $\{1, \ldots, n\}$, with $O_{m+1} \neq \emptyset$, the natural extension is the strongest least-committal lower expectation functional $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ that logically follows from the conditional model $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$. What we prove is that $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is nothing else but the limit of $\underline{P}_{m+1}^\epsilon(X_{O_{m+1}}|X_{I_{m+1}})$ when ϵ goes to zero.

A useful way to interpret these results is to regard the conditional model as actually equivalent to the sequence $\underline{P}_\epsilon(X_1, \ldots, X_n)$, $\epsilon \in \mathbb{R}^+$. In a sense, this means that we could forget about conditional models and just focus on a more familiar framework made of (sets of) joint mass functions, although we need to consider limits.

We should mention that ours is not the first work in this direction. A very interesting paper by Walley, Pelessoni and Vicig [29] has introduced these ideas originally in a more limited setup made of lower and upper probability functionals. This is equivalent to focusing on special types of closed and convex sets of mass functions, as lower and upper probability functionals are not as informative as lower and upper expectation functionals. Moreover, the sets that correspond to lower and upper probabilities can be represented as polytopes in the space that have a *finite* number of vertices. This has allowed the authors of the mentioned paper to use duality tools from linear programming to obtain their results.

In our framework this has not been possible, because the sets we deal with may have an infinite number of vertices. This has required us to use mathematical tools other than those in [29], such as a separation hyperplane theorem, which has made the technical treatment somewhat more involved. We have opted then for clarity to gather all the proofs of the statements in this paper in an Appendix.

2. An overview of the theory of coherent lower previsions

In this section we give an overview of the theory of coherent lower previsions in the case of finite spaces of possibilities.[5] This is a theory of probability generalised to handle imprecisely specified probabilities through sets of mass functions. Despite being a theory of probability, its formulation may look unusual to the reader familiar with more traditional ways to present probability, and this can make the theory somewhat uneasy to access. Because of this fact, we shall point out here informally some of the differences in the formulations, in order to help the reader have a smoother start into the theory. In the remainder of the section we shall rigorously state all the notions introduced here. We shall also use from now on the terminology from Walley's theory.

Probability theory is most often defined, after Kolmogorov, using a triple made of a sample space, a sigma algebra, and a probability function P. The functions from the sample space into the real numbers that are *measurable* with respect to the sigma algebra are called *random variables*. The expectation of a random variable is *defined* on the basis of the probability P. Conditional probability is also *defined* using P but only when the conditioning event is assigned positive probability by P.

The theory of coherent lower previsions has its focus on expectation rather than probability. We still have the sample space (which is usually referred to as the *possibility space*). We also have a set of random variables, which are called *gambles*: these are *bounded* functions from the possibility space to the real numbers. The set of gambles does not need to be concerned with measurability questions, that is, it can be chosen arbitrarily. Finally, a *coherent lower prevision* is defined as a functional, from the set of gambles to the real numbers, that satisfies some rationality criteria. This function is conjugate to another that is called a *coherent upper prevision*. The intuition behind the notions of coherent lower and upper previsions is that of lower and upper expectation functionals.

In the theory of coherent lower previsions, probability is a derived notion: a *coherent lower probability* is a coherent lower prevision defined on a set of indicator functions of events. The conjugate function is called a *coherent upper probability*. When a coherent lower prevision coincides with its conjugate coherent upper prevision, we call it a *linear prevision*. In case it is applied to a set of indicator functions, we are back to *precise* probability, which is indeed a special case of coherent lower previsions. Moreover, it can be shown that a linear prevision is in one-to-one correspondence with a finitely additive probability, and that a lower prevision is in one-to-one correspondence with a set of linear previsions. In other words, a coherent lower prevision can be equivalently regarded as a set of finitely additive probabilities.

In the conditional framework, the differences between the theories are even more marked. In fact, a conditional lower prevision can be defined without any reference to an unconditional one. It can even be defined when the conditioning event has (lower or upper) probability equal to zero. In a sense, the notion of conditional lower prevision is the fundamental one, and the unconditional notion is regarded as a special case. This change of perspective originates an issue that is not perceived

[5] We refer to [27] for an in-depth study of the theory in the general case, and to [19] for a survey. We can also find in these works a detailed interpretation of the different notions we shall introduce in terms of desirable buying and selling prices.

in the theories that regard conditional probability as a derived notion: that when it is specified a set of conditional lower previsions, it is not guaranteed that those conditionals are automatically self-consistent. This needs be imposed by a rationality notion called *joint* (or *strong*) *coherence*. Strong coherence can be regarded as the unique axiom in the theory; all the properties of coherent lower previsions can be derived from it, included the axioms of precise probability.

It has been shown that strong coherence is especially needed to rule out the inconsistencies that may arise on top of events of zero probability (Example 2 is just an instance of these problems). This aspect is perhaps one of the most peculiar of the theory of coherent lower previsions, just because in other theories it is not allowed to condition on events of probability zero. Even more, precise-probability theories tend to regard such events as negligible. In contrast, this view is difficult to justify in a theory of imprecise probability. This is because, for instance, the fact that an event has zero lower probability only means that it is consistent with zero probability, not that is must have probability equal to zero. Therefore those events have to be admitted in the theory, and properly dealt with, which creates some complications.

A final remark is on notation. A coherent lower prevision is usually denoted by $\underline{P}$, and a linear prevision by P. This can be confusing to the readers used to reserve the symbol P for probability. We stress that, in the theory of coherent lower prevision, whether or not P is referring to a probability is understood by considering the gamble to which P is applied. If it is an indicator function, then P expresses a probability. Analogous considerations hold for lower probabilities. Also, in order to simplify the notation we shall identify any set A with the gamble given by its indicator function.

2.1. **Conditional lower previsions.** Consider variables $X_1, \ldots, X_n$, taking values in respective *finite* sets $\mathcal{X}_1, \ldots, \mathcal{X}_n$. For any non-empty subset $J \subseteq \{1, \ldots, n\}$ we shall denote by X_J the (new) variable

$$X_J := (X_j)_{j \in J},$$

which takes values in the product space

$$\mathcal{X}_J := \times_{j \in J} \mathcal{X}_j.$$

We shall also use the notation $\mathcal{X}^n$ for $\mathcal{X}_{\{1,\ldots,n\}}$. This is going to be our possibility space for the rest of the paper.

Remark 1. In the present formulation the possibility space is given some structure using smaller possibility spaces related to variables $X_1, \ldots, X_n$. It is useful to point out that the theory of coherent lower previsions can be formulated using a possibility space defined directly without reference to an underlying set of variables, as it is done by Walley in his seminal book, for instance. Such a formulation is somewhat more elegant, and also more expressive as the possibility space does not need to be a product space. On the other hand, the formulation made of variables makes it easier to connect the theory with the several probabilistic and statistical models that are naturally formulated using variables (e.g., see [23, 33]), and hence can be more easily exploited by many common models and applications. ⧫

For any $J \subseteq \{1,\dots,n\}$, a *gamble* on $\mathcal{X}_J$ is a bounded real-valued function on $\mathcal{X}_J$ [6].

Definition 1. Let J be a subset of $\{1,\dots,n\}$, and let $\pi_J : \mathcal{X}^n \to \mathcal{X}_J$ be the so-called *projection operator*, i.e., the operator that drops the elements of a vector in $\mathcal{X}^n$ that do not correspond to indexes in J. A gamble f on $\mathcal{X}^n$ is called *$\mathcal{X}_J$-measurable* when for all $x, y \in \mathcal{X}^n$, $\pi_J(x) = \pi_J(y)$ implies that $f(x) = f(y)$.

This notion means that the value f takes depends only on the components of $x \in \mathcal{X}^n$ that belong to the set J. There is a one-to-one correspondence between the gambles on $\mathcal{X}^n$ that are $\mathcal{X}_J$-measurable and the gambles on $\mathcal{X}_J$. For the aims of this paper there is no restriction of generality in identifying a gamble on $\mathcal{X}_J$ with the corresponding $\mathcal{X}_J$-measurable gamble. Therefore *all the gambles we shall deal with will be defined on $\mathcal{X}^n$.*[7] In particular, we shall denote by $\mathcal{K}_J$ the set of $\mathcal{X}_J$-measurable gambles.

Consider two disjoint subsets O, I of $\{1,\dots,n\}$, with $O \neq \emptyset$. A *conditional lower prevision* $\underline{P}(X_O|X_I)$ is a functional defined on a subset $\mathcal{H}_{O\cup I}$ of the set of $\mathcal{X}_{O\cup I}$-measurable gambles $\mathcal{K}_{O\cup I}$, such that for every gamble $f \in \mathcal{H}_{O\cup I}$, $\underline{P}(f|X_I)$ is a mapping between $\mathcal{X}_I$ and the reals. In particular, for every element $z \in \mathcal{X}_I$, $\underline{P}(f|X_I = z)$ is understood as the lower expectation of f conditional on $X_I = z$. Note that we are not placing restrictions on the set $\mathcal{H}_{O\cup I}$ of gambles. We shall also use the notations

$$G(f|z) := \pi_I^{-1}(z)(f - \underline{P}(f|z)),\ G(f|X_I) := \sum_{z\in\mathcal{X}_I} G(f|z) = f - \underline{P}(f|X_I)$$

for all $f \in \mathcal{K}_{O\cup I}$ and all $z \in \mathcal{X}_I$. These are $\mathcal{X}_{O\cup I}$-measurable gambles, too.

2.2. **Consistency notions.** Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be conditional lower previsions with respective domains $\mathcal{H}^1,\dots,\mathcal{H}^m \subseteq \mathcal{L}(\mathcal{X}^n)$, where $\mathcal{H}^j$ is a subset of the set $\mathcal{K}^j$ of $\mathcal{X}_{O_j\cup I_j}$-measurable gambles,[8] for $j = 1,\dots,m$. As discussed at the beginning of Section 2, we need to impose the requirement that these conditional lower previsions are self-consistent.

Definition 2. $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are *(strongly) coherent* when for every $f_j^k \in \mathcal{H}^j, j = 1,\dots,m, k = 1,\dots,n_j$, and for every $j_0 \in \{1,\dots,m\}, f_0 \in \mathcal{H}^{j_0}, z_{j_0} \in \mathcal{X}_{I_{j_0}}$,

$$\max_{x\in\pi^{-1}_{I_{j_0}}(z_{j_0})\cup\mathbb{S}(f_j^k)} \left[\sum_{j=1}^m \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) - G_{j_0}(f_0|z_{j_0})\right](x) \geq 0. \tag{1}$$

Here, $\mathbb{S}(f_j^k) := \cup_{j=1}^m \cup_{k=1}^{n_j} S_j(f_j^k)$ denotes the elements $x \in \mathcal{X}^n$ in the union of the supports. The *$\mathcal{X}_I$-support* $S(f)$ of a gamble f in $\mathcal{K}_{O\cup I}$ is given by

$$S(f) := \{\pi_I^{-1}(z) : z \in \mathcal{X}_I, f\pi_I^{-1}(z) \neq 0\},$$

i.e., it is the set of conditioning events for which the restriction of f is not identically zero.

[6]Since in this paper we assume that the sets $\mathcal{X}_J$ are finite for all J, a gamble on $\mathcal{X}_J$ will be simply a real-valued function on $\mathcal{X}_J$.

[7]This will also be the case when, by an abuse of notation, we shall consider gambles on $\mathcal{X}_J$.

[8]We use $\mathcal{K}^j$ instead of $\mathcal{K}_{O_j\cup I_j}$ in order to alleviate the notation when no confusion is possible about the variables involved.

In some cases we shall focus on the set of $z \in \mathcal{X}_I$ s.t. $\pi_I^{-1}(z) \in S(f)$. To keep things simple, we shall make an abuse of notation by writing them as $z \in S(f)$, despite $S(f)$ contains subsets of $\mathcal{X}^n$, not elements of $\mathcal{X}_I$.

The notion of coherence is the strongest consistency notion we shall use in this paper. Detailing the motivations behind coherence is out of the scope of this paper; we instead refer to [27, Section 7.1.4(b)] for a detailed description and justification of this notion. As we have already mentioned, coherence can be taken as the unique axiom in the theory of coherent lower previsions: all the more familiar properties of probability follow from it, as well as many others. Below we list some weaker consistency criteria implied by coherence.

In the particular case where we have only one conditional lower prevision, the notion of strong coherence is called separate coherence:

Definition 3. A conditional lower prevision $\underline{P}(X_O|X_I)$ with domain $\mathcal{H}_{O\cup I}$ is *separately coherent* if for every $z \in \mathcal{X}_I$, the gamble $\pi_I^{-1}(z)$ belongs to $\mathcal{H}_{O\cup I}$ and $\underline{P}(\pi_I^{-1}(z)|z) = 1$, and moreover

$$\max_{x\in\pi_I^{-1}(z)} \left[\sum_{j=1}^{\ell} \lambda_j G(f_j|z) - G(f_0|z)\right](x) \geq 0$$

for every $\ell \in \mathbb{N}, f_j \in \mathcal{H}_{O\cup I}, \lambda_j \geq 0, j = 1, \ldots, \ell, f_0 \in \mathcal{H}_{O\cup I}$.

Theorem 1. [27, Theorem 6.2.7] *When the domain $\mathcal{H}_{O\cup I}$ is a linear set of gambles (i.e., closed under addition and multiplication by real numbers), separate coherence holds if and only if the following conditions are satisfied for all $z \in \mathcal{X}_I, f, g \in \mathcal{H}_{O\cup I}$, and $\lambda > 0$:*

$$\underline{P}(f|z) \geq \min_{x\in\pi_I^{-1}(z)} f(x) \tag{SC1}$$

$$\underline{P}(\lambda f|z) = \lambda\underline{P}(f|z) \tag{SC2}$$

$$\underline{P}(f+g|z) \geq \underline{P}(f|z) + \underline{P}(g|z). \tag{SC3}$$

Separate coherence is necessary and sufficient to deduce that for all $z \in \mathcal{X}_I$, $\underline{P}(X_O|X_I = z)$ is in one-to-one correspondence with a set of conditional mass functions. We detail this aspect in Section 2.4.

Remark 2. As a side comment, which is however useful for our future developments, it is possible to deduce from Definition 3 that given a separately coherent conditional lower prevision $\underline{P}(X_O|X_I)$, we may assume without loss of generality that its domain $\mathcal{H}_{O\cup I}$ contains all the gambles $\lambda f - \mu$ for every $f \in \mathcal{H}_{O\cup I}$, $\lambda \geq 0$ and $\mu \in \mathbb{R}$, and moreover that for each $z \in \mathcal{X}_I$, $f_z \in \mathcal{H}_{O\cup I}$, also the gamble $\sum_{z\in\mathcal{X}_I} f_z\pi_I^{-1}(z)$ belongs to $\mathcal{H}_{O\cup I}$ (see, for example, [27, Section 6.2.4]). We shall hold these assumptions about $\mathcal{H}_{O\cup I}$ throughout the paper, unless stated otherwise. They imply that the $\mathcal{X}_I$-measurable gambles are in $\mathcal{H}_{O\cup I}$, and moreover that

- for all $f \in \mathcal{H}_{O\cup I}, z \in \mathcal{X}_I$, both $G(f|z)$ and $G(f|X_I)$ belong to $\mathcal{H}_{O\cup I}$;
- for all $f \in \mathcal{H}_{O\cup I}, z \in \mathcal{X}_I, \lambda \geq 0$, $\lambda G(f|z) = G(\lambda f|z)$ and $\lambda G(f|X_I) = G(\lambda f|X_I)$.

The second point in particular will allow us to simplify the notation by removing the λ-coefficients from many formulae. ⧫

In general, a number of conditional lower previsions can be separately coherent but not strongly coherent. Three intermediate consistency conditions, implied by strong coherence as well, are introduced in the following definition (see [27, Section 7.1] for more details). We need these intermediate definitions in the paper because it is through them that we shall be able to make connections between known concepts in conditional models and coherent lower previsions.

Definition 4. Let $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ be a number of separately coherent conditional lower previsions with domains $\mathcal{H}^j \subseteq \mathcal{K}^j$ for $j = 1, \dots, m$.

(1) They *avoid uniform sure loss* if for every $f_j^k \in \mathcal{H}^j, j = 1, \dots, m, k = 1, \dots, n_j$,

$$\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k | X_{I_j}) \right](x) \geq 0.$$

(2) They *avoid partial loss* if for every $f_j^k \in \mathcal{H}^j, j = 1, \dots, m, k = 1, \dots, n_j$ such that not all the f_j^k are zero gambles,

$$\max_{x \in \mathbb{S}(f_j^k)} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k | X_{I_j}) \right](x) \geq 0,$$

where $\mathbb{S}(f_j^k) = \cup_{j=1}^{m} \cup_{k=1}^{n_j} S_j(f_j^k)$, as in Definition 2.

(3) They are *weakly coherent* if for every $f_j^k \in \mathcal{H}^j, j = 1, \dots, m, k = 1, \dots, n_j$, and for every $j_0 \in \{1, \dots, m\}, f_0 \in \mathcal{H}^{j_0}, z_{j_0} \in \mathcal{X}_{I_{j_0}}$,

$$\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k | X_{I_j}) - G_{j_0}(f_0 | z_{j_0}) \right](x) \geq 0. \quad (2)$$

Condition 1 prevents one from expressing (a severe type of) contradictions using coherent lower previsions.[9] It is equivalent to the non-emptiness of a certain set of probabilities, as shown in Proposition 2 in Section 3. Condition 2 strengthens the previous one by taking the maximum over the union of the supports. It can be shown that this permits excluding the contradictions that can be originated through events of probability zero, and that are not handled properly by the previous condition. Still, a collection of conditional lower previsions may not give rise to contradictions, and hence can avoid uniform sure loss, but be inconsistent, in the sense that the logical implications of some of them can be used to force a change in one of the others. This is avoided using Condition 3. However, a number of weakly coherent conditional lower previsions can still present some forms of inconsistency with one other. See [27, Chapter 7], [20] and [29] for some discussion. This is again related to zero probabilities, as weak coherence inherits from avoiding uniform sure loss the inability to cope with them effectively. To avoid also those, one needs strong coherence.

