THE LAW OF ITERATED EXPECTATION AND IMPRECISE PROBABILITIES

ENRIQUE MIRANDA AND ARTHUR VAN CAMP

Abstract. The law of iterated expectation tells us how to combine hierarchical pieces of information when our uncertainty is modelled by means of probability measures. It has been extended to the imprecise case through Walley's marginal extension theorem for coherent lower previsions. In this paper, we investigate the extent to which a similar result can be established for other imprecise probability models that are either more general (choice functions) or more particular (possibility measures, belief functions) than coherent lower previsions. By doing this, we also establish links with other results established in the literature in the context of imprecise versions of Jeffrey's rule.
Keywords: Law of iterated expectation, imprecise probabilities, non-additive measures, belief functions, marginal extension.

1. Introduction

Consider two variables X_1, X_2 taking values in respective spaces Ω_1, Ω_2. Assume that our uncertainty about the value that X_1 takes can be modelled by means of a probability measure P_1 on Ω_1 and that, for each $\omega_1 \in \Omega_1$, we have a conditional probability measure $P_2(\cdot|\{\omega_1\})$ that models our uncertainty about X_2 if we know that the variable X_1 has taken the value ω_1. Then we can model our uncertainty about the vector (X_1, X_2) by means of the concatenation of the two models, so that we have for instance, if Ω_1 and Ω_2 are finite, that

$$P(\{(\omega_1, \omega_2)\}) = P_1(\{\omega_1\}) \cdot P_2(\{\omega_2\}|\{\omega_1\});$$

and from the joint model on (X_1, X_2) we can obtain the marginal on X_2. This is an example of the *law of total probability*, which can also be expressed in terms of expectations: for any function $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$, we have that

$$E(f) = \sum_{\omega_1 \in \Omega_1} P_1(\{\omega_1\}) E_2(f|\{\omega_1\}) = E_1(E_2(f|X_1)) \tag{1}$$

where $E_2(\cdot|\{\omega_1\})$ is the expectation operator associated with the probability measure $P_2(\cdot|\{\omega_1\})$. This would allow us to determine the expectation operator for the variable X_2, giving rise to the *law of iterated expectation*.

The above procedure eases the computation of the joint distribution of (X_1, X_2) and is useful in contexts where the information is of a hierarchical nature. It is thus relevant in the context of stochastic processes where a sequence of variables $X_1, X_2, \ldots$ is observed. It can also be formulated more generally in contexts where we have information conditional on the observation of some event B in a partition $\mathcal{B}$ of the possibility space, as well as marginal information on the partition. And finally, it is also formally linked to Jeffrey's updating rule (Jeffrey, 1965, 1988), which is of interest when we update the marginal information on our model while keeping the conditional intact.

In spite of these advantages, there will be scenarios where the available information does not allow us to elicit a precise marginal or conditional probability measure. This may happen for instance when we have vague or ambiguous information, if there is a conflict between the opinions of several experts or in the presence of missing data. In order to deal more effectively with those situations, a number of alternatives to probability measures have been proposed; these are nowadays usually encompassed with the term *imprecise probabilities* (Augustin et al., 2014). They include in particular belief functions (Shafer, 1976), possibility measures (Dubois and Prade, 1988), sets of probability measures (Levi, 1980) or coherent lower previsions (Walley, 1991). Some of these models appear under the name *non-additive* or *fuzzy* measures in the literature.

Our goal in this paper is to determine the extent to which the law of iterated expectation can be generalised to the case where the marginal and the conditional models belong to some particular family of imprecise probability models. More specifically, we shall investigate if it is possible to determine a joint model that belongs to the same family and that is also *compatible* with the marginal and conditional models. What we shall understand by compatibility will be key in the mathematical developments in the paper. Note also that, unlike what is usually done in the formulation of the law of iterated expectation for variables, we shall focus on the global model, and not take the immediate step of marginalising it in each of the variables.

Two important assumptions we shall consider throughout are the following: on the one hand, we shall focus on *finite* possibility spaces. This will simplify in some cases the mathematical developments, and in particular it will allow us to dispose with considerations of conglomerability (Dubins, 1975) and with the distinction between finitely and countably additive probability measures. Secondly, we shall focus on the case where we have one marginal and one conditional model; nevertheless, most of our developments can be generalised to a finite number of hierarchical conditional models.

Before we proceed, we should remark that we are not the first to tackle this problem: it has been considered in the context of coherent lower previsions by Walley (1991). In his celebrated *marginal extension theorem*, he proved that there is a smallest, or most conservative, joint coherent lower prevision that is coherent with any given marginal and conditional coherent lower previsions. Moreover, this joint lower prevision can be obtained as the lower envelope of the probability models determined applying the law of iterated expectation to the precise probability models that are compatible with the marginal and conditional lower previsions, respectively. The marginal extension theorem was later generalised to a finite number of variables by Miranda and de Cooman (2007) and to coherent sets of desirable gambles by de Cooman and Hermans (2008). On the other hand, the problem has also been considered for belief functions in a number of references (Ma et al., 2011; Smets, 1993; Spies, 1994; Zhou and Cuzzolin, 2017; Zhou et al., 2014). However, as we shall show, these references consider a relationship between the joint and the conditional model that does not comply with coherence in general.

The paper is organised as follows: after giving some preliminary notions about imprecise probabilities and introducing the problem in Section 2, we shall investigate the generalisation of the law for possibility measures (Section 3), belief functions (Section 4), distortion models (Section 5), sets of desirable gambles (Section 6) and choice functions (Section 7). Some additional discussion will be given in Section 8. In order to ease the reading, we have gathered the proofs of the results in an Appendix.

A preliminary version of this paper was presented at the ISIPTA 2023 conference (Miranda and Van Camp, 2023); the current paper includes additional results, examples and discussion.

2. Imprecise probability models

Consider a finite possibility space Ω, and let us denote $\mathcal{L}(\Omega) := \{f : \Omega \to \mathbb{R}\}$ the set of *gambles* on Ω[1]. A function $\underline{P} : \mathcal{K} \to \mathbb{R}$ defined on a subset $\mathcal{K}$ of $\mathcal{L}$ is called a *lower prevision*, and it may be understood as a model of a subject's behavioural dispositions; specifically, $\underline{P}(f)$ may be understood as the subject's supremum acceptable buying price for μ, in the sense that he considers desirable the transaction $f - \mu$ for any $\mu < \underline{P}(f)$. The rationality of these behavioural dispositions leads to the notion of lower previsions that *avoid sure loss* and are *coherent*.

A lower prevision $\underline{P} : \mathcal{L}(\Omega) \to \mathbb{R}$ is called a *coherent lower prevision* when it satisfies

(C1) $(\forall f \in \mathcal{L}(\Omega))\ \underline{P}(f) \geq \min f$.
(C2) $(\forall f \in \mathcal{L}(\Omega), \forall \lambda > 0)\ \underline{P}(\lambda f) = \lambda \underline{P}(f)$.
(C3) $(\forall f, g \in \mathcal{L}(\Omega))\ \underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$.

When it moreover satisfies condition (C3) with equality for any pair of gambles f, g, it is called a *linear prevision*, and it corresponds to the expectation operator with respect to a finitely additive probability.

Linear previsions can be used to characterise the coherence of a lower prevision $\underline{P}$ with domain $\mathcal{L}(\Omega)$: it holds that $\underline{P}$ is coherent if and only if it is the lower envelope of its associated *credal set*, which is given by

$$\mathcal{M}(\underline{P}) := \{P \text{ linear prevision} : P(f) \geq \underline{P}(f)\ \forall f \in \mathcal{L}\}, \tag{2}$$

in the sense that

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

On the other hand, a weaker notion than coherence is that of *avoiding sure loss*: a lower prevision $\underline{P}$ on $\mathcal{L}$ is said to avoid sure loss when the credal set it determines by means of Eq. (2) is non-empty.

The above notions can be easily extended to lower previsions whose domain $\mathcal{K}$ is a proper subset of $\mathcal{L}$: in that case, we say that $\underline{P}$ is coherent when there exists a coherent lower prevision on $\mathcal{L}$ that agrees with $\underline{P}$ on $\mathcal{K}$; the smallest[2] such lower prevision is called the *natural extension* of $\underline{P}$, and it is the lower envelope of

$$\mathcal{M}(\underline{P}) := \{P \text{ linear prevision} : P(f) \geq \underline{P}(f)\ \forall f \in \mathcal{K}\}. \tag{3}$$

We also say that a lower prevision $\underline{P}$ with domain $\mathcal{K}$ avoids sure loss when the credal set it determines by means of Eq. (3) is non-empty.

---

[1] When it is clear what the possibility space Ω is, we shall sometimes simply use $\mathcal{L}$ to indicate $\mathcal{L}(\Omega)$.

[2] Throughout the paper, when we say that a lower prevision $\underline{P}$ is the smallest member of a family to satisfy certain properties, we are using the order determined by pointwise dominance; i.e., we mean that $\underline{P} \leq \underline{Q}$ for any other member of the same family with the same properties. On the other hand, in the case of sets of desirable gambles the term 'smallest' will mean with respect to the order determined by set inclusion.

In this paper we shall consider two uncertainty models that are more general than coherent lower previsions: coherent sets of desirable gambles and coherent choice functions.

The first of these two models can be used to represent a subject's binary preference relations over the set of gambles $\mathcal{L}(\Omega)$. If we denote by D the set of gambles that a subject considers preferable over the status quo, this set is called a *coherent set of desirable gambles* when it satisfies the following rationality conditions (de Cooman and Quaeghebeur, 2012; Quaeghebeur, 2014; Seidenfeld et al., 1990; Walley, 1991) for all f and g in $\mathcal{L}(\Omega)$ and λ in $\mathbb{R}_{>0}$

D1. $0 \notin D$; [avoiding non-positivity]
D2. $\mathcal{L}_{>0} \subseteq D$; [accepting partial gain]
D3. if $f \in D$ then $\lambda f \in D$; [scaling]
D4. if $f, g \in D$ then $f + g \in D$. [combination]

In axiom D2, we use $\mathcal{L}_{>0}$ to denote the set of positive gambles: $\{f \in \mathcal{L} : f \geq 0, f \neq 0\}$.

We shall denote by $\overline{\mathbf{D}}(\Omega)$, or simply $\overline{\mathbf{D}}$, the collection of all coherent[3] sets of desirable gambles on Ω. A coherent set of desirable gambles D induces a coherent lower prevision by means of the formula

$$\underline{P}(f) := \sup\{\alpha \in \mathbb{R} : f - \alpha \in D\}, \tag{4}$$

and this allows to give the value $\underline{P}(f)$ a behavioural interpretation as the subject's supremum acceptable buying price for f. However, two different coherent sets of desirable gambles $D_1 \neq D_2$ may induce the same coherent lower prevision via Eq. (4). Thus, coherent sets of desirable gambles are a more expressive model than coherent lower previsions.

A set of desirable gambles D is an equivalent representation of a binary preference relation $\prec$ between gambles: $g \prec f \Leftrightarrow f - g \in D$, for all gambles f and g. This indicates a limitation of working with them: they can only capture beliefs based on binary preferences—preferences between two gambles. In order to overcome this, Kadane et al. (2004) introduced imprecise-probabilistic choice functions, which were further developed by Seidenfeld et al. (2010) and Van Camp et. al. (2018). A choice function C identifies from any finite decision problem $F \in \mathcal{Q}(\Omega) := \{G \subseteq \mathcal{L}(\Omega) \colon |G| \in \mathbb{N}\}$,[4] the subset $C(F)$ of admissible, or non-rejected, gambles. Similarly, the corresponding rejection function $R(F) := F \setminus C(F)$ identifies the rejected gambles from F. In order to make the connection with a useful equivalent model in the following paragraph, we will impose compatibility with the vector addition: $f \in R(F) \Leftrightarrow f + g \in R(F + \{g\})$, for all $f, g \in \mathcal{L}$ and $F \in \mathcal{Q}$, where we defined the addition $G + G'$ of sets of gambles G and G' as $G + G' := \{f + g \colon f \in G, g \in G'\}$, for any $G, G' \subseteq \mathcal{L}$.

Choice functions may be equivalently represented by means of *sets of desirable gamble sets* (De Bock and de Cooman, 2018; De Bock and de Cooman, 2019; de Cooman, 2022). The idea is to lift the qualification 'desirable' from gambles to finite sets of gambles $F \in \mathcal{Q}(\Omega)$—called 'gamble sets (on Ω)'. A gamble set F is called *desirable* when F contains a gamble that our subject prefers to 0, and

[3]Note that here the overline is not understood as a topological closure, but rather as a closure with respect to the rationality axioms D1–D4.

[4]When it is clear what the possibility space Ω is, we shall sometimes simply use $\mathcal{Q}$ to indicate $\mathcal{Q}(\Omega)$.

a *set of desirable gamble sets on* Ω As such, K generalises binary preferences: a gamble set $\{f, g\}$ may be desirable—an element of K—because f is desirable or g is desirable—so either f or g is preferred to 0—without being able to identify which one is. The set of all sets of desirable gamble sets will be denoted by $\mathbf{K}$.

Sets of desirable gamble sets K are related to rejection functions R, and therefore also to choice functions. Let $F \ominus f \coloneqq (F \setminus \{f\}) - \{f\}$. Then $f \in R(F) \Leftrightarrow 0 \in R(F - \{f\}) \Leftrightarrow (\exists g \in F \ominus f)$ g is desirable $\Leftrightarrow F \ominus f \in K$, so we see that R and K are equivalent representations of the same information.

A set of desirable gamble sets $K \subseteq \mathcal{Q}$ is called *coherent* if for all F and G in $\mathcal{Q}$, all $\{\lambda_{f,g}, \mu_{f,g} \colon f \in F, g \in G\} \subseteq \mathbb{R}$, and all f in $\mathcal{L}$:

K1. $\emptyset \notin K$;
K2. if $F \in K$ then $F \setminus \{0\} \in K$;
K3. if $f \in \mathcal{L}_{>0}$ then $\{f\} \in K$;
K4. if $F, G \in K$ and if, for all f in F and g in G, $(\lambda_{f,g}, \mu_{f,g}) > 0$[5], then $\{\lambda_{f,g} f + \mu_{f,g} g \colon f \in F, g \in G\} \in K$;
K5. if $F_1 \in K$ and $F_1 \subseteq F_2$, then $F_2 \in K$.

We collect all the coherent sets of desirable gamble sets on Ω in the collection $\overline{\mathbf{K}}(\Omega)$, often simply denoted by $\overline{\mathbf{K}}$.

While coherent sets of desirable gambles and coherent sets of desirable gambles sets are more expressive uncertainty models than coherent lower previsions, in this paper we shall also consider particular cases of coherent lower previsions whose domain are the indicator functions, $\mathbb{I}_A$, given by

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

for some $A \subseteq \Omega$. Since there is clearly a one-to-one correspondence between the sets $\{\mathbb{I}_A : A \subseteq \Omega\}$ and $\mathcal{P}(\Omega)$, we shall sometimes use $\underline{P}(A)$ to denote $\underline{P}(\mathbb{I}_A)$. The restriction to $\mathcal{P}(\Omega)$ of a coherent lower prevision is called a *coherent lower probability*. As particular cases of coherent lower probabilities, we have the following:

- We say that $\underline{P}$ is 2-*monotone* when
$$(\forall A, B \subseteq \Omega)\ \underline{P}(A \cup B) + \underline{P}(A \cap B) \geq \underline{P}(A) + \underline{P}(B).$$
- We say that $\underline{P}$ is a *belief function* when
$$(\forall k \in \mathbb{N}, \forall A_1, \dots, A_k \subseteq \Omega)\ \underline{P}\Big(\bigcup_{i=1}^{k} A_i\Big) \geq \sum_{\emptyset \neq I \subseteq \{1,\dots,k\}} (-1)^{|I|+1} \underline{P}\Big(\bigcap_{i \in I} A_i\Big).$$
- We say that $\underline{P}$ is *minitive* when
$$(\forall A, B \subseteq \Omega)\ \underline{P}(A \cap B) = \min\{\underline{P}(A), \underline{P}(B)\}.$$

Any probability measure is a particular instance of a belief function, which is in particular 2-monotone, which is in particular coherent. On the other hand, a minitive lower probability is also a belief function. We shall denote by $\mathbb{P}_\Omega$, $\mathtt{N}_\Omega$, $\mathtt{Bel}_\Omega$ and $\mathcal{C}_\Omega$ the classes of probability measures, minitive measures, belief functions and coherent lower probabilities on Ω.

While a precise probability measure P on Ω uniquely determines its corresponding expectation operator E_P, such a one-to-one correspondence does not hold in

[5]For any sequence $(\mu_1, \dots, \mu_k)$ of real numbers, by $(\mu_1, \dots, \mu_k) > 0$ we mean '$\mu_\ell \geq 0$ for all ℓ in $\{1, \dots, k\}$ and $\mu_\ell > 0$ for some ℓ in $\{1, \dots, k\}$'.

general for coherent lower probabilities: two different coherent lower previsions may have the same restriction to indicators of events. In other words, coherent lower previsions are a *more expressive* uncertainty model than coherent lower probabilities.

