A Study of Jeffrey's Rule With Imprecise Probability Models

Enrique Miranda MIRANDAENRIQUE@UNIOVI.ES
Statistics and Operations Research, University of Oviedo, Spain

Arthur Van Camp ARTHUR.VANCAMP@BRISTOL.AC.UK
Department of Philosophy, University of Bristol, UK

Abstract

Jeffrey's rule tells us how to update our beliefs about a probability measure when we have updated information conditional on some partition of the possibility space, while keeping the original marginal information on this partition. It is linked to the law of total probability, and is therefore connected to the notion of marginal extension of coherent lower previsions. In this paper, we investigate its formulation for some other imprecise probability models that are either more general (choice functions) or more particular (possibility measures, distortion models) than coherent lower previsions.

Keywords: Jeffrey's rule, marginal extension, coherent lower previsions, sets of desirable gambles, non-additive measures, choice functions

1. Introduction

Consider a finite possibility space Ω, and a partition $\mathcal{B}$ of Ω. Given a probability measure P on $\mathcal{P}(\Omega)$,[1] it is possible to relate the probability $P(A)$ of any event A to the probabilities $P(B)$ of the events B in the partition $\mathcal{B}$ and the probabilities $P(A|B)$ conditional on events B in $\mathcal{B}$,[2] using the *law of total probability*:

$$P(A) = \sum_{B\in\mathcal{B}} P(B)P(A|B).$$

Suppose that we "observe" a new 'input' probability measure $\check{P}$ on $\mathcal{B}$. If we now want to obtain a new probability measure $\widehat{P}$ on Ω that satisfies the constraints

- $\widehat{P}(B) = \check{P}(B)$ for all B in $\mathcal{B}$; [agreeing on $\mathcal{B}$]
- $\widehat{P}(A|B) = P(A|B)$ for all B in $\mathcal{B}$ and $A \subseteq \Omega$, [rigidity]

then by the law of total probability

$$\widehat{P}(A) = \sum_{B\in\mathcal{B}} \widehat{P}(A|B)\widehat{P}(B) = \sum_{B\in\mathcal{B}} P(A|B)\check{P}(B) \qquad \text{(Jeffrey's rule)}$$

for all $A \subseteq \Omega$, so the two constraints above are a unique description of $\widehat{P}$ [18, 19].

The equivalent expectation operator version of Jeffrey's rule is given by

$$\widehat{E}(f) = \check{E}(E(f|\mathcal{B})) \quad \text{for all } f \text{ in } \mathcal{L}(\Omega) \qquad \text{(Jeffrey's rule for expectations)}$$

where $\mathcal{L}(\Omega)$ is the set of all real-valued maps on Ω—called *gambles*—and sometimes denoted by $\mathcal{L}$ when it is clear from the context what the domain Ω is.

In this paper, we investigate the formulation of Jeffrey's rule in the context of imprecise probabilities. One prominent such connection was already established by Peter Walley in his celebrated *marginal extension theorem* in [41]. He showed that the law of total probability can be extended to coherent lower previsions in the following manner: given a coherent lower prevision $\underline{P}_{\mathcal{B}}$ on the class of $\mathcal{B}$-measurable gambles and a separately coherent conditional lower prevision $\underline{P}(\bullet|\mathcal{B})$ on $\mathcal{L}(\Omega)$, the smallest coherent lower prevision that is coherent with $\underline{P}_{\mathcal{B}}, \underline{P}(\bullet|\mathcal{B})$ is given by

$$\underline{P} \coloneqq \underline{P}_{\mathcal{B}}(\underline{P}(\bullet|\mathcal{B})).$$

This result was later extended to a finite number of conditional lower previsions in [24] and to sets of desirable gambles in [10, Thm. 3]. In this paper, we shall look at the formulation of this result for other different families of imprecise probability models, be it more general or more specific ones than coherent lower previsions. Given one such family C, we consider thus a marginal and a conditional model within this family, and look for a global model that

(a) belongs to the same class C of uncertainty models as the marginal and conditional, and

(b) is *compatible* with them in the manner we shall specify later on.

In case there exists such as model, we may then analyse whether this model is unique, or, if it is not, whether it is possible to characterise the smallest such model. This would be a sort of natural extension, with the additional assumption of compatibility (that perhaps may be stronger

[1] We let $\mathcal{P}(\mathcal{X})$ be the power set of its input set $\mathcal{X}$. Elements of $\mathcal{P}(\Omega)$ are called events.

[2] We assume in this introductory section, for simplicity, that $P(B) > 0$ for every B in $\mathcal{B}$, making sure that $P(A|B)$ is well defined.

than usual coherence) and the structural assumption of being an element of C. When there is no such model, we may investigate if it is possible to characterise the closest approximation, for instance using inner approximations. One minor technical assumption throughout shall be that every element B of $\mathcal{B}$ contains at least two elements; otherwise the conditioning is trivial.