2.3. **Unconditional models.** It is useful for this paper to consider the particular case where $I = \emptyset$. We have then an *(unconditional) lower prevision* $\underline{P}(X_O)$ on a subset $\mathcal{H}_O$ of the set $\mathcal{K}_O$ of $\mathcal{X}_O$-measurable gambles. Separate coherence is called in that case coherence:

[9]Those contradictions could in principle be exploited in a betting system to give rise to a sure loss. This explains the origin of the condition's name.

Definition 5. A lower prevision $\underline{P}(X_O)$ with domain $\mathcal{H}_O$ is *coherent* if for every $\ell \in \mathbb{N}, f_0, f_1, \dots, f_\ell \in \mathcal{H}_O$ and $\lambda_1, \dots, \lambda_\ell \geq 0$, it holds that

$$\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{\ell} \lambda_j G(f_j) - G(f_0) \right] (x) \geq 0.$$

Theorem 2. [27, Theorem 2.5.5] *When the domain $\mathcal{H}_O$ is a linear set of gambles, coherence holds if and only if the following conditions hold for all $f, g \in \mathcal{H}_O$, and $\lambda > 0$:*

$$\underline{P}(f) \geq \min f \tag{C1}$$

$$\underline{P}(\lambda f) = \lambda \underline{P}(f) \tag{C2}$$

$$\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g). \tag{C3}$$

An example of a coherent lower prevision is the *vacuous* lower prevision $\underline{P}(X_O)$ given by $\underline{P}(f) := \min_{x \in \mathcal{X}_O} f(x)$ for every $f \in \mathcal{K}_O$. This is a model for ignorance as it only states that X_O belongs to $\mathcal{X}_O$, and can indeed be shown to be equivalent to the set of all mass functions for X_O.

2.4. **Linear previsions and envelope theorems.** Let us focus on a very important special case of lower previsions, that of linear previsions.

Definition 6. We say that a conditional lower prevision $\underline{P}(X_O|X_I)$ on the set $\mathcal{K}_{O \cup I}$ [10] is *linear* if and only if it is separately coherent and moreover $\underline{P}(f+g|z) = \underline{P}(f|z) + \underline{P}(g|z)$ for all $z \in \mathcal{X}_I$ and $f, g \in \mathcal{K}_{O \cup I}$.

When a separately coherent conditional lower prevision $\underline{P}(X_O|X_I)$ is linear we shall denote it by $P(X_O|X_I)$; in the unconditional case, we shall use the notation $P(X_O)$ when the domain is the set $\mathcal{K}_O$ of $\mathcal{X}_O$-measurable gambles and simply P when the domain is the set $\mathcal{L}(\mathcal{X}^n)$ of all gambles.

With linear previsions, the consistency notions can be formulated entirely in terms of losses, as we state in the following proposition.

Proposition 1. [27, Section 7.1.4] *A number of conditional linear previsions are coherent if and only if they avoid partial loss; and they are weakly coherent if and only if they avoid uniform sure loss.*

Conditional linear previsions correspond to expectations with respect to a conditional probability. They are the *precise* models, in contradistinction with the imprecise models represented by conditional lower previsions. In particular, an unconditional linear prevision P is the expectation with respect to the probability which is the restriction of P to events. For conditional linear previsions there is no difference between working with events (probabilities) or gambles (expectations), as one model is uniquely determined by the other (see [27, Section 2.8]). Such an equivalence does not hold any longer when we have conditional lower previsions, and this is the reason why the theory is formulated in general in terms of gambles.

The following result gives a characterisation of the coherence of a linear conditional and a linear unconditional prevision:

[10] We shall always assume for mathematical convenience in this paper that the domain of a conditional linear prevision $P(X_O|X_I)$ is the whole set $\mathcal{K}_{O \cup I}$ of $\mathcal{X}_{O \cup I}$-measurable gambles. This will be sufficient for the results we develop.

Theorem 3. [27, Theorem 6.5.7] *If we have a linear prevision P and a linear conditional prevision $P(X_O|X_I)$, they are coherent if and only if for all $\mathcal{X}_{O\cup I}$-measurable f, $P(f) = P(P(f|X_I))$. This is equivalent to requiring that $P(f|z) = \frac{P(f\pi_I^{-1}(z))}{P(z)}$ for all $f \in \mathcal{K}_{O\cup I}$ and all $z \in \mathcal{X}_I$ with $P(z) > 0$.*

One of the nice features of the notion of coherence is that it can be given a Bayesian sensitivity analysis interpretation. Given an unconditional lower prevision $\underline{P}$ with domain $\mathcal{H}$, we shall denote the set of *dominating* linear previsions by

$$\mathcal{M}(\underline{P}) := \{P : P(f) \geq \underline{P}(f)\ \forall f \in \mathcal{H}\}.$$

A closed[11] and convex[12] set of linear previsions is also called a *credal set* [18]; an instance is the set $\mathcal{M}(\underline{P})$, that we shall call the *credal set associated to* $\underline{P}$. Given that a linear prevision is in one-to-one correspondence with a probability mass function, the credal set $\mathcal{M}(\underline{P})$ is precisely the set of mass functions that we mentioned in the Introduction and that is associated to $\underline{P}$. Similarly, for a conditional lower prevision $\underline{P}(X_O|X_I)$ with domain $\mathcal{H}_{O\cup I}$, we define the credal set associated to $\underline{P}(X_O|X_I)$ as

$$\mathcal{M}(\underline{P}(X_O|X_I)) := \{P(X_O|X_I) : P(f|z) \geq \underline{P}(f|z)\ \forall f \in \mathcal{H}_{O\cup I}, z \in \mathcal{X}_I\}.$$

The following result allows us to use lower previsions in the place of sets of mass functions, and vice versa.

Theorem 4. [27, Theorem 3.3] *An unconditional lower prevision $\underline{P}$ is coherent if and only if it is the lower envelope of $\mathcal{M}(\underline{P})$. A conditional lower prevision $\underline{P}(X_O|X_I)$ is separately coherent if and only if it is the lower envelope of $\mathcal{M}(\underline{P}(X_O|X_I))$.*[13]

The situation when we have more than one conditional lower prevision is as follows.

Theorem 5. [27, Theorem 8.1.10] *When the referential spaces are finite and the domains are linear spaces, coherent $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are the envelope of a set $\{P_\gamma(X_{O_1}|X_{I_1}),\dots,P_\gamma(X_{O_m}|X_{I_m}) : \gamma \in \Gamma\}$ of dominating coherent conditional linear previsions.*

In [20], a similar result was established for weak coherence. In Section 4 we shall generalise this second property to arbitrary domains. Taking this into account, we can see conditional lower previsions as a form of representing imprecisely specified conditional models, by taking the lower envelopes of the set of possible models. In the rest of this paper we shall give representations of the consistency notions introduced in Definitions 2 and 4 in terms of sets of mass functions.

2.5. **Generalised Bayes Rule and compatibility.** An unconditional lower prevision can be updated through a generalisation of Bayes' rule.

Theorem 6. [27, Theorems 6.4.1 and 6.5.7] *When we are given only an unconditional lower prevision $\underline{P}$ on $\mathcal{L}(\mathcal{X}^n)$ and a conditional lower prevision $\underline{P}(X_O|X_I)$ on*

[11]In the *weak* topology*, which is the smallest topology for which all the evaluation functionals given by $f(P) := P(f)$, where $f \in \mathcal{L}(\mathcal{X}^n)$, are continuous.

[12]That is, for all linear previsions P_1, P_2 in the set and all $\alpha \in (0,1)$, the linear prevision $\alpha P_1 + (1-\alpha)P_2$ also belongs to this set.

[13]This is a bit of an abuse of notation, since actually for every $z \in \mathcal{X}_I$ the set $\mathcal{M}(\underline{P}(X_O|z))$ is a set of linear previsions.

$\mathcal{K}_{O\cup I}$, weak and strong coherence are equivalent, and they hold if and only if, for all $\mathcal{X}_{O\cup I}$-measurable gambles f and all $z \in \mathcal{X}_I$,

$$\underline{P}(G(f|z)) = 0. \tag{GBR}$$

This is called the Generalised Bayes Rule (GBR). When $\underline{P}(z) > 0$, GBR can be used to determine the value $\underline{P}(f|z)$, as it is the unique *value for which $\underline{P}(G(f|z)) = \underline{P}(\pi_I^{-1}(z)(f - \underline{P}(f|z))) = 0$ holds.*

GBR can be given a sensitivity analysis interpretation, thanks to the next result.

Theorem 7. [27, Theorem 6.4.2] *Using GBR when $\underline{P}(z) > 0$ to determine the value $\underline{P}(f|z)$ is equivalent to applying Bayes' rule to each element of $\mathcal{M}(\underline{P})$ in order to produce the set of posteriors (note that when $\underline{P}(z) > 0$ the conditioning event has positive probability for all the elements of $\mathcal{M}(\underline{P})$).*

This rule allows us to introduce the notion of compatibility of a conditional model:[14]

Definition 7. Let $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ be a number of separately coherent conditional lower previsions with domains $\mathcal{H}^j \subseteq \mathcal{K}^j$ for $j = 1, \dots, m$, and let $\underline{P}$ be a coherent lower prevision on $\mathcal{L}(\mathcal{X}^n)$. We say that $\underline{P}$ is a *compatible* joint lower prevision (or a *compatible joint*, for short) for $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ when for every $j \in \{1, \dots, m\}, z_j \in \mathcal{X}_{I_j}$ s.t. $\underline{P}(X_{I_j} = z_j) > 0$ and $f_j \in \mathcal{H}^j$, $\underline{P}_j(f_j|z_j)$ can be derived from $\underline{P}$ by the Generalised Bayes Rule.

It is interesting for the purposes of this paper to note that the weak coherence of a number of conditional lower previsions is related to the existence of a compatible joint. This matter has been investigated in [23, Section 8], and we shall come back to this in Section 4 of this paper.

In this paper, we shall also use another possibility for defining conditional lower previsions in a coherent way, called the regular extension.

Definition 8. Given a set $\mathcal{M}$ of linear previsions and disjoint O, I, $O \neq \emptyset$, the *regular extension* $\underline{R}(X_O|X_I)$ is given by

$$\underline{R}(f|z) := \inf \left\{ \frac{P(f\pi_I^{-1}(z))}{P(z)} : P \in \mathcal{M}, P(z) > 0 \right\} \tag{3}$$

for every $z \in \mathcal{X}_I, f \in \mathcal{K}_{O\cup I}$. This amounts to applying Bayes' rule to the dominating linear previsions whenever possible (i.e., disregarding the linear previsions that assign zero probability to the conditioning event).

The regular extension has been proposed and used a number of times in the literature as an updating rule [8, 9, 11, 14, 27, 28]. A comparison with natural extension (and hence with GBR) in the finite case has been made in [20].

2.6. **Extensions to larger domains.** We focus now on extending conditional lower previsions to larger domains using only their coherent implications.

Definition 9. Let $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions with domains $\mathcal{H}^j \subseteq \mathcal{K}^j$ for $j = 1, \dots, m$ and avoiding partial

[14]This notion coincides with that anticipated in the Introduction, despite the slightly different form.

loss. Their *natural extensions* to $\mathcal{K}^1,\dots,\mathcal{K}^m$ are defined, for every $f \in \mathcal{K}^j$ and every $z_j \in \mathcal{X}^j$, by

$$\underline{E}_j(f|z_j) := \sup\{\alpha : \exists f_i^k \in \mathcal{H}^i, i=1,\dots,m, k=1,\dots,n_i \text{ s.t.} \left[\sum_{i=1}^{m}\sum_{k=1}^{n_i} G_i(f_i^k|X_{I_i}) - \pi_{I_j}^{-1}(z_j)(f-\alpha)\right] < 0 \text{ on } \mathbb{S}(f_i^k)\cup\pi_{I_j}^{-1}(z_j)\}, \quad (4)$$

where $\mathbb{S}(f_j^k) = \cup_{j=1}^m \cup_{k=1}^{n_j} S_j(f_j^k)$, as usual.

Theorem 8. [21, Proposition 11] *When all the conditioning spaces are finite, the natural extensions are the smallest conditional lower previsions which are coherent and dominate $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ on their domains. Moreover, they coincide with the initial assessments if and only if $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are themselves coherent. Otherwise, they 'correct' the initial assessments taking into account the implications of the notion of coherence.*

In other words, the natural extension is a procedure that yields the strongest least-committal implications of the original assessments that follow from considerations of coherence alone. In Section 4 we shall investigate the counterpart of this notion when we focus instead on the property of weak coherence.

3. Characterising avoiding uniform sure loss and avoiding partial loss

At this point, we start providing results in the theory of coherent lower previsions that help us connect it to conditional models. We focus initially on the two consistency notions called avoiding uniform sure loss and avoiding partial loss.

Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions with respective domains $\mathcal{H}^1,\dots,\mathcal{H}^m$, where $\mathcal{H}^j$ is a (not necessarily linear) subset of the class $\mathcal{K}^j$ of $\mathcal{X}_{O_j\cup I_j}$-measurable gambles.

Our first result shows that the notion of avoiding uniform sure loss is equivalent to the existence of dominating weakly coherent linear previsions. It is an extension of [23, Proposition 5] to arbitrary domains. This result is particularly important for the traditional approaches to conditional models, because, as it will become clearer when we characterise weak coherence in Section 4, it shows that avoiding uniform sure loss is equivalent to the existence of a set of joint mass functions compatible with the credal sets of conditional mass functions associated to the conditional lower previsions.

Proposition 2. *$\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ avoid uniform sure loss if and only if there are dominating weakly coherent conditional linear previsions with domains $\mathcal{K}^1,\dots,\mathcal{K}^m$.*

This result will be used in Section 4 when we study the smallest dominating weakly coherent lower previsions (which will be showed to be equivalent to the set of joint mass functions compatible with $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$). It follows that avoiding uniform sure loss is a necessary and sufficient condition for the existence of such lower previsions. Since moreover we shall prove in Lemma 2 that weak coherence is preserved by taking lower envelopes when all the referential spaces are finite (as it is always the case in this paper), we deduce that a way of computing the smallest dominating weakly coherent lower previsions is to take the lower envelopes of the (non-empty) sets of weakly coherent dominating conditional linear previsions.

We have already argued that avoiding uniform sure loss (and its counterpart, that is, weak coherence) is not able to properly deal with all inconsistencies. For this reason, we focus next on the stronger notion of avoiding partial loss (and hence on the related notion of strong coherence). It follows from [27, Section 8.1] that when all the referential spaces are finite and the domains are linear spaces, the notion of avoiding partial loss is equivalent to the existence of dominating coherent linear conditional previsions. For the sake of completeness, we give an explicit proof of this result and generalise it to non-linear domains. First, we introduce a lemma.

Lemma 1. *Assume that* $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *avoid partial loss, and let* $\underline{E}_1(X_{O_1}|X_{I_1}),\dots,\underline{E}_m(X_{O_m}|X_{I_m})$ *be their natural extensions to* $\mathcal{K}^1,\dots,\mathcal{K}^m$*. Then* $\underline{E}_1(X_{O_1}|X_{I_1}),\dots,\underline{E}_m(X_{O_m}|X_{I_m})$ *are coherent.*

From this lemma, we deduce the following:

Proposition 3. $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *avoid partial loss if and only if there are dominating coherent conditional linear previsions with respective domains* $\mathcal{K}^1,\dots,\mathcal{K}^m$*.*

This result is similar in spirit to that in Proposition 2, and they both show similar connections between their respective loss and coherence notions. On the other hand, the connection of this result with conditional models will be made in Section 5, where we characterise coherence through sequences of sets of joint mass functions.

We provide next another characterisation, which has more of a technical flavor, and where we can find some of the ideas we shall use in our approximation of the natural extension in Section 5.

Proposition 4. $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *avoid partial loss if and only if for any* $\epsilon>0$*,* $f_j^k\in\mathcal{H}^j, j=1,\dots,m$*,* $k=1,\dots,n_j$ *such that not all* f_j^k *are zero gambles, it holds that*

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+\epsilon R(f_j^k)S_j(f_j^k)\right](x)>0, \tag{5}$$

where $R(f_j^k):=\max f_j^k-\min f_j^k$ *is the* range *of the gamble* f_j^k*.*

Hence, by introducing these ϵ-terms, we can replace the maximum on the union of the supports with a maximum on $\mathcal{X}^n$.

Note also that Equation (5) is equivalent to requiring that

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j}\lambda_j^k(G_j(f_j^k|X_{I_j})+\epsilon R(f_j^k)S_j(f_j^k))\right](x)>0, \tag{6}$$

for any $\epsilon>0$, $f_j^k\in\mathcal{H}^j,\lambda_j^k\geq 0, j=1,\dots,m$, $k=1,\dots,n_j$ such that not all $\lambda_j^k f_j^k$ are zero gambles: it suffices to note that the gambles $\lambda_j^k f_j^k$ belong to $\mathcal{H}^j$ for all $j=1,\dots,m$, $k=1,\dots,n_j$ because of Remark 2, and that

$$\lambda_j^k(G_j(f_j^k|X_{I_j})+\epsilon R(f_j^k)S_j(f_j^k))=G_j(\lambda_j^k f_j^k|X_{I_j})+\epsilon R(\lambda_j^k f_j^k)S_j(\lambda_j^k f_j^k)).$$

Expression (6) will be useful later (to prove Proposition 10, stated in Section 5) to relate Proposition 4 to the weak coherence of some approximations of our conditional lower previsions.

4. Extensions of weakly coherent conditionals

We focus next on the notion of weak coherence of a number of conditional lower previsions. The results in this section are particularly important to relate coherent lower previsions to the traditional approaches to conditional models. The reason is, as it follows from Theorem 9, that weak coherence is equivalent to the existence of a set of joint mass functions from which the conditional mass functions can be recovered through Generalised Bayes Rule. Moreover, using this set of joint mass functions to make new inferences is equivalent to a procedure called weak natural extension in Theorem 10. That this procedure is sub-optimal will be shown in Section 5.

We begin by providing a characterisation of weak coherence and determining the smallest (unconditional) coherent lower prevision which is weakly coherent with a number of conditionals. This is an extension of [20, Theorem 3] to arbitrary domains:

Theorem 9. *Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions with domains $\mathcal{H}^1,\dots,\mathcal{H}^m$. The following are equivalent:*

(WC1) They are weakly coherent.

(WC2) There is a coherent lower prevision $\underline{P}$ on $\mathcal{L}(\mathcal{X}^n)$ which is weakly coherent with them.