One way to have a one-to-one correspondence is to consider the property of 2-monotonicity for lower previsions: a lower prevision $\underline{P}$ satisfying

$$(\forall f, g \in \mathcal{L}(\Omega))\underline{P}(f \vee g) + \underline{P}(f \wedge g) \geq \underline{P}(f) + \underline{P}(g), \tag{5}$$

is called a 2-*monotone lower prevision*; in this equation, $\vee$ and $\wedge$ denote the point-wise maximum and point-wise minimum respectively, so

$$(f \vee g)(\omega) \coloneqq \max\{f(\omega), g(\omega)\} \text{ and } (f \wedge g)(\omega) \coloneqq \min\{f(\omega), g(\omega)\} \text{ for all } \omega \text{ in } \Omega.$$

Any 2-monotone lower prevision is coherent, and it determines a 2-monotone lower probability as its restriction to events; conversely, a 2-monotone lower probability $\underline{P}$ has a unique extension $\underline{E}_{\underline{P}}$ to $\mathcal{L}(\Omega)$ satisfying Eq. (5) (see Walley (1981) and also (de Cooman, Troffaes and Miranda, 2008, Corollary 10)): its *Choquet integral*, given by

$$\underline{E}_{\underline{P}}(f) = \min f + \int_{\min f}^{\max f} \underline{P}(f \geq t)dt, \tag{6}$$

that coincides moreover with the natural extension of $\underline{P}$.

On the other hand, a coherent lower prevision is said to be *minitive on gambles* when

$$(\forall f, g \in \mathcal{L})\, \underline{P}(f \wedge g) = \min\{\underline{P}(f), \underline{P}(g)\}. \tag{7}$$

In that case, it is also 2-monotone and its restriction to indicators of events is a minitive lower probability; however, and in contradistinction with the situation for 2-monotonicity, the natural extension of a minitive measure need not be minitive on gambles; it will only be so when it is $\{0, 1\}$-valued on events, as showed by de Cooman and Miranda (2014, Prop. 7). As an example of a minitive lower prevision, we have the *vacuous* lower prevision, given by $\underline{P}(f) = \min_{\omega\in\Omega} f(\omega)$ for every $f \in \mathcal{L}$.

Figure 1 summarises the relationships between the different models we have introduced; an arrow between two nodes means that the parent is a particular case (and hence a less general model) of the child.

2.1. **Conditional models.** So far, we have considered unconditional or marginal models. Next we shall discuss conditional models.

Consider a partition $\mathcal{B}$ of Ω, and for each $B \in \mathcal{B}$, a coherent lower prevision $\underline{P}(\cdot|B)$ that models our uncertainty on the outcome of the experiment conditional on the observation that this outcome belongs to the event B. The conditional lower prevision $\underline{P}(\cdot|\mathcal{B})$ is defined, for any gamble f on Ω, by

$$\underline{P}(f|\mathcal{B}) = \sum_{B\in\mathcal{B}} \mathbb{I}_B \underline{P}(f|B). \tag{8}$$

Here, we are using the notation $\mathbb{I}_B g$ to denote the gamble given by

$$\mathbb{I}_B g(\omega) = \begin{cases} g(\omega) & \text{if } \omega \in B \\ 0 & \text{otherwise.} \end{cases}$$

In other words, $\mathbb{I}_B g$ is a gamble that is called off unless we observe that the outcome of the experiment belongs to B. Note that for any gamble f, $\underline{P}(f|\mathcal{B})$ belongs to

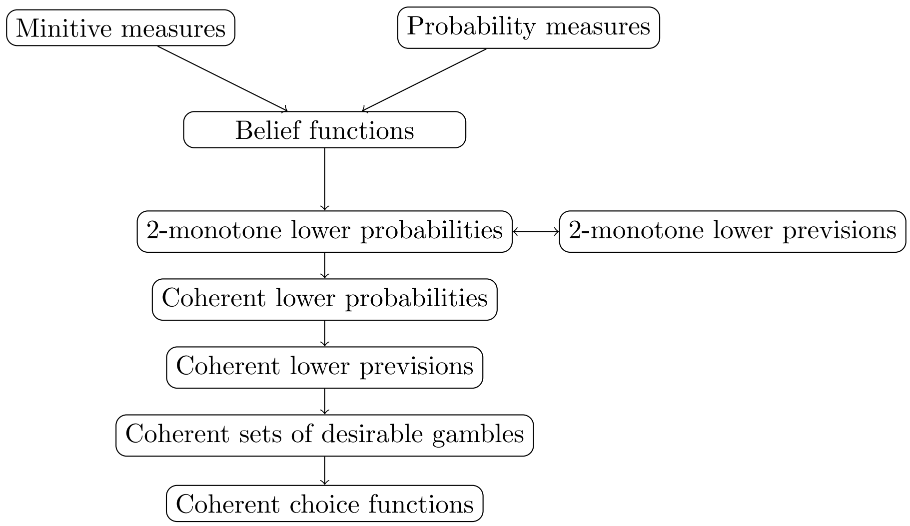

FIGURE 1. Relationships between different models.

the class $\mathcal{L}_{\mathcal{B}}$ of $\mathcal{B}$-measurable gambles, which are the gambles that are constant on the elements of the partition $\mathcal{B}$.

We then say that a conditional lower prevision is *separately coherent* when for each $B \in \mathcal{B}$, the lower prevision $\underline{P}(\cdot|B)$ on $\mathcal{L}(\Omega)$ is coherent[6], and denote by $\mathcal{C}(\cdot|\mathcal{B})$ the set of separately coherent conditional lower previsions. Similarly, we shall denote $\mathtt{Bel}(\cdot|\mathcal{B})$ (resp., $\mathtt{N}(\cdot|\mathcal{B})$, $\mathbb{P}(\cdot|\mathcal{B})$) the particular cases where the conditional model is a belief function (resp., a minitive measure, a probability measure) for any $B \in \mathcal{B}$.

Likewise, we can consider a conditional set of desirable gambles $D\rfloor\mathcal{B}$ by putting together coherent sets of desirable gambles $D\rfloor B$ for $B \in \mathcal{B}$, where $D\rfloor B$ is a coherent subset of $\mathcal{L}(B)$; and also a conditional coherent choice function $C\rfloor\mathcal{B}$ by putting together a family of coherent choice functions $C\rfloor B$ on $\mathcal{L}(B)$, one for each B.

2.2. **Formulation of the problem.** Assume now that for a given partition $\mathcal{B}$ of Ω we have a marginal uncertainty model on $\mathcal{B}$ and that, for each $B \in \mathcal{B}$, we consider a conditional model on B. Then we may want to aggregate these two models into some global uncertainty model on Ω. As we mentioned in the introduction, when the marginal and conditional models are given by probability measures, it is immediate to determine the joint model by means of the law of iterated expectation.

When the models are imprecise and are given by coherent lower previsions $\breve{\underline{P}}$ on $\mathcal{L}_{\mathcal{B}}$ and $\underline{P}(\cdot|\mathcal{B})$ on $\mathcal{L}(\Omega)$, we must specify what entails the compatibility of the output model $\widehat{\underline{P}}$ with the input models. In the case of the marginal model, this is not too controversial: we should require that

$$(\forall f \in \mathcal{L}_{\mathcal{B}})\widehat{\underline{P}}(f) = \breve{\underline{P}}(f). \tag{9}$$

[6]While at first it may seem confusing that $\underline{P}(\cdot|B)$ is defined on $\mathcal{L}(\Omega)$ instead of $\mathcal{L}(B)$, the reason for this is to have a common domain for all $\underline{P}(\cdot|B)$ so as to be able to use Eq. (8). Note also that it is a consequence of coherence nonetheless that $\underline{P}(f|B) = \underline{P}(g|B)$ whenever $\mathbb{I}_B f = \mathbb{I}_B g$, so only the values of a gamble on B matter for its conditional lower prevision on B. See Walley (1991, Sect. 6.2) for more details.

With respect to the conditional model, the extension of Bayes' rule to the conditional case means that $\underline{P}(f|B)$ should satisfy

$$\underline{P}(\mathbb{I}_B(f - \underline{P}(f|B))) = 0. \tag{10}$$

In the finitary context we are considering in this paper, conditions (9) and (10) are equivalent to the *coherence* of the global model with the marginal $\check{\underline{P}}$ and the conditional $\underline{P}(\cdot|\mathcal{B})$. The interpretation of this condition is that the behavioural implications of the supremum acceptable buying prices encompassed by $\check{\underline{P}}$ and $\underline{P}(\cdot|\mathcal{B})$ should not lead to a sure loss, and also that they should not imply a raise on the previously assessed supremum buying price for a gamble.

On the other hand, when $\underline{P}(B) > 0$, there is only one value α^* for which $\underline{P}(\mathbb{I}_B(f-\alpha^*)) = 0$, and it can also be computed as

$$\underline{P}(f|B) = \inf\{P(f|B) : P \geq \underline{P}\}. \tag{11}$$

We refer to Walley (1991, Ch. 6) and Miranda (2009) for more details about the problem of updating coherent lower previsions.

From the considerations of coherence it follows the *marginal extension theorem*:

Theorem 1. *(Walley, 1991) Consider a marginal coherent lower prevision $\check{\underline{P}}$ on $\mathcal{L}_{\mathcal{B}}$ and a separately coherent conditional lower prevision $\underline{P}(\cdot|\mathcal{B})$ on $\mathcal{L}(\Omega)$. The smallest coherent lower prevision on $\mathcal{L}(\Omega)$ that is coherent with $\check{\underline{P}}, \underline{P}(\cdot|\mathcal{B})$ is*

$$\widehat{\underline{P}} := \check{\underline{P}}(\underline{P}(\cdot|\mathcal{B})). \tag{12}$$

That $\widehat{\underline{P}}$ is the smallest coherent lower prevision that is coherent with $\check{\underline{P}}, \underline{P}(\cdot|\mathcal{B})$ means that there may be other coherent lower previsions satisfying Eqs. (9), (10); these will encompass behavioural assessments that are not implied by those present in $\check{\underline{P}}, \underline{P}(\cdot|\mathcal{B})$ and the rationality requirements of coherence, which is something that we may want to avoid if possible (but see also the discussion in the paragraphs below). In particular, when the conditional model $\underline{P}(\cdot|\mathcal{B})$ is precise, Eq. (12) gives the unique coherent lower prevision on $\mathcal{L}(\Omega)$ that is coherent with $\check{\underline{P}}, \underline{P}(\cdot|\mathcal{B})$.

Assume now that the marginal $\check{\underline{P}}$ and the conditional $\underline{P}(\cdot|\mathcal{B})$ belong to some particular subfamily of interest (belief functions, minitive measures, etcetera). We may then wonder if:

(a) The marginal extension $\widehat{\underline{P}}$ also belongs to that subfamily of interest;
(b) If it does not, if we can make a minimal correction so that it does.

In this second case, our correction shall always be in the form of a more informative model. The idea here is that any desirability assessment present in the marginal extension should be considered in any correction that is implemented; but we also open the possibility of adding other desirability assessments that are not included in the minimal implications modelled by the marginal extension. This leads to the consideration, if needed, of *inner approximations* (Miranda et al., 2023) of the marginal extension, i.e., coherent lower previsions $\underline{P}' \geq \widehat{\underline{P}}$.

In other words, the problem of determining a global model from marginal and conditional pieces of information has two parts: first, we have the implications of coherence, that lead us to considering the marginal extension as a minimal global model that is compatible with the sources of information; and then we have the structural assessment of this global model belonging to the family of interest. When the marginal extension does not, we should make some correction, and, in order not to erase the implications of coherence, we may end up with a more informative model. In that case, the behavioural assessments present in the resulting model will not only be implied by coherence but also by its membership to the family of interest.

In what follows, we shall analyse these problems (a) and (b) for the imprecise probability models we have introduced earlier in this section. In the case of capacities, we start from a coherent lower probability $\underline{\breve{P}}$ on $\mathcal{P}(\mathcal{B})$, the events that are finite unions of elements from the partition $\mathcal{B}$, and also assume that for each $B \in \mathcal{B}$, we have a coherent lower probability $\underline{P}(\cdot|B)$ on $\mathcal{P}(\Omega)$. Then we can consider their respective natural extensions $\underline{\breve{E}}$ on $\mathcal{L}_{\mathcal{B}}$ and $\underline{E}(\cdot|\mathcal{B})$ on $\mathcal{L}(\Omega)$.

We shall denote by $M(\underline{\breve{P}}, \underline{P}(\cdot|\mathcal{B}))$, or by $\underline{\widehat{E}}$ if no confusion is possible, the *marginal extension of* $\underline{\breve{E}}$ *and* $\underline{E}(\cdot|\mathcal{B})$, given by

$$\underline{\widehat{E}} = \underline{\breve{E}}(\underline{E}(\cdot|\mathcal{B})). \tag{13}$$

The credal set associated with $\underline{\widehat{E}}$ is given in the following proposition in the case when all conditioning events have strictly positive lower probability.

Proposition 2. *If* $\underline{\breve{P}}(B) > 0$ *for every* $B \in \mathcal{B}$, *then*

$$\mathcal{M}(\underline{\widehat{E}}) = \{P \colon (\forall A \in \mathcal{P}(\mathcal{B}))P(A) \geq \underline{P}(A) \text{ and } (\forall A' \subseteq \Omega, B \in \mathcal{B})P(A'|B) \geq \underline{P}(A'|B)\}.$$

Now, let $\mathcal{H}$ be a subfamily of coherent lower probabilities, and assume that $\underline{\breve{P}}$ and $\underline{P}(\cdot|B)$ belong to $\mathcal{H}$ for every $B \in \mathcal{B}$. Considering our earlier discussion, we look for the joint models $\underline{\widehat{P}}$ on $\mathcal{P}(\Omega)$ such that

(TP1) $\underline{\widehat{P}}(A) \geq \underline{\breve{P}}(A)$ for every $A \in \mathcal{P}(\mathcal{B})$; [agreeing on $\mathcal{B}$]
(TP2) $\underline{\widehat{P}}(A'|B) \geq \underline{P}(A'|B)$ for every $B \in \mathcal{B}$ such that $\underline{\widehat{P}}(B) > 0$ and every $A' \subseteq \Omega$; [rigidity]
(TP3) $\underline{\widehat{P}} \in \mathcal{H}$, [closure]

and in particular, for the smallest such model, if it exists.

Trivially, if the marginal extension $\underline{\widehat{E}}$ belongs to $\mathcal{H}$, then it is the smallest model satisfying conditions (TP1)–(TP3) above; however, it is not hard to see that properties such as complete monotonicity are not generally preserved by marginal extension:

Example 1. Let $\Omega = \{\omega_1, \dots, \omega_8\}, B = \{\omega_1, \dots, \omega_4\}$ and the partition $\mathcal{B} = \{B, B^c\}$. Let $P(\cdot|B)$ the precise prevision associated with $(0.3, 0.15, 0.15, 0.4)$, and $P(\cdot|B^c)$ the precise prevision associated with $(0.2, 0.25, 0.25, 0.3)$. Consider on the other hand the vacuous lower probability $\underline{\breve{P}}$ on $\{B, B^c\}$, given by $\underline{\breve{P}}(B) = \underline{\breve{P}}(B^c) = 0, \underline{\breve{P}}(\Omega) = 1$. Given $A_1 = \{\omega_1, \omega_2, \omega_5, \omega_6\}$ and $A_2 = \{\omega_1, \omega_3, \omega_5, \omega_7\}$, we get:

- $P(A_1|B) = 0.45 = P(A_1|B^c) \Rightarrow \underline{\widehat{E}}(A_1) = 0.45$;
- $P(A_2|B) = 0.45 = P(A_2|B^c) \Rightarrow \underline{\widehat{E}}(A_2) = 0.45$;
- $P(A_1 \cup A_2|B) = 0.6, P(A_1 \cup A_2|B^c) = 0.7 \Rightarrow \underline{\widehat{E}}(A_1 \cup A_2) = 0.6$;
- $P(A_1 \cap A_2|B) = 0.3, P(A_1 \cap A_2|B^c) = 0.2 \Rightarrow \underline{\widehat{E}}(A_1 \cap A_2) = 0.2$;

This implies that $\underline{\widehat{E}}(A_1 \cup A_2) + \underline{\widehat{E}}(A_1 \cap A_2) = 0.8 < 0.9 = \underline{\widehat{E}}(A_1) + \underline{\widehat{E}}(A_2)$, whence $\underline{\widehat{E}}$ is not 2-monotone. ♢

On the other hand, any model satisfying conditions (TP1)–(TP3) is an *inner approximation* of the marginal extension:

Proposition 3. *Let $\widehat{\underline{P}} \in \mathcal{H}$ be a coherent lower prevision. Then $\widehat{\underline{P}}$ satisfies conditions (TP1) and (TP2) if and only if it is an inner approximation of the marginal extension $\breve{\underline{E}}(\underline{E}(\cdot|\mathcal{B}))$.*

In next sections, we shall consider a number of imprecise probability models $\mathcal{H}$ and look for an aggregation procedure T that satisfies condition (TP1)–(TP3). Specifically, we shall study in which cases the marginal extension $M(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ satisfies these three conditions, and, if it does not, whether it is possible (a) to characterise for which cases of $\underline{P}, \underline{P}(\cdot|\mathcal{B})$ it does; and (b) to determine a unique inner approximation of $M(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ satisfying (TP1)–(TP3). We will consider first of all some subfamilies of coherent lower probabilities: minitive measures (Section 3), belief functions (Section 4) and distortion models (Section 5). In the second half of the paper, we will investigate if the marginal extension theorem can be generalised to coherent sets of desirable gambles (Section 6) or choice functions (Section 7). The formulation of the rule in those two cases will require us to specify how to derive a conditional model from an unconditional one, generalising Eq. (10) to those contexts; details will be given in the corresponding sections.