After recalling briefly the formulation of the marginal extension theorem for sets of desirable gambles in Section 2, we consider the cases of choice functions (Section 3), possibility measures (Section 4.1) and distortion models (Section 4.2). Some additional comments and discussion will be given in Section 5. Due to a lack of space, we've had to exclude the proofs for our results; we refer the interested reader to the supplementary material.

2. Jeffrey's Rule for Sets of Desirable Gambles

In this section we review the marginal extension theorem [10, Thm. 3] for sets of desirable gambles, and show how it implies Jeffrey's rule for expectations. We will at the same time establish some of the notation we will need later in Section 3.

We interpret the set of gambles $\mathcal{L}$ as uncertain rewards: after observing which outcome ω in Ω occurs, having a gamble $f \in \mathcal{L}$ changes your capital by $f(\omega)$, expressed in a predetermined linear utility scale. Gambles f are risky transactions: $f(\omega)$ may be negative, in which case you lose capital.

Coherent Sets of Desirable Gambles A set of desirable gambles D is a subset of $\mathcal{L}$, which contains the gambles that a subject—we will refer to the subject as 'you'—prefers over the status quo indicated by 0, which is the constant gamble that leaves your capital unchanged regardless of the outcome ω in Ω. If $f \leq 0$—by which we mean $f(\omega) \leq 0$ for every ω in Ω—then f is a gamble that never yields positive utility. The gambles in $\mathcal{L}(\Omega)_{\leq 0} := \{f \in \mathcal{L} : f \leq 0\}$—also called $\mathcal{L}_{\leq 0}$ when it is clear what the domain Ω is—should therefore never be desirable. On the other hand, if $f > 0$—by which we mean $f \geq 0$ and $f \neq 0$—then f can never make you lose utility, and there is some outcome in which f yields a (strictly) positive amount of utility. The gambles in $\mathcal{L}(\Omega)_{>0} := \{f \in \mathcal{L} : f > 0\}$—also called $\mathcal{L}_{>0}$ when it clear what the domain Ω is— should therefore always be desirable, regardless of your beliefs.

We call a set of desirable gambles D on Ω *coherent* when [13, 31, 32, 41] when for all f and g in $\mathcal{L}$ and λ in $\mathbb{R}_{>0}$

D1. $0 \notin D$; [avoiding non-positivity]

D2. $\mathcal{L}_{>0}(\Omega) \subseteq D$; [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]

Cylindrical Extension Just as we did for the precise case, we consider an initial coherent set of desirable gambles D on Ω, and an 'input' coherent set of desirable gambles $\breve{D}$ on $\mathcal{B}$, and will look for suitable reformulation of 'agreeing on $\mathcal{B}$' and 'rigidity'. 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:

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

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 $\breve{D}$ consists of gambles on $\mathcal{B}$, using this device we may interpret $\breve{D}$ as a subset of $\mathcal{L}(\Omega)$. Also, using it we can say that any f on $\mathcal{B}$ is constant on the elements of $\mathcal{B}$, by which we mean that for all B in $\mathcal{B}$, it holds that $f^\star(\omega) = f^\star(\omega')$ for all ω and ω' in B.

Conditioning Sets of Desirable Gambles Given an event $B \subseteq \Omega$, we let $\mathbb{I}_B$ be B's indicator (gamble), defined as $\mathbb{I}_B(\omega) := 1$ if $\omega \in B$, and 0 otherwise, for every ω in Ω. When $f \in \mathcal{L}(\Omega)$, the gamble $\mathbb{I}_B f$ is a called-off version of f: if B obtains, the transaction goes through as described by f, but if B does not obtain, the transaction gets cancelled. We will also use the following notation: given a gamble f on B, we let $\mathbb{I}_B f$ be the gamble on Ω that takes the value f on B and 0 elsewhere. Similarly, given a set F of gambles on B, we let $\mathbb{I}_B F := \{\mathbb{I}_B f : f \in F\}$ be a set of gambles on Ω whose elements agree with the elements of F on B, and are 0 elsewhere. We use this to define the procedure of conditioning a set of desirable gambles $D \subseteq \mathcal{L}(\Omega)$ on an non-empty event $B \subseteq \Omega$: the set $D\rfloor B := \{f \in \mathcal{L}(B) : \mathbb{I}_B f \in D\}$ on B contains the called-off gambles that are desirable. This conditioning rule preserves coherence; we refer to [13] for more details.