(WC3) There is a coherent lower prevision $\underline{P}$ on $\mathcal{L}(\mathcal{X}^n)$ which is pairwise coherent with them.

Moreover, the smallest coherent lower prevision in (WC2) and (WC3) is given by

$$\underline{P}(f) := \sup\{\alpha : \exists f_j^k \in \mathcal{H}^j, j=1,\dots,m, k=1,\dots,n_j \text{ s.t. } \max_{x\in\mathcal{X}^n}[\sum_{j=1}^m\sum_{k=1}^{n_j} G(f_j^k|X_{I_j}) - (f-\alpha)](x) < 0\} \quad (7)$$

for any gamble f on $\mathcal{X}^n$.

The connection between weak coherence and conditional models is given in this theorem by (WC3). In fact, we know from Theorem 6 that an unconditional lower prevision, i.e., a set of unconditional joint mass functions, is coherent with a conditional one if and only if they are related to each other via the Generalised Bayes Rule. Therefore what (WC3) is saying is that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent if and only if there is an unconditional lower prevision from which they can be obtained through GBR. Moreover, the least-committal such a lower prevision is given by Expression (7).

We can also give the following characterisation, which extends the envelope theorem in [20, Theorem 2] to arbitrary domains.

Lemma 2. *$\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent if and only if they are the lower envelopes of a class of weakly coherent conditional linear previsions $\{P_1^\lambda(X_{O_1}|X_{I_1}),\dots,P_m^\lambda(X_{O_m}|X_{I_m}) : \lambda\in\Lambda\}$ with domains $\mathcal{K}^1,\dots,\mathcal{K}^m$.*

This means that if we take a number of precise-probability conditional models, each one admitting a compatible joint, by taking their lower envelope we obtain an imprecise-probability conditional model that has a compatible joint. And conversely, an imprecise-probability conditional model that has a compatible joint, must be the lower envelope of precise-probability conditional models that admit a compatible joint.

We summarise the relationships between the different consistency conditions analysed so far in the following figure.

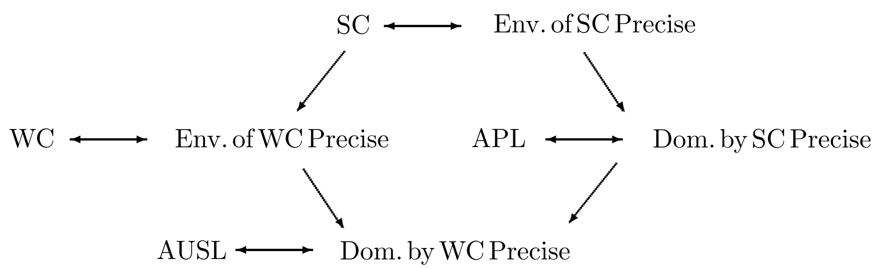

FIGURE 1. Equivalences and implications between consistency concepts analysed in the paper. Keys: SC = strongly coherent; WC = weakly coherent; AUSL = avoiding uniform sure loss; APL = avoiding partial loss; Env. = envelope; Dom. = dominated.

It is useful at this point to compare the functional $\underline{P}$ defined in Equation (7) with the unconditional natural extension $\underline{E}$ of $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$, that we should define using Equation (4). In order to do this, we should consider $O_{m+1} := \{1,\dots,n\}, I_{m+1} := \emptyset$ and add $\underline{P}(X_{O_{m+1}})$ to our set of gambles with the trivial domain given by the constant gambles. For this discussion to make sense, we are going to assume also that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ avoid partial loss and are weakly coherent.

We see from [21, Theorem 12] that in that case the functionals $\underline{P}$ and $\underline{E}$ coincide. In a sense, then, weak and strong coherence lead to the same result when the focus is on deducing an unconditional lower prevision: the unconditional natural extension $\underline{E}$ is the smallest unconditional lower prevision which is weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$; and it is coherent with them if and only if the initial assessments are coherent. This is stated in the following result, which is a trivial consequence of [21, Theorem 4]; its proof is immediate and therefore omitted.

Corollary 1. *Assume that* $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *avoid partial loss and are weakly coherent. The functional* $\underline{P}$ *given by Equation* (7) *is the smallest coherent lower prevision which is coherent with* $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *if and only if these conditional previsions are coherent.*

Given $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ with domains $\mathcal{K}^1,\dots,\mathcal{K}^m$, a sufficient condition for their coherence is that $\underline{P}(z_j) > 0$ for all $z_j \in \mathcal{X}_{I_j}$ and for all $j = 1,\dots,m$ [20, Theorem 11].[15] This leads to an important consideration. Suppose that the unconditional lower prevision $\underline{P}$ in Expression (7) assigns positive

[15]On the other hand, in [20, Example 2] we can find an example of assessments which avoid partial loss and are weakly coherent, but are not coherent.

lower probability to every event. Then there is a unique way to deduce conditionals from it through GBR. Given (WC3) in Theorem 9, this means that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ can be obtained exactly as they are through $\underline{P}$.

In other words, in this case the credal sets of conditional mass functions associated to the conditional lower previsions are equivalent in every respect to the credal set associated to an unconditional lower prevision. This is true even for inference, as it follows from Theorem 10 below: the new conditional mass functions that we can deduce applying Bayes' rule to the credal set associated to the joint have to coincide with those that can be obtained from the conditional model through coherence. This is stated more strongly also in Proposition 11, which we shall establish in Section 5. All of this means, in a sense, that one could forget in this case about the sets of conditional mass functions associated to the conditional lower previsions and just focus on the set of mass functions associated to their compatible joint. This shows that the key difference between conditional models and models based on joint mass functions is due to the presence of zero probabilities.

Assume that we have a number of weakly coherent conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$, and that given disjoint O_{m+1},I_{m+1}, we want to determine the smallest conditional lower prevision $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ which is weakly coherent with the rest. Our next result shows that it suffices to go through the unconditional lower prevision given by Equation (7):

Theorem 10. *Assume that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ avoid partial loss and are weakly coherent. The smallest $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ with domain $\mathcal{K}^{m+1}$ which is weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ is given by*

$$\underline{P}_{m+1}(f|z_{m+1}) := \begin{cases} \min_{x\in\pi^{-1}_{I_{m+1}}(z_{m+1})} f(x) & \text{if } \underline{P}(z_{m+1})=0 \\ \min\{P(f|z_{m+1}) : P\in\mathcal{M}(\underline{P})\} & \text{otherwise,} \end{cases} \tag{8}$$

where $\underline{P}$ is the coherent lower prevision defined in Equation (7).

Recall that $\mathcal{M}(\underline{P})$ is the set of linear previsions that dominate $\underline{P}$ in its domain (in this case, all of $\mathcal{L}(\mathcal{X}^n)$). The conditional prevision $P(f|z_{m+1})$ is given by $P(f|z_{m+1}) = \frac{P(f\pi^{-1}_{I_{m+1}}(z_{m+1}))}{P(z_{m+1})}$: note that it is possible to do so because in that part of Equation (8) we are assuming that $\underline{P}(z_{m+1})>0$, which guarantees that $P(z_{m+1})$ is also greater than 0 for every linear prevision P that dominates $\underline{P}$. We shall refer to the functional in Equation (8) as the *weak natural extension* of $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$.

Despite the apparent unfamiliar form, the weak natural extension is perhaps one of the most used procedures for probabilistic inference. It consists in building a joint out of a number of probabilistic assessments and updating it whenever the conditioning event has positive (lower) probability. In case of zero probability, the weak natural extension leads to a vacuous conditional, while it is perhaps more common in probability to simply let the conditional be undefined. The difference is not so much if we consider that the vacuous lower previsions is a model of ignorance, whence what the weak natural extension states in this case is that we have no information a posteriori. On the other hand, the vacuous model creates conditionals that are fully within the theory of coherent lower previsions, and thus it does not lead to undefined objects. All of this shows in particular that deducing new conditional probabilities or expectations from conditional models through the set of compatible joints and GBR, is basically what we call the weak natural extension.

Our result stresses in addition that the weak coherence of a number of conditional lower previsions can be summarised by an unconditional lower prevision.

Now, let us consider what happens if instead of deducing a new conditional through the weak natural extension above, we use the procedure of natural extension.

Proposition 5. *Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be weakly coherent conditional lower previsions with domains $\mathcal{H}^1,\dots,\mathcal{H}^m$, and let $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ be defined on $\mathcal{K}^{m+1}$ by Equation (4). Then*

$\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m}),\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ *are weakly coherent.*

In particular, if $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are coherent, it follows from the results in [21] that $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is the smallest conditional lower prevision that is coherent with them.

However, $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ may be strictly greater than the conditional lower prevision $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ derived in Equation (8). In other words, the smallest conditional lower prevision which is *weakly coherent* with the initial assessments may be strictly smaller than the smallest conditional lower prevision which is *coherent* with the initial assessments. This is a crucial point: it means that the inferences that we do from a conditional model through the set of compatible joint mass functions are generally not strong enough; we can obtain stronger conclusions, while still working on the side of least-committal inference, by the procedure of natural extension.

In the next example we show that the natural and the weak natural extensions may differ completely also when the initial lower previsions are coherent, and that this may happen even when we consider linear previsions.

Example 3. Consider X_1, X_2 taking values in $\mathcal{X} := \{1, 2\}$. Define linear previsions $P_1(X_1), P_2(X_2|X_1)$ through the assessments $P_1(X_1 = 1) := 1, P_2(X_2 = 1|X_1 = 1) := 0.5, P_2(X_2 = 1|X_1 = 2) := 1$. These two previsions are coherent because they satisfy the hypotheses of the Marginal Extension Theorem (see [27, Theorem 6.7.2] and [22, Theorem 4] for its formulation in terms of variables). Moreover, it follows from [27, Theorem 6.7.3] that there is a unique joint coherent with them: the linear prevision P given by $P := P_1(P_2(X_2|X_1))$. Since in this case $P_1(X_1), P_2(X_2|X_1)$ are weakly coherent if and only if they are coherent, we deduce that P coincides with the functional defined in Theorem 9. This joint satisfies $P(X_1 = 2) = P_1(X_1 = 2) = 0$.

As a consequence, the weak natural extension $\underline{P}(X_2|X_1 = 2)$ that we can derive using Theorem 10 is the vacuous lower prevision for X_2 conditional on $X_1 = 2$. On the other hand, it follows from the coherence of $P_1(X_1), P_2(X_2|X_1)$ that the procedure of natural extension yields back the original linear prevision $P_2(X_2|X_1 = 2)$. Hence, in the former case we are left with no information at all about X_2 conditional on $X_1 = 2$. In the latter, we are certain that $X_2 = 1$. ⧫

Note that in this example we have the counter-intuitive property that two different conditional lower previsions $P_2(X_2|X_1), \underline{P}(X_2|X_1)$ can be weakly coherent; this cannot happen with the stronger notion of coherence.

This example shows on the one hand that the notion of weak coherence is indeed too weak to fully capture the implications of our assessments, and on the other that the natural extension cannot be derived in general from the unconditional lower prevision $\underline{P}$. In the following section, we get around this problem by showing: (i) that we can instead derive it using a *sequence* of unconditional lower previsions that converges to $\underline{P}$ and (ii) that in some cases it coincides with the weakly coherent natural extension.

5. Natural extension as a limit of regular extensions

Let $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be separately coherent conditional lower previsions with domains $\mathcal{H}^j \subseteq \mathcal{K}^j$ for $j = 1,\dots,m$. For the time being we shall assume that they are weakly coherent and avoid partial loss, but they are not necessarily coherent. Our goal in this section is to characterise their natural extension $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ to the set $\mathcal{K}^{m+1}$ of $\mathcal{X}_{O_{m+1}\cup I_{m+1}}$-measurable gambles, given by Equation (4) (in order to use this equation it suffices to include among the original assessments a conditional lower prevision $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ defined on a trivial domain, such as the set of gambles piece-wise constant on the elements of the partition induced by the conditioning events). This is important because we have seen in Example 3 that we need to focus on natural extensions to obtain strongest least-committal inferences. But at present, we do not have an expression of the natural extension in terms of a set of joint mass functions. And we need this to tightly connect coherent lower previsions with models based on conditional mass functions. This is what we set out to do in the following.

In particular, we shall prove that the natural extension can be computed as a limit of regular extensions. Remember that the regular extension is the application of Bayes' rule, whenever possible, to the set of mass functions in the credal set associated to the joint unconditional lower prevision. Therefore, what we are going to prove is that the inferences done from a conditional model can always be regarded as done through a sequence of sets of unconditional joint mass functions through Bayes' rule. In order to do this, we are going to consider a sequence of sets of mass functions (credal sets) associated to conditional lower previsions that converge pointwise to $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$. How this is done is detailed in the following.

5.1. Definition of the approximating sets.

For every $\epsilon > 0$, let $\mathcal{M}(\epsilon)$ be the set of linear previsions satisfying the constraints[16]

$$P(f_j|z_j) \geq \underline{P}_j(f_j|z_j) - \epsilon R(f_j) \tag{9}$$

for every $f_j \in \mathcal{H}^j, z_j \in \mathcal{X}_{I_j}, j = 1,\dots,m$ such that $P(z_j) > 0$, where $R(f_j) := \max f_j - \min f_j$ is the range of the gamble f_j. Let us also consider the set of gambles

$$\mathcal{V}_\epsilon := \{f \geq \sum_{j=1}^{m}\sum_{k=1}^{n_j} \lambda_j^k (G_j(f_j^k|X_{I_j}) + \epsilon R(f_j^k) S_j(f_j^k)) \text{ for some } f_j^k \in \mathcal{H}^j, \lambda_j^k \geq 0, j = 1,\dots,m, k = 1,\dots,n_j\}, \tag{10}$$

where, with a certain abuse of notation, $S_j(f_j^k)$ is used to denote also the indicator function of the set $S_j(f_j^k)$.

[16] An equivalent formulation of the constraints, which we shall sometimes use in the proofs, is the following: $P(f_j \pi_{I_j}^{-1}(z_j)) \geq P(z_j)(\underline{P}_j(f_j|z_j) - \epsilon R(f_j))$ for every $f_j \in \mathcal{H}^j, z_j \in \mathcal{X}_{I_j}, j = 1,\dots,m$.

For $\epsilon = 0$ we obtain the set $\mathcal{M}(0)$ of linear previsions satisfying

$$P(f_j|z_j) \geq \underline{P}_j(f_j|z_j)\ \forall f_j \in \mathcal{H}^j, z_j \in \mathcal{X}_{I_j} \text{ s.t. } P(z_j) > 0, j = 1, \ldots, m \tag{11}$$

and the set of gambles

$$\mathcal{V} := \{f \geq \sum_{j=1}^{m} \sum_{k=1}^{n_j} \lambda_j^k G_j(f_j^k|X_{I_j}) \quad \text{for some } f_j^k \in \mathcal{H}^j, \lambda_j^k \geq 0, j = 1, \ldots, m, k = 1, \ldots, n_j\}.$$

It follows from their definition that $\mathcal{V}_\epsilon \subseteq \mathcal{V}$ and $\mathcal{M}(0) \subseteq \mathcal{M}(\epsilon)$ for any $\epsilon > 0$. Since the gamble constant on 0 belongs to $\mathcal{V}_\epsilon$ for all $\epsilon > 0$, we deduce that these sets of gambles are non-empty. On the other hand, $\mathcal{M}(\epsilon)$ is a convex set of linear previsions for all $\epsilon > 0$. $\mathcal{M}(0)$ (and therefore also $\mathcal{M}(\epsilon)$ for all $\epsilon > 0$) is non-empty because the conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent. This follows from the following proposition, where we show that a linear prevision P belongs to $\mathcal{M}(0)$ if and only if it dominates the coherent lower prevision $\underline{P}$ given by Equation (7). Let $\underline{P}_\epsilon$ be the lower envelope of $\mathcal{M}(\epsilon)$, and $\underline{P}_0$ the lower envelope of $\mathcal{M}(0)$.

Proposition 6. *Let $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$ be weakly coherent conditional lower previsions with respective domains $\mathcal{H}^1, \ldots, \mathcal{H}^m$ and avoiding partial loss.*

(a) For any $\epsilon \geq 0$, $\mathcal{M}(\epsilon) = \{P : P(f) \geq 0\ \forall f \in \mathcal{V}_\epsilon\}$.
(b) $\mathcal{M}(0) = \cap_{\epsilon>0}\mathcal{M}(\epsilon) = \mathcal{M}(\underline{P})$, where $\underline{P}$ is the coherent lower prevision given by Equation (7).
(c) $\underline{P}_0 = \sup_{\epsilon>0} \underline{P}_\epsilon = \underline{P}$.

Some comments here can clarify the situation with respect to conditional models. We said already that the coherent lower prevision $\underline{P}$ given by Equation (7) represents the set of all joint mass functions compatible with the conditional model. This is made clear by point (b) of the proposition, as it shows that $\underline{P}$ is just the set of mass functions determined by the linear constraints in (11). Moreover, point (c) shows that such a lower prevision can be obtained as a limit of the ϵ-approximations $\underline{P}_\epsilon$.

Now we focus on the ϵ-approximations by providing a number of technical results that we shall need to prove our main result in Theorem 11. Let us first establish a one-to-one correspondence between the set $\mathcal{M}(\epsilon)$ of linear previsions and the closure of the set of gambles $\mathcal{V}_\epsilon$:

Proposition 7. *For every $\epsilon \geq 0$, $\{f \in \mathcal{L}(\mathcal{X}^n) : P(f) \geq 0\ \forall P \in \mathcal{M}(\epsilon)\} = \{f : f + \delta \in \mathcal{V}_\epsilon\ \forall \delta > 0\} = \overline{\mathcal{V}}_\epsilon$, where the closure is taken in the topology of uniform convergence.*

In the particular case of precise assessments (i.e., conditional linear previsions) we can go a bit further. In this case, and in analogy with the situation in the unconditional case, we can show that events provide all the information we need. Note also that in the linear case the notion of avoiding partial loss is equivalent to coherence (and implies therefore weak coherence); this is why it is our only requirement in the following proposition:

Proposition 8. *Let $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ be coherent conditional linear previsions with domains $\mathcal{K}^1,\dots,\mathcal{K}^m$. Define $\mathcal{M}(\epsilon),\mathcal{V}_\epsilon$ by Equations (9), (10), respectively. Let $\mathcal{M}^A_\epsilon,\mathcal{V}^A_\epsilon$ be the corresponding sets determined by the restrictions of $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ to events.*

(1) *For every $\epsilon>0$, $\overline{\mathcal{V}}_\epsilon\subseteq\mathcal{V}^A_{\epsilon_1}$, where $\epsilon_1:=\frac{\epsilon}{\max_j|\mathcal{X}_{O_j}|}$, and as a consequence $\cup_\epsilon\mathcal{V}_\epsilon=\cup_\epsilon\mathcal{V}^A_\epsilon=\cup_\epsilon\overline{\mathcal{V}}_\epsilon$.*
(2) *$\mathcal{M}(\epsilon)\supseteq\mathcal{M}^A_{\epsilon_1}$, whence $\mathcal{M}(0)=\mathcal{M}^A_0$.*

This result will be very useful for us because it allows us to connect our results with the ones established in [29] for the particular case of conditional lower previsions defined on events. The case of events is also interesting because the corresponding sets of gambles are *finitely generated* (i.e., they can be summarised by a finite number of gambles), and this makes it easier to apply separation results.