3. Minitive Models

We begin by considering the class of *minitive measures*, which are those coherent lower probabilities satisfying $\underline{P}(A_1 \cap A_2) = \min\{\underline{P}(A_1), \underline{P}(A_2)\}$ for all $A_1, A_2 \subseteq \Omega$. The conjugate upper probability, given by $\overline{P}(A) = 1 - \underline{P}(A^c)$ for every $A \subseteq \Omega$ is called a *maxitive* or *possibility measure*.

The law of iterated expectation has been investigated in the context of possibility measures by Benferhat et al. (2011); the main difference lies in the use of a conditioning rule that differs from the Generalised Bayes Rule (in the case of Benferhat et al. (2011), the product- and min-based rules).

As we mentioned in Section 2, minitive measures $\underline{P}$ are a particular case of belief functions; as a consequence, their natural extension to $\mathcal{L}(\Omega)$ is given by Eq. (6). However, given minitive measures $\breve{\underline{P}}$ on $\mathcal{P}(\mathcal{B})$ and $\underline{P}(\cdot|B)$ on $\mathcal{P}(\Omega)$ for each $B \in \mathcal{B}$, the coherent lower previsions $\breve{\underline{E}}$ on $\mathcal{L}(\mathcal{B})$ and $\underline{E}(\cdot|\mathcal{B})$ defined by natural extension need not be minitive: as showed by de Cooman and Miranda (2014, Prop. 7), this is only the case then the minitive measures are $\{0,1\}$-valued on events, and in that case they are associated with filters of subsets of the possibility space.

Let $\widehat{\underline{E}}$ be the marginal extension of $\breve{\underline{E}}, \underline{E}(\cdot|\mathcal{B})$. The following result gives sufficient conditions for $\widehat{\underline{E}}$ to be minitive on gambles, i.e., for it to satisfy Eq. (7), and also for $\widehat{\underline{E}}$ to be minitive on events.

Proposition 4. *Consider $\underline{P} \in \mathtt{N}_{\mathcal{B}}, \underline{P}(\cdot|\mathcal{B}) \in \mathtt{N}(\cdot|\mathcal{B})$ and let $\breve{\underline{E}}, \underline{E}(\cdot|\mathcal{B})$ be their natural extensions to $\mathcal{L}_{\mathcal{B}}, \mathcal{L}(\Omega)$. Let $\widehat{\underline{E}}$ be given by Eq.* (13)*.*

(1) If $\breve{\underline{E}}, \underline{E}(\cdot|\mathcal{B})$ are minitive on gambles, then so is $\widehat{\underline{E}}$.
(2) If either $\breve{\underline{E}}$ or $\underline{E}(\cdot|\mathcal{B})$ is minitive on gambles, then $\widehat{\underline{E}}$ is minitive on events.
(3) If both $\breve{\underline{E}}, \underline{E}(\cdot|\mathcal{B})$ are minitive on events but not on gambles, then $\widehat{\underline{E}}$ may not be minitive on events.

The reasoning in the proof of the last item, that may be found in the Appendix, allows us to build examples where both $\breve{\underline{E}}, \underline{E}(\cdot|\mathcal{B})$ are not minitive on gambles and yet $\widehat{\underline{E}}$ is minitive on events. In other words, the sufficient condition in the second item of Proposition 4 is not necessary.

Example 2. Consider $\Omega = \{a,b,c,d\}, B = \{a,b\}$ and $\mathcal{B} = \{B, B^c\}$. Let our marginal lower probability be given by $\underline{\check{P}}(B) = 0.5, \underline{\check{P}}(B^c) = 0$; it is minitive on events (any coherent lower probability on a binary space is) but not on gambles, because it is not 0-1-valued. Consider next the conditional lower previsions $\underline{P}(\cdot|B), \underline{P}(\cdot|B^c)$ given by $\underline{P}(\{a\}|B) = 1, \underline{P}(\{b\}|B) = 0$ and $\underline{P}(\{c\}|B^c) = 0, \underline{P}(\{d\}|B^c) = 0.5$. Then it can be checked that $\underline{M}(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ determines the following lower probability $\widehat{\underline{E}}$:

A	$\widehat{\underline{E}}(A)$
$\{a\}$	0.5
$\{b\}$	0
$\{c\}$	0
$\{d\}$	0
$\{a,b\}$	0.5
$\{a,c\}$	0.5
$\{a,d\}$	0.75
$\{b,c\}$	0
$\{b,d\}$	0
$\{c,d\}$	0
$\{a,b,c\}$	0.5
$\{a,b,d\}$	0.75
$\{a,c,d\}$	1
$\{b,c,d\}$	0
$\{a,b,c,d\}$	1

It follows that this lower probability is minitive on events. ◊

On the other hand, it may also happen that $\widehat{\underline{E}}$ is minitive on gambles even if $\underline{E}(\cdot|B)$ is only minitive on events (i.e., not $\{0,1\}$-valued); a sufficient condition for this is, after denoting $\mathcal{E}_{\mathcal{B}} := \{B \in \mathcal{B} \colon \underline{P}(\cdot|B) \text{ not minitive on gambles}\}$, that $\overline{\check{E}}(\bigcup_{B\in\mathcal{E}_{\mathcal{B}}} B) = 0$. However, it is a consequence of (de Cooman and Miranda, 2014, Prop. 6) that the following statements are equivalent:

(a) $\underline{P}(\cdot|B)$ is $\{0,1\}$-valued on events for every $B \in \mathcal{B}$;
(b) $\underline{\check{P}}(\underline{P}(\cdot|\mathcal{B}))$ is minitive on gambles for every coherent lower prevision $\underline{\check{P}}$ that is minitive on gambles.

When $\widehat{\underline{E}}$ is not minitive on events, it is not hard to find the smallest coherent lower probability that dominates it and is minitive, provided at least one such inner approximation exists. This is a consequence of the following result. Its proof is immediate and therefore omitted.

Proposition 5. *Let $\underline{P}$ be a coherent lower probability on $\mathcal{P}(\Omega)$. Then there exists a minitive lower probability $\underline{P}'$ on $\mathcal{P}(\Omega)$ satisfying $\underline{P}'(A) \geq \underline{P}(A)$ for every $A \subseteq \Omega$ if and only if $\max_{\omega\in\Omega} \overline{P}(\{\omega\}) = 1$. In that case, the smallest such minitive lower probability is given by*

$$\underline{P}'(A) = 1 - \max_{\omega\notin A} \overline{P}(\{\omega\}). \tag{14}$$

We may summarise the results in this section as follows:

- If the class $\mathcal{C}$ is that of coherent lower previsions that are minitive on gambles, then the marginal extension is the smallest model satisfying conditions (TP1)–(TP3).

- If the class $\mathcal{C}$ is that of coherent lower previsions that are minitive on events, then the marginal extension is the smallest model satisfying conditions (TP1)–(TP3) if either the marginal or the conditional models are also minitive on gambles; otherwise, the smallest such model is determined by Eq. (14).

Therefore, in this case we give a full answer to the questions stated at the end of Section 2.

4. Belief functions

Next we consider the possible extensions of the law for belief functions. This problem, be it under the guise of the law of total probability or under Jeffrey's rule, has already been investigated in the context of belief functions in (Ma et al., 2011; Smets, 1993; Spies, 1994; Zhou and Cuzzolin, 2017; Zhou et al., 2014). In particular, the combination of marginal and conditional belief functions has been considered in the context of decision theory under objective ambiguity (Petturiti and Vantaggi, 2020, 2022b) and of statistical matching (Petturiti and Vantaggi, 2022a).

One important difference with our approach is that this literature uses a different conditioning rule than ours: while we use the GBR to obtain a conditional model $\underline{P}(\cdot|\mathcal{B})$, works in the framework of belief functions typically use Dempster's conditioning (Smets, 1993; Zhou and Cuzzolin, 2017), coarsening conditioning (Zhou et al., 2014) or geometric conditioning (Smets, 1993). However, these rules do not guarantee coherence between the joint and the conditional models, which lies at the basis of our approach. This is a consequence of the following result:

Proposition 6. *Let $\mathcal{B}$ be a partition of Ω and consider an aggregation procedure*

$$T : \mathtt{Bel}_{\mathcal{B}} \times \mathtt{Bel}(\cdot|\mathcal{B}) \to \mathtt{Bel}_{\Omega}.$$

If T satisfies the following two conditions:

(i) Given $\underline{P}^1_{\mathcal{B}} \leq \underline{P}^2_{\mathcal{B}} \in \mathtt{Bel}_{\mathcal{B}}$ and $\underline{P}^1(\cdot|\mathcal{B}) \leq \underline{P}^2(\cdot|\mathcal{B}) \in \mathtt{Bel}(\cdot|\mathcal{B})$ it holds that $T(\underline{P}^1_{\mathcal{B}}, \underline{P}^1(\cdot|\mathcal{B})) \leq T(\underline{P}^2_{\mathcal{B}}, \underline{P}^2(\cdot|\mathcal{B}))$; [monotonicity]

(ii) If $P_{\mathcal{B}} \in \mathbb{P}_{\mathcal{B}}$ and $P(\cdot|\mathcal{B}) \in \mathbb{P}(\cdot|\mathcal{B})$, then $T(P_{\mathcal{B}}, P(\cdot|\mathcal{B})) = P_{\mathcal{B}}(P(\cdot|\mathcal{B}))$ [extension]

then $T \leq M$. As a consequence, whenever $T(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ does not coincide with $M(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ it follows that $T(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ is not coherent with $\underline{P}, \underline{P}(\cdot|\mathcal{B})$.

Note that conditions (i) and (ii) in Proposition 6 are very mild and apply to basically any aggregation procedure for belief functions in the literature: they simply entail that the output of the aggregation procedure becomes more imprecise when its inputs do, and also that it is truly a generalisation of the law of iterated expectation. Proposition 6 then means that these procedures will produce an *outer* approximation of the marginal extension in general, and as a consequence that they will not satisfy either condition (TP1) or (TP2), due to Theorem 1. Usually the aggregation procedure is designed to satisfy (TP1) (agreement with the marginal), and what will happen is that the conditional model it determines by means of GBR will be less informative then the one we started with, due to the use in these procedures of a different conditioning rule. This is illustrated in the following example:

Example 3. Let us consider $\Omega_1 = \{a,b\}, \Omega_2 = \{c,d\}$ and the belief functions $Bel, Bel(\cdot|\Omega_1)$ given by[7]

$$\begin{array}{lll} Bel(\{a\}) = 0.2, & Bel(\{b\}) = 0.6, & Bel(\{a,b\}) = 1 \\ Bel(\{c\}|\{a\}) = 0.3, & Bel(\{d\}|\{a\}) = 0.5, & Bel(\{c,d\}|\{a\}) = 1 \\ Bel(\{c\}|\{b\}) = 0.1, & Bel(\{d\}|\{b\}) = 0.4, & Bel(\{c,d\}|\{b\}) = 1 \end{array}$$

Let T denote the aggregation procedure described by Zhou and Cuzzolin (2017), that determines a belief function on $\Omega_1 \times \Omega_2$ through the following steps:

- First, a conditional embedding of $Bel(\cdot|\{a\})$, $Bel(\cdot|\{b\})$ is made, giving rise to two belief functions $\underline{P}^a, \underline{P}^b$.
- Next, we combine $\underline{P}^a$ and $\underline{P}^b$ via Dempster's rule of combination and obtain a belief function $\underline{P}$.
- In a third step, we take the vacuous extension of Bel to $\Omega_1 \times \Omega_2$, producing another belief function $\underline{P}'$.
- And finally we define $T(Bel, Bel(\cdot|\Omega_1))$ as Dempster's combination of $\underline{P}$ and $\underline{P}'$.

In order to simplify the notation, let $x_1 = \{(a,c)\}, x_2 = \{(a,d)\}, x_3 = \{(b,c)\}, x_4 = \{(b,d)\}$. Then it follows that $T(Bel, Bel(\cdot|\Omega_1))$ and $M(Bel, Bel(\cdot|\Omega_1))$ are given by:

A	T(A)	M(A)
$\{x_1\}$	0.06	0.06
$\{x_2\}$	0.1	0.1
$\{x_3\}$	0.06	0.06
$\{x_4\}$	0.24	0.24
$\{x_1, x_2\}$	0.2	0.2
$\{x_1, x_3\}$	0.126	0.14
$\{x_1, x_4\}$	0.324	0.36
$\{x_2, x_3\}$	0.17	0.18
$\{x_2, x_4\}$	0.38	0.42
$\{x_3, x_4\}$	0.6	0.6
$\{x_1, x_2, x_3\}$	0.28	0.28
$\{x_1, x_2, x_4\}$	0.52	0.52
$\{x_1, x_3, x_4\}$	0.72	0.72
$\{x_2, x_3, x_4\}$	0.8	0.8
$\{x_1, x_2, x_3, x_4\}$	1	1

We observe that $T(A) < M(A)$ for $A = \{x_1, x_3\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_4\}$ and as a consequence that T is not coherent with $Bel, Bel(\cdot|\Omega_1)$. ◊

We also see in Example 3 that the lower and upper probabilities of the singletons coincide in both the model by Zhou and Cuzzolin and in the marginal extension; this leads us to wonder whether this equality holds for arbitrary belief functions, and as a consequence whether the probability intervals induced by both models agree. A deeper study into this matter is left as future research.

[7] We are making a small abuse of notation here, in that we are using $Bel(\{c\}|\{a\})$ to denote what should be $Bel(I_{\Omega_1 \times \{c\}}|\{a\} \times \Omega_2)$, and similarly for other conditional lower probabilities. The same comment applies to Examples 4 and 5 later on.

Our proposal thus in this paper is to use if possible the marginal extension, that will correspond to the lower envelope of the probability measures that can be obtained, using the law of iterated expectation, as a combination of marginal and conditional probability measures that are compatible with $Bel_{\mathcal{B}}, Bel(\cdot|\mathcal{B})$. We would thus be making tractable the approach already discussed by Wagner (1992).

As Example 1 shows, if we apply M to a marginal and a conditional belief function the aggregated model may not be a belief function. In those cases, we may use the results by Miranda et al. (2023) and look for inner approximations by means of a linear optimization problem. These look for the model that minimises the distance defined by Baroni and Vicig (2005) with respect to the original model:

$$d(\underline{P}, \underline{Q}) = \sum_{A \subseteq \Omega} |\underline{P}(A) - \underline{Q}(A)|. \tag{15}$$

As our next example shows, this approach does not give a unique solution in general:

Example 4. Let us consider $\Omega_1 = \{a, b\}, \Omega_2 = \{c, d\}$ and the belief functions $Bel, Bel(\cdot|\Omega_1)$ given by

$$\begin{array}{lll} Bel(\{a\}) = 0, & Bel(\{b\}) = 0, & Bel(\{a,b\}) = 1 \\ Bel(\{c\}|\{a\}) = 0.4, & Bel(\{d\}|\{a\}) = 0.4, & Bel(\{c,d\}|\{a\}) = 1 \\ Bel(\{c\}|\{b\}) = 0.4, & Bel(\{d\}|\{b\}) = 0.4, & Bel(\{c,d\}|\{b\}) = 1 \end{array}$$

Then $M(Bel, Bel(\cdot|\Omega_1))$ is given in the following table:

A	M(A)	$Bel_1(A)$	$Bel_2(A)$
$\{x_1\}$	0	0.2	0
$\{x_2\}$	0	0.4	0
$\{x_3\}$	0	0	0.4
$\{x_4\}$	0	0	0.2
$\{x_1, x_2\}$	0	0.6	0
$\{x_1, x_3\}$	0.4	0.4	0.4
$\{x_1, x_4\}$	0.4	0.4	0.4
$\{x_2, x_3\}$	0.4	0.4	0.4
$\{x_2, x_4\}$	0.4	0.4	0.4
$\{x_3, x_4\}$	0	0	0.6
$\{x_1, x_2, x_3\}$	0.4	0.8	0.48
$\{x_1, x_2, x_4\}$	0.4	0.8	0.6
$\{x_1, x_3, x_4\}$	0.4	0.6	0.8
$\{x_2, x_3, x_4\}$	0.4	0.4	0.8
$\{x_1, x_2, x_3, x_4\}$	1	1	1

Then it can be checked that both Bel_1, Bel_2 are two different inner approximations in $\mathtt{Bel}_{\Omega_1 \times \Omega_2}$ that minimise the distance defined by Eq. (15) with respect to M. $\Diamond$

Nevertheless, it is possible to obtain a unique solution by means of quadratic programming, applying (Miranda et al., 2023, Prop. 3(ii)).