Jeffrey's Rule We are now in a position to state versions of 'agreeing on $\mathcal{B}$' and 'rigidity' in a desirability context, thereby generalising Jeffrey's rule to this framework. We want the result of Jeffrey's rule to be an 'output' coherent set of desirable gambles $D^\star$ on Ω that satisfies the following constraints:

- $D^\star \supseteq \breve{D}$; [agreeing on $\mathcal{B}$]
- $D^\star\rfloor B \supseteq D\rfloor B$ for all B in $\mathcal{B}$. [rigidity]

"Agreeing on $\mathcal{B}$" here means that we preserve all the assessments about $\mathcal{B}$ made by $\breve{D}$. "Rigidity" means that we preserve all the conditional (on elements of $\mathcal{B}$) assessments present in the original D. There may be multiple $D^\star$ satisfying these constraints, but there will be a unique smallest one. The following result is an immediate consequence of [10, Thm. 3], which is a more general result that holds even for arbitrary possibility spaces Ω.

Theorem 1 *The unique smallest $D^\star$ satisfying "agreeing on $\mathcal{B}$" and "rigidity" is given by*

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

(Jeffrey's rule for desirability)

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

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

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 1.

Lemma 2 *We have that*

$$\widehat{D} = \operatorname{posi}\Big(\breve{D} \cup \bigcup_{B\in\mathcal{B}} \mathbb{I}_B(D\rfloor B)\Big). \tag{1}$$

Example 1 *Let us derive Jeffrey's rule for expectations as a special case of Jeffrey's rule for desirability. Consider an original expectation operator E on $\mathcal{L}(\Omega)$ and an input expectation operator $\breve{E}$ on $\mathcal{L}(\mathcal{B})$. We use them to define the coherent sets of desirable D and $\breve{D}$, as*

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

and

$$\breve{D} \coloneqq \{f \in \mathcal{L}(\mathcal{B}) \colon \breve{E}(f) > 0 \text{ or } f > 0\}.$$

Jeffrey's rule for desirability 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

$$\underline{P}_{\widehat{D}}(f) \coloneqq \sup\{\alpha \in \mathbb{R} \colon f - \alpha \in \widehat{D}\} \quad \text{for all } f \text{ in } \mathcal{L}(\Omega),$$

satisfies Jeffrey's rule for expectations: we will show that $\underline{P}_{\widehat{D}}(f) = \breve{E}(E(f|\mathcal{B})) \eqqcolon \widehat{E}(f)$ 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 $\alpha \coloneqq \frac{\widehat{E}(f)}{2} > 0$. For every B in $\mathcal{B}$, consider the gambles $f_B \colon B \to \mathbb{R} \colon \omega \mapsto f(\omega)$ and $g_B \coloneqq f_B - E(f|B) + \alpha$. 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 Jeffrey's rule for expectations to infer that $\breve{E}(E(f|\mathcal{B}) - \alpha) = \breve{E}(E(f|\mathcal{B})) - \alpha = \widehat{E}(f) - \alpha = \frac{\widehat{E}(f)}{2} > 0$, whence $E(f|\mathcal{B}) - \alpha \in \breve{D}$. So we conclude that

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

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

For the second implication, consider any $f \in \widehat{D}$. Then, using Eq. (1) and taking into account the coherence of $\breve{D}$, there are g in $\breve{D}$, n in $\mathbb{N}$,[3] $B_1, \dots, B_n$ in $\mathcal{B}$, h_1 in $D\rfloor B_1$, $\dots$, h_n in $D\rfloor B_n$ and $(\mu, \lambda_1, \dots, \lambda_n) > 0$[4] such that $f = \mu g + \sum_{k=1}^{n} \lambda_k \mathbb{I}_{B_k} h_k$, which implies that

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

where we used Jeffrey's rule for expectations. 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,

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

[3] We let $\mathbb{N}$ be the natural numbers $\{1, 2, 3, \dots\}$, so we consider 0 not a natural number.

[4] 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 $\{1, \dots, k\}$'.

$$+ \sum_{B \in \mathcal{B} \setminus \mathcal{E}} \breve{E}(\mathbb{I}_B)\mu g(B)$$
$$= \sum_{B \in \mathcal{B}} \breve{E}(\mathbb{I}_B)\mu g(B) + \sum_{\ell=1}^{n} \breve{E}(\mathbb{I}_{B_\ell})\lambda_\ell E(h_\ell|B_\ell)$$
$$= \mu \breve{E}(g) + \sum_{\ell=1}^{n} \lambda_\ell \breve{E}(\mathbb{I}_{B_\ell}) E(h_\ell|B_\ell).$$