5.2. **Convergence result.** Now that we have clarified a bit the structure of the sets $\mathcal{M}(\epsilon),\mathcal{V}_\epsilon$, we explore how they can be used to characterise the conditional natural extension. A first result is given in the following proposition:

Proposition 9. *Consider $f\in\mathcal{K}^{m+1}$ and $z_{m+1}\in\mathcal{X}_{I_{m+1}}$. Then*

$$\underline{E}_{m+1}(f|z_{m+1})=\sup\{\mu:\pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu)\in\cup_\epsilon\mathcal{V}_\epsilon\}$$
$$\leq\sup\{\mu:\pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu)\in\mathcal{V}\}.$$

Now, let us define the functional that we obtain by applying the regular extension to the generic ϵ-approximating credal set $\mathcal{M}(\epsilon)$, with $\epsilon>0$:

$$\underline{P}^\epsilon_{m+1}(f|z_{m+1}):=\inf\{P(f|z_{m+1}):P\in\mathcal{M}(\epsilon),P(z_{m+1})>0\}.\tag{12}$$

What we should like to show is that the limit of these functionals when ϵ goes to zero coincides with $\underline{E}_{m+1}(f|z_{m+1})$. To this end, the first thing we have to prove is that the definition of $\underline{P}^\epsilon_{m+1}(f|z_{m+1})$ makes sense, i.e., that for every $\epsilon>0$ and every $z_{m+1}\in\mathcal{X}_{I_{m+1}}$ there is some $P\in\mathcal{M}(\epsilon)$ such that $P(z_{m+1})>0$. This is established in the following proposition:

Proposition 10. *For every $z_{m+1}\in\mathcal{X}_{I_{m+1}}$ and every $\epsilon>0$, there is some $P\in\mathcal{M}(\epsilon)$ such that $P(z_{m+1})>0$.*

Since the set $\mathcal{M}(\epsilon)$ does not increase as ϵ converges to zero, we deduce that the conditional lower previsions $\underline{P}^\epsilon_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ given by Equation (12) do not decrease as ϵ goes to zero. We can thus consider

$$\underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}):=\lim_{\epsilon\to0}\underline{P}^\epsilon_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}),$$

the limit of these conditional lower previsions. In analogy with Proposition 9, we can characterise this functional in terms of sets of gambles:

Lemma 3. *$\underline{F}_{m+1}(f|z_{m+1})=\sup\{\mu:\pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu)\in\cup_\epsilon\overline{\mathcal{V}}_\epsilon\}$ for every gamble f in $\mathcal{K}^{m+1}$ and every $z_{m+1}\in\mathcal{X}_{I_{m+1}}$. As a consequence, $\underline{F}(f|z_{m+1})\geq\underline{E}(f|z_{m+1})$.*

Since the sets $\mathcal{V}_\epsilon$ are not necessarily closed, we may wonder if the functional $\underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ defined as a limit of regular extensions is actually more precise that the natural extension $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. In our next result, we show that this is not the case. The proof is based on applying Proposition 8 to obtain the result for linear previsions, and then use envelope results.

Theorem 11. *Assume that* $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *are weakly coherent and avoid partial loss. Their natural extension* $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ *coincides with* $\underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$.

This theorem is our main result. It shows that strongest least-committal inferences (i.e., natural extensions) from a conditional model can be made equivalently through conditioning, whenever possible, the set $\mathcal{M}(\epsilon)$, and taking the limit.

This result is valid in particular if $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are coherent. We can also determine, as a corollary, that the conditional lower prevision derived from an unconditional by natural extension is also the limit of conditional lower previsions obtained by regular extension. In this particular case our sets $\mathcal{M}(\epsilon),\epsilon\geq 0$ would be

$$\mathcal{M}(\epsilon):=\{P:P(f)\geq\underline{P}(f)-\epsilon R(f)\ \forall f\in\mathcal{H}\}.$$

Note also that in this case where we have a conditional and an unconditional lower prevision only, weak and strong coherence are equivalent, and therefore we do not need to distinguish between the weak natural extension and the natural extension:

Corollary 2. *Let* $\underline{P}$ *be a coherent lower prevision with domain* $\mathcal{H}$, *and consider disjoint* O,I, *with* $O\neq\emptyset$. *For every* $\epsilon>0$, *let* $\mathcal{M}(\epsilon)$ *be given by Equation* (9) *and let* $\underline{P}_\epsilon(X_O|X_I)$ *be the conditional lower prevision defined from* $\mathcal{M}(\epsilon)$ *using regular extension. Then* $\lim_{\epsilon\to 0}\underline{P}_\epsilon(X_O|X_I)$ *coincides with the conditional natural extension* $\underline{E}(X_O|X_I)$.

At this point we may still be wondering if going through the sets $\mathcal{M}(\epsilon)$ is really necessary, or if we could have applied the regular extension to the set $\mathcal{M}(0)$ given by Equation (11) and use it to approximate $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. This is not possible in general, because Proposition 10 does not necessarily hold for $\epsilon=0$, i.e., there may not be any $P\in\mathcal{M}(0)$ such that $P(z_{m+1})>0$, and therefore we may not be able to use the regular extension in that case; this is easy to see with precise assessments. Moreover, even if we can apply regular extension in $\mathcal{M}(0)$, we do not necessarily have the equality $\underline{E}_{m+1}(f|z_{m+1})=\inf\{P(f|z_{m+1}):P\in\mathcal{M}(0),P(z_{m+1})>0\}$. This is discussed for the particular case of lower probabilities in [29, Sections 3.7,3.8], and some illustrative examples are provided.

Hence, the inequality given in Proposition 9 is not necessarily an equality. In the following result, we show that a sufficient condition for the equality to hold is that the lower probability of the conditioning event is positive; see also [27, Theorem 8.1.4]:

Proposition 11. *Let* $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ *be weakly coherent conditional lower previsions that avoid partial loss. Let* $\underline{P}$ *be their unconditional natural extension, given by Equation* (7), *and consider the conditional lower prevision* $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ *given by Equation* (8). *If* $\underline{P}(z_{m+1})>0$, *then*

$$\underline{E}_{m+1}(f|z_{m+1})=\underline{P}_{m+1}(f|z_{m+1})=\sup\{\mu:\pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu)\in\mathcal{V}\}.$$

Hence, in this case the natural extension is also the smallest conditional lower prevision that is weakly coherent with the initial assessments. In particular, if $\underline{P}(z_{m+1})>0$ for all $z_{m+1}\in\mathcal{X}_{I_{m+1}}$, we deduce the equality

$$\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})=\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}).$$

The intuition here is that in that case $\underline{P}_\epsilon(z_{m+1}) > 0$ for all $z_{m+1} \in \mathcal{X}_{I_{m+1}}$ and for ϵ small enough, and then the regular extension from $\mathcal{M}(\epsilon)$ coincides with the natural extension. From here it suffices then to apply a limit result.

This result complements the discussion we have made in Section 4 about the positivity of the conditioning events. What we have, essentially, is that the differences between weak and strong coherence, as well as between the weak and strong natural extensions, vanish in the case of positivity of the conditioning events involved in the inference.

Finally, we are going to show that our results allow to derive a characterisation of the notion of coherence for conditional lower previsions on finite spaces. In order to this, we need to establish first the following lemma:

Lemma 4. *Let* $\{\underline{P}_1^k(X_{O_1}|X_{I_1}), \dots, \underline{P}_m^k(X_{O_m}|X_{I_m})\}_{k\in\mathbb{N}}$ *be a sequence of conditional lower previsions with domains* $\mathcal{H}^1, \dots, \mathcal{H}^m$. *Assume their pointwise limits* $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ *exist.*

(1) *If* $\underline{P}_1^k(X_{O_1}|X_{I_1}), \dots, \underline{P}_m^k(X_{O_m}|X_{I_m})$ *are weakly coherent for all* k*, then so are* $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$.
(2) *If* $\underline{P}_1^k(X_{O_1}|X_{I_1}), \dots, \underline{P}_m^k(X_{O_m}|X_{I_m})$ *are coherent for all* k*, then so are* $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$.

We deduce the following:

Theorem 12. *Let* $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ *be separately coherent conditional lower previsions. They are coherent if and only if they are the pointwise limits of a sequence of coherent conditional lower previsions defined by regular extension.*

Hence, in the case of finite spaces the notion of coherence, which, as we have argued, is the central consistency notion in Walley's theory, is equivalent to the approximation by means of regular extensions. This result summarises the core of our research in this paper: it states that a conditional model is strongly coherent if and only if it can be regarded as the output of a sequence of sets of unconditional mass functions obtained using Bayes' rule. This, together with Theorem 11, helps clarify the tight relationship between conditional models and the more traditional models based on joint mass functions.

5.3. **An example.** It is useful to remark that the tools developed in this paper can be used operationally to check the consistency of a conditional model. The next example shows how this can be done in practice.

Example 4. Consider two random variables X_1, X_2 taking values in the finite space $\mathcal{X} := \{1, 2, 3\}$. We focus on a conditional model made of two sets of conditional mass functions represented through the conditional lower previsions $\underline{P}(X_2|X_1)$ and $\underline{P}(X_1|X_2)$.

We define such mass functions as follows. We let $P(X_2|X_1 = 1), P(X_2|X_1 = 2)$ be precise and determined by the respective assessments $P(X_2 = 1|X_1 = 1) := 1$ and $P(X_2 = 3|X_1 = 2) := 1$. In case $X_1 = 3$, we take $\underline{P}(X_2|X_1 = 3)$ to be determined by the set of mass functions that assign probability one to the subset $\{2, 3\}$ of $\mathcal{X}$. As for $\underline{P}(X_1|X_2)$, we let $P(X_1|X_2 = 1)$ be precise and determined by $P(X_1 = 2|X_2 = 1) := 1$. Each of the remaining two cases, $P(X_1|X_2 = 2)$ and $P(X_1|X_2 = 3)$, is taken to correspond to the set of all mass functions for X_1.

Our goal is to check whether or not $\underline{P}(X_2|X_1)$ and $\underline{P}(X_1|X_2)$ are (strongly) coherent. To this end, let us first re-formulate $\underline{P}(X_2|X_1)$ and $\underline{P}(X_1|X_2)$ in the formalism of coherent lower previsions. It is enough to write that for any gamble f in $\mathcal{L}(\mathcal{X}^2)$, we let

$$\begin{aligned}
\underline{P}(f|X_1 = 1) &:= f(1,1)\\
\underline{P}(f|X_1 = 2) &:= f(2,3)\\
\underline{P}(f|X_1 = 3) &:= \min\{f(3,2), f(3,3)\}\\
\underline{P}(f|X_2 = 1) &:= f(2,1)\\
\underline{P}(f|X_2 = 2) &:= \min\{f(1,2), f(2,2), f(3,2)\}\\
\underline{P}(f|X_2 = 3) &:= \min\{f(1,3), f(2,3), f(3,3)\}.
\end{aligned}$$

Now, let us consider the unconditional lower prevision $\underline{P}$ on $\mathcal{L}(\mathcal{X}^2)$ given by $\underline{P}(f) := \min\{f(3,2), f(3,3)\}$. This corresponds to the set of mass functions that assign probability one to the subset $\{(3,2),(3,3)\}$ of $\mathcal{X}^2$. Using Theorem 9, $\underline{P}, \underline{P}(X_1|X_2)$ and $\underline{P}(X_2|X_1)$ are weakly coherent. In other words, the lower prevision $\underline{P}$ is compatible with $\underline{P}(X_2|X_1), \underline{P}(X_1|X_2)$ through GBR.

To see that $\underline{P}(X_1|X_2)$ and $\underline{P}(X_2|X_1)$ avoid partial loss, we apply Proposition 3 and consider the dominating conditional linear previsions $P(X_1|X_2), P(X_2|X_1)$ given by

$$\begin{aligned}
P(f|X_1 = 1) &:= f(1,1)\\
P(f|X_1 = 2) &:= f(2,3)\\
P(f|X_1 = 3) &:= f(3,3)\\
P(f|X_2 = 1) &:= f(2,1)\\
P(f|X_2 = 2) &:= f(3,2)\\
P(f|X_2 = 3) &:= f(3,3),
\end{aligned}$$

for any gamble f in $\mathcal{L}(\mathcal{X}^2)$. It is not difficult to show that these conditional previsions are coherent.

Let us use now our results to show that $\underline{P}(X_1|X_2), \underline{P}(X_2|X_1)$ are not coherent. Let f be the indicator function of $\{(2,3),(3,3)\}$. In order to show that $\underline{E}(f|X_2 = 3) > \underline{P}(f|X_2 = 3) = 0$, we are going to show that there is some $\epsilon > 0$ such that $\underline{P}^\epsilon(f|X_2 = 3) > 0$. The inequality will follow then from Theorem 11.

Consider $\epsilon := \frac{1}{4}$. Then $\mathcal{M}(\epsilon)$ is the set of linear previsions satisfying the following inequalities for every gamble g on $\{1,2,3\} \times \{1,2,3\}$:

$$\begin{aligned}
P(g|X_1 = 1) &\geq g(1,1) - \epsilon R(g) && \text{if } P(X_1 = 1) > 0\\
P(g|X_1 = 2) &\geq g(2,3) - \epsilon R(g) && \text{if } P(X_1 = 2) > 0\\
P(g|X_1 = 3) &\geq \min\{g(3,2), g(3,3)\} - \epsilon R(g) && \text{if } P(X_1 = 3) > 0\\
P(g|X_2 = 1) &\geq g(2,1) - \epsilon R(g) && \text{if } P(X_2 = 1) > 0
\end{aligned}$$

since the other constraints follow trivially from separate coherence.

Let P be an element of $\mathcal{M}(\epsilon)$ satisfying $P(X_2 = 3) > 0$. If $P(f|X_2 = 3) < \epsilon$, this implies that

$$P(2,3) + P(3,3) < \frac{\epsilon}{1-\epsilon} P(1,3),$$

whence $P(1,3) > 0$. Applying the first of the above constraints to the indicator function of $(1,1)$, we deduce that

$$P(1,2) + P(1,3) \leq \frac{\epsilon}{1-\epsilon} P(1,1),$$

whence $P(1,1) > 0$. Applying the fourth of the above constraints to the indicator function of $(2,1)$, we deduce that

$$P(1,1) + P(3,1) \leq \frac{\epsilon}{1-\epsilon} P(2,1),$$

whence $P(2,1) > 0$. Applying the second of the above constraints to the indicator function of $(2,3)$, we deduce that

$$P(2,1) + P(2,2) \leq \frac{\epsilon}{1-\epsilon} P(2,3),$$

whence $P(2,3) > 0$. But if we consider all these inequalities altogether we observe that

$$0 < P(2,3) \leq \frac{\epsilon}{1-\epsilon} P(1,3) \leq (\frac{\epsilon}{1-\epsilon})^2 P(1,1) \leq (\frac{\epsilon}{1-\epsilon})^3 P(2,1) \leq (\frac{\epsilon}{1-\epsilon})^4 P(2,3),$$

whence $1 \leq \frac{1}{81}$, a contradiction. Hence, there is no $P \in \mathcal{M}(\epsilon)$ such that $P(X_2 = 3) > 0, P(f|X_2 = 3) < \epsilon$, and as a consequence $\underline{P}^\epsilon(f|X_2 = 3) \geq \epsilon > 0 = \underline{P}(f|X_2 = 3)$. Since $\underline{P}^\epsilon(f|X_2 = 3)$ is not decreasing when ϵ goes to zero (because $\mathcal{M}(\epsilon)$ does not increase), we deduce that $\underline{P}(X_1|X_2), \underline{P}(X_2|X_1)$ do not coincide with their natural extensions and therefore they are not coherent. Therefore, despite $\underline{P}(X_1|X_2), \underline{P}(X_2|X_1)$ admit the compatible joint $\underline{P}$ previously defined, they are still incoherent as they cannot be regarded as the outcome of any sequence of sets of joint mass functions obtained through Bayes' rule. ♦

Note that this example also provides an instance where the natural extension $\underline{E}(f|X_2 = 3)$ does not coincide with the weak natural extension given by Equation (8), which in this case is vacuous.

6. Conclusions

In this paper, we have focused on conditional probabilistic models as an alternative to more traditional models based on unconditional joint mass functions. We have expressed these models using the theory of lower previsions by Walley [27], and employed the tools from this theory to investigate two kinds of problems.

The first one is the self-consistency of the assessments. We have approached this problem by studying the different consistency notions for conditional lower previsions in Walley's theory. The most important ones are called weak coherence, which turns out to be equivalent to the existence of a compatible joint, and the stronger notion of coherence, which rules out some inconsistencies that can appear because of the conditioning on sets of zero probability.

The second problem is how to make inferences from a self-consistent conditional model. Again here we have two paths: one is to use the notion of weak coherence and define the so-called weak natural extension, which turns out to be too conservative in some cases. The other one is based on the notion of strong coherence and produces the natural extension which is more informative.

The main result of the paper is that the notion of coherence, and its consequences, can be related to the approximation of our model by a sequence of (possibly sets of) unconditional mass functions: on the one hand, we have showed that the coherence of a number of conditional lower previsions is equivalent to being the limit of a sequence of conditional lower previsions derived applying regular extension to a sequence of unconditional lower previsions. And moreover, this family can also be used to determine the natural extension of the conditional lower previsions to bigger domains.