On the other hand, the marginal extension produces a belief function when the conditional model is vacuous, as showed in the following proposition:

Proposition 7. *Let $\mathcal{B}$ be a partition of Ω; consider a belief function $Bel_{\mathcal{B}}$ in $\mathtt{Bel}_{\mathcal{B}}$ and for each $B \in \mathcal{B}$ let $A_B \subseteq B$ and let $Bel(\cdot|B)$ be the vacuous belief function on A_B, given by*

$$Bel(C|B) = \begin{cases} 1 & \text{if } A_B \subseteq C \\ 0 & \text{otherwise.} \end{cases}$$

Denote by $Bel(\cdot|\mathcal{B})$ the conditional belief function determined by $Bel(\cdot|B), B \in \mathcal{B}$. Then $M(Bel_{\mathcal{B}}, Bel(\cdot|\mathcal{B}))$ is also a belief function on events.

This proposition shows that we can use the marginal extension in one of the contexts where belief functions arise naturally: *random sets* (Dempster, 1967; Nguyen, 1978).

5. Distortion Models

Next we consider the family of *distortion* models. These refer to those imprecise probability models that originate by transforming a probability measure P_0 using a distortion function and for some distorting factor δ. They are particularly relevant in the context of robust statistics (Huber, 1981). There are several models within this family, such as the pari-mutuel, linear-vacuous or Kolmogorov models. We refer to Montes et al. (2020a,b) for a comparison of the properties of some of the most important distortion models. A study of the connection between Jeffrey's rule and convex and bi-elastic distortion models was made by Škulj (2006). Here, we shall focus on the linear-vacuous and pari-mutuel models.

5.1. **Linear-Vacuous Mixture.** We begin our study with the family of linear-vacuous mixtures, also referred to as *contamination models* in the literature.

Definition 1. Let P_0 be a probability measure on Ω and consider a distortion factor $\delta \in (0, 1)$. The associated *linear-vacuous mixture* is given by the lower probability

$$\underline{P}(A) = \begin{cases} (1-\delta)P_0(A) & \text{if } A \neq \Omega \\ 1 & \text{otherwise} \end{cases} \qquad \text{for all } A \subseteq \Omega.$$

We say then that $\underline{P}$ is determined by (P_0, δ).

Linear-vacuous (LV) mixtures have been studied in the context of robust statistics (Huber, 1981): the set of probability measures that dominate $\underline{P}$ are the convex combinations of P_0 with any other probability measure Q, with respective weights $(1 - \delta)$ and δ. They have also been applied in the context of dynamic portfolio choice by Petturiti and Vantaggi (2024). We shall denote by $\mathcal{C}_{\text{LV}}$ the class of linear-vacuous mixtures.

It follows from the definition above that the lower probability $\underline{P}$ associated with a linear-vacuous mixture is always additive on proper subsets of Ω: given $A \subset \Omega$, it holds that

$$\underline{P}(A) = \sum_{\omega \in A} \underline{P}(\{\omega\}). \tag{16}$$

On the other hand, the natural extension from events to gambles of an LV mixture $\underline{P}$ determined by $(\underline{P}_0, \delta)$ is given by

$$\underline{P}(f) = (1-\delta)P_0(f) + \delta \min f \qquad \text{for all } f \in \mathcal{L}.$$

With these two properties we can establish necessary and sufficient conditions for the existence of an LV mixture that is coherent with the marginal and conditional models, provided no zero lower probabilities are involved. In that case, it follows from the introduction to Section 2.2 that this is the smallest LV model satisfying conditions (TP1)–(TP3).

Proposition 8. *Consider an LV mixture $\underline{\breve{E}}$ on $\mathcal{L}(\mathcal{B})$ determined by $(P_{\mathcal{B}}, \delta_{\mathcal{B}})$ and, for each $B \in \mathcal{B}$, let $\underline{E}(\cdot|B)$ be an LV mixture on $\mathcal{L}(\mathcal{B})$ determined by (P_B, δ_B). Assume that $\underline{\breve{E}}(B) > 0$ for every $B \in \mathcal{B}$, and that $\underline{E}(A|B) > 0$ for every $B \in \mathcal{B}$ and non-empty $A \subseteq B$. Let $\underline{P}$ be the LV mixture on Ω determined by (P, δ). Then the following are equivalent:*

(i) $\underline{P}$ *is coherent with* $\underline{\breve{E}}, \underline{E}(\cdot|\mathcal{B})$.
(ii) $\underline{P}(f) = \underline{\breve{E}}(f)$ *for every* $f \in \mathcal{L}(\mathcal{B})$ *and* $\underline{P}(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) = 0$ *for every* $B \in \mathcal{B}$ *and* $\omega \in B$.
(iii) $\delta = \delta_{\mathcal{B}}, P(\{\omega\}) = \frac{\underline{E}(\{\omega\}|B)(\delta+P(B))}{1-\delta}$ *and* $\delta_B = \frac{\delta+\delta P(B)}{\delta+P(B)}$ *for every* $B \in \mathcal{B}$.

It also follows from this proposition that there will be situations where there is no LV mixture inducing the same marginal and conditional models we started with, taking into account the necessary relationship between the distortion factors $\delta_{\mathcal{B}}, \delta_B, B \in \mathcal{B}$ established in the third item.

Interestingly, the marginal extension $M(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ of two LV mixtures does *never* belong to the class $\mathcal{C}_{LV}$:

Proposition 9. *Let $\underline{P}$ be an LV mixture on $\mathcal{L}(\mathcal{B})$ determined by $(P_{\mathcal{B}}, \delta_{\mathcal{B}})$, and for every $B \in \mathcal{B}$ let $\underline{P}(\cdot|B)$ be an LV mixture on Ω determined by (P_B, δ_B). Then the marginal extension $\underline{\widehat{E}} = \underline{\breve{E}}(\underline{E}(\cdot|\mathcal{B}))$ does not belong to $\mathcal{C}_{\mathrm{LV}}$.*

Taking into account Prop. 3, we may consider then the inner approximations of the marginal extension $\underline{\widehat{E}}$ in the class of LV mixtures. As we mentioned before, inner approximations of coherent lower probabilities were investigated by Miranda et al. (2023); in the case of distortion models, they are moreover linked with the notion of *centroids* of credal sets (Miranda and Montes, 2023). In (Miranda et al., 2023, Sect. 4.1) it was established that the optimal inner approximations, in that they minimise the distance defined by Eq. (15) with respect to the original model, can be determined by considering the maximum value of δ such that the lower probability defined by $\underline{Q}(A) := \frac{\underline{P}(A)}{1-\delta}$ for all $A \subset \Omega$, and $\underline{Q}(\Omega) := 1$, avoids sure loss. This result is applicable when the original lower probability is non-zero on any non-trivial event, which is also an assumption in our Propositions 2 and 8.

A word of caution here, though: the set of inner approximations of the coherent lower prevision $\underline{\widehat{E}}$ does not coincide with the set of inner approximations of the coherent lower probability that we obtain by restricting $\underline{\widehat{E}}$ to events. The reason is that $\underline{\widehat{E}}$ will *not* be in general the natural extension of the coherent lower probability that is its restriction to events, as the following example shows:

Example 5. Consider $\Omega = \Omega_1 \times \Omega_2$, where $\Omega_1 = \{a, b\}, \Omega_2 = \{c, d\}$ and let P_0 be the probability measure with mass function $P_0(\{(a, c)\}) = 0.15$, $P_0(\{(a, d)\}) = 0.35$, $P_0(\{(b, c)\}) = 0.3$ and $P_0(\{(b, d)\}) = 0.2$. Consider $\delta = 0.1$ and consider the marginal and conditional LV models $\underline{P}(\cdot|\Omega_1)$ and $\underline{P}_{\Omega_1}$ it determines, that satisfy $\underline{P}(\{a\}) = \underline{P}(\{b\}) = 0.45, \underline{P}(\{c\}|\{a\}) = 0.27, \underline{P}(\{d\}|\{a\}) = 0.63, \underline{P}(\{c\}|\{b\}) = 0.54, \underline{P}(\{d\}|\{b\}) = 0.36$. Given the marginal extension $\underline{\widehat{E}}$ of these models, its restriction to events satisfies, letting $x_1 = \{(a, c)\}, x_2 = \{(a, d)\}, x_3 = \{(b, c)\}, x_4 = \{(b, d)\}$:

A	$\underline{\widehat{E}}(A)$
$\{x_1\}$	0.1215
$\{x_2\}$	0.2835
$\{x_3\}$	0.243
$\{x_4\}$	0.162
$\{x_1, x_2\}$	0.45
$\{x_1, x_3\}$	0.3915
$\{x_1, x_4\}$	0.3105
$\{x_2, x_3\}$	0.5805
$\{x_2, x_4\}$	0.4815
$\{x_3, x_4\}$	0.45
$\{x_1, x_2, x_3\}$	0.747
$\{x_1, x_2, x_4\}$	0.648
$\{x_1, x_3, x_4\}$	0.5985
$\{x_2, x_3, x_4\}$	0.7965
$\{x_1, x_2, x_3, x_4\}$	1

Then we observe that the probability measure P_1 with mass function

$$P_1(\{(a,c)\}) = P_1(\{(b,d)\}) = 0.2, \quad P_1(\{(a,d)\}) = P_1(\{(b,c)\}) = 0.3$$

satisfies $P_1(A) \geq \underline{\widehat{E}}(A)$ for all A. However, it does not belong to $\mathcal{M}(\underline{\widehat{E}})$, because for instance $P_1(\{d\}|\{a\}) = 0.6 < \underline{P}(\{d\}|\{a\})$. $\lozenge$

This means that if our goal is to inner approximate the coherent lower prevision $\underline{\widehat{E}}$ the results by Miranda et al. (2023) are not immediately applicable, and should be generalised appropriately. We conjecture that there will not be in general a unique optimal inner approximation of $\underline{\widehat{E}}$ in the class of LV mixtures. Interestingly, this issue does not arise if we considered *outer* approximations of the coherent lower prevision $\underline{\widehat{E}}$; see (Montes et al., 2018, Sect. 6) for more details.

We also deduce from Proposition 8 that in some cases there will be a LV mixture that is coherent with $\underline{P}, \underline{P}(\cdot|\mathcal{B})$; this model will be an inner approximation of the marginal extension $\underline{\widehat{E}}$, and in the conditions of the proposition it will induce the same marginal and conditional; thus, in those cases there will be a unique solution to our problem. When the conditions in Proposition 8 do not hold, we should look for a/the smallest inner approximation of $\underline{\widehat{E}}$ in the family $\mathcal{C}_{LV}$, and this inner approximation will in general induce marginal and conditional models that are more informative than the ones we started with.

5.2. **Pari-Mutuel Model.** The second distortion model we consider in this paper is the pari-mutuel model (PMM) (Walley, 1991). It is usually formulated in terms of its upper probability.

Definition 2. Given a probability measure P_0 and a distortion factor $\delta > 0$, the associated *pari-mutuel model* is given by $\overline{P}(A) = \min\{1, (1+\delta)P_0(A)\}$ for every $A \subseteq \Omega$.

We shall denote by $\mathcal{C}_{\text{PMM}}$ the family of pari-mutuel models, and we refer to (Montes et al., 2019; Pelessoni et al., 2010) for a study of their mathematical properties. It follows from the definition above that any pari-mutuel model $\overline{P}$ satisfies the following additivity property:

$$\overline{P}(A) < 1 \Rightarrow \overline{P}(A) = \sum_{\omega \in A} \overline{P}(\{\omega\}).$$

We begin by showing that, in contradistinction with LV mixtures, given a marginal and a conditional pari-mutuel model, there is *never* a PMM that is coherent with both of them.

Proposition 10. *Consider a PMM $\overline{P}_{\mathcal{B}}$ on $\mathcal{L}(\mathcal{B})$ determined by $(P_0, \delta_{\mathcal{B}})$ and for each $B \in \mathcal{B}$ let $\overline{P}(\cdot|B)$ be a PMM on $\mathcal{L}(\mathcal{B})$ determined by (P_B, δ). Assume that $\overline{P}_{\mathcal{B}}(A) < 1$ for every $A \subset \Omega$ and $\overline{P}(A|B) < 1$ for every $B \in \mathcal{B}$ and $A \subseteq B$. Then there is no PMM $\overline{P}$ on $\mathcal{L}(\Omega)$ that is coherent with $\overline{P}_{\mathcal{B}}, \overline{P}(\cdot|\mathcal{B})$.*

Again, we should then look at the inner approximations of the marginal extension in order to find a suitable formulation of the law for PMMs; we expect that the results in (Miranda et al., 2023, Sect. 4.2) should be of interest, provided they are suitably extended from coherent lower probabilities to coherent lower previsions. If we focus on the restriction to events, we can see that there may be more than one optimal inner approximation:

Example 6. Consider the same marginal and conditional lower probabilities as in Example 4. Since they are defined on binary spaces, they are particular cases of pari-mutuel models. The conjugate upper probability of the marginal extension $M(Bel, Bel(\cdot|\Omega_1))$ is given in the following table:

A	$\overline{M}(A)$	$\overline{P}_1(A)$	$\overline{P}_2(A)$
$\{x_1\}$	0.6	0.3	0.36
$\{x_2\}$	0.6	0.3	0.36
$\{x_3\}$	0.6	0.3	0.24
$\{x_4\}$	0.6	0.3	0.24
$\{x_1, x_2\}$	1	0.6	0.72
$\{x_1, x_3\}$	0.6	0.6	0.6
$\{x_1, x_4\}$	0.6	0.6	0.6
$\{x_2, x_3\}$	0.6	0.6	0.6
$\{x_2, x_4\}$	0.6	0.6	0.6
$\{x_3, x_4\}$	1	0.6	0.48
$\{x_1, x_2, x_3\}$	1	0.9	0.96
$\{x_1, x_2, x_4\}$	1	0.9	0.96
$\{x_1, x_3, x_4\}$	1	0.9	0.84
$\{x_2, x_3, x_4\}$	1	0.9	0.84
$\{x_1, x_2, x_3, x_4\}$	1	1	1

Then it can be checked that both $\overline{P}_1, \overline{P}_2$ are two different pari-mutuel models that minimise the distance defined by Eq. (15) with respect to the original model; they are determined by $\delta = 0.2$ and the mass functions $P_0^1 = (0.25, 0.25, 0.25, 0.25)$ and $P_0^2 = (0.3, 0.3, 0.2, 0.2)$, respectively. ♢

6. The marginal extension theorem for Sets of Desirable Gambles

We begin now with the second part of this paper, where we look at the formulation of the marginal extension theorem for imprecise probability models that are more general than coherent lower previsions: coherent sets of desirable gambles and coherent choice functions. In this section, we review the marginal extension theorem (de Cooman and Hermans, 2008, Thm. 3) for sets of desirable gambles, and show how it implies Eq. (1). We will at the same time establish some of the notation we will need later in Section 7.

Just as we did for the precise case, we consider an initial coherent set of desirable gambles $\check{D}$ on $\mathcal{B}$, and a conditional coherent set of desirable gambles $D\rfloor\mathcal{B}$, and will look for a coherent set of desirable gambles $\widehat{D}$ on Ω that extends both of them, in the sense that $\widehat{D}$ agrees on $\mathcal{B}$ with $\check{D}$, and is rigid in that it agrees with $D\rfloor\mathcal{B}$ about its conditional information. Before we can start this, we must look for a suitable reformulation of conditions (TP1) and (TP2). In order to do so, we will need to be able to relate a set of desirable gambles on Ω with one on $\mathcal{B}$. To this end, we will use the simplifying device of *equating a gamble f on $\mathcal{B}$ with its cylindrical extension $f^\star$ on* Ω, given by:

$$f^\star(\omega) \coloneqq f(B) \text{ for the (unique) } B \text{ in } \mathcal{B} \text{ such that } \omega \in B$$

for any ω in Ω. Our not notationally distinguishing between f and its cylindrical extension $f^\star$ will mostly be harmless for the reader. As $\check{D}$ consists of gambles on $\mathcal{B}$, using this device we may interpret $\check{D}$ as a subset of $\mathcal{L}_{\mathcal{B}}$.

We are now in a position to formulate (TP1) (agreeing on $\mathcal{B}$) in this context:

$$(\forall f \in \check{D}) f \in \widehat{D} \text{ or, in other words, } \check{D} \subseteq \widehat{D}.$$

In words, "agreeing on $\mathcal{B}$" means that we should preserve all the assessments about $\mathcal{B}$ made by $\check{D}$.

In order to formulate (TP2), we need to explain how to condition a coherent set of desirable gambles. Given a set F of gambles on B, we let $\mathbb{I}_B F \coloneqq \{\mathbb{I}_B f \colon f \in F\}$ be a set of gambles on Ω whose elements agree with the elements of F on B, and are 0 elsewhere. If we consider then a coherent set of desirable gambles $D \subseteq \mathcal{L}(\Omega)$ and a non-empty event $B \subseteq \Omega$, the set $D\rfloor B \coloneqq \{f \in \mathcal{L}(B) \colon \mathbb{I}_B f \in D\}$ on B contains the called-off gambles that are desirable. This conditioning rule preserves coherence; we refer to de Cooman and Quaeghebeur (2012) for more details.