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

3. Jeffrey's Rule for Choice Functions

A set of desirable gambles D uniquely determines a binary preference relation $\prec$ between gambles: $g \prec f \Leftrightarrow f - g \in D$, for all gambles f and g [13, 31]. If D is coherent, then $\prec$ is a strict partial order that is also a vector ordering and is compatible with $<$.[5] Conversely, if $\prec$ is a strict partial order that is also a vector ordering compatible with $<$, then the set $D := \{f \in \mathcal{L} : 0 \prec f\}$ is a coherent set of desirable gambles [31, Sect. 1.4.1]. So a set of desirable gambles is an equivalent representation of a binary preference relation. 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. [20] have introduced imprecise-probabilistic choice functions, which were further developed by Seidenfeld et al. [33]. A choice function C identifies from any finite decision problem $F \in \mathcal{Q}(\Omega) := \{G \subseteq \mathcal{L}(\Omega) : |G| \in \mathbb{N}\}$—the set $\mathcal{Q}(\Omega)$ is the set of all finite but non-empty subsets of gambles, which we will also indicate by $\mathcal{Q}$ when it is clear from the context what the domain Ω is—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. Rejection functions may be interpreted as follows: rejecting a gamble $f \in R(F)$ means that F contains another gamble that you prefer to 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 of sets of gambles as $G + G' := \{f + g : f \in G, g \in G'\}$, for any $G, G' \subseteq \mathcal{L}$.

Sets of Desirable Gamble Sets In this paper, we will work with the equivalent representation of *sets of desirable gamble sets* [6, 7, 9]. The idea is to lift the qualification 'desirable' from gambles on Ω to finite sets of gambles $F \in \mathcal{Q}(\Omega)$—called 'gamble sets (on Ω)'. A gamble set F is called *desirable* set when F contains a gamble that you prefer to 0, and a *set of desirable gamble sets on Ω*[6] is the collection of all gamble sets that are desirable: it is a subset $K \subseteq \mathcal{Q}(\Omega)$. As such, K generalises binary preferences: a gamble set $\{f, g\}$ may be desirable—an element of K—because you know that f is desirable or g is desirable—so you prefer f to 0 or you prefer g to 0—without being able to identify which one is. So we see that sets of desirable gamble sets can disjunctively combine statements of the type 'this gamble is desirable'. The set of all sets of desirable gamble sets—all subsets of $\mathcal{Q}$—will be denoted by $\mathbf{K}$.

Sets of desirable gamble sets K are related to rejection functions R, and therefore also to choice functions. In order to describe this relation elegantly, let us follow de Cooman [9] in defining $F \ominus f := (F \setminus \{f\}) - \{f\}$. 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. We use the account of coherence introduced by [6] and further developed in [7].

Definition 3 (Coherent set of desirable gamble sets) *A set of desirable gamble sets $K \subseteq \mathcal{Q}$ is called* coherent *if for all F and G in $\mathcal{Q}$, all $\{\lambda_{f,g}, \mu_{f,g} : f \in F, g \in G\} \subseteq \mathbb{R}$, 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, $(\lambda_{f,g}, \mu_{f,g}) > 0$, then $\{\lambda_{f,g} f + \mu_{f,g} g : f \in F, g \in G\} \in K$;*

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 $\overline{\mathbf{K}}$.

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 [6] 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

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

and

$$\mathrm{Posi}(\mathcal{F}) := \left\{ \left\{ \sum_{k=1}^{n} \lambda_k^{f_{1:n}} f_k : f_{1:n} \in \bigtimes_{k=1}^{n} F_k \right\} : \right.$$

[5] A strict partial order $\prec$ is an order that is irreflexive and transitive. Adding the requirement that $\prec$ is a vector ordering, amounts to imposing that $f \prec g \Leftrightarrow \lambda f + h \prec \lambda g + h$ for all $f, g, h \in \mathcal{L}$ and $\lambda \in \mathbb{R}_{>0}$. We call $\prec$ compatible with $<$ when $f < g \Rightarrow f \prec g$ for all $f, g \in \mathcal{L}$.

[6] We sometimes refrain from using 'on Ω' when it is clear what the domain Ω of the gambles is.

$$n \in \mathbb{N}, F_1, \ldots, F_n \in \mathcal{F}, \left(\forall f_{1:n} \in \bigtimes_{k=1}^{n} F_k\right) \lambda_{1:n}^{f_{1:n}} > 0\Big\}$$

for all $\mathcal{F}$ in $\mathbf{K}$, and the set $\mathcal{L}^{\mathrm{s}}(\Omega)_{>0} := \{\{f\} : f \in \mathcal{L}(\Omega)_{>0}\}$—often denoted simply by $\mathcal{L}^{\mathrm{s}}_{>0}$.

Representation in Terms of Desirability Given a set of desirable gamble sets K, its *binary* part $D_K := \{f \in \mathcal{L} : \{f\} \in K\}$ 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 [6, 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 $\{K \in \mathbf{K} : D_K = D\}$ may contain more than one element. However, if D is coherent, it always contains one unique smallest element [37, Prop. 5] $K_D := \{F \in \mathcal{Q} : F \cap D \neq \emptyset\}$, which is then equal to $\bigcap\{K \in \overline{\mathbf{K}} : D_K = D\}$. If we generalise K_D's definition above to arbitrary subsets $D \subseteq \mathcal{L}$, then De Bock and de Cooman [7, Prop. 8] have established that K_D is coherent if and only if D is.