One important comment here is that throughout we have made no assumptions about the domains of the conditional lower previsions, meaning that they are sets of gambles which are not necessarily linear. Because of this, our results extend the ones established in [29] for conditional lower probabilities, and more generally they can be used when the domains are finite sets of gambles.

These results show that, after all, conditional models are tightly related to the traditional view of probability based on joint mass functions, and that also in these models it is still Bayes' rule that plays a central role to guarantee self-consistency; on the other, they imply that conditional models are more expressive that the latter models, because a sequence of joint mass functions is more informative than a single one.

With respect to future work, we should like to point out three avenues: one is the obvious possibility to try to extend the results presented here to the case of infinite spaces of possibilities. We envisage that most of them will not be immediately extendable because in our proofs we have used a number of separation theorems and envelope results that do not apply directly to the infinite case. Another aspect worth investigating would be how the results presented here generalise in case structural judgments, such as independence, are introduced in a model. Finally, the idea of using a certain sequence of previsions to check coherence and compute extensions, although different from the one presented here, is present also in other works [7, 25] which have a common root in the work of Krauss [17]. It is an open problem to investigate the relationship between the two types of approaches.

Acknowledgements

This work has been partially supported by the projects TIN2008-06796-C04-01, MTM2007-61193 and by the Swiss NSF grants n. 200020-116674/1 and 200020-121785/1. We should also like to thank the referees for their helpful comments.

Appendix: proofs

In this appendix, we have gathered the proofs of the results mentioned in the paper, as well as an auxiliary lemma, that we proceed to establish:

Lemma 5. *Let $\underline{P}, \underline{P}(X_O|X_I)$ be coherent lower previsions with respective domains $\mathcal{L}(\mathcal{X}^n), \mathcal{H}_{O\cup I}$. For every $P \in \mathcal{M}(\underline{P})$ there is some conditional linear prevision $P(X_O|X_I)$ in $\mathcal{M}(\underline{P}(X_O|X_I))$ such that $P, P(X_O|X_I)$ are coherent.*

Proof. Consider $z \in \mathcal{X}_I$. If $P(z) = 0$, we select an arbitrary $P(X_O|z)$ in the set $\mathcal{M}(\underline{P}(X_O|z))$ (that this set is non-empty follows from the separate coherence of $\underline{P}(X_O|X_I)$). If $P(z) > 0$, then we define $P(X_O|z)$ using GBR.

It follows from Theorem 3 in Section 2 that the conditional prevision $P(X_O|X_I)$ thus defined is coherent with P. Let us prove that it belongs to $\mathcal{M}(\underline{P}(X_O|X_I))$. For this, it suffices to show that $P(f|z) \geq \underline{P}(f|z)$ for every $f \in \mathcal{H}_{O\cup I}$ and every $z \in \mathcal{X}_I$. This is trivial if $f = 0$, therefore let us consider $f \neq 0$. Assume ex-absurdo that $P(f|z) < \underline{P}(f|z)$. Then

$$0 = P(\pi_I^{-1}(z)(f - P(f|z)) \geq P(\pi_I^{-1}(z)(f - \underline{P}(f|z))) \geq \underline{P}(\pi_I^{-1}(z)(f - \underline{P}(f|z))) = 0,$$

where the first equality is due to GBR and the following inequality to the monotonicity of P; to see the last equality, apply Equation (1) twice and the coherence of $\underline{P}, \underline{P}(X_O|X_I)$ to deduce that

$$\max_{x\in\mathcal{X}^n}[G(G(f|z)) - G(f|z)](x) = -\underline{P}(G(f|z)) \geq 0$$

on the one hand, and

$$\max_{x\in\mathcal{X}^n}[G(f|z) - G(G(f|z))](x) = \underline{P}(G(f|z)) \geq 0$$

on the other, where $G(G(f|z)) = G(f|z) - \underline{P}(G(f|z))$. But this means that there are two different values of μ such that $P(\pi_I^{-1}(z)(f-\mu)) = 0$, a contradiction with Theorem 6. Hence, $P(f|z) \geq \underline{P}(f|z)$ and thus $P(X_O|X_I) \geq \underline{P}(X_O|X_I)$. □

From the proof of the lemma and the super-additivity of coherent lower previsions, we deduce also the following result, whose proof is omitted:

Corollary 3. *Given coherent $\underline{P}, \underline{P}(X_O|X_I)$ with domains $\mathcal{L}(\mathcal{X}^n), \mathcal{H}_{O\cup I}$, it holds that*

$$\underline{P}(G(f|z)) = 0 \text{ and } \underline{P}(G(f|X_I)) \geq 0$$

for every gamble $f \in \mathcal{H}_{O\cup I}$ and every $z \in \mathcal{X}_I$.

Proof of Proposition 2. Assume there are dominating weakly coherent linear conditional previsions $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$, and let $f_j^k \in \mathcal{H}^j$ for $j = 1,\dots,m$, $k = 1,\dots,n_j$. Then

$$\max_{x\in\mathcal{X}^n}\sum_{j=1}^{m}\sum_{k=1}^{n_j}(f_j^k - \underline{P}_j(f_j^k|X_{I_j})) \geq \max_{x\in\mathcal{X}^n}\sum_{j=1}^{m}\sum_{k=1}^{n_j}(f_j^k - P_j(f_j^k|X_{I_j})) \geq 0,$$

where the second inequality holds because $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ are weakly coherent. Hence, $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ avoid uniform sure loss.

Conversely, assume that $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ avoid uniform sure loss. Take $\mathcal{D} := \{G_j(f_j|X_{I_j}) : j = 1,\dots,m, f_j \in \mathcal{H}^j\}$. Applying [27, Lemma 3.3.2], there is a linear prevision P s.t. $P(G_j(f_j|X_{I_j})) \geq 0$ for all $f_j \in \mathcal{H}^j, j = 1,\dots,m$; applying this property to $g_j := f_j\pi_{I_j}^{-1}(z_j) \in \mathcal{H}^j$, we deduce that $P(G_j(f_j|z_j)) \geq 0$ for all $f_j \in \mathcal{H}^j, z_j \in \mathcal{X}_{I_j}, j = 1,\dots,m$.

Consider $j \in \{1,\dots,m\}, z_j \in \mathcal{X}_{I_j}$, and let us prove the existence of a linear conditional prevision $P_j(X_{O_j}|z_j)$ dominating $\underline{P}_j(X_{O_j}|z_j)$ such that $P(\pi_{I_j}^{-1}(z_j)(f - P_j(f|z_j))) = 0$ for all $f \in \mathcal{K}^j$. There are two possibilities:

- If $P(z_j) > 0$, then $P_j(X_{O_j}|z_j)$ is uniquely determined by P using Bayes' rule, and it is the unique value μ such that $P(\pi_{I_j}^{-1}(z_j)(f-\mu)) = 0$. To see that it dominates $\underline{P}_j(X_{O_j}|z_j)$, assume ex-absurdo the existence of some gamble $f_j \in \mathcal{H}^j$ for which $\underline{P}_j(f_j|z_j) > P_j(f_j|z_j)$. Then it follows that

$$0 \leq P(\pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j))) \leq P(\pi_{I_j}^{-1}(z_j)(f_j - P_j(f|z_j))) = 0,$$

  where the first inequality holds because $P(G(f_j|z_j)) \geq 0$. This is a contradiction with the uniqueness mentioned above.
- If $P(z_j) = 0$ we simply consider any linear conditional prevision $P_j(X_{O_j}|z_j)$ that dominates $\underline{P}_j(X_{O_j}|z_j)$ and it automatically satisfies GBR with P.

By doing this for every $j \in \{1, \dots, m\}$ and every $z_j \in \mathcal{X}_{I_j}$, we obtain a conditional linear prevision $P_j(X_{O_j}|X_{I_j})$ that dominates $\underline{P}_j(X_{O_j}|X_{I_j})$ and that is coherent with P. Applying [23, Theorem 1], we deduce the existence of weakly coherent $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ dominating our conditional lower previsions. □

Proof of Lemma 1. If $\underline{E}_1(X_{O_1}|X_{I_1}), \dots, \underline{E}_m(X_{O_m}|X_{I_m})$ are not coherent, there are $\delta > 0$, $f_j \in \mathcal{K}^j$ for $j = 1, \dots, m$, $j_0 \in \{1, \dots, m\}$, $f_0 \in \mathcal{K}^{j_0}, z_{j_0} \in \mathcal{X}_{I_{j_0}}$ such that

$$\left[\sum_{j=1}^{m} f_j - \underline{E}_j(f_j|X_{I_j}) - \pi^{-1}_{I_{j_0}}(z_{j_0})(f_0 - \underline{E}_{j_0}(f_0|z_{j_0}))\right](x) \leq -\delta < 0$$

for every $x \in \mathbb{S}(f_j) \cup \pi^{-1}_{I_{j_0}}(z_{j_0})$. Let us assume that f_j is not zero for all $j = 1, \dots, m$ (otherwise it is enough to drop the terms in the sum related to the zero gambles).

By definition of natural extension (see Definition 9 in Section 2), for every $j = 1, \dots, m$, and every $z_j \in S_j(f_j)$,[17] there are $f^k_{i,z_j} \in \mathcal{H}^i, i = 1, \dots, m$, $k = 1, \dots, n_i$ s.t.

$$\left[\sum_{i=1}^{m}\sum_{k=1}^{n_i} G_i(f^k_{i,z_j}|X_{I_i}) - \pi^{-1}_{I_j}(z_j)\left(f_j - \underline{E}_j(f_j|z_j) + \frac{\delta}{2m}\right)\right](x) < -\delta_{z_j} < 0$$

for every $x \in \mathbb{S}(f^k_{i,z_j}) \cup \pi^{-1}_{I_j}(z_j)$, and for some $\delta_{z_j} > 0$.

By making the (finite) sum on $z_j \in S_j(f_j)$, we deduce that

$$\left[\sum_{z_j \in S_j(f_j)} \sum_{i=1}^{m}\sum_{k=1}^{n_i} G_i(f^k_{i,z_j}|X_{I_i}) - (f_j - \underline{E}_j(f_j|X_{I_j}))\right](x) < -\min_{z_j \in S_j(f_j)} \delta_{z_j} + \frac{\delta}{2m} S_j(f_j)(x)$$

for every $x \in \mathbb{S}(f^k_{i,z_j}) \cup S_j(f_j)$.

It follows that

$$\begin{aligned}
&\left[\sum_{j=1}^{m}\sum_{z_j \in S_j(f_j)} \sum_{i=1}^{m}\sum_{k=1}^{n_i} G_i(f^k_{i,z_j}|X_{I_i}) - \pi^{-1}_{I_{j_0}}(z_{j_0})\left(f_0 - (\underline{E}_{j_0}(f_0|z_{j_0}) + \frac{\delta}{4})\right)\right](x) \\
&= \left[\sum_{j=1}^{m}\sum_{z_j \in S_j(f_j)} \sum_{i=1}^{m}\sum_{k=1}^{n_i} G_i(f^k_{i,z_j}|X_{I_i}) - \sum_{j=1}^{m} f_j - \underline{E}_j(f_j|X_{I_j})\right](x) \\
&+ \left[\sum_{j=1}^{m} f_j - \underline{E}_j(f_j|X_{I_j}) - \pi^{-1}_{I_{j_0}}(z_{j_0})\left(f_0 - \underline{E}_{j_0}(f_0|z_{j_0})\right)\right](x) + \frac{\delta}{4}\pi^{-1}_{I_{j_0}}(z_{j_0})(x) \\
&=: A(x) + B(x) + \frac{\delta}{4}\pi^{-1}_{I_{j_0}}(z_{j_0})(x)
\end{aligned}$$

for every $x \in \mathcal{X}^n$. Now, given $x \in \pi^{-1}_{I_{j_0}}(z_{j_0}) \cup \mathbb{S}(f^k_{i,z_j})$, there are three possibilities:

- If $x \in \mathbb{S}(f_j)$, then $A(x) \leq \frac{\delta}{2} - \min_{z_j \in S_j(f_j), j=1,\dots,m} \delta_{z_j}$ and $B(x) \leq -\delta$, whence $A(x) + B(x) + \pi^{-1}_{I_{j_0}}(z_{j_0})\frac{\delta}{4} < -\frac{\delta}{4} < 0$.

[17] Remember that this is a bit of an abuse of notation, as explained in Section 2.2.

- If $x \in \pi_{I_{j_0}}^{-1}(z_0)$, $x \notin \mathbb{S}(f_j)$, there are two possibilities: either x belongs to $\mathbb{S}(f_{i,z_j}^k)$, and then $A(x) \leq -\min_{z_j \in S_j(f_j), j=1,\dots,m} \delta_{z_j}$ and $B(x) \leq -\delta$, whence $A(x)+B(x)+\pi_{I_{j_0}}^{-1}(z_{j_0})\frac{\delta}{4} < -\frac{3}{4}\delta < 0$; or it is not in $\mathbb{S}(f_j) \cup \mathbb{S}(f_{i,z_j}^k)$, and then $A(x) = 0$, whence $A(x) + B(x) + \frac{\delta}{4}\pi_{I_{j_0}}^{-1}(z_{j_0})(x) \leq -\frac{3}{4}\delta < 0$.
- Finally, if $x \in \mathbb{S}(f_{i,z_j}^k)$, $x \notin \mathbb{S}(f_j) \cup \pi_{I_{j_0}}^{-1}(z_{j_0})$, it follows that $A(x) \leq -\min_{z_j \in S_j(f_j), j=1,\dots,m} \delta_{z_j}$, $B(x) = 0$, whence $A(x)+B(x)+\frac{\delta}{4}\pi_{I_{j_0}}^{-1}(z_{j_0})(x) \leq -\min_{z_j \in S_j(f_j)} \delta_{z_j} < 0$.

This means that we can increase the value of $\underline{E}_{j_0}(f_0|z_{j_0})$ in $\frac{\delta}{4}$, which contradicts the definition of the natural extension. Hence, $\underline{E}_1(X_{O_1}|X_{I_1}), \dots, \underline{E}_m(X_{O_m}|X_{I_m})$ are coherent. □

Proof of Proposition 3. Consider dominating coherent linear conditional previsions $P_1(X_{O_1}|X_{I_1})$,…,$P_m(X_{O_m}|X_{I_m})$, and let $f_j^k \in \mathcal{H}^j$ for $j = 1, \dots, m, k = 1, \dots, n_j$ such that not all the f_j^k are zero gambles. Then

$$\max_{x \in \mathbb{S}(f_j^k)} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} (f_j^k - \underline{P}_j(f_j^k|X_{I_j})) \right](x) \geq \max_{x \in \mathbb{S}(f_j^k)} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} (f_j^k - P_j(f_j^k|X_{I_j})) \right](x) \geq 0,$$

where the second inequality follows because $P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are coherent. Hence, $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ avoid partial loss.

Conversely, assume that $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ avoid partial loss. From Lemma 1 their natural extensions $\underline{E}_1(X_{O_1}|X_{I_1}), \dots, \underline{E}_m(X_{O_m}|X_{I_m})$ are coherent, and therefore they are the lower envelopes of the class of dominating coherent linear previsions (Theorem 5). Since from Theorem 8 the natural extensions dominate the initial assessments, we deduce the existence of dominating coherent linear conditional previsions. □

Proof of Proposition 4. Assume Equation (6) fails. Then there are $\epsilon > 0$, $f_j^k \in \mathcal{H}^j$, $j = 1, \dots, m$, $k = 1, \dots, n_j$, such that for all $x \in \mathcal{X}^n$,

$$\left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + \epsilon R(f_j^k) S_j(f_j^k) \right](x) \leq 0,$$

and hence

$$\left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) \right](x) \leq -\epsilon \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} R(f_j^k) S_j(f_j^k) \right](x) \leq 0$$

for every $x \in \mathcal{X}^n$. Since not all the f_j^k are zero gambles, this implies that the conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ incur partial loss.

Conversely, assume our conditional lower previsions incur partial loss. Then there are $\delta > 0$, $f_j^k \in \mathcal{H}^j$, $j = 1, \dots, m$, such that not all the f_j^k are zero gambles, which lead to

$$\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j})\right](x) \leq -\delta$$

for all $x \in \mathbb{S}(f_j^k)$. Therefore, we can define $\epsilon := \frac{\delta}{1+\sum_{j=1}^{m}\sum_{k=1}^{n_j} R(f_j^k)}$, and obtain that for all $x \in \mathbb{S}(f_j^k)$,

$$\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + \epsilon R(f_j^k)S_j(f_j^k)\right](x) =$$
$$\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j})\right](x) + \epsilon\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j} R(f_j^k)S_j(f_j^k)\right](x) \leq$$
$$-\delta + \delta \cdot \frac{\sum_{j=1}^{m}\sum_{k=1}^{n_j} R(f_j^k)}{1+\sum_{j=1}^{m}\sum_{k=1}^{n_j} R(f_j^k)} < 0.$$

This proves that Expression (5) fails, also considered that $\sum_{j=1}^{m}\sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + \epsilon R(f_j^k)S_j(f_j^k)$ is zero outside $\mathbb{S}(f_j^k)$. □

Proof of Theorem 9. Let us prove that (WC1) implies (WC2). Assume the weak coherence of $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$, and let $\underline{P}$ be given by Equation (7). Note that $\underline{P}(f)$ is well-defined as it is bounded. In particular, it holds that $\min f \leq \underline{P}(f) \leq \max f$ for any gamble f: given $\alpha > \max f$, there are no gambles f_j^k satisfying Equation (7) or we contradict the weak coherence of the conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$; and for any $\alpha < \min f$ we can take $n_j = 0\ \forall j$. It is also easy to see that $\underline{P}$ satisfies conditions (C1)–(C3) from Section 2, whence it is a coherent lower prevision. Let us show that $\underline{P}, \underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent. Consider $f \in \mathcal{L}(\mathcal{X}^n)$, $f_j^k \in \mathcal{H}^j, j = 1, \dots, m, k = 1, \dots, n_j$ and some other gamble f_0 in one of the domains $\mathcal{H}^{j_0}$, and let us show that Equation (2) holds. For every $\epsilon > 0$, the definition of $\underline{P}$ implies that there are $g_j^k \in \mathcal{H}^j, j = 1, \dots, m, k = 1, \dots, l_j$, s.t.