The reformulation of (TP2) (rigidity) in this context is then:

$$(\forall B \in \mathcal{B}, f \in D\rfloor B) \mathbb{I}_B f \in \widehat{D} \text{ or, in other words, } (\forall B \in \mathcal{B}) D\rfloor B \in \widehat{D}\rfloor B.$$

In words, "rigidity" means that we should preserve all the conditional (on elements of $\mathcal{B}$) assessments present in the original D.

To summarise, in this section we look for a joint coherent set of desirable gambles $\widehat{D}$ on Ω such that

- $\widehat{D} \supseteq \check{D}$; [agreeing on $\mathcal{B}$]
- $\widehat{D}\rfloor B \supseteq D\rfloor B$ for all B in $\mathcal{B}$, [rigidity]

and in particular, for the smallest such set, if it exists.

The following result is an immediate consequence of (de Cooman and Hermans, 2008, Thm. 3), which is a more general result that holds even for arbitrary possibility spaces Ω.

Theorem 11. *The unique smallest coherent set of desirable gambles on Ω satisfying "agreeing on $\mathcal{B}$" and "rigidity" is given by*

$$\widehat{D} \coloneqq \operatorname{posi}\Big(\check{D} \cup \bigcup_{B \in \mathcal{B}} \mathbb{I}_B (D\rfloor B) \cup \mathcal{L}_{>0}\Big).$$

Here, posi is the operator that returns the smallest convex cone that includes its input set:

$$\operatorname{posi}(F) := \Big\{\sum_{k=1}^{n} \lambda_k f_k \colon n \in \mathbb{N}, \lambda_k \in \mathbb{R}_{>0}, f_k \in F\Big\}$$

for all $F \subseteq \mathcal{L}$. It will be useful later on to establish the following alternative expression for $\widehat{D}$, showing that the union with $\mathcal{L}_{>0}$ is superfluous in Theorem 11.

Proposition 12. *We have that*

$$\widehat{D} = \operatorname{posi}\Big(\check{D} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)\Big). \tag{17}$$

It is worth noting that Proposition 12 guarantees that the set of desirable gambles $\widehat{D}$ in Eq. (17) satisfies 'rigidity' with equality:

$$\widehat{D}\rfloor B = D\rfloor B \text{ for all } B \text{ in } \mathcal{B}.$$

Moreover, there is a sense in which it also satisfies to most tight form of 'agreeing on $\mathcal{B}$' possible, namely that any $\mathcal{B}$-measurable gamble belongs to $\widehat{D}$ if and only if it belongs to $\check{D}$:

$$(\forall f \in \mathcal{L}_{\mathcal{B}}) f \in \widehat{D} \Leftrightarrow f \in \check{D}.$$

Example 7. In this example we will start with an expectation operator E on $\mathcal{L}(\Omega)$ and an input expectation operator $\check{E}$ on $\mathcal{L}_{\mathcal{B}}$, and use them to define the coherent sets of desirable D and $\check{D}$, as

$$D := \{f \in \mathcal{L}(\Omega) \colon E(f) > 0 \text{ or } f \in \mathcal{L}_{>0}\}$$

and

$$\check{D} := \{f \in \mathcal{L}(\mathcal{B}) \colon \check{E}(f) > 0 \text{ or } f \in \mathcal{L}_{>0}\}.$$

Our goal is to use Proposition 12 to derive the law of iterated expectation

$$\widehat{E}(f) = \check{E}(E(f|\mathcal{B})) \quad \text{for all } f \text{ in } \mathcal{L}(\Omega). \tag{18}$$

Eq. (17) yields a coherent set of desirable gambles $\widehat{D}$ on $\mathcal{L}(\Omega)$. We will show that the lower prevision $\underline{P}_{\widehat{D}}$ associated with $\widehat{D}$, defined by Eq. (4) satisfies Eq. (18): in other words, that $\underline{P}_{\widehat{D}}(f) = \check{E}(E(f|\mathcal{B})) =: \widehat{E}(f)$ for every f in $\mathcal{L}$. To this end, it suffices to show that $\widehat{E}(f) > 0 \Rightarrow f \in \widehat{D}$ and $f \in \widehat{D} \Rightarrow \widehat{E}(f) \geq 0$, for every f in $\mathcal{L}$.

For the first implication, consider any gamble f for which $\widehat{E}(f) > 0$, and let $\alpha := \frac{\widehat{E}(f)}{2} > 0$. For every B in $\mathcal{B}$, consider the gambles $f_B \colon B \to \mathbb{R} \colon \omega \mapsto f(\omega)$ and $g_B := f_B - E(f|B) + \alpha$. The gamble f_B is the restriction of f to B, so $E(f|B) = E(f_B|B)$, and therefore $E(g_B|B) = E(f_B|B) - E(f|B) + \alpha = \alpha > 0$, whence $g_B \in D\rfloor B$. Note that

$$\begin{aligned} f &= E(f|\mathcal{B}) - \alpha + f - E(f|\mathcal{B}) + \alpha \\ &= E(f|\mathcal{B}) - \alpha + \sum_{B\in\mathcal{B}} \mathbb{I}_B(f_B - E(f|B) + \alpha) \\ &= E(f|\mathcal{B}) - \alpha + \sum_{B\in\mathcal{B}} \mathbb{I}_B g_B. \end{aligned}$$

Also, use (18) to infer that $\check{E}(E(f|\mathcal{B})-\alpha) = \check{E}(E(f|\mathcal{B}))-\alpha = \widehat{E}(f)-\alpha = \frac{\widehat{E}(f)}{2} > 0$, whence $E(f|\mathcal{B}) - \alpha \in \check{D}$. So we conclude that

$$f = \underbrace{E(f|\mathcal{B}) - \alpha}_{\in \check{D}} + \sum_{B\in\mathcal{B}} \mathbb{I}_B \underbrace{g_B}_{\in D\rfloor B},$$

whence, indeed, $f \in \operatorname{posi}(\check{D} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)) = \widehat{D}$.

For the second implication, consider any $f \in \widehat{D}$. Then, using Eq. (17) and taking into account the coherence of $\check{D}$, there are g in $\check{D}$, n in $\mathbb{N}$,$B_1, \dots, B_n$ in $\mathcal{B}$, h_1 in $D\rfloor B_1, \dots, h_n$ in $D\rfloor B_n$ and $(\mu, \lambda_1, \dots, \lambda_n) > 0$ such that $f = \mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k$, which implies that

$$\widehat{E}(f) = \check{E}\Big(E\Big(\mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k \Big| \mathcal{B}\Big)\Big),$$

where we used (18). Using the coherence of D, we may assume that all the $B_1, \dots, B_n$ are different. Let $\mathcal{E} \coloneqq \{B_k \colon k \in \{1, \dots, n\}\} \subseteq \mathcal{B}$, and note that for any B_ℓ in $\mathcal{E}$

$$E\Big(\mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k \Big| B_\ell\Big) = \mu g(B_\ell) + \lambda_\ell E(h_\ell|B_\ell)$$

and for any B in $\mathcal{B} \setminus \mathcal{E}$

$$E\Big(\mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k \Big| B\Big) = \mu g(B).$$

Hence,

$$\begin{aligned}
\widehat{E}(f) &= \check{E}\Big(E\Big(\mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k \Big| \mathcal{B}\Big)\Big) \\
&= \sum_{\ell=1}^{n} \check{E}(\mathbb{I}_{B_\ell}) E\Big(\mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k \Big| B_\ell\Big) + \sum_{B\in\mathcal{B}\setminus\mathcal{E}} \check{E}(\mathbb{I}_B) E\Big(\mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k \Big| B\Big) \\
&= \sum_{\ell=1}^{n} \check{E}(\mathbb{I}_{B_\ell}) \big(\mu g(B_\ell) + \lambda_\ell E(h_\ell|B_\ell)\big) + \sum_{B\in\mathcal{B}\setminus\mathcal{E}} \check{E}(\mathbb{I}_B) \mu g(B) \\
&= \sum_{B\in\mathcal{B}} \check{E}(\mathbb{I}_B) \mu g(B) + \sum_{\ell=1}^{n} \check{E}(\mathbb{I}_{B_\ell}) \lambda_\ell E(h_\ell|B_\ell) \\
&= \mu \check{E}(g) + \sum_{\ell=1}^{n} \lambda_\ell \check{E}(\mathbb{I}_{B_\ell}) E(h_\ell|B_\ell).
\end{aligned}$$

Since $\mu \geq 0$, $\check{E}(g) \geq 0$, $\lambda_1 \geq 0, \dots, \lambda_n \geq 0$, $E(h_1|B_1) \geq 0, \dots, E(h_n|B_n) \geq 0$, we find that, indeed, $\widehat{E}(f) \geq 0$. ◊

7. The marginal extension theorem for Choice Functions

In this section, we shall investigate how to formulate the marginal extension theorem for choice functions or, taking into account their equivalent representation, for sets of desirable gambles sets from Section 2. In order to give our formulation, we must first of all recall some additional features of this theory, related to the coherent extension of assessments and to the connection with desirability. Before we do this, in order to present the results succinctly, we first make a small digression explaining *natural extension* of an assessment, and *representation* of a set of desirable gamble sets.

Assessments may be given in the form of a subset $\mathcal{F} \subseteq \mathcal{Q}$, which contains gamble sets $F \in \mathcal{F}$ that you think desirable. If an assessment $\mathcal{F} \subseteq \mathcal{Q}$ has a coherent extension $K \supseteq \mathcal{F}$, then we call $\mathcal{F}$ *consistent.* If this is the case, De Bock and de Cooman (2018) have established that there is a unique smallest coherent extension—called *natural extension*—which is given by $\mathrm{Rs}(\mathrm{Posi}(\mathcal{F} \cup \mathcal{L}^{\mathrm{s}}(\Omega)_{>0}))$, where the two operators Rs and Posi are defined by

$$\mathrm{Rs}(\mathcal{F}) \coloneqq \{F \in \mathcal{Q} \colon (\exists G \in \mathcal{F}) G \setminus \mathcal{L}_{\leq 0} \subseteq F\}$$

and

$$\mathrm{Posi}(\mathcal{F}) \coloneqq \Bigg\{\Bigg\{\sum_{k=1}^{n} \lambda_k^{f_{1:n}} f_k \colon f_{1:n} \in \bigtimes_{k=1}^{n} F_k\Bigg\} \colon \\ n \in \mathbb{N}, F_1, \dots, F_n \in \mathcal{F}, \Big(\forall f_{1:n} \in \bigtimes_{k=1}^{n} F_k\Big) \lambda_{1:n}^{f_{1:n}} > 0\Bigg\}$$

for all $\mathcal{F}$ in $\mathbf{K}$, and where $\mathcal{L}^{\mathrm{s}}(\Omega)_{>0} \coloneqq \{\{f\} \colon f \in \mathcal{L}_{>0}\}$, $\mathcal{L}_{\leq 0} \coloneqq \{f \in \mathcal{L} : f \leq 0\}$.

Given a set of desirable gamble sets K, its *binary* part $D_K \coloneqq \{f \in \mathcal{L} \colon \{f\} \in K\}$ summarises all the binary preferences present in K: D_K collects the gambles f that form desirable gamble sets $\{f\}$. If K is coherent, then so is D_K (De Bock and de Cooman, 2018, Lem. 18).

Conversely, given a set of desirable gambles D, there may be multiple sets of desirable gamble sets K that are compatible with it, in the sense that $D_K = D$: the non-empty set $\{K \in \mathbf{K} \colon D_K = D\}$ may contain more than one element. However, if D is coherent, it always contains one unique smallest element (Van Camp and Miranda, 2020, Prop. 5) $K_D \coloneqq \{F \in \mathcal{Q} \colon F \cap D \neq \emptyset\}$, which is then equal to $\bigcap\{K \in \overline{\mathbf{K}} \colon D_K = D\}$, where $\overline{\mathbf{K}}$ denotes the collection of all coherent sets of desirable gamble sets. If we generalise K_D's definition above to arbitrary subsets $D \subseteq \mathcal{L}$, then De Bock and de Cooman (2019, Prop. 8) have established that K_D is coherent if and only if D is.

In that same paper, they establish the following useful representation result; recall that $\overline{\mathbf{D}}$ denotes the collection of all coherent sets of desirable gambles.

Theorem 13. *(De Bock and de Cooman, 2019, Thm. 9) Any set of desirable gamble sets K is coherent if and only if there is a non-empty set $\mathbf{D} \subseteq \overline{\mathbf{D}}$ such that $K = \bigcap_{D \in \mathbf{D}} K_D$. We then say that $\mathbf{D}$ represents K. Moreover, K's largest representing set is $\mathbf{D}(K) \coloneqq \{D \in \overline{\mathbf{D}} \colon K \subseteq K_D\}$.*

Let us focus next on the formulation of the marginal extension. As before, we consider an initial coherent set of desirable gamble sets $\breve{K}$ on $\mathcal{B}$, and a conditional coherent set of desirable gamble sets $K\rfloor\mathcal{B}$, and will look for a coherent set of desirable gamble sets $\widehat{K}$ on Ω that extends both of them. To do so, we must again suitably reformulate (TP1) and (TP2) to the current context. The idea is same as for sets of desirable gambles: we use the simplifying device of *equating a gamble set F on $\mathcal{B}$ with its cylindrical extension $F^\star$ on* Ω, given by $F^\star \coloneqq \{f^\star \colon f \in F\}$. As $\breve{K}$ consists of gamble sets on $\mathcal{B}$, using this device we may interpret $\breve{K}$ as a subset of $\mathcal{Q}(\Omega)$.

We are now in a position to formulate (TP1) (agreeing on $\mathcal{B}$) in the current context:

$$(\forall F \in \check{K})F \in \widehat{K} \text{ or, in other words, } \check{K} \subseteq \widehat{K}.$$

In order to formulate (TP2), we must specify how to condition a set of desirable gamble sets. Given a set of desirable gamble sets K on Ω and a (non-empty) conditioning event $B \subseteq \Omega$, we define the conditioned set of desirable gamble set $K\rfloor B := \{F \in \mathcal{Q}(B) \colon \mathbb{I}_B F \in K\} \subseteq \mathcal{Q}(B)$ as the collection of the called-off versions of gamble sets present in K. This conditioning rule preserves coherence (Van Camp and Miranda, 2020, Prop. 7). It furthermore is compatible with the conditioning rule for sets of desirable gambles: conditioning a coherent K yields a conditioned coherent set of desirable gamble sets $K\rfloor B$ that is represented by $\{D\rfloor B \colon D \in \mathbf{D}(K)\}$ (Van Camp et al., 2023, Prop. 7).

The reformulation of (TP2) (agreeing on $\mathcal{B}$) in this context is then:

$$(\forall B \in \mathcal{B}, F \in K\rfloor \mathcal{B})\mathbb{I}_B F \in \widehat{K} \text{ or, in other words, } (\forall B \in \mathcal{B})K\rfloor F \subseteq \widehat{K}\rfloor F.$$

To summarise, in this section we look for a joint coherent set of desirable gamble sets $\widehat{K}$ on Ω such that

- $\widehat{K} \supseteq \check{K}$; [agreeing on $\mathcal{B}$]
- $\widehat{K}\rfloor B \supseteq K\rfloor B$ for all B in $\mathcal{B}$. [rigidity]

and in particular, for the smallest such set, if it exists.

Theorem 14. *The unique smallest coherent set of desirable gamble sets $\widehat{K}$ on Ω satisfying "agreeing on $\mathcal{B}$" and "rigidity" is given by*

$$\widehat{K} := \operatorname{Rs}\Big(\operatorname{Posi}\Big(\check{K} \cup \bigcup_{B \in \mathcal{B}} \mathbb{I}_B(K\rfloor B) \cup \mathcal{L}^{\mathrm{s}}(\Omega)_{>0}\Big)\Big) = \operatorname{Rs}\Big(\operatorname{Posi}\Big(\check{K} \cup \bigcup_{B \in \mathcal{B}} \mathbb{I}_B(K\rfloor B)\Big)\Big).$$

Moreover, $\widehat{K}$ is represented by

$$\widehat{\mathbf{D}} := \Big\{\operatorname{posi}\Big(\check{D} \cup \bigcup_{B \in \mathcal{B}} \mathbb{I}_B(D\rfloor B)\Big) \colon \check{D} \in \mathbf{D}(\check{K}), D \in \mathbf{D}(K)\Big\}.$$

8. Conclusions

In an imprecise-probabilistic context, the well-known marginal extension theorem shows how to combine a marginal model with conditional ones. We have shown how it also naturally generalises the law of total probability to sets of desirable gambles, and have extended it to the even more versatile framework of sets of desirable gambles, and therefore to choice functions, too. In addition, we have also focused on more specific frameworks, and studied the specific forms the marginal extension can take in these models. We obtained a characterisation of the marginal extension for minitive measures, and showed that, perhaps not entirely unexpected, the two classes of distortion models we considered do not allow for an expression of the marginal extension within its class. Moreover, we have showed that the majority of the rules extending the law to the framework of belief functions give an outer approximation of the marginal extension, and do not satisfy thus the property of coherence.