In their same paper, De Bock and de Cooman [7, Thm. 9] establish the following useful representation result:[7]

Theorem 4 (Representation [7, 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 $\mathbf{D}(K) := \{D \in \overline{\mathbf{D}} : K \subseteq K_D\}$.*

The representation result above plays an important role in the development of the theory of sets of desirable gamble sets: sometimes, it is easier to find a representing set $\mathbf{D}$ of a coherent set of desirable gamble sets K, rather than K itself. An instance of a result where the representation proved to be crucial is the independent natural extension [38, Thm. 15]. That is the reason why we will be also after a representation of Jeffrey's rule for choice functions, at the end of this section.

Cylindrical Extension As before, we will consider an initial coherent set of desirable gamble sets K on Ω, and an 'input' coherent set of desirable gamble sets $\check{K}$ on $\mathcal{B}$, and will look for a suitable reformulation of 'agreeing on $\mathcal{B}$' and 'rigidity'. Before we can do so, we need to relate a set of desirable gamble sets on Ω with one on $\mathcal{B}$. The idea will be the same as for sets of desirable gambles: we will use the simplifying device of *equating a gamble set F on $\mathcal{B}$ with its cylindrical extension $F^\star$ on Ω*, given by $F^\star := \{f^\star : f \in F\}$. So $F^\star$ collects all the cylindrical extensions of gambles in F. As $\check{K}$ consists of gamble sets on $\mathcal{B}$, using this device we may interpret $\check{K}$ as a subset of $\mathcal{Q}(\Omega)$.

[7]This result first appeared in [6, Thm. 9] but we prefer their later formulation.

Conditioning Sets 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 $K\rfloor B := \{F \in \mathcal{Q}(B) : \mathbb{I}_B F \in K\} \subseteq \mathcal{Q}(B)$ as the collection of the called-off versions of gamble sets present in K. This conditioning rule preserves coherence [37, 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 $\{D\rfloor B : D \in \mathbf{D}(K)\}$ [38, Prop. 7].

For further notational streamlining, we define the multiplication of an indicator $\mathbb{I}_B$ with a subset $\mathcal{F} \subseteq \mathcal{Q}(B)$ as $\mathbb{I}_B\mathcal{F} := \{\mathbb{I}_B F : F \in \mathcal{F}\}$: it is the collection of called-off versions of the gamble sets in $\mathcal{F}$.

Jeffrey's Rule We are now in a position to define 'agreeing on $\mathcal{B}$' and 'rigidity' for sets of desirable gamble sets, thereby generalising Jeffrey's rule to sets of desirable gamble sets, and therefore to choice functions too, after a suitable translation of the properties and result. We want the result of Jeffrey's rule to be an 'output' coherent set of desirable gamble sets $K^\star$ on Ω that satisfies the following constraints:

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

Here too, "agreeing on $\mathcal{B}$" means that we preserve all the assessments about $\mathcal{B}$ made by $\check{K}$. "Rigidity" means that we preserve all the conditional (on elements of $\mathcal{B}$) assessments present in the original K. There may be multiple $K^\star$ satisfying these constraints, but there will be a unique smallest one.

Theorem 5 *The unique smallest $K^\star$ satisfying "agreeing on $\mathcal{B}$" and "rigidity" is given by*

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

(Jeffrey's rule for choice functions)

$$= \mathrm{Rs}\Big(\mathrm{Posi}\Big(\check{K} \cup \bigcup_{B \in \mathcal{B}} \mathbb{I}_B(K\rfloor B)\Big)\Big).$$

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

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

4. Jeffrey's Rule for Non-Additive Measures

We shift our attention to a number of models that constitute particular cases of coherent lower previsions: *non-additive measures*. For this, recall that a coherent lower prevision may also be defined on a subset $\mathcal{K}$ of the set of all gambles

$\mathcal{L}(\Omega)$; when this subset contains only indicators of events, it is called a *coherent lower probability*. On the other hand, given a coherent lower probability $\underline{P}$ with domain $\mathcal{K} \subseteq \mathcal{P}(\Omega)$, the smallest coherent lower prevision $\underline{E}$ on $\mathcal{K}' \supseteq \mathcal{K}$ that agrees with $\underline{P}$ on $\mathcal{K}$ is given by the lower envelope of

$$\mathcal{M}(\underline{P}) := \{P\colon (\forall A \in \mathcal{K}) P(\mathbb{I}_A) \geq \underline{P}(A)\},$$

i.e., $\underline{E}(f) = \min\{P(f)\colon P \in \mathcal{M}(\underline{P})\}$ for every $f \in \mathcal{K}'$. It is called the *natural extension* of $\underline{P}$ to $\mathcal{K}'$.