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{l_j} G_j(g_j^k|X_{I_j}) - G(f) - \frac{\epsilon}{2}\right](x) < 0, \tag{13}$$

where $G(f) = f - \underline{P}(f)$ (just consider $\alpha = \underline{P}(f) - \frac{\epsilon}{2}$ in Equation (7)). Hence, $G(f) > \sum_{j=1}^{m}\sum_{k=1}^{l_j} G_j(g_j^k|X_{I_j}) - \frac{\epsilon}{2}$.

There are two possible cases in Equation (2): that j_0 belongs to $\{1, \dots, m\}$, i.e., that we focus on one of the conditional assessments (case (a) below) or that it does not, i.e., that we focus on one of the unconditional assessments (case (b)).

(a) Consider $f_0 \in \mathcal{H}^{j_0}, z_{j_0} \in \mathcal{X}_{I_{j_0}}$ for some j_0 in $\{1,\dots,m\}$. Then, using Equation (13) and the weak coherence of $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$,

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+G(f)-G_{j_0}(f_0|z_{j_0})\right](x)$$
$$\geq \max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+\sum_{j=1}^m\sum_{k=1}^{l_j}G_j(g_j^k|X_{I_j})-\frac{\epsilon}{2}-G_{j_0}(f_0|z_{j_0})\right](x)\geq -\frac{\epsilon}{2}.$$

Since this holds for any $\epsilon>0$, we deduce that

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+G(f)-G_{j_0}(f_0|z_{j_0})\right](x)\geq 0.$$

(b) Take $f_0\in\mathcal{L}(\mathcal{X}^n)$. Using Equation (13),

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+G(f)-G(f_0)\right](x)\geq -\epsilon;$$

otherwise, we should have

$$0>\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+G(f)-G(f_0)+\epsilon\right](x)$$
$$\geq \max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+\sum_{j=1}^m\sum_{k=1}^{l_j}G_j(g_j^k|X_{I_j})-G(f_0)+\frac{\epsilon}{2}\right](x),$$

where the second inequality follows from Equation (13). But this means that we can raise the value $\underline{P}(f_0)$ by $\frac{\epsilon}{2}$, which contradicts the definition of $\underline{P}$. Since this holds for any $\epsilon>0$, $\max_{x\in\mathcal{X}^n}[\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})+G(f)-G(f_0)](x)\geq 0$.

Hence, $\underline{P},\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent.

Let us show that (WC2) implies (WC3). If $\underline{P},\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent, then for any $j=1,\dots,m$, $\underline{P},\underline{P}_j(X_{O_j}|X_{I_j})$ are weakly coherent, and in the case of a conditional and an unconditional lower prevision, weak coherence is equivalent to coherence.

We prove next that (WC3) implies (WC1). Let $\underline{P}$ be pairwise coherent with $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$. Consider $f_j^k\in\mathcal{H}^j, j=1,\dots,m, k=1,\dots,n_j$, $j_0\in\{1,\dots,m\}, f_0\in\mathcal{H}^{j_0}, z_{j_0}\in\mathcal{X}_{I_{j_0}}$, and let us prove that Equation (2) holds. Since $\underline{P},\underline{P}_j(X_{O_j}|X_{I_j})$ are coherent, Corollary 3 implies that $\underline{P}(G_j(f_j^k|X_{I_j}))\geq 0$ for every $f_j^k\in\mathcal{H}^j$, and that $\underline{P}(G_{j_0}(f_0|z_{j_0}))=0$. If we let $g_j^k:=G_j(f_j^k|X_{I_j})$ and $g_0:=G_{j_0}(f_0|z_{j_0})$, we deduce that

$$\sum_{j=1}^m\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j})-G_{j_0}(f_0|z_{j_0})\geq\sum_{j=1}^m\sum_{k=1}^{n_j}(g_j^k-\underline{P}(g_j^k))-(g_0-\underline{P}(g_0)),$$

and the coherence of $\underline{P}$ implies the existence of $x \in \mathcal{X}^n$ such that

$$\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j}(g_j^k - \underline{P}(g_j^k)) - (g_0 - \underline{P}(g_0))\right](x) \geq 0.$$

This implies that Equation (2) holds.

Finally, assume that $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent, and let $\underline{P}$ be the functional given by Equation (7). Let $\underline{Q}$ be also weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$, and assume that there is some gamble f such that $\underline{Q}(f) = \underline{P}(f) - \delta$ for some $\delta > 0$. It follows from the definition of $\underline{P}$ that there are $f_j^k \in \mathcal{K}^j$ for $j = 1, \dots, m, k = 1, \dots, n_j$ such that

$$\max_{x \in \mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j} G(f_j^k|X_{I_j}) - \left(f - (\underline{Q}(f) + \frac{\delta}{2})\right)\right](x) < 0,$$

whence

$$\max_{x \in \mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j} G(f_j^k|X_{I_j}) - (f - \underline{Q}(f))\right](x) < -\frac{\delta}{2},$$

contradicting the weak coherence of $\underline{Q}, \underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$. Hence, $\underline{P}$ is the smallest coherent lower prevision which is weakly coherent with the conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$. □

Proof of Lemma 2. Assume that $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent. Since weak coherence is stronger than avoiding uniform sure loss, Proposition 2 implies that the class

$$\mathcal{J} := \{P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m}) \text{ weakly coherent } : \\ P_j(X_{O_j}|X_{I_j}) \in \mathcal{M}(\underline{P}_j(X_{O_j}|X_{I_j})), j = 1, \dots, m\}$$

is non-empty. Let us prove that $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are the lower envelopes of this class. Consider $f \in \mathcal{H}^j$ for some $j \in \{1, \dots, m\}$ and $z_j \in \mathcal{X}_{I_j}$, and let $\underline{P}$ be the coherent lower prevision defined in Equation (7). There are two possibilities:

(a) If $\underline{P}(z_j) > 0$, we deduce from the coherence of $\underline{P}$ and from Corollary 3 that there is some $P \in \mathcal{M}(\underline{P})$ such that $P(G(f|z_j)) = \underline{P}(G(f|z_j)) = 0$. Since $P(z_j) \geq \underline{P}(z_j) > 0$, $P_j(f|z_j)$ is the unique real value μ such that $P(\pi_{I_j}^{-1}(z_j)(f - \mu)) = 0$. As a consequence, $P_j(f|z_j) = \underline{P}_j(f|z_j)$.

(b) If $\underline{P}(z_j) = 0$, we take $P \in \mathcal{M}(\underline{P})$ such that $P(z_j) = 0$, and we select $P_j(X_{O_j}|z_j)$ in $\mathcal{M}(\underline{P}_j(X_{O_j}|z_j))$ such that $P_j(f|z_j) = \underline{P}_j(f|z_j)$ (there is one such linear prevision because $\underline{P}_j(X_{O_j}|X_{I_j})$ is separately coherent).

Apply now Lemma 5 and define a conditional linear prevision $Q_i(X_{O_i}|X_{I_i})$ which is coherent with the linear prevision P selected above and dominates $\underline{P}_i(X_{O_i}|X_{I_i})$ for every $i = 1, \dots, m$. Let us define then conditional linear previsions $P_i^{'}(X_{O_i}|X_{I_i})$ for $i = 1, \dots, m$ by

$$P_i^{'}(f|z_i') := \begin{cases} Q_i(f|z_i') \text{ if } i \neq j \\ Q_j(f|z_j') \text{ if } i = j, z_j' \neq z_j \\ P_j(f|z_j) \text{ if } i = j, z_j' = z_j. \end{cases}$$

It follows from Theorem 3 that $P_i^{'}(X_{O_i}|X_{I_i})$ is coherent with P for $i = 1, \dots, m$: note that in case (b) above $P(G(g|z_j)) = 0$ for every $g \in \mathcal{K}^j$ because $P(z_j) = 0$. Since $P_i^{'}(X_{O_i}|X_{I_i})$ dominates by construction $\underline{P}_i(X_{O_i}|X_{I_i})$ for $i = 1, \dots, m$, we deduce applying Theorem 9 that $P_1^{'}(X_{O_1}|X_{I_1}), \dots, P_m^{'}(X_{O_m}|X_{I_m})$ belong to $\mathcal{J}$. Hence, $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are the lower envelopes of $\mathcal{J}$.

Conversely, we need to show that the lower envelopes of a class

$$\{P_1^{\lambda}(X_{O_1}|X_{I_1}), \dots, P_m^{\lambda}(X_{O_m}|X_{I_m}) : \lambda \in \Lambda\}$$

of weakly coherent linear previsions are again weakly coherent. Consider $f_j^k \in \mathcal{H}^j$ for $j = 1, \dots, m$, $j_0 \in \{1, \dots, m\}, f_0 \in \mathcal{H}^{j_0}, z_{j_0} \in \mathcal{X}_{I_{j_0}}$. Then for every $\epsilon > 0$ there is some $\lambda \in \Lambda$ such that $P_{j_0}^{\lambda}(f_0|z_{j_0}) - \epsilon \leq \underline{P}_{j_0}(f_0|z_{j_0})$. As a consequence,

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j}(f_j^k - \underline{P}_j(f_j^k|X_{I_j})) - \pi_{I_{j_0}}^{-1}(z_{j_0})(f_0 - \underline{P}_{j_0}(f_0|z_{j_0}))\right](x)$$
$$\geq \max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j}(f_j^k - P_j^{\lambda}(f_j^k|X_{I_j})) - \pi_{I_{j_0}}^{-1}(z_{j_0})(f_0 - P_{j_0}^{\lambda}(f_0|z_{j_0}) + \epsilon)\right](x) \geq -\epsilon,$$

using the weak coherence of $P_1^{\lambda}(X_{O_1}|X_{I_1}), \dots, P_m^{\lambda}(X_{O_m}|X_{I_m})$. Since we can do this for every $\epsilon > 0$, we deduce that $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent. □

Proof of Theorem 10. It follows from [27, Theorem 6.8.2(c)] that the conditional lower prevision in Equation (8) is defined from $\underline{P}$ using natural extension, and is therefore coherent with $\underline{P}$. Applying Theorem 9, we deduce that the conditional previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ are weakly coherent.

Let now $\underline{Q}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ be a conditional lower prevision which is weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$. From Theorem 9, there is a coherent lower prevision $\underline{Q}$ which is weakly coherent with the conditional previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m}), \underline{Q}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. Applying the same theorem, we deduce that $\underline{Q}(f) \geq \underline{P}(f)$ for any gamble f on $\mathcal{X}^n$. As a consequence, given $z_{m+1} \in \mathcal{X}_{I_{m+1}}$ such that $\underline{P}(z_{m+1}) > 0$, it holds that

$$\begin{aligned}\underline{Q}_{m+1}(f|z_{m+1}) &= \min\{P(f|z_{m+1}) : P \geq \underline{Q}\}\\ &\geq \min\{P(f|z_{m+1}) : P \geq \underline{P}\} = \underline{P}_{m+1}(f|z_{m+1})\end{aligned}$$

for any gamble $f \in \mathcal{K}^{m+1}$, where the first equality holds from Theorem 7. Since on the other hand the separate coherence of $\underline{Q}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ implies that $\underline{Q}_{m+1}(f|z_{m+1}) \geq \min_{x\in\pi_{I_{m+1}}^{-1}(z_{m+1})} f(x)$ for every $z_{m+1} \in \mathcal{X}_{I_{m+1}}$ and $f \in \mathcal{K}^{m+1}$, we deduce that $\underline{Q}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ dominates $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. □

Proof of Proposition 5. Consider $f_j^k \in \mathcal{H}^j$ for $j = 1, \dots, m, k = 1, \dots, n_j$. Let $f_{m+1} \in \mathcal{K}^{m+1}$, $j_0 \in \{1, \dots, m+1\}, z_0 \in \mathcal{X}_{I_{j_0}}, f_0 \in \mathcal{H}^{j_0}$, and let us prove that

$$\max_{x\in\mathcal{X}^n}\left[\sum_{j=1}^{m}\sum_{k=1}^{n_j}G_j(f_j^k|X_{I_j}) + G_{m+1}(f_{m+1}|X_{I_{m+1}}) - G_{j_0}(f_0|z_0)\right](x) \geq 0. \quad (14)$$

For every $\delta > 0$ and every $z_{m+1} \in \mathcal{X}_{I_{m+1}}$, it follows from Equation (4) that there are $g_j^{k,z_{m+1}} \in \mathcal{H}^j$ such that

$$\pi^{-1}_{I_{m+1}}(z_{m+1})(f_{m+1} - \underline{E}_{m+1}(f|z_{m+1}) + \delta) > \sum_{j=1}^{m} \sum_{k=1}^{l_{j,z_{m+1}}} G_j(g_j^{k,z_{m+1}}|X_{I_j}),$$

in $\mathbb{S}(g_j^{k,z_{m+1}}) \cup \pi^{-1}_{I_{m+1}}(z_{m+1})$, whence

$$G_{m+1}(f_{m+1}|X_{I_{m+1}}) + \delta \geq \sum_{z_{m+1} \in \mathcal{X}_{I_{m+1}}} \sum_{j=1}^{m} \sum_{k=1}^{l_{j,z_{m+1}}} G_j(g_j^{k,z_{m+1}}|X_{I_j}),$$

and therefore

$$\begin{aligned} &\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + G_{m+1}(f_{m+1}|X_{I_{m+1}}) - G_{j_0}(f_0|z_0) \right](x) \\ &\geq \max_{x \in \mathcal{X}^n} [\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + \sum_{z_{m+1} \in \mathcal{X}_{I_{m+1}}} \sum_{j=1}^{m} \sum_{k=1}^{l_{j,z_{m+1}}} G_j(g_j^{k,z_{m+1}}|X_{I_j}) \\ &- G_{j_0}(f_0|z_0)](x) - \delta \end{aligned} \tag{15}$$

There are two possibilities in Equation (14): either $j_0 \in \{1, \ldots, m\}$ or $j_0 = m+1$. If $j_0 \in \{1, \ldots, m\}$, we deduce from the above equation that

$$\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + G_{m+1}(f_{m+1}|X_{I_{m+1}}) - G_{j_0}(f_0|z_0) \right](x) \geq -\delta,$$

taking into account that $\underline{P}_1(X_{O_1}|X_{I_1}), \ldots, \underline{P}_m(X_{O_m}|X_{I_m})$ are weakly coherent. Since we can do this for every $\delta > 0$, we conclude that Equation (14) holds.

Assume now that $j_0 = m+1$, and that Equation (14) does not hold. Then there is some $\alpha > 0$ such that

$$\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + G_{m+1}(f_{m+1}|X_{I_{m+1}}) - G_{m+1}(f_0|z_0) \right](x) = -\alpha < 0.$$

Given $\delta := \frac{\alpha}{2}$, we can apply Equation (15) and deduce that

$$\begin{aligned} - \alpha \geq \max_{x \in \mathcal{X}^n} [&\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) \\ &+ \sum_{z_{m+1} \in \mathcal{X}_{I_{m+1}}} \sum_{j=1}^{m} \sum_{k=1}^{l_{j,z_{m+1}}} G_j(g_j^{k,z_{m+1}}|X_{I_j}) - G_{m+1}(f_0|z_0)](x) - \delta, \end{aligned}$$

and this means that

$$\begin{aligned} \max_{x \in \mathcal{X}^n} [\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k|X_{I_j}) + \sum_{z_{m+1} \in \mathcal{X}_{I_{m+1}}} \sum_{j=1}^{m} \sum_{k=1}^{l_{j,z_{m+1}}} G_j(g_j^{k,z_{m+1}}|X_{I_j}) \\ - G_{m+1}(f_0|z_0)](x) + \frac{\alpha}{4} < 0, \end{aligned}$$

contradicting the definition of $\underline{E}_{m+1}(f_0|z_0)$ via Equation (4). Hence, Equation (14) holds and $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m}),\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ are weakly coherent. □

Proof of Proposition 6. (1) Consider $\epsilon \geq 0$. We begin with the direct inclusion. Consider $P \in \mathcal{M}(\epsilon)$, $f_j \in \mathcal{H}^j$ for some $j \in \{1,\dots,m\}$ and $z_j \in \mathcal{X}_{I_j}$. We can express Equation (9) as

$$P(\pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j) + \epsilon R(f_j))) \geq 0, \tag{16}$$

whence, by making the (finite) sum over the elements z_j of the class $S_j(f_j)$,

$$P(G_j(f_j|X_{I_j}) + \epsilon R(f_j)S_j(f_j)) \geq 0.$$

Applying now the linearity and the monotonicity of the linear prevision P, we deduce that $P(f) \geq 0$ for every $f \in \mathcal{V}_\epsilon$.

Conversely, let P be a linear prevision such that $P(f) \geq 0$ for all $f \in \mathcal{V}_\epsilon$. Consider $f_j \in \mathcal{H}^j, z_j \in \mathcal{X}_{I_j}$, and let us prove that Equation (16) holds. Since this holds trivially if $P(z_j) = 0$, we assume therefore that $P(z_j) > 0$; we can also assume without loss of generality that $\min f_j = 0$, taking into account that given $g_j := f_j - \min f_j$,

$$\pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j) + \epsilon R(f_j)) = \pi_{I_j}^{-1}(z_j)(g_j - \underline{P}_j(g_j|z_j) + \epsilon R(g_j)).$$

Since in this paper we have assumed (by separate coherence) without loss of generality that $f_j\pi_{I_j}^{-1}(z_j)$ also belongs to $\mathcal{H}^j$ (see Remark 2 in Section 2), and $R(f_j\pi_{I_j}^{-1}(z_j)) \leq R(f_j)$ because $\min f_j = 0$,

$$\begin{aligned}\pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j) + \epsilon R(f_j)) &\geq \pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j) + \epsilon R(f_j\pi_{I_j}^{-1}(z_j)))\\ &= G_j(f_j\pi_{I_j}^{-1}(z_j)|z_j) + \epsilon R(f_j\pi_{I_j}^{-1}(z_j))\pi_{I_j}^{-1}(z_j)\\ &= G_j(f_j\pi_{I_j}^{-1}(z_j)|X_{I_j}) + \epsilon R(f_j\pi_{I_j}^{-1}(z_j))S_j(f_j\pi_{I_j}^{-1}(z_j)),\end{aligned}$$

and therefore $\pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j) + \epsilon R(f_j))$ belongs to $\mathcal{V}_\epsilon$. As a consequence, $P(\pi_{I_j}^{-1}(z_j)(f_j - \underline{P}_j(f_j|z_j) + \epsilon R(f_j))) \geq 0$ and from Equation (16) we deduce that P belongs to $\mathcal{M}(\epsilon)$.