Our results can be summarised in the following table:

Family $\mathcal{H}$	$M(\underline{P}, \underline{P}(\cdot\vert\mathcal{B})) \in \mathcal{H}$?	If not, $\exists!$ inner approximation?
Minitive	Sometimes (Prop. 4)	Yes (Prop. 5)
Belief	Sometimes (Prop. 7)	No (Ex. 4)
LV	NO (Prop. 9)	Sometimes (Prop. 8)
PMM	NO (Prop. 10)	No (Ex. 6)
Sets of desir. gambles	YES (Thm. 11)	
Choice functions	YES (Thm. 14)	

While in this paper we have only dealt with one marginal and one conditional model, we may more generally consider the case of a finite number of conditional models on nested partitions, in which case we expect that an iterative application of the marginal extension should provide the global model. While the marginal extension theorem has been generalised to a finite number of partitions in (de Cooman and Miranda, 2009; Miranda and de Cooman, 2007) and the results on inner approximations from Miranda et al. (2023) would also be applicable to the resulting lower probability, we should be careful in our analysis in that Jeffrey's rule does not comply with commutativity in general (Diaconis and Zabell, 1982; Wagner, 2002).

Concerning the computation of the model, in the case of the marginal extension of coherent lower previsions, a representation in terms of the extreme points of the associated credal set was established by Miranda and de Cooman (2007); when we are interested in obtaining an optimal inner approximation, the case of minitive measures has been characterised in Sect. 3, while in the case of distortion models the connection with centroids of credal sets may allow us to use the results from Miranda and Montes (2023). It may also be interesting to analyse the advantages we obtain in the particular case where the we have precise information in the marginal model, as is done by Petturiti and Vantaggi (2022a, Prop. 2) in the context of belief functions.

Note also that in our compatibility study we have required that any assessment present in the original models shall also be present in the updated ones; as such, it is sort of reminiscent of the ideas behind the *temporal coherence* considered by Zaffalon and Miranda (2013). It would also be interesting to consider this problem from the point of view of belief revision, taking into account the discussions in (Chan and Darwiche, 2003; Couso and Dubois, 2016; Marchetti and Antonucci, 2018) in the precise case. For this, the work by de Cooman (2005) and Ma et al. (2011) would be particularly relevant.

Finally, even if the results above provide some analysis of the formulation of the law under imprecision, space and time limitations have prevented us from discussing a number of interesting side topics, such as (a) a deeper study of the connection with the approaches established in the context of belief functions under other conditioning rules; (b) sufficient conditions for the (non)-uniqueness of the optimal inner approximations; (c) the study for other imprecise probability models, such as probability intervals or 2-monotone capacities; and (d) the extension to infinite spaces. We intend to address these problems in future work.

Acknowledgements

This research was partially conducted during a research visit of Enrique Miranda to the University of Bristol; the kind invitation by Jason Konek is gratefully acknowledged. We also express our thanks to the helpful comments and suggestions from the participants of the ISIPTA'2023 conference, where a preliminary version of this work was presented. This contribution is part of grant PID2022-140585NB-I00 funded by MICIU/ AEI/10.13039/501100011033 and "FEDER/UE".

References

T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors. *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

P. Baroni and P. Vicig. An uncertainty interchange format with imprecise probabilities. *International Journal of Approximate Reasoning*, 40:147–180, 2005.

S. Benferhat, K. Tabia, and K. Sedki. Jeffrey's rule of conditioning in a possibilistic framework. *Annals of Mathematics and Artificial Intelligence*, 61:185–202, 2011.

H. Chan and A. Darwiche. On the revision of probabilistic beliefs using uncertain evidence. *Artificial Intelligence*, 163:67–90, 2003.

I. Couso and D. Dubois. Belief revision and the EM algorithm. In *Proceedings of IPMU'2016*, pages 279–290, 2016.

J. De Bock and G. de Cooman. A desirability-based axiomatisation for coherent choice functions. In *Uncertainty Modelling in Data Science (Proceedings of SMPS 2018)*, pages 46–53, 2018.

J. De Bock and G. de Cooman. Interpreting, axiomatising and representing coherent choice functions in terms of desirability. In *Proceedings of ISIPTA'2019*, pages 125–134, 2019.

G. de Cooman. Belief models: an order-theoretic investigation. *Annals of Mathematics and Artificial Intelligence*, 45:5–34, 2005.

G. de Cooman. Coherent and Archimedean choice in general Banach spaces. *International Journal of Approximate Reasoning*, 140:255–281, 2022.

G. de Cooman and F. Hermans. Imprecise probability trees: Bridging two theories of imprecise probability. *Artificial Intelligence*, 172(11):1400–1427, 2008.

G. de Cooman and E. Miranda. Forward irrelevance. *Journal of Statistical Planning and Inference*, 139:256–276, 2009.

G. de Cooman and E. Miranda. Lower previsions induced by filter maps. *Journal of Mathematical Analysis and Applications*, 410(1):101–116, 2014.

G. de Cooman and E. Quaeghebeur. Exchangeability and sets of desirable gambles. *International Journal of Approximate Reasoning*, 53(3):363–395, 2012. Special issue in honour of Henry E. Kyburg, Jr.

G. de Cooman, M. Troffaes and E. Miranda. n-Monotone exact functionals. *Journal of Mathematical Analysis and Applications*, 347(1):143–156, 2008.

A. Dempster. Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38(2):325–339, 1967.

P. Diaconis and S. Zabell. Updating subjective probability. *Journal of the American Statistical Association*, 77:822–830, 1982.

L. Dubins. Finitely additive conditional probabilities, conglomerability and disintegrations. *The Annals of Probability*, 3(1):89–99, 1975.

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

P. J. Huber. *Robust Statistics*. John Wiley & Sons, New York, 1981.

R. Jeffrey. *The logic of decision*. McGraw, 1965.

R. Jeffrey. Conditioning, kinematics and exchangeability. *Causation, chance and credence*, 1:221–255, 1988.

I. Levi. *The Enterprise of Knowledge*. MIT Press, 1980.

J. Kadane, M. Schervish, and T. Seidenfeld. A Rubinesque Theory of Decision. In *A Festschrift for Herman Rubin*, volume 45 of *Lecture Notes- Monograph Series*, pages 45–55. Institute of Mathematical Statistics, 2004.

J. Ma, W. Liu, D. Dubois, and H. Prade. Bridging Jeffrey's rule, AGM revision and Dempster conditioning in the theory of evidence. *International Journal of Artificial Intelligence Tools*, 20(4):691–720, 2011.

S. Marchetti and A. Antonucci. Imaginary kinematics. In *Proceedings of UAI'2018*, pages 104–113, 2018.

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

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.

E. Miranda, G. de Cooman, and I. Couso. Lower previsions induced by multi-valued mappings. *Journal of Statistical Planning and Inference*, 133(1) 173–197, 2005.

E. Miranda and I. Montes. Centroids of the core of exact capacities: a comparative study. *Annals of Operations Research*, 321(1-2):409–449, 2023.

E. Miranda and A. Van Camp. A study of Jeffrey's Rule with imprecise probability models. In *Proceedings of ISIPTA'2023*. PMLR 215:344–355, 2023.

E. Miranda, I. Montes, and A. Presa. Inner approximations of coherent lower probabilities and their application to decision making problems. *Annals of Operations Research*, 2023. In press. https://doi.org/10.1007/s10479-023-05577-y

I. Montes, E. Miranda, P. Vicig. 2-monotone outer approximations of coherent lower probabilities. *International Journal of Approximate Reasoning*, 101:181–205, 2018.

I. Montes, E. Miranda, and S. Destercke. Pari-mutuel probabilities as an uncertainty model. *Information Sciences*, 481:550–573, 2019.

I. Montes, E. Miranda, and S. Destercke. Unifying neighbourhood and distortion models. Part I: new results on old models. *International Journal of General Systems*, 49(6):606–635, 2020a.

I. Montes, E. Miranda, and S. Destercke. Unifying neighbourhood and distortion models. Part II: new models and synthesis. *International Journal of General Systems*, 49(6):636–674, 2020b.

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

R. Pelessoni, P. Vicig, and M. Zaffalon. Inference and risk measurement with the pari-mutuel model. *International Journal of Approximate Reasoning*, 51(9): 1145–1158, 2010.

D. Petturiti and B. Vantaggi. Modeling agent's conditional preferences under objective ambiguity in Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 119:151–176, 2020.

D. Petturiti and B. Vantaggi. Probability envelopes and their Dempster-Shafer approximations in statistical matching. *International Journal of Approximate Reasoning*, 150:199–222, 2022a.

D. Petturiti and B. Vantaggi. Conditional decisions under objective and subjective ambiguity in Dempster-Shafer theory. *Fuzzy Sets and Systems*, 447:155–181, 2022b.

D. Petturiti and B. Vantaggi. The impact of ambiguity on dynamic portfolio selection in the epsilon-contaminated binomial market model. *European Journal of Operational Research*, 314(3):1029–1039, 2024.

E. Quaeghebeur. Desirability. In T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors, *Introduction to Imprecise Probabilities*, chapter 1, pages 1–27. John Wiley & Sons, 2014.

T. Seidenfeld, M. Schervish, and J. Kadane. Decisions without ordering. In W. Sieg, editor, *Acting and reflecting*, volume 211 of *Synthese Library*, pages 143–170. Kluwer, Dordrecht, 1990.

T. Seidenfeld, M. Schervish, and J. Kadane. Coherent choice functions under uncertainty. *Synthese*, 172(1):157–176, 2010.

G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, 1976.

D. Škulj. Jeffrey's conditioning rule in neighbourhood models. *International Journal of Approximate Reasoning*, 42:192–211, 2006.

P. Smets. Jeffrey's rule of conditioning generalized to belief functions. In *Proceedings of UAI'1993*, pages 500–505, 1993.

M. Spies. Conditional events, conditioning, and random sets. *IEEE Transactions on Systems, Man and Cybernetics*, 24:1755–1763, 1994.

A. Van Camp, G.de Cooman, E. Miranda, and E. Quaeghebeur. Coherent choice functions, desirability and indifference. *Fuzzy Sets and Systems*, 341:, 1-36, 2018.

A. Van Camp and E. Miranda. Modelling epistemic irrelevance with choice functions. *International Journal of Approximate Reasoning*, 125:49–72, 2020.

A. Van Camp, K. Blackwell, and J. Konek. Independent natural extension for choice functions. *International Journal of Approximate Reasoning*, 152:390–413, 2023.

C. Wagner. Generalizing Jeffrey conditionalization. In *Proceedings of UAI'1992*, pages 331–335, 1992.

C. Wagner. Probability kinematics and commutativity. *Philosophy of Science*, 69: 266–278, 2002.

P. Walley. *Coherent lower (and upper) probabilities*. Technical report, University of Warwick, Coventry, 1981, Statistics research report 22.

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

M. Zaffalon and E. Miranda. Probability and time. *Artificial Intelligence*, 198:1–51, 2013.

C. Zhou and F. Cuzzolin. The total belief theorem. In *Proceedings of UAI'2017*, 2017.

C. Zhou, M. Wang, and B. Qin. Belief-kinematics Jeffrey's rules in the theory of evidence. In *Proceedings of UAI'2014*, pages 917–926, 2014.

Appendix: proofs

Proof of Proposition 2. We will show (i) $\mathcal{M}(\underline{\widehat{E}}) \subseteq \{P\colon (\forall A \in \mathcal{P}(\mathcal{B}))P(A) \geq \underline{P}(A)$ and $(\forall A' \subseteq \Omega, B \in \mathcal{B})P(A'|B) \geq \underline{P}(A'|B)\}$ and (ii) $\mathcal{M}(\underline{\widehat{E}}) \supseteq \{P\colon (\forall A \in \mathcal{P}(\mathcal{B}))P(A) \geq \underline{P}(A)$ and $(\forall A' \subseteq \Omega, B \in \mathcal{B})P(A'|B) \geq \underline{P}(A'|B)\}$. For (i), consider any P in $\mathcal{M}(\underline{\widehat{E}})$, which implies that $P \geq \underline{\widehat{E}}$. Since $\underline{\widehat{E}} \geq \underline{\breve{P}}$ as $\underline{\widehat{E}}$ extends $\underline{\breve{E}}$ [and $\underline{\breve{E}}$ is $\breve{P}$'s natural extension] and $\underline{\breve{P}}(B) > 0$ for every B in $\mathcal{B}$, we find that also $P(B) > 0$ for all $B \in \mathcal{B}$, whence $P = P(P(\cdot|\mathcal{B}))$, so $P(\cdot|B)$ is determined uniquely by Bayes' rule. This implies that $P(f) \geq \underline{\breve{E}}(f)$ for every $f \in \mathcal{L}(\mathcal{B})$, and in particular that $P(A) \geq \underline{P}(A)$ for every $A \in \mathcal{P}(\mathcal{B})$. Moreover, this also implies that $P(f|B) \geq \underline{E}(f|B)$ for every $B \in \mathcal{B}$ and every $f \in \mathcal{L}$, whence $P(A'|B) \geq \underline{P}(A'|B)$ for all $A' \subseteq \Omega$ and $B \in \mathcal{B}$.

To show (ii), the inverse inclusion, consider any P such that $P(A) \geq \underline{P}(A)$ for all $A \in \mathcal{P}(\mathcal{B})$ and $P(A'|B) \geq \underline{P}(A'|B)$ for all $A' \subseteq \Omega$ and $B \in \mathcal{B}$. It follows by natural extension that $P(f|B) \geq \underline{E}(f|B)$ for every f in $\mathcal{L}(B)$ and B in $\mathcal{B}$, and similarly, that $P(g) \geq \underline{\breve{E}}(g)$ for every g in $\mathcal{L}_{\mathcal{B}}$. consequence,

$$P(f) = P(P(f|\mathcal{B})) \geq \underline{\breve{E}}(\underline{E}(f|\mathcal{B})) = \underline{\widehat{E}}(f)$$

for every $f \in \mathcal{L}(\Omega)$, which completes the proof.

Proof of Proposition 3. Assume first of all that $\underline{\widehat{P}} \geq \underline{\widehat{E}} = \underline{\breve{E}}(\underline{E}(\cdot|\mathcal{B}))$. Then for any gamble $f \in \mathcal{L}_{\mathcal{B}}$ we infer $\underline{\widehat{P}}(f) \geq \underline{\widehat{E}}(f) = \underline{\breve{E}}(f)$, whence (TP1) holds. With respect to (TP2), for any $B \in \mathcal{B}$ such that $\underline{\widehat{P}}(B) > 0$, the conditional $\underline{\widehat{P}}(\cdot|B)$ coincides with the model $\underline{\widehat{P}}$ induces by Eq. (11); since $\mathcal{M}(\underline{\widehat{P}}) \supseteq \mathcal{M}(\underline{\widehat{E}})$, this in turn dominates the conditional induced by $\underline{\widehat{E}}$ by Eq. (11), which must then dominate $\underline{E}(\cdot|B)$, that satisfies GBR with respect to $\underline{\widehat{E}}$, using (Miranda, 2009, Lem. 2). Thus (TP2) holds.

Conversely, if (TP1) and (TP2) holds but there is some gamble such that $\underline{\widehat{P}}(f) < \underline{\widehat{E}}(f)$, then it cannot be $f \in \mathcal{L}_{\mathcal{B}}$ by (TP1); consider then the conditional lower prevision $\underline{\widehat{P}}(\cdot|\mathcal{B})$ where $\underline{\widehat{P}}(\cdot|B)$ is defined by Eq. (11) if $\underline{\widehat{P}}(B) > 0$ and $\underline{\widehat{P}}(\cdot|B) = \underline{P}(\cdot|B)$ if $\underline{\widehat{P}}(B) = 0$. Then $\underline{\widehat{P}}$ is coherent with $\underline{\widehat{P}}(\cdot|\mathcal{B})$, whence $\underline{\widehat{P}}(f) \geq \underline{\widehat{P}}(\underline{\widehat{P}}(f|\mathcal{B}))$. As a consequence, there must be some $B \in \mathcal{B}$ such that $\underline{\widehat{P}}(f|B) < \underline{P}(f|B)$. But then can neither be $\underline{\widehat{P}}(B) > 0$ (by (TP2)) nor $\underline{\widehat{P}}(B) = 0$ (by definition), which leads to a contradiction.

Proof of Proposition 4.

(1) Consider two gambles f_1, f_2 on Ω. Then $\underline{\widehat{E}}(f_1 \wedge f_2) = \underline{\breve{E}}(\underline{E}(f_1 \wedge f_2|\mathcal{B})) = \underline{\breve{E}}(g_1 \wedge g_2) = \min\{\underline{\breve{E}}(g_1), \underline{\breve{E}}(g_2)\}$, where $g_1 = \underline{E}(f_1|\mathcal{B}), g_2 = \underline{E}(f_2|\mathcal{B})$. Thus, $\underline{\widehat{E}}$ is minitive.