Our starting point in this section shall be a coherent lower probability $\underline{\breve{P}}$ on $\mathcal{P}(\mathcal{B})$, the events that are finite unions of elements from the partition $\mathcal{B}$. We also assume that for each $B \in \mathcal{B}$, we have a coherent lower probability $\underline{P}(\bullet|B)$ on $\mathcal{P}(\Omega)$. By considering the above procedure, we can obtain their respective natural extensions $\underline{\breve{E}}$ on $\mathcal{L}(\mathcal{B})$ and $\underline{E}(\bullet|\mathcal{B})$ on $\mathcal{L}(\Omega)$. Let $\underline{\widehat{E}}$ be the marginal extension of $\underline{\breve{E}}$ and $\underline{E}(\bullet|\mathcal{B})$, given by $\underline{\widehat{E}}(f) = \underline{\breve{E}}(\underline{E}(f|\mathcal{B}))$ for all f in $\mathcal{L}(\Omega)$. This yields the following proposition.

Proposition 6 *If $\underline{\breve{P}}(B) > 0$ for every $B \in \mathcal{B}$, then*

$$\begin{aligned}\mathcal{M}(\underline{\widehat{E}}) = \{P\colon & (\forall A \in \mathcal{P}(\mathcal{B})) P(\mathbb{I}_A) \geq \underline{P}(A) \text{ and} \\ & (\forall A' \subseteq \Omega, B \in \mathcal{B}) P(A'|B) \geq \underline{P}(A'|B)\}.\end{aligned}$$

Now, let C be a subfamily of coherent lower probabilities, and assume that $\underline{\breve{P}}$ and $\underline{P}(\bullet|B)$ belong to C for every $B \in \mathcal{B}$. We look then for the joint models $\underline{\widehat{P}}$ on $\mathcal{P}(\Omega)$ such that

(a) $\underline{\widehat{P}}(A) \geq \underline{\breve{P}}(A)$ for every $A \in \mathcal{P}(\mathcal{B})$; [agreeing on $\mathcal{B}$]

(b) $\underline{\widehat{P}}(A'|B) \geq \underline{P}(A'|B)$ for every $B \in \mathcal{B}$ such that $\underline{\widehat{P}}(B) > 0$ and every $A' \subseteq \Omega$; [rigidity]

(c) $\underline{\widehat{P}} \in C$,

and in particular, for the smallest such model, if it exists. For the conditional lower prevision $\underline{\widehat{P}}(\bullet|\mathcal{B})$ in condition (b), when $\underline{\widehat{P}}(B) > 0$ we use the one determined by Generalised Bayes Rule (GBR) [41, Sect. 6.4]; when $\underline{\widehat{P}}(B) = 0$, given that the conditional lower prevision determined by GBR is vacuous and $\underline{\widehat{P}}$ satisfies coherence with any conditional, we make $\underline{\widehat{P}}(\bullet|B) = \underline{P}(\bullet|B)$.

Trivially, if the marginal extension $\underline{\widehat{E}}$ belongs to C, it follows that it is the smallest model satisfying conditions (a)–(c) above; but, as we shall see, it does not need to be so in general. In fact, it is not hard to see that properties such as complete monotonicity are not preserved by marginal extension in general:

Example 2 *Let $\Omega = \{\omega_1, \dots, \omega_8\}$, $B = \{\omega_1, \dots, \omega_4\}$ and the partition $\mathcal{B} = \{B, B^c\}$. Let $P(\bullet|B)$ the precise prevision associated with the mass function* $(0.3, 0.15, 0.15, 0.4)$*, and $P(\bullet|B^c)$ the precise prevision associated with the mass function* $(0.2, 0.25, 0.25, 0.3)$. *Consider on the other hand the vacuous lower probability $\underline{\breve{P}}$ on $\{B, B^c\}$. Given the events $A_1 = \{\omega_1, \omega_2, \omega_5, \omega_6\}$ 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 (a)–(c) is an *inner approximation* of the marginal extension:

Proposition 7 *Let $\underline{\widehat{P}}$ be a coherent lower prevision in C. Then $\underline{\widehat{P}}$ satisfies conditions (a) and (b) if and only if it is an inner approximation of the marginal extension $\underline{\breve{E}}(\underline{E}(\bullet|\mathcal{B}))$.*

Next, we shall investigate the above problem for different families of non-additive measures. In this respect, we should mention that Jeffrey's rule was already investigated in the context of belief functions in Refs. [21, 35, 36, 43, 44]. One important difference with our approach is that this literature uses a different conditioning rule in condition (b) than ours: while we use the GBR to obtain a conditional model $\underline{P}(\bullet|\mathcal{B})$, works in the framework of belief functions typically use Dempster's conditioning [35, 43], coarsening conditioning [44] or geometric conditioning [35]. However, these rules do not guarantee coherence between the joint and the conditional models, which lies at the basis of our approach. We would thus be making tractable the approach already discussed by Wagner [39].