(2) Since $\mathcal{M}(0) \subseteq \mathcal{M}(\epsilon)$ for any $\epsilon > 0$, we deduce that $\mathcal{M}(0) \subseteq \cap_{\epsilon>0}\mathcal{M}(\epsilon)$. Conversely, let P be a linear prevision in $\cap_{\epsilon>0}\mathcal{M}(\epsilon)$. Then for any $j \in \{1,\dots,m\}, z_j \in \mathcal{X}_{I_j}$ s.t. $P(z_j) > 0, f_j \in \mathcal{K}^j$ and $\epsilon > 0$, $P(f|z_j) \geq \underline{P}_j(f_j|z_j) - \epsilon R(f_j)$, whence $P(f|z_j) \geq \underline{P}_j(f_j|z_j)$ and as a consequence $P \in \mathcal{M}(0)$.

Let us prove now that $\mathcal{M}(0) = \mathcal{M}(\underline{P})$. Let P be a linear prevision that dominates $\underline{P}$. Consider $j \in \{1,\dots,m\}$, $f_j \in \mathcal{H}^j$, and $z_j \in \mathcal{X}_{I_j}$ s.t. $P(z_j) > 0$, and let us prove that

$$P(f_j|z_j) \geq \underline{P}_j(f_j|z_j). \tag{17}$$

This is a consequence of Lemma 5, taking into account that $P_j(f_j|z_j)$ is uniquely determined by GBR (see Theorem 6 in Section 2).

To see the converse, consider a linear prevision P in $\mathcal{M}(0)$, and assume that $P(f) < \underline{P}(f)$ for some gamble f. For every $j = 1,\dots,m$, consider a conditional prevision $P(X_{O_j}|X_{I_j})$ which is coherent with P and dominates $\underline{P}_j(X_{O_j}|X_{I_j})$; we can do so because $P \in \mathcal{M}(0)$ and because when a conditioning event z_j has probability zero any conditional prevision $P(X_{O_j}|z_j)$ satisfies the Generalised Bayes Rule with P. From Theorem 9, $P, P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})$ are weakly coherent. Applying Lemma 2, the lower envelope of $\{\underline{P}, \underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})\}$ and $\{P, P_1(X_{O_1}|X_{I_1}), \dots, P_m(X_{O_m}|X_{I_m})\}$ is again weakly coherent. But this means that the coherent lower prevision $\underline{Q} := \min\{P, \underline{P}\}$ is weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ and is strictly dominated by $\underline{P}$. This is a contradiction with the second part of Theorem 9. As a consequence, $P \in \mathcal{M}(\underline{P})$.

(3) This is an immediate consequence of the second statement, taking into account that the sets $\mathcal{M}(\epsilon)$ form a nested sequence.

□

Proof of Proposition 7. Let us define $\mathcal{V}_\epsilon^1 := \{f : f+\delta \in \mathcal{V}_\epsilon \ \forall \delta > 0\}$, and let us prove that this set is a *coherent set of almost-desirable gambles* with respect to $\mathcal{L}(\mathcal{X}^n)$. For this, we are going to show that it satisfies the axioms in [27, Section 3.7.3]:

(1) Let f be a gamble such that $\max f < 0$. Then there is some $\delta > 0$ such that $\max f + \delta < 0$. As a consequence, $f + \delta \notin \mathcal{V}_\epsilon$, or by Proposition 4 we contradict that $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ avoid partial loss. This implies that $f \notin \mathcal{V}_\epsilon^1$.

(2) Since $\lambda f \in \mathcal{V}_\epsilon$ for every $f \in \mathcal{V}_\epsilon, \lambda \geq 0$, for every $\delta > 0$ there is some $f \in \mathcal{V}_\epsilon$ such that $\max f < \delta$. Since $\mathcal{V}_\epsilon$ includes all positive constants, this implies that any gamble g with $\min g > 0$ belongs to $\mathcal{V}_\epsilon$, and as a consequence also to $\mathcal{V}_\epsilon^1$.

(3) Let $f \in \mathcal{V}_\epsilon^1$, $\lambda > 0$. Then $\lambda(f + \delta) = \lambda f + \lambda\delta$ belongs to $\mathcal{V}_\epsilon$ for every $\delta > 0$. As a consequence, $\lambda f \in \mathcal{V}_\epsilon^1$.

(4) For every $f, g \in \mathcal{V}_\epsilon$ it follows immediately that $f + g \in \mathcal{V}_\epsilon$. As a consequence, given $f_1, g_1 \in \mathcal{V}_\epsilon^1$ and $\delta > 0$, $f_1 + g_1 + \delta = (f_1 + \frac{\delta}{2}) + (g_1 + \frac{\delta}{2}) \in \mathcal{V}_\epsilon$. Hence, $f_1 + g_1 \in \mathcal{V}_\epsilon^1$.

(5) Finally, if $f + \delta \in \mathcal{V}_\epsilon^1$ for every $\delta > 0$, we deduce that $f + \delta' \in \mathcal{V}_\epsilon$ for all $\delta' > 0$, and as a consequence $f \in \mathcal{V}_\epsilon^1$.

Applying [27, Theorem 3.8.5], $\mathcal{V}_\epsilon^1$ is equal to

$$\{f : P(f) \geq 0 \ \forall P \in \mathcal{M}(\mathcal{V}_\epsilon^1)\},$$

where

$$\mathcal{M}(\mathcal{V}_\epsilon^1) = \{P : P(f) \geq 0 \ \forall f \in \mathcal{V}_\epsilon^1\} = \{P : P(f) \geq 0 \ \forall f \in \mathcal{V}_\epsilon\} = \mathcal{M}(\epsilon);$$

to see the second equality, consider a linear prevision P such that $P(f) \geq 0$ for all $f \in \mathcal{V}_\epsilon$, and let $f \in \mathcal{V}_\epsilon^1$. Then, $f + \delta \in \mathcal{V}_\epsilon$ for all $\delta > 0$, whence $P(f + \delta) = P(f) + \delta \geq 0$ for all $\delta > 0$. Hence, $P(f) \geq 0$. The third equality follows from the first statement in Proposition 6.

Hence,

$$\{f : P(f) \geq 0 \ \forall P \in \mathcal{M}(\epsilon)\} = \mathcal{V}_\epsilon^1 = \{f : f + \delta \in \mathcal{V}_\epsilon \ \forall \delta > 0\}.$$

It remains to prove that $\mathcal{V}_\epsilon^1$ is the closure of $\mathcal{V}_\epsilon$ in the topology of uniform convergence. To see this, note that for any gamble f in $\mathcal{V}_\epsilon^1$, f is the uniform limit of the sequence $\{f + \frac{1}{n} : n \in \mathbb{N}\}$, and as a consequence it belongs to $\overline{\mathcal{V}}_\epsilon$. Conversely, let $(f_n)_n$ be a sequence of elements in $\mathcal{V}_\epsilon^1$ that converges uniformly to f. Then for every $\delta > 0$, there is some $n_\delta \in \mathbb{N}$ such that $\|f_n - f\| < \delta \ \forall n \geq n_\delta$, whence

$f + \delta \geq f_n\ \forall n \geq n_\delta$ and therefore $f + \delta \in \mathcal{V}_\epsilon$. This implies that $f \in \mathcal{V}_\epsilon^1$ and as a consequence $\mathcal{V}_\epsilon^1 = \overline{\mathcal{V}}_\epsilon$. □

Proof of Proposition 8. Let us begin with the first statement. It is immediate that for every $\epsilon > 0$, $\mathcal{V}_\epsilon^A \subseteq \mathcal{V}_\epsilon \subseteq \overline{\mathcal{V}}_\epsilon$, whence $\cup_\epsilon \mathcal{V}_\epsilon^A \subseteq \cup_\epsilon \mathcal{V}_\epsilon \subseteq \cup_\epsilon \overline{\mathcal{V}}_\epsilon$.

Consider now $\epsilon > 0, f \in \mathcal{V}_\epsilon$. Then there are gambles $f_j^k \in \mathcal{K}^j$, $j = 1, \dots, m, k = 1, \dots, n_j$ s.t.

$$f \geq \sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k | X_{I_j}) + \epsilon R(f_j^k) S_j(f_j^k).$$

We can assume without loss of generality that $\min f_j^k = 0$ for $j = 1, \dots, m, k = 1, \dots, n_j$; otherwise, it suffices to take $g_j^k := (f_j^k - \min f_j^k) S_j(f_j^k)$, which belongs to $\mathcal{H}^j$ by Remark 2 (in Section 2), and moreover satisfies $G_j(g_j^k | X_{I_j}) = G_j(f_j^k | X_{I_j})$, $R(g_j^k) \leq R(f_j^k)$ and whose support $S_j(g_j^k)$ is included in $S_j(f_j^k)$. Then

$$\begin{aligned}
&\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f_j^k | X_{I_j}) + \epsilon R(f_j^k) S_j(f_j^k) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left[\sum_{z_j \in S_j(f_j^k)} G_j(f_j^k | z_j) \right] + \epsilon R(f_j^k) S_j(f_j^k) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left[\sum_{z_j \in S_j(f_j^k)} \pi_{I_j}^{-1}(z_j) \left(\sum_{x_j \in \mathcal{X}_{O_j}} f_j^k(x_j, z_j) (\pi_{O_j}^{-1}(x_j) - P_j(x_j | z_j)) \right) \right] \\
&\quad + \epsilon R(f_j^k) S_j(f_j^k) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left[\sum_{z_j \in S_j(f_j^k)} \sum_{x_j \in \mathcal{X}_{O_j}} \left(f_j^k(x_j, z_j) \pi_{I_j}^{-1}(z_j) (\pi_{O_j}^{-1}(x_j) - P_j(x_j | z_j)) \right) \right] \\
&\quad + \epsilon R(f_j^k) S_j(f_j^k) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left[\sum_{z_j \in S_j(f_j^k)} \sum_{x_j \in \mathcal{X}_{O_j}} f_j^k(x_j, z_j) G_j(x_j | z_j) \right] + \epsilon R(f_j^k) S_j(f_j^k) \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left[\sum_{z_j \in S_j(f_j^k)} \left(\epsilon R(f_j^k) \pi_{I_j}^{-1}(z_j) + \sum_{x_j \in \mathcal{X}_{O_j}} f_j^k(x_j, z_j) G_j(x_j | z_j) \right) \right] \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n_j} \left[\sum_{z_j \in S_j(f_j^k)} \sum_{x_j \in \mathcal{X}_{O_j}} \left(f_j^k(x_j, z_j) G_j(x_j | z_j) + \epsilon \frac{R(f_j^k) \pi_{I_j}^{-1}(z_j)}{|\mathcal{X}_{O_j}|} \right) \right] \\
&\geq \sum_{j=1}^{m} \sum_{k=1}^{n_j} \sum_{z_j \in S_j(f_j^k)} \sum_{x_j \in \mathcal{X}_{O_j}} f_j^k(x_j, z_j) \left(G_j(x_j | z_j) + \frac{\epsilon \pi_{I_j}^{-1}(z_j)}{|\mathcal{X}_{O_j}|} \right),
\end{aligned}$$

whence $f \in \mathcal{V}_{\epsilon_1}^A$. To see this, consider that in the last expression $f_j^k(x_j, z_j)$ can be regarded as λ_j^k in the definition of $\mathcal{V}_{\epsilon_1}^A$ (see (10)), and moreover that $G_j(x_j | z_j) +$

$\frac{\epsilon\pi^{-1}_{I_j}(z_j)}{|\mathcal{X}_{O_j}|} = G_j(f_j|X_{I_j}) + \frac{\epsilon}{|\mathcal{X}_{O_j}|}R(f_j)S_j(f_j)$, if we take f_j to be the indicator function of $\pi^{-1}_{O_j \cup I_j}(\{(x_j, z_j)\})$.

This implies that $\mathcal{V}_\epsilon \subseteq \mathcal{V}^A_{\epsilon_1}$, and as a consequence $\overline{\mathcal{V}}_\epsilon \subseteq \overline{\mathcal{V}}^A_{\epsilon_1} = \mathcal{V}^A_{\epsilon_1}$, taking into account that the set of gambles $\mathcal{V}^A_{\epsilon_1}$ is closed because it is finitely generated [27, Section 4.2.1]. As a consequence, $\cup_\epsilon \overline{\mathcal{V}}_\epsilon \subseteq \cup_\epsilon \mathcal{V}^A_\epsilon \subseteq \cup_\epsilon \mathcal{V}_\epsilon$, and therefore the three sets coincide.

We turn now to the second statement. Since for every $\epsilon > 0$ the set $\mathcal{V}^A_\epsilon$ is included in $\mathcal{V}_\epsilon$, we deduce that $\mathcal{M}^A_\epsilon \supseteq \mathcal{M}(\epsilon)$, and as a consequence $\mathcal{M}^A_0 = \cap_\epsilon \mathcal{M}^A_\epsilon \supseteq \cap_\epsilon \mathcal{M}(\epsilon) = \mathcal{M}(0)$, where the equalities follow from Proposition 6. Conversely, given $\epsilon > 0$ it follows from Proposition 7 that

$$\mathcal{M}(\epsilon) = \{P : P(f) \geq 0 \ \forall f \in \overline{\mathcal{V}}_\epsilon\} \supseteq \{P : P(f) \geq 0 \ \forall f \in \mathcal{V}^A_{\epsilon_1}\} = \mathcal{M}^A_{\epsilon_1},$$

where the inclusion follows from the first part of the current proof. As a consequence, $\cap_\epsilon \mathcal{M}^A_\epsilon \subseteq \cap_\epsilon \mathcal{M}(\epsilon \max_j |\mathcal{X}_{O_j}|)$, and since $\mathcal{M}(\epsilon)$ does not increase as ϵ goes to zero, $\cap_\epsilon \mathcal{M}(\epsilon \max_j |\mathcal{X}_{O_j}|) = \cap_\epsilon \mathcal{M}(\epsilon) = \mathcal{M}(0)$. We conclude then that $\mathcal{M}(0) = \cap_\epsilon \mathcal{M}(\epsilon) = \cap_\epsilon \mathcal{M}^A_\epsilon = \mathcal{M}^A_0$. □

Proof of Proposition 9. For every $\epsilon > 0$, $f^k_j \in \mathcal{H}^j$ for $j = 1, \dots, m, k = 1, \dots, n_j$,

$$\pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu) \geq \sum_{j=1}^{m} \sum_{k=1}^{n_j} [G_j(f^k_j|X_{I_j}) + \epsilon R(f^k_j)S_j(f^k_j)]$$

implies that

$$\left[\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f^k_j|X_{I_j}) - \pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu + \delta)\right](x) < 0$$

for every $x \in \pi^{-1}_{I_{m+1}}(z_{m+1}) \cup \mathbb{S}(f^k_j)$ and every $\delta > 0$. Hence, $\sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu) \in \cup_\epsilon \mathcal{V}_\epsilon\} \leq \underline{E}_{m+1}(f|z_{m+1})$.

To see the converse, note that, if for $\mu \in \mathbb{R}$ there are $f^k_j \in \mathcal{H}^j, j = 1, \dots, m, k = 1, \dots, n_j$, such that

$$\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f^k_j|X_{I_j}) - \pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu) < 0$$

on $\pi^{-1}_{I_{m+1}}(z_{m+1}) \cup \mathbb{S}(f^k_j)$, then there is some $\epsilon > 0$ such that

$$\sum_{j=1}^{m} \sum_{k=1}^{n_j} G_j(f^k_j|X_{I_j}) - \pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu) + \sum_{j=1}^{m} \sum_{k=1}^{n_j} \epsilon R(f_j)S_j(f^k_j) \leq 0,$$

and as a consequence $\pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu) \in \mathcal{V}_\epsilon$.

Finally, the inequality $\underline{E}_{m+1}(f|z_{m+1}) \leq \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f - \mu) \in \mathcal{V}\}$ follows from the inclusion $\cup_\epsilon \mathcal{V}_\epsilon \subseteq \mathcal{V}$. □

Proof of Proposition 10. Since $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ avoid partial loss, Proposition 3 implies that there are dominating coherent conditional linear previsions $P^\lambda_1(X_{O_1}|X_{I_1}), \dots, P^\lambda_m(X_{O_m}|X_{I_m})$. Let us consider the vacuous conditional lower prevision $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ on $\mathcal{K}^{m+1}$. It follows that the conditional previsions $P^\lambda_1(X_{O_1}|X_{I_1}), \dots, P^\lambda_m(X_{O_m}|X_{I_m}), \underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ avoid partial loss.

Let us consider the finite sets of gambles

$$\mathcal{W}_1 := \{G_j(A_j|z_j) + \epsilon\pi_{I_j}^{-1}(z_j) : j = 1, \dots, m+1, z_j \in \mathcal{X}_{I_j}, A_j \subseteq \mathcal{X}_{O_j}\}, \mathcal{W}_2 := \{0\};$$

note that

$$G_j(A_j|z_j) + \epsilon\pi_{I_j}^{-1}(z_j) = G_j(f_j|X_{I_j}) + \epsilon R(f_j) S_j(f_j),$$

where f_j is the indicator function of $\pi_{O_j \cup I_j}^{-1}(A_j \times \{z_j\})$.

Since $P_1^\lambda(X_{O_1}|X_{I_1}), \dots, P_m^\lambda(X_{O_m}|X_{I_m}), \underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ avoid partial loss, we can apply Proposition 4 (Equation (6)) to deduce that for every $\lambda_{A_j,z_j} \geq 0$, $A_j \subseteq \mathcal{X}_{O_j}, z_j \in \mathcal{X}_{I_j}, j = 1, \dots, m+1$ such that not all the λ_{A_j,z_j} are equal to zero,

$$\max_{x \in \mathcal{X}^n} \left[\sum_{j=1}^{m+1} \sum_{z_j \in \mathcal{X}_{I_j}, A_j \subseteq \mathcal{X}_{O_j}} \lambda_{A_j,z_j} \left(G_j(A_j|z_j) + \epsilon\pi_{I_j}^{-1}(z_j) \right) \right](x) > 0.$$

Hence, we can apply Gale's theorem of the alternative ([12]; [27, p. 612]) to deduce the existence of a continuous linear functional P such that $P(v) > 0$ for every $v \in \mathcal{W}_1$ and $P(0) = 0$. Now, the functional $Q = \frac{P}{P(1)}$ satisfies $Q(1) = 1$ and is therefore a linear prevision such that $Q(v) > 0$ for every $v \in \mathcal{W}_1$.