(2) Assume first of all that $\underline{\breve{E}}$ is minitive on gambles. Given two events A_1, A_2, infer that $\underline{\widehat{E}}(A_1 \cap A_2) = \underline{\breve{E}}(\underline{E}(A_1 \cap A_2|\mathcal{B})) = \underline{\breve{E}}(g_1 \wedge g_2) = \min\{\underline{\breve{E}}(g_1), \underline{\breve{E}}(g_2)\} = \min\{\underline{\widehat{E}}(A_1), \underline{\widehat{E}}(A_2)\}$, where $g_1 = \underline{E}(A_1|\mathcal{B}), g_2 = \underline{E}(A_2|\mathcal{B})$.

Next, if $\underline{E}(\cdot|B)$ is minitive on gambles, it is $\{0,1\}$-valued on events by (de Cooman and Miranda, 2014, Prop. 7). As a consequence, there exists a filter $\mathcal{F}_B$ such that

$$\underline{P}(A_1 \cap A_2|B) = \begin{cases} 1 & \text{if } A_1 \cap A_2 \in \mathcal{F}_B \\ 0 & \text{otherwise.} \end{cases}$$

This implies that $\underline{\widehat{E}}(A_1 \cap A_2) = \underline{\breve{E}}(H)$ for $H \coloneqq \bigcup\{B \colon A_1 \cap A_2 \in \mathcal{F}_B\}$. But since $H = H_1 \cap H_2$ for $H_1 \coloneqq \bigcup\{B \colon A_1 \in \mathcal{F}_B\}$ and $H_2 \coloneqq \bigcup\{B \colon A_2 \in \mathcal{F}_B\}$ since filters are closed under finite intersections, we deduce that $\underline{\widehat{E}}(A_1) = \underline{\breve{E}}(H_1)$ and $\underline{\widehat{E}}(A_2) = \underline{\breve{E}}(H_2)$, and therefore $\underline{\widehat{E}}(A_1 \cap A_2) = \min\{\underline{\widehat{E}}(A_1), \underline{\widehat{E}}(A_2)\}$.

(3) To see this, we need to find some $B \in \mathcal{B}$ such that $\underline{\breve{P}}(B) \in (0,1)$, which always exists because $\underline{\breve{E}}$ is not minitive on gambles. Similarly, there is some $A_1 \subset B$

such that $\underline{P}(A_1|B) \in (0,1)$. By defining the events $H_1 := A_1 \cup B^c$ and $H_2 := B$, we infer that:

$$\begin{aligned}\widehat{\underline{E}}(H_1 \cap H_2) &= \widehat{\underline{E}}(A_1) = \check{\underline{P}}(B) \cdot \underline{P}(A_1|B)\\ \widehat{\underline{E}}(H_2) &= \check{\underline{P}}(B)\\ \widehat{\underline{E}}(H_1) &= \underline{P}(A_1|B) + (1 - \underline{P}(A_1|B)) \cdot \underline{P}(B^c),\end{aligned}$$

whence $\widehat{\underline{E}}(H_1 \cap H_2) < \min\{\widehat{\underline{E}}(H_1), \widehat{\underline{E}}(H_2)\}$.

Proof of Proposition 6. Consider $\underline{P} \in \mathtt{Bel}_{\mathcal{B}}$ and $\underline{P}(\cdot|\mathcal{B}) \in \mathtt{Bel}(\cdot|\mathcal{B})$. Then given $P \geq \underline{P}$ in $\mathbb{P}_{\mathcal{B}}$ and $P(\cdot|\mathcal{B})$ in $\mathbb{P}(\cdot|\mathcal{B})$, it follows from (i)and (ii) that

$$T(\underline{P}, \underline{P}(\cdot|\mathcal{B})) \leq T(P, P(\cdot|\mathcal{B})) = P(P(\cdot|\mathcal{B})).$$

As a consequence,

$$T(\underline{P}, \underline{P}(\cdot|\mathcal{B})) \leq \inf\{P(P(\cdot|\mathcal{B})) : P \geq \underline{P}, P(\cdot|\mathcal{B}) \geq \underline{P}(\cdot|\mathcal{B})\} = M(\underline{P}, \underline{P}(\cdot|\mathcal{B})),$$

where the equality follows from (Walley, 1991, Thm.6.7.4). Since by (Walley, 1991, Thm. 6.7.2) $M(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ is the smallest coherent lower prevision that satisfies coherence with $\underline{P}, \underline{P}(\cdot|\mathcal{B})$, we deduce that whenever $T \neq M$ then $T(\underline{P}, \underline{P}(\cdot|\mathcal{B}))$ is not coherent with $\underline{P}, \underline{P}(\cdot|\mathcal{B})$.

Proof of Proposition 7. The restriction to events of $M(Bel_{\mathcal{B}}, Bel(\cdot|\mathcal{B}))$ coincides with the lower probability of the random set $\Gamma : \mathcal{B} \to \mathcal{P}(\Omega)$ given by $\Gamma(B) = A_B$, where in the initial space we take $Bel_{\mathcal{B}}$ as a lower probability. Applying (Miranda et al., 2005, Thm. 1), we deduce that the restriction to events of $M(Bel_{\mathcal{B}}, Bel(\cdot|\mathcal{B}))$ is a belief function.

Proof of Proposition 8. That the first statement implies the second is trivial. To see the converse, let us establish first of all the implication

$$(\forall \omega \in B)\underline{P}(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) = 0 \Rightarrow (\forall f \in \mathcal{L})\underline{P}(\mathbb{I}_B(f - \underline{E}(f|B))) = 0. \quad (19)$$

To this end, consider first of all any event $A \subseteq B$, and we will show that $\underline{P}(\mathbb{I}_B(\mathbb{I}_A - \underline{E}(A|B))) = 0$. If $A = B$ then we have $\underline{P}(\mathbb{I}_B(\mathbb{I}_B - \underline{E}(B|B))) = \underline{P}(\mathbb{I}_B - \mathbb{I}_B\underline{E}(B|B)) = \underline{P}(\mathbb{I}_B - \mathbb{I}_B 1) = 0$, so we assume that $A \subset B$. Then indeed

$$\begin{aligned}\underline{P}(\mathbb{I}_B(\mathbb{I}_A - \underline{E}(A|B))) &= \underline{P}\left(\sum_{\omega \in A} \mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))\right)\\ &= \sum_{\omega \in A} \underline{P}(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) = 0,\end{aligned}$$

where the first equality follows from applying Eq. (16) twice, taking into account that $\underline{E}(A|B)$ is a constant, and that coherent lower previsions satisfy constant additivity. The second equality follows once we realise that $\min \mathbb{I}_B(\mathbb{I}_A - \underline{E}(A|B)) = -\underline{E}(A|B) = -\sum_{\omega \in A} \underline{E}(\{\omega\}|B) = \sum_{\omega \in A} \min \mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))$, using that $A \subset B$, and the third one by left hand side in Eq. (19).

Next, consider a gamble f on Ω such that $f = Bf$, and let us express it as $f = \sum_{i=1}^n x_i \mathbb{I}_{A_i}$, for $x_1 > x_2 > \cdots > x_n$ and a partition $\{A_1, \ldots, A_n\}$ of B. Since a coherent lower prevision always satisfies constant additivity, we can assume without loss of generality that $x_n = 0$. Then

$$\begin{aligned}\underline{P}\left(\mathbb{I}_B(\sum_{i=1}^n x_i \mathbb{I}_{A_i} - \underline{E}(f|B))\right) &= \underline{P}\left(\sum_{i=1}^n x_i \mathbb{I}_{A_i} - \sum_{i=1}^n x_i \underline{E}(A_i|B)\right) \\ &= \underline{P}\left(\sum_{i=1}^n x_i(\mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B)))\right) \\ &= \sum_{i=1}^n x_i \underline{P}(\mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))) = 0.\end{aligned}$$

Here the first equality follows from

$$\begin{aligned}\underline{E}(f|B) &= (1-\delta_B)P_B\left(\sum_{i=1}^n x_i \mathbb{I}_{A_i}\right) + \delta_B \min_B \left(\sum_{i=1}^n x_i \mathbb{I}_{A_i}\right) \\ &= (1-\delta_B)P_B\left(\sum_{i=1}^n x_i \mathbb{I}_{A_i}\right) = \sum_{i=1}^n x_i \underline{E}(A_i|B),\end{aligned}$$

the third from

$$\begin{aligned}&\underline{P}\left(\sum_{i=1}^n x_i \mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))\right) \\ &= (1-\delta)P\left(\sum_{i=1}^n x_i \mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))\right) + \delta \min\left(\sum_{i=1}^n x_i \mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))\right) \\ &= (1-\delta)\sum_{i=1}^n x_i P(\mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))) - \delta \sum_{i=1}^n x_i \underline{E}(A_i|B) \\ &= \sum_{i=1}^n x_i \left[(1-\delta)P(\mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))) + \delta \min \mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))\right] \\ &= \sum_{i=1}^n x_i \underline{P}(\mathbb{I}_B(\mathbb{I}_{A_i} - \underline{E}(A_i|B))),\end{aligned}$$

taking $\omega \in A_n$ for the minimum in the second equality, and the fourth one by the assumption in Eq. (19). This establishes Eq. (19). Since $\underline{P}$ is coherent with $\underline{\breve{E}}, \underline{E}(\cdot|\mathcal{B})$ if and only if $\underline{P}(f) = \underline{\breve{E}}(f)$ and $\underline{P}(\mathbb{I}_B(f - \underline{E}(f|B)) = 0$ for every $f \in \mathcal{L}$ and every $B \in \mathcal{B}$, we deduce (a) from (b).

Let us now prove the equivalence between the second and third statements. To this end, note already that two LV models determined by (P_1, δ_1) and (P_2, δ_2) are equal if and only if $P_1 = P_2$ and $\delta_1 = \delta_2$. Therefore, we see that $\underline{P} = \underline{\breve{E}}$ is equivalent to (i) $(B \in \mathcal{B})P(B) = P_{\mathcal{B}}(B)$, whence $\delta = 1 - \sum_{B\in\mathcal{B}} \underline{\breve{E}}(B) = 1 - \sum_{B\in\mathcal{B}} \underline{P}_{\mathcal{B}}(B) = \delta_{\mathcal{B}}$; and (ii) $\underline{P}(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) = 0$.

Next, given $B \in \mathcal{B}$ and $\omega \in B$,

$$\begin{aligned}&\underline{P}(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) \\ &= (1-\delta)P(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) + \delta \min(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) \\ &= (1-\delta)(P(\{\omega\}) - P(B)\underline{E}(\{\omega\}|B)) + \delta(-\underline{E}(\{\omega\}|B)),\end{aligned}$$

whence $\underline{P}(\mathbb{I}_B(\mathbb{I}_{\{\omega\}} - \underline{E}(\{\omega\}|B))) = 0$ if and only if $P(\{\omega\}) = \frac{\underline{E}(\{\omega\}|B)(\delta+P(B))}{1-\delta}$. Moreover, since $P(B) = \sum_{\omega\in B} P(\{\omega\})$, we infer that

$$\sum_{\omega\in B} \frac{\underline{E}(\{\omega\}|B)(\delta + P(B))}{1-\delta} = P(B).$$

The left hand side is equal to

$$\sum_{\omega\in B} \frac{(1-\delta_B)P_B(\{\omega\}|B)(\delta+P(B))}{1-\delta} = \frac{(1-\delta_B)(\delta+P(B))}{1-\delta},$$

so this is equal to $P(B)$ if and only if

$$\delta_B = \frac{\delta+\delta P(B)}{\delta+P(B)}.$$

This completes the proof.

Proof of Proposition 9. Consider the conditional probability measure $P(\cdot|\mathcal{B})$ given by $P(A|B) = P_B(A)$ for every B in $\mathcal{B}$ and $A \subseteq B$, and let P_0 denote the probability measure on Ω determined by $P_{\mathcal{B}}, P(\cdot|\mathcal{B})$. Then it holds that, for any $B \in \mathcal{B}$ and $\omega \in B$,

$$\widehat{\underline{E}}(\{\omega\}) = \check{\underline{E}}((1-\delta_B)P(\{\omega\}|B)\mathbb{I}_B) = (1-\delta_{\mathcal{B}})(1-\delta_B)P_0(\{\omega\}),$$

whence

$$\sum_{\omega\in B} \widehat{\underline{E}}(\{\omega\}) = (1-\delta_{\mathcal{B}})(1-\delta_B)\sum_{\omega\in B} P_0(\{\omega\}) = (1-\delta_{\mathcal{B}})(1-\delta_B)P_0(B),$$

while

$$\widehat{\underline{E}}(B) = \check{\underline{E}}(\mathbb{I}_B) = (1-\delta_{\mathcal{B}})P_{\mathcal{B}}(B).$$

Thus, Eq. (16) is not satisfied and therefore $\widehat{\underline{E}} \notin \mathcal{C}_{\mathrm{LV}}$, since δ_B and $\delta_{\mathcal{B}}$ belong to the open interval $(0,1)$.

Proof of Proposition 10. In order to establish this, we shall need to use the expression of the natural extension of a PMM from events to gambles. It is given by (Pelessoni et al., 2010; Walley, 1991):

$$\overline{E}(f) = f_\tau + (1+\delta)P((f-f_\tau)^+), \tag{20}$$

where $\tau = \frac{\delta}{1+\delta}$, $f_\tau = \sup\{x \in \mathbb{R} \colon P(\{f \le x\}) \le \tau\}$ and $(f-f_\tau)^+ = \max\{f-f_\tau, 0\}$.

Assume *ex absurdo* that there is some such PMM $\overline{P}$, and let (P,δ) be its associated parameters. Consider a gamble f on B given by $f = \sum_{i=1}^n x_i\mathbb{I}_{A_i}$ for $x_1 = 1 > x_2 > \cdots > x_n = 0$ and for a partition $\{A_1,\ldots,A_n\}$ of B, and let us characterise under which conditions we have that $\overline{P}(\overline{P}(f|B) - Bf) = 0$. Also, by coherence we get that

$$\underline{P}(A) \ge \underline{P}_{\mathcal{B}}(\underline{P}(A|\mathcal{B})) > 0 \tag{21}$$

for any event A.

First of all, taking into account that for any $x > x_n$ it holds that

$$P(\{f \le x\}) \ge P(A_n) > \frac{\delta}{1+\delta},$$

since $(1+\delta)P(A_n) - \delta = \underline{P}(A_n|B) > 0$ by assumption, we deduce that if we apply Eq. (20) to compute $\overline{P}(f|B)$ we obtain

$$\overline{P}(f|B) = x_n + (1+\delta)P_B((f-x_n)^+) = (1+\delta)P_B(f),$$

whence $\overline{P}(f|B) - Bf = (1+\delta)P_B(f) - Bf$.

On the other hand, it follows from Eq. (21) that $\underline{P}(\{\omega\}) > 0$ for every $\omega \in B$. As a consequence, defining $g := \overline{P}(f|B) - Bf$, for any value $x > \min g = \overline{P}(f|B) - x_1 = \overline{P}(f|B) - 1$ it holds that

$$P(\{g \leq x\}) \geq P(A_1) > \frac{\delta}{1+\delta},$$

since $(1+\delta)P(A_1) - \delta = \underline{P}(A_1) > 0$ by Eq. (21). Thus, Eq. (20) gives

$$\begin{aligned}\overline{E}(g) &= \overline{P}(f|B) - 1 + (1+\delta)P((g - \min g)^+) \\ &= (1+\delta)P_B(f) - 1 + (1+\delta)P(\mathbb{I}_B(1-f)).\end{aligned}$$

Therefore,

$$\begin{aligned}\overline{E}(g) = 0 &\Leftrightarrow (1+\delta)P_B(f) - 1 + (1+\delta)P(\mathbb{I}_B(1-f)) = 0 \\ &\Leftrightarrow P(f) = P_B(f) + P_0(B) - \frac{1}{1+\delta}.\end{aligned}$$

Applying this to $f = \mathbb{I}_{\{\omega\}}$ for some $\omega \in B$, we obtain that P should satisfy

$$P(\{\omega\}) = P_B(\{\omega\}) + P_0(B) - \frac{1}{1+\delta}. \tag{22}$$

This means that

$$\sum_{\omega \in B} P(\{\omega\}) = 1 + |B|P_0(B) - \frac{|B|}{1+\delta} = P_0(B) \Leftrightarrow P_0(B) = \frac{|B| - 1 - \delta}{(|B|-1)(1+\delta)},$$

and this for every $B \in \mathcal{B}$. If we consider B with more than two elements and take both $\omega_1, \omega_2 \in B$ with $\omega_1 \neq \omega_2$, then $\overline{E}(\overline{P}(\mathbb{I}_{\{\omega_1,\omega_2\}}|B) - B\mathbb{I}_{\{\omega_1,\omega_2\}}) = 0$ if and only if

$$P(\{\omega_1,\omega_2\}) = P_B(\{\omega_1,\omega_2\}) + P_0(B) - \frac{1}{1+\delta};$$

but by Eq. (22) it is

$$P(\{\omega_1,\omega_2\}) = P_B(\{\omega_1,\omega_2\}) + 2P_0(B) - 2\frac{1}{1+\delta};$$

and this can only be if $P_0(B) = \frac{1}{1+\delta}$. Since $\frac{1}{1+\delta} \neq \frac{|B|-1-\delta}{(|B|-1)(1+\delta)}$, we obtain a contradiction.