4.1. Minitive Measures

We begin by considering the class of *minitive measures*, which are those coherent lower probabilities satisfying

$$\underline{P}(A_1 \cap A_2) = \min\{\underline{P}(A_1), \underline{P}(A_2)\} \quad \text{for all } A_1, A_2 \subseteq \Omega.$$

The conjugate upper probability is a *maxitive measure* [15].

Similar to our previous comments, Jeffrey's rule has been investigated in the context of possibility measures in [3]; the main difference lies in the use of another conditioning rule (in the case of [3], the product- and min-based rules).

Minitive measures $\underline{P}$ are a particular case of belief functions, which means in particular that they are 2-monotone: for every pair of events A_1, A_2, they satisfy

$$\underline{P}(A_1 \cup A_2) + \underline{P}(A_1 \cap A_2) \geq \underline{P}(A_1) + \underline{P}(A_2).$$

As a consequence, given a minitive lower probability on $\mathcal{P}(\Omega)$, it follows [see [1] for an overview of this and related results] from its 2-monotonicity that its natural extension to $\mathcal{L}(\Omega)$ is given by its *Choquet integral*: if $f \in \mathcal{L}(\Omega)$, then

$$\underline{E}(f) = \inf f + \int_{\inf f}^{\sup f} \underline{P}(\{f > t\})dt,$$

where we denoted the level set $\{f > t\} := \{\omega \in \Omega \colon f(\omega) > t\}$, and we will later on do similarly for $\{f \leq t\} := \{\omega \in \Omega \colon f(\omega) \leq t\}$.

However, given minitive measures $\underline{\breve{P}}$ on $\mathcal{P}(\mathcal{B})$ and $\underline{P}(\bullet|B)$ on $\mathcal{P}(\Omega)$ for each $B \in \mathcal{B}$, the coherent lower previsions $\underline{\breve{E}}$ on $\mathcal{L}(\mathcal{B})$ and $\underline{E}(\bullet|\mathcal{B})$ defined by natural extension need not be minitive: as showed in [12, 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.

Let $\underline{\widehat{E}}$ be the marginal extension of $\underline{\breve{E}}, \underline{E}(\bullet|\mathcal{B})$. The following result gives sufficient conditions for $\underline{\widehat{E}}$ to be minitive on gambles (that is, $\underline{\widehat{E}}(f \wedge g) = \min\{\underline{\widehat{E}}(f), \underline{\widehat{E}}(g)\} = \underline{\widehat{E}}(f) \wedge \underline{\widehat{E}}(g)$ for all gambles f and g, so $\underline{\widehat{E}}$ is a $\wedge$-homomorphism, where we use $\wedge$ to denote the point-wise minimum of two gambles: $f \wedge g$ is the gamble whose values are $(f \wedge g)(\omega) := \min\{f(\omega), g(\omega)\}$ for every ω in Ω) or on events.

Proposition 8

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

The reasoning in the proof of the last item makes it easy to find examples where both $\underline{\breve{E}}, \underline{E}(\bullet|\mathcal{B})$ are not minitive on gambles and yet $\underline{\widehat{E}}$ is minitive on events: take for instance $\Omega = \{a,b,c,d\}, B = \{a,b\}, \mathcal{B} = \{B, B^c\}, \underline{\breve{P}}(B) = 0.5, \underline{\breve{P}}(B^c) = 0, \underline{P}(\bullet|B)$ 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 $\underline{\widehat{E}}$ is minitive on events, with focal elements $\{a\}, \{a,d\}, \{a,c,d\}$.

On the other hand, it may also happen that $\underline{\widehat{E}}$ is minitive on gambles even if $\underline{E}(\bullet|B)$ is only minitive on events (i.e., not $\{0,1\}$-valued); a sufficient condition for this is, after denoting $\mathcal{E}_{\mathcal{B}} := \{B \in \mathcal{B} \colon \underline{P}(\bullet|B)$ not minitive on gambles$\}$, that $\underline{\breve{E}}(\bigcup_{B \in \mathcal{E}_{\mathcal{B}}} B) = 0$. However, it is a consequence of [12, Prop. 6] that the following statements are equivalent:

(a) $\underline{P}(\bullet|B)$ is $\{0,1\}$-valued on events for every $B \in \mathcal{B}$;

(b) $\underline{\breve{P}}(\underline{P}(\bullet|\mathcal{B}))$ is minitive on gambles for every coherent lower prevision $\underline{\breve{P}}$ that is minitive on gambles.

When $\underline{\widehat{E}}$ is not minitive on events, it is not hard to find the smallest coherent lower probability that dominates it and is minitive. This is a consequence of the following result.