If $Q(z_j) = 0$ for some $z_j \in \mathcal{X}_{I_j}$, it follows that $Q(G_j(A_j|z_j) + \epsilon\pi_{I_j}^{-1}(z_j)) = 0$ for every $A_j \subseteq \mathcal{X}_{O_j}$, which is a contradiction with the previous statement. Hence, it must be $Q(z_j) > 0$ for every $z_j \in \mathcal{X}_{I_j}, j = 1, \dots, m+1$, and in particular $Q(z_{m+1}) > 0$.

Let $\mathcal{V}_\epsilon^{\lambda,A}$ be defined by Equation (10) relative to the linear conditional previsions $P_1^\lambda(X_{O_1}|X_{I_1}), \dots, P_m^\lambda(X_{O_m}|X_{I_m})$, once we restrict their domains to events. Take f in $\mathcal{V}_\epsilon^{\lambda,A}$. It follows from Equation (10) that there are $v_1, \dots, v_n \in \mathcal{W}_1$, $\lambda_1, \dots, \lambda_n \geq 0$ such that $f \geq \sum_{i=1}^n \lambda_i v_i$. We deduce from this and the linearity and monotonicity of Q that $Q(f) > 0$ for every $f \in \mathcal{V}_\epsilon^{\lambda,A}$, and from the first statement in Proposition 6 we deduce that Q belongs to the set $\mathcal{M}_\epsilon^{\lambda,A}$ associated to the restrictions of $P_1^\lambda(X_{O_1}|X_{I_1}), \dots, P_m^\lambda(X_{O_m}|X_{I_m})$ to events. Applying Proposition 8, Q also belongs to $\mathcal{M}^\lambda(\epsilon \max_j |\mathcal{X}_{O_j}|)$.

Since moreover $P_j^\lambda(f|z_j) \geq \underline{P}_j(f|z_j)$, we deduce that Q belongs to the set $\mathcal{M}(\epsilon \max_j |\mathcal{X}_{O_j}|)$ associated to $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ by Equation (9): note that, since $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is vacuous, the set of linear previsions determined by the assessments $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$ coincides with the one determined if we also add $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$.

The above considerations show that for every $\epsilon > 0$ there is $P \in \mathcal{M}(\epsilon \max_j |\mathcal{X}_{O_j}|)$ such that $P(z_{m+1}) > 0$. Since we can do this for all $\epsilon > 0$, we deduce that the result holds. □

Proof of Lemma 3. For any $\epsilon > 0$,

$$\begin{aligned}
\underline{P}_{m+1}^\epsilon(f|z_{m+1}) &= \inf\{P(f|z_{m+1}) : P \in \mathcal{M}(\epsilon), P(z_{m+1}) > 0\} \\
&= \sup\{\mu : P(f|z_{m+1}) \geq \mu \ \forall P \in \mathcal{M}(\epsilon), P(z_{m+1}) > 0\} \\
&= \max\{\mu : P(\pi_{I_{m+1}}^{-1}(z_{m+1})f) \geq \mu P(z_{m+1}) \ \forall P \in \mathcal{M}(\epsilon)\} \\
&= \max\{\mu : P(\pi_{I_{m+1}}^{-1}(z_{m+1})(f - \mu)) \geq 0 \ \forall P \in \mathcal{M}(\epsilon)\} \\
&= \max\{\mu : \pi_{I_{m+1}}^{-1}(z_{m+1})(f - \mu) \in \overline{\mathcal{V}}_\epsilon\},
\end{aligned}$$

where the last equality follows from Proposition 7. As a consequence,

$$\underline{F}_{m+1}(f|z_{m+1}) = \lim_{\epsilon\to 0} \underline{P}^{\epsilon}_{m+1}(f|z_{m+1}) = \lim_{\epsilon\to 0} \max\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \overline{\mathcal{V}}_{\epsilon}\}$$
$$= \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \cup_{\epsilon}\overline{\mathcal{V}}_{\epsilon}\}.$$

This completes the proof of the first statement. For the second, it suffices to note that $\cup_{\epsilon}\mathcal{V}_{\epsilon} \subseteq \cup_{\epsilon}\overline{\mathcal{V}}_{\epsilon}$ and to apply Proposition 9. □

Proof of Theorem 11. Let us prove first of all the equality for linear assessments. Consider thus $P_1(X_{O_1}|X_{I_1}),\dots,P_m(X_{O_m}|X_{I_m})$ avoiding partial loss (i.e., coherent) conditional linear previsions with domains $\mathcal{K}^1,\dots,\mathcal{K}^m$. Then

$$\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) = \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \cup_{\epsilon}\mathcal{V}_{\epsilon}\}$$
$$= \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \cup_{\epsilon}\overline{\mathcal{V}}_{\epsilon}\} = \underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}),$$

where the first equality follows from Proposition 9, the second from Proposition 8, and the last one from Lemma 3.

Let now $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_m(X_{O_m}|X_{I_m})$ be conditional lower previsions avoiding partial loss. Then they also avoid partial loss with the conditional prevision $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ defined in the trivial way on the set of gambles which are constant on the sets $\{\pi^{-1}_{I_{m+1}}(z_{m+1}) : z_{m+1} \in \mathcal{X}_{I_{m+1}}\}$. From Lemma 1 and [21, Corollary 16] their natural extensions $\underline{E}_1(X_{O_1}|X_{I_1}),\dots,\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ are coherent and are the lower envelopes of a family of coherent conditional linear previsions

$$\{P^{\lambda}_1(X_{O_1}|X_{I_1}),\dots,P^{\lambda}_m(X_{O_m}|X_{I_m}),P^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) : \lambda\in\Lambda\}.$$

For every $\lambda \in \Lambda$, let $\mathcal{V}^{\lambda}_{\epsilon}$ be the set of gambles given by Equation (10) with respect to $P^{\lambda}_1(X_{O_1}|X_{I_1}),\dots,P^{\lambda}_m(X_{O_m}|X_{I_m})$, and derive $\underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ and $\underline{F}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ from $P^{\lambda}_1(X_{O_1}|X_{I_1}),\dots,P^{\lambda}_m(X_{O_m}|X_{I_m})$ by natural extension and as a limit of regular extensions, respectively. It follows from the first part of the proof that $\underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) = \underline{F}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. Since the natural extension $\underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is the smallest conditional lower prevision which is coherent with $P^{\lambda}_1(X_{O_1}|X_{I_1}),\dots,P^{\lambda}_m(X_{O_m}|X_{I_m})$, it follows that $\underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is dominated by $P^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$, and consequently

$$\inf_{\lambda\in\Lambda} \underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) \le \inf_{\lambda\in\Lambda} P^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) = \underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}).$$

Let us show that $\underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) \ge \underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$: from Theorem 5, $\underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is the lower envelope of the conditional previsions $P_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ which are coherent with $P^{\lambda}_1(X_{O_1}|X_{I_1}),\dots,P^{\lambda}_m(X_{O_m}|X_{I_m})$. For any such $P_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$, it follows that the conditional previsions

$$P^{\lambda}_1(X_{O_1}|X_{I_1}),\dots,P^{\lambda}_m(X_{O_m}|X_{I_m}),P_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$$

dominate $\underline{P}_1(X_{O_1}|X_{I_1}),\dots,\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ on their domains. Applying Theorem 8, we deduce that $P_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) \ge \underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. As a consequence,

$$\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) = \inf_{\lambda\in\Lambda} \underline{E}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$$
$$= \inf_{\lambda\in\Lambda} \underline{F}^{\lambda}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) \ge \underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}),$$

where the inequality follows because for every $\epsilon > 0$ and every $\lambda \in \Lambda$, $\mathcal{V}_\epsilon^\lambda \supseteq \mathcal{V}_\epsilon$, and as a consequence for every gamble $f \in \mathcal{K}^{m+1}$ and every $z_{m+1} \in \mathcal{X}_{I_{m+1}}$,

$$\begin{aligned}\underline{F}^\lambda_{m+1}(f|z_{m+1}) &= \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \cup_\epsilon \overline{\mathcal{V}}^\lambda_\epsilon\} \\ &\geq \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \cup_\epsilon \overline{\mathcal{V}}_\epsilon\} = \underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}),\end{aligned}$$

where the last equality follows from Lemma 3. Since the same lemma also shows that $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}}) \leq \underline{F}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$, we deduce that the two are equal. □

Proof of Proposition 11. From Proposition 6, $\mathcal{M}(\underline{P}) = \{P : P(f) \geq 0 \ \forall f \in \mathcal{V}\}$, whence $\mathcal{V} \subseteq \{f : \underline{P}(f) \geq 0\}$. If $\underline{P}(z_{m+1}) > 0$, there is a unique value μ^* such that $\underline{P}(\pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu^*)) = 0$, and, taking into account that $\underline{P}(\pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu))$ decreases as μ increases, we deduce that

$$\underline{P}_{m+1}(f|z_{m+1}) = \mu^* \geq \sup\{\mu : \pi^{-1}_{I_{m+1}}(z_{m+1})(f-\mu) \in \mathcal{V}\} \geq \underline{E}_{m+1}(f|z_{m+1}),$$

where the last inequality follows from Proposition 9.

Since on the other hand Proposition 5 implies that $\underline{E}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$ is weakly coherent with $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$, it dominates the smallest conditional lower prevision with this property, given by $\underline{P}_{m+1}(X_{O_{m+1}}|X_{I_{m+1}})$. Therefore $\underline{E}_{m+1}(f|z_{m+1}) = \underline{P}(f|z_{m+1})$ for every $f \in \mathcal{K}^{m+1}$. □

Proof of Lemma 4. Let $j \in \{1, \dots, m\}$, $z_j \in \mathcal{X}_{I_j}$ and $f \in \mathcal{H}^j$. Since $\underline{P}_j(f|z_j)$ is the limit of $\underline{P}^n_j(f|z_j)$, for every $\epsilon > 0$ there is some $n_\epsilon \in \mathbb{N}$ such that $\|G_j(f|z_j) - G^n_j(f|z_j)\| < \epsilon$ for all $n \geq n_\epsilon$, where $\|\cdot\|$ denotes the supremum norm. Since $\mathcal{X}_{I_j}$ is finite, we deduce the existence of $n_{\epsilon,1} \in \mathbb{N}$ such that $\|G_j(f|X_{I_j}) - G^n_j(f|X_{I_j})\| < \epsilon$ for all $n \geq n_{\epsilon,1}$.

Consider then $f^k_j \in \mathcal{H}^j$ for $j = 1, \dots, m, k = 1, \dots, n_j$, and $j_0 \in \{1, \dots, m\}, z_{j_0} \in \mathcal{X}_{I_{j_0}}, f_0 \in \mathcal{K}^{j_0}$. There is some $n_{\epsilon,2} \in \mathbb{N}$ such that

$$\left\| \left(\sum_{j=1}^m \sum_{k=1}^{n_j} G_j(f^k_j|X_{I_j}) - G_{j_0}(f_0|z_{j_0}) \right) - \left(\sum_{j=1}^m \sum_{k=1}^{n_j} G^n_j(f^k_j|X_{I_j}) - G^n_{j_0}(f_0|z_{j_0}) \right) \right\| < \epsilon$$

for all $n \geq n_{\epsilon,2}$. As a consequence, for every subset A of $\mathcal{X}^n$ we deduce that

$$\begin{aligned}\max_{x \in A} &\left[\sum_{j=1}^m \sum_{k=1}^{n_j} G_j(f^k_j|X_{I_j}) - G_{j_0}(f_0|z_{j_0}) \right](x) \\ &\geq \max_{x \in A} \left[\sum_{j=1}^m \sum_{k=1}^{n_j} G^n_j(f^k_j|X_{I_j}) - G^n_{j_0}(f_0|z_{j_0}) \right](x) - \epsilon\end{aligned}$$

for every $n \geq n_{\epsilon,2}$. Since we can moreover do this reasoning for any $\epsilon > 0$, we deduce that the limit of a sequence of weakly coherent (resp., coherent) conditional lower previsions is also weakly coherent (resp., coherent). □

Proof of Theorem 12. We start with the direct implication. Consider coherent conditional lower previsions $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$. For every $\epsilon > 0$, let $\mathcal{M}(\epsilon)$ be the set of linear previsions given by Equation (9). We deduce from Proposition 10 that for every $z_j \in \mathcal{X}_{I_j}$, $j = 1, \dots, m$, there is some $P \in \mathcal{M}(\epsilon)$

such that $P(z_j) > 0$. As a consequence, we can define the conditional lower previsions $\underline{P}_1^\epsilon(X_{O_1}|X_{I_1}), \dots, \underline{P}_m^\epsilon(X_{O_m}|X_{I_m})$ by regular extension. From [33, Theorem 3], $\underline{P}_1^\epsilon(X_{O_1}|X_{I_1}), \dots, \underline{P}_m^\epsilon(X_{O_m}|X_{I_m})$ are coherent.

For every $j \in \{1, \dots, m\}$, we can apply Theorem 11 to deduce that the natural extension $\underline{E}_j(X_{O_j}|X_{I_j})$ of $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$, which coincides with $\underline{P}_j(X_{O_j}|X_{I_j})$ on $\mathcal{H}^j$, is equal to $\lim_{\epsilon\to 0} \underline{P}_j^\epsilon(X_{O_j}|X_{I_j})$. As a consequence, the sequence of coherent conditional lower previsions

$$\{\underline{P}_1^\epsilon(X_{O_1}|X_{I_1}), \dots, \underline{P}_m^\epsilon(X_{O_m}|X_{I_m})\}_{\epsilon>0},$$

which are defined from $\mathcal{M}(\epsilon)$ using regular extension, converges pointwise to the initial assessments $\underline{P}_1(X_{O_1}|X_{I_1}), \dots, \underline{P}_m(X_{O_m}|X_{I_m})$.

The converse implication follows immediately from Lemma 4. $\square$

References

[1] E. W. Adams. *A Primer of Probability Logic*. CSLI Publications, Stanford, CA, 2002.

[2] B. Arnold, E. Castillo, and J.M. Sarabia. *Conditional specification of statistical models*. Springer, 1999.

[3] B. C. Arnold, E. Castillo, and J. M. Sarabia. Compatibility of partial or complete conditional probabilities specifications. *Journal of Statistical Planning and Infernce*, 123:133–159, 2004.

[4] B. C. Arnold and S. Press. Compatible conditional distributions. *Journal of the American Statistical Association*, 84:152–156, 1989.

[5] V. Biazzo, A. Gilio, T. Lukasiewicz, and G. Sanfilippo. Probabilistic logic under coherence, model-theoretic probabilistic logic, and default reasoning in system P. *Journal of Applied Non-Classical Logics*, 12(2):189–213, 2002.

[6] L Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and equilibrium refinements. *Econometrica*, 58(1):81–98, 1991.

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

[8] L. M. de Campos, M. T. Lamata, and S. Moral. The concept of conditional fuzzy measures. *International Journal of Intelligent Systems*, 5:237–246, 1990.

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

[10] L. E. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3:88–99, 1975.

[11] R. Fagin and J. Y. Halpern. A new approach to updating beliefs. In P. P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, volume 6, pages 347–374. North-Holland, Amsterdam, 1991.

[12] D. Gale. *The theory of linear economic models*. McGraw Hill, New York, 1960.

[13] E. H. Ip and Y. J. Wang. Canonical representation of conditionally specified multivariate discrete distributions. *Journal of Multivariate Analysis*, 100:1282–1290, 2009.

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

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

[16] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 14(1):167–207, 1990.

[17] P. H. Krauss. Representation of conditional probability measures on boolean algebras. *Acta Math. Acad. Sci. Hungar*, 19:229–241, 1968.

[18] I. Levi. Potential surprise: its role in inference and decision making. In L. J. Cohen and M. Hesse, editors, *Applications of Inductive Logic*, pages 1–27. Clarendon Press, Oxford, 1980.

[19] E. Miranda. A survey of the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 48(2):628–658, 2008.

[20] E. Miranda. Updating coherent previsions on finite spaces. *Fuzzy Sets and Systems*, 160(9):1286–1307, 2009.

[21] E. Miranda and G. de Cooman. Coherence and independence in non-linear spaces. Technical report, 2005. Downloadable at http://bellman.ciencias.uniovi.es/ emiranda/wp05-10.pdf.

[22] 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.

[23] E. Miranda and M. Zaffalon. Coherence graphs. *Artificial Intelligence*, 173(1):104–144, 2009.

[24] E. Miranda and M. Zaffalon. Natural extension as a limit of regular extensions. In T. Augustin, F. Coolen, S. Moral, and M. Troffaes, editors, *Proceedings of ISIPTA '09: 6th International Symposium on Imprecise Probability: Theories and Applications*, Durham, England, 2009.

[25] R. Pelessoni and P. Vicig. A consistency problem for imprecise conditional probability assessments. In *Proceedings of the Seventh International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 1998)*, pages 1478–1485. EDK, Paris, 1998.

[26] P. Vicig, M. Zaffalon, and F. Cozman. Notes on "notes on conditional previsions". *International Journal of Approximate Reasoning*, 44:358–365, 2007.

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

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

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

[30] Y. J. Wang and E. H. Ip. Conditionally specified continuous distributions. *Biometrika*, 95(3):735–746, 2008.

[31] P. M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975. Reprinted in [32].

[32] P. M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

[33] M. Zaffalon and E. Miranda. Conservative inference rule for uncertain reasoning under incompleteness. *Journal of Artificial Intelligence Research*, 34:757–821, 2009.

University of Oviedo, Department of Statistics and Operations Research, C-Calvo Sotelo, s/n, 33007 Oviedo, Spain

E-mail address: `mirandaenrique@uniovi.es`

IDSIA, Galleria 2, CH-6928 Manno (Lugano), Switzerland

E-mail address: `zaffalon@idsia.ch`