Finally, if $|B| = 2$ for all B then it must be $|\mathcal{B}| = \frac{n}{2}$. We get on the one hand $P_0(B) = \frac{1-\delta}{1+\delta}$ for all B, and the equality $1 = \sum_B P_0(B) = \frac{n}{2}P_0(B)$ implies that $\delta = \frac{n-2}{n+2}$; but on the other hand for $\underline{P}(B) > 0$ we should have then $n < 4$; this means that $n = 2$ and that $\mathcal{B}$ has only one element.

Proof of Proposition 12. We will abbreviate $F := \bigcup_{B \in \mathcal{B}} \mathbb{I}_B(D \rfloor B)$, and show that $\mathcal{L}_{>0} \subseteq \text{posi}(F)$. This will imply the desired result that $\text{posi}(\check{D} \cup F \cup \mathcal{L}_{>0}) = \text{posi}(\check{D} \cup F)$: indeed, since posi is a closure operator, we infer that $\text{posi}(\check{D} \cup F \cup \mathcal{L}_{>0}) \subseteq \text{posi}(\check{D} \cup \text{posi}(F) \cup \mathcal{L}_{>0}) = \text{posi}(\check{D} \cup \text{posi}(F)) \subseteq \text{posi}(\text{posi}(\check{D} \cup F)) = \text{posi}(\check{D} \cup F) \subseteq \text{posi}(\check{D} \cup F \cup \mathcal{L}_{>0})$, where the first equality follows once we establish that $\mathcal{L}_{>0} \subseteq \text{posi}(F)$. So consider any $f \in \mathcal{L}_{>0}$. For any B in $\mathcal{B}$, let $f_B \colon B \to \mathbb{R} \colon x \mapsto f(x)$ be f's restriction to B, so that $f = \sum_{B \in \mathcal{B}} \mathbb{I}_B f_B$. Collect in $\mathcal{E} := \{B \in \mathcal{B} \colon f_B \in \mathcal{L}_{>0}\} = \{B \in \mathcal{B} \colon f(x) > 0 \text{ for some } x \text{ in } B\} \subseteq \mathcal{B}$ the events in $\mathcal{B}$ on which f attains a positive value. That $f \in \mathcal{L}_{>0}$ implies that $\mathcal{E}$ is non-empty. For every B in $\mathcal{B} \setminus \mathcal{E}$ it follows that $f_B = 0$, and hence $f = \sum_{B \in \mathcal{E}} \mathbb{I}_B f_B$. Note that, for every B in $\mathcal{E}$, the gamble $f_B \in \mathcal{L}_{>0}$ belongs to $D\rfloor B$ by its coherence whence $\mathbb{I}_B f \in \mathbb{I}_B D\rfloor B$, and therefore, indeed, $f = \sum_{B\in\mathcal{B}} \mathbb{I}_B f_{\mathcal{B}} \in \operatorname{posi}(\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)) = \operatorname{posi}(F)$.

Lemma 15. *For any $F^\star \subseteq \mathcal{L}$ such that $\operatorname{posi}(F^\star) \cap \mathcal{L}_{<0} = \emptyset$, we have*

$$K_{\operatorname{posi}(F^\star)} = \operatorname{Rs}(\operatorname{Posi}(K_{F^\star})).$$

Proof. We will show that (i) $K_{\operatorname{posi}(F^\star)} \subseteq \operatorname{Rs}(\operatorname{Posi}(K_{F^\star}))$ and (ii) $K_{\operatorname{posi}(F^\star)} \supseteq \operatorname{Rs}(\operatorname{Posi}(K_{F^\star}))$.

For (i), consider any F in $K_{\operatorname{posi}(F^\star)}$, implying that there are n in $\mathbb{N}$, real coefficients $\lambda_{1:n} > 0$ and $g_1, \ldots, g_n$ in $F^\star$ such that $g \coloneqq \sum_{k=1}^n \lambda_k g_k \in F$. Note that the requirement $\operatorname{posi}(F^\star) \cap \mathcal{L}_{<0} = \emptyset$ implies that $g \notin \mathcal{L}_{<0}$. By letting $F_1 \coloneqq \{g_1\} \in K_{F^\star}, \ldots, F_n \coloneqq \{g_n\} \in K_{F^\star}$, and $\lambda_{1:n}^{g_{1:n}} \coloneqq \lambda_{1:n} > 0$, we find that $\{\sum_{k=1}^n \lambda_k^{f_{1:n}} f_k \colon f_{1:n} \in \times_{k=1}^n F_k\} = \{\sum_{k=1}^n \lambda_k g_k\} = \{g\}$ belongs to $\operatorname{Posi}(K_{F^\star})$, whence, indeed, $F \in \operatorname{Rs}(\operatorname{Posi}(K_{F^\star}))$ since $g \notin \mathcal{L}_{<0}$.

Conversely, for (ii), consider any F in $\operatorname{Rs}(\operatorname{Posi}(K_{F^\star}))$, so $F \supseteq F' \setminus \mathcal{L}_{>0}$ for some F' in $\operatorname{Posi}(K_{F^\star})$. This implies that $F' = \{\sum_{k=1}^n \lambda_k^{f_{1:n}} f_k \colon f_{1:n} \in \times_{k=1}^n F_k\}$ for some n in $\mathbb{N}$, $F_1, \ldots, F_n$ in $K_{F^\star}$ and real coefficients $\lambda_{1:n}^{f_{1:n}} > 0$ for every $f_{1:n}$ in $\times_{k=1}^n F_k$. That all of $F_1, \ldots, F_n$ belong to $K_{F^\star}$ means that $F_1 \cap F^\star \neq \emptyset, \ldots, F_n \cap F^\star \neq \emptyset$, so there are $g_1 \in F_1 \cap F^\star, \ldots, g_n \in F_n \cap F^\star$. Then the specific $g \coloneqq \sum_{k=1}^n \lambda_k^{g_{1:n}} g_k \in F'$ belongs to $\operatorname{posi}(F^\star)$, which tells us that $g \notin \mathcal{L}_{<0}$, and hence g belongs to F, whence $F \cap \operatorname{posi}(F^\star) \neq \emptyset$. Therefore indeed $F \in K_{\operatorname{posi}(F^\star)}$. □

Lemma 16. *For any $\mathcal{F} \subseteq \mathcal{P}(\mathcal{L})$ we have*

$$K_{\bigcup \mathcal{F}} = \bigcup_{F\in\mathcal{F}} K_F.$$

Proof. Consider any $F^\star$ in $\mathcal{Q}$, and infer that, indeed,

$$F^\star \in K_{\bigcup \mathcal{F}} \Leftrightarrow F^\star \cap \bigcup \mathcal{F} \neq \emptyset \Leftrightarrow (\exists F \in \mathcal{F}) F^\star \cap F \neq \emptyset$$
$$\Leftrightarrow (\exists F \in \mathcal{F}) F^\star \in K_F \Leftrightarrow F^\star \in \bigcup_{F\in\mathcal{F}} K_F,$$

which establishes the desired equality. □

Proof of Theorem 14. We start with the first statement. We will first show that $\mathcal{L}^{\mathrm{s}}(\Omega)_{>0} \subseteq \operatorname{Posi}(\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B))$, which will establish the second equality: indeed, abbreviating $\mathcal{F} \coloneqq \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B)$, since K and Posi are closure operators, we infer that $\operatorname{Rs}(\operatorname{Posi}(\check{K} \cup \mathcal{F} \cup \mathcal{L}^{\mathrm{s}}_{>0})) \subseteq \operatorname{Rs}(\operatorname{Posi}(\check{K} \cup \operatorname{Posi}(\mathcal{F}) \cup \mathcal{L}^{\mathrm{s}}_{>0})) = \operatorname{Rs}(\operatorname{Posi}(\check{K} \cup \operatorname{Posi}(\mathcal{F}))) \subseteq \operatorname{Rs}(\operatorname{Posi}(\operatorname{Posi}(\check{K} \cup \mathcal{F}))) = \operatorname{Rs}(\operatorname{Posi}(\check{K} \cup \mathcal{F})) \subseteq \operatorname{Rs}(\operatorname{Posi}(\check{K} \cup \mathcal{F} \cup \mathcal{L}^{\mathrm{s}}_{>0}))$, where the first equality follows once we establish that $\mathcal{L}^{\mathrm{s}}_{>0} \subseteq \operatorname{Posi}(\mathcal{F})$. So consider any $\{f\}$ in $\mathcal{L}^{\mathrm{s}}(\Omega)_{>0}$—which implies that $f \in \mathcal{L}(\Omega)_{>0}$—and any D in $\mathbf{D}(K)$. Then using the same argument as in the proof of Proposition 12 we infer that $f \in \operatorname{posi}(\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B))$, or, in other words, that $\{f\} \in K_{\operatorname{posi}(\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B))}$. Now use Lemma 15 to infer that then $\{f\} \in \operatorname{Rs}(\operatorname{Posi}(K_{\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)}))$ and hence $\{f\} \in \operatorname{Posi}(K_{\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)})$ since $f \in \mathcal{L}_{>0}$ and therefore $f \notin \mathcal{L}_{<0}$, and use subsequently Lemma 16 to infer that then $\{f\} \in \operatorname{Posi}(\bigcup_{B\in\mathcal{B}} K_{\mathbb{I}_B D\rfloor B}) = \operatorname{Posi}(\bigcup_{B\in\mathcal{B}} \mathbb{I}_B K_{D\rfloor B}) = \operatorname{Posi}(\bigcup_{B\in\mathcal{B}} \mathbb{I}_B K_D\rfloor \mathcal{B})$.

Next we show that $\widehat{K}$ satisfies "agreeing on $\mathcal{B}$" and "rigidity". To this end, note that $\widehat{K}$ satisfies "agreeing on $\mathcal{B}$" by its definition and the fact that Rs and

Posi are closure operators. Moreover, for any B in $\mathcal{B}$, we have that $\widehat{K}\rfloor B = \{F \in \mathcal{Q}(B) \colon \mathbb{I}_B F \in \widehat{K}\} \supseteq \{F \in \mathcal{Q}(B) \colon \mathbb{I}_B F \in \mathbb{I}_B(K\rfloor B)\} = K\rfloor B$ again using that Rs and Posi are closure operators, so $\widehat{K}$ satisfies "rigidity" also.

We now turn to showing that $\widehat{K}$ is coherent. To this end, we infer from (De Bock and de Cooman, 2018, Thm. 10) that if $\check{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B)$ is consistent, then $\widehat{K}$ is the expression for its natural extension, which then is guaranteed to be coherent. We verify that $\check{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B)$ is consistent by considering any $\widehat{D}$ in the non-empty $\widehat{\mathbf{D}} \subseteq \overline{\mathbf{D}}$,[8] and showing that $\check{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B)$ is a subset of $K_{\widehat{D}}$, which is a coherent set of desirable gamble sets by (De Bock and de Cooman, 2018, Lem. 12). This will prove in one fell swoop that $\widehat{K} \subseteq \bigcap_{\widehat{D}\in\widehat{\mathbf{D}}} K_{\widehat{D}}$, a useful property that we will use later on in this proof when establishing the second statement .

In order to do so, note that $\widehat{D} = \operatorname{posi}(\check{D} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B))$ for some $\check{D}$ in $\mathbf{D}(\check{K})$ and D in $\mathbf{D}(K)$. Consider any F in $\widehat{K}$, meaning that $F \supseteq F' \setminus \mathcal{L}_{<0}$ for some n in $\mathbb{N}$, $F_1, \dots, F_n$ in $\check{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B)$, and, for every $f_{1:n}$ in $\bigtimes_{k=1}^n F_k$, real coefficients $\lambda_{1:n}^{f_{1:n}} > 0$ such that $F' = \{\sum_{k=1}^n \lambda_k^{f_{1:n}} f_k \colon f_{1:n} \in \bigtimes_{k=1}^n F_k\}$. So any F_k belongs to $\check{K}$—in which case it also belongs to $K_{\check{D}}$ as $\check{D} \in \mathbf{D}(\check{K})$, and hence F_k contains a gamble $g_k \in \check{D}$—or F_k belongs to $\mathbb{I}_B(K\rfloor B)$ for some B in $\mathcal{B}$—in which case it also belongs to $\mathbb{I}_B(K_D\rfloor B) = \mathbb{I}_B(K_{D\rfloor B})$ as $D \in \mathbf{D}(K)$, and hence F contains a gamble $\mathbb{I}_B g_k$ where $g_k \in D\rfloor B$. In any case, we find that $\sum_{k=1}^n \lambda_k^{g_{1:n}} g_k \in F'$ belongs to $\operatorname{posi}(\check{D} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)) = \widehat{D}$, and hence $F' \in K_{\widehat{D}}$. This implies that, indeed, $F \in K_{\widehat{D}}$.

So we have established that $\widehat{K}$ satisfies "agreeing on $\mathcal{B}$", "rigidity" and "coherence". To complete the proof for the first statement, we show that $\widehat{K}$ is the smallest such set of desirable gamble sets. To this end, consider any set of desirable gamble sets $K^\star$ satisfying "agreeing on $\mathcal{B}$", "rigidity" and "coherence". Note that $K^\star$ must include $\check{K}$ by "agreeing on $\mathcal{B}$" and $\bigcup_{B\in\mathcal{B}} \mathbb{I}_B K\rfloor B$ by "rigidity". By "coherence" it must therefore include $\operatorname{Rs}\big(\operatorname{Posi}\big(\check{D}\cup\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)\big)\big) = \widehat{K}$, whence $K^\star \supseteq \widehat{K}$, showing that, indeed, $\widehat{K}$ is the smallest set of desirable gambles that satisfies "agreeing on $\mathcal{B}$", "rigidity" and "coherence". This also establishes that the smallest set of desirable gamble sets that satisfies "agreeing on $\mathcal{B}$", "rigidity" and "coherence" is necessarily unique.

Now we turn to the second statement. We need to show that $\widehat{K} = \bigcap_{\widehat{D}\in\widehat{\mathbf{D}}} K_{\widehat{D}}$. Recall from the proof of the first statement that $\widehat{K} \subseteq \bigcap_{\widehat{D}\in\widehat{\mathbf{D}}} K_{\widehat{D}}$, so it suffices to prove the converse set inclusion $\bigcap_{\widehat{D}\in\widehat{\mathbf{D}}} K_{\widehat{D}} \subseteq \widehat{K}$. To this end, we use that, from (Van Camp et al., 2023, Thm. 6), the natural extension of a consistent assessment $\mathcal{F}$ is given by $\bigcap\{K_D \colon D \in \overline{\mathbf{D}}, \mathcal{F} \subseteq K_D\}$. Applied to the current case, as the assessment $\check{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B)$ is already known to be consistent from the proof for the first statement, we infer that

$$\widehat{K} = \bigcap\Big\{K_D \colon D \in \overline{\mathbf{D}}, \check{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B) \subseteq K_D\Big\},$$

[8] It is a consequence of Thm. 11, which is a direct consequence of (de Cooman and Hermans, 2008, Thm. 3), that every element of $\widehat{\mathbf{D}}$ is a coherent set of desirable gambles.

and hence to establish that $\bigcap_{\widehat{D}\in\widehat{\mathbf{D}}} K_{\widehat{D}} \subseteq \widehat{K}$ it suffices to show that any $D^\star$ in $\overline{\mathbf{D}}$ such that $\breve{K} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B) \subseteq K_{D^\star}$ belongs to $\widehat{\mathbf{D}}$. So consider such a $D^\star$, which implies that $\widehat{K} \subseteq K_{D^\star}$—meaning that $D^\star \in \mathbf{D}(\widehat{K})$—and $\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(K\rfloor B) \subseteq K_{D^\star}$—meaning that $\mathbb{I}_B(K\rfloor B) \subseteq K_{D^\star}$ and hence $K\rfloor B \subseteq K_{D^\star}\rfloor B = K_{D^\star\rfloor B}$ whence $D^\star\rfloor B \in \mathbf{D}(K\rfloor B)$ for all B in $\mathcal{B}$. As $\bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D^\star\rfloor B) = \bigcup_{B\in\mathcal{B}}\{\mathbb{I}_B f\colon f \in D^\star\rfloor B\} = \bigcup_{B\in\mathcal{B}}\{\mathbb{I}_B f\colon \mathbb{I}_B f \in D^\star\} \subseteq D^\star$, we find, taking into account its coherence, that $D^\star = \operatorname{posi}(D^\star \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D^\star\rfloor B))$, whence indeed $D^\star \in \widehat{\mathbf{D}}$.

University of Oviedo, Department of Statistics and Operations Research
Email address: mirandaenrique@uniovi.es

Eindhoven University of Technology, Artificial Intelligence Group
Email address: a.van.camp@tue.nl