Proposition 9 *Let $\underline{P}$ be a coherent lower probability on $\mathcal{P}(\Omega)$. Then the smallest minitive lower probability $\underline{P}'$ on $\mathcal{P}(\Omega)$ satisfying $\underline{P}'(A) \geq \underline{P}(A)$ for every $A \subseteq \Omega$ is given by*

$$\underline{P}'(A) = 1 - \max_{\omega \notin A} \overline{P}(\{\omega\}). \tag{2}$$

To conclude this section, we summarise its results as follows:

- If the class C is that of coherent lower previsions that are minitive on gambles (that is, $\wedge$-homomorphisms), then the marginal extension is the smallest model satisfying conditions (a)–(c).
- If the class C is that of coherent lower previsions that are minitive on events, then the marginal extension is the smallest model satisfying conditions (a)–(c) if either the marginal or the conditional models are also minitive on gambles; otherwise, the smallest such model is determined by Eq. (2).

4.2. Distortion Models

Next we consider the family of distortion models. These refer to those imprecise probability models that originate by distorting 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 [17]. There are several distortion models, such as the pari-mutuel, linear-vacuous or Kolmogorov models. We refer to [28, 29] to 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 in [34]. Here, we shall focus on the linear-vacuous, pari-mutuel and total variation models.

4.2.1. Linear-Vacuous Model

We begin by considering the family of linear-vacuous models, also referred to as contamination models in the literature.

Definition 10 *Let P_0 be a probability measure on Ω and consider a distortion factor $\delta \in (0,1)$. The associated* linear-vacuous model *is given by the lower probability*

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

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

Linear-vacuous models (LV models) have been studied in the context of robust statistics [17]: 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. We shall denote by $\mathcal{C}_{\mathrm{LV}}$ the class of linear-vacuous mixtures.

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

$$\underline{P}(A) = \sum_{\omega \in A} \underline{P}(\{\omega\}). \tag{3}$$

On the other hand, the natural extension from events to gambles of an LV model $\underline{P}$ 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 model 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 4 that this LV model is the smallest model satisfying conditions (a)–(c).

Proposition 11 *Consider an LV model $\breve{\underline{E}}$ on $\mathcal{L}(\mathcal{B})$ determined by $(P_{\mathcal{B}}, \delta_{\mathcal{B}})$ and, for each $B \in \mathcal{B}$, let $\underline{E}(\bullet|B)$ be an LV model on $\mathcal{L}(\mathcal{B})$ determined by (P_B, δ_B). Assume that $\breve{\underline{E}}(B) > 0$ 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 model on Ω determined by (P, δ). Then the following are equivalent:*

(a) $\underline{P}$ is coherent with $\breve{\underline{E}}, \underline{E}(\bullet|\mathcal{B})$.

(b) $\underline{P}(f) = \breve{\underline{E}}(f)$ 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 $\omega \in B$.

(c) $\delta = \delta_{\mathcal{B}}, P(\{\omega\}) = \frac{E(\{\omega\}|B)(\delta+P(B))}{1-\delta}$ 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 model inducing the same marginal and conditional models we started with. It is not hard to show that the class of LV models is not closed under marginal extension:

Proposition 12 *Let $\underline{P}$ be an LV model on $\mathcal{L}(\mathcal{B})$ determined by $(P_{\mathcal{B}}, \delta_{\mathcal{B}})$, and for every $B \in \mathcal{B}$ let $\underline{P}(\bullet|B)$ be an LV model on Ω determined by (P_B, δ_B). Then the marginal extension $\widehat{\underline{E}} = \breve{\underline{E}}(\underline{E}(\bullet|\mathcal{B}))$ does not belong to $\mathcal{C}_{\mathrm{LV}}$.*

Taking into account Prop. 7, we may consider then the inner approximations of the marginal extension $\widehat{\underline{E}}$ in the class of LV models. Inner approximations of coherent lower probabilities were investigated in [26]; in the case of distortion models, they are moreover linked with the notion of *centroids* of credal sets [25]. In [26, Sect. 3.1] it was established that the optimal inner approximations, in that they minimise the distance defined by Baroni and Vicig in [2] 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 Prop. 6.

A word of caution here, though: the marginal extension $\widehat{\underline{E}}$ 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 3 *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 $\underline{P}(\bullet|\Omega_1)$ 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 $\widehat{\underline{E}}$ of these models, it can be checked that the probability measure P_1 with mass function*

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

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

This means that the results in [26] are not immediately applicable, and should be generalised to the context of this paper. Considering the results in [26], we conjecture that there will not be in general a unique optimal inner approximation of $\widehat{\underline{E}}$ in the class of LV models.

4.2.2. Pari-Mutuel Model

The second distortion model we consider in this paper is the pari-mutuel model (PMM) [41].

Definition 13 *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 $\mathcal{C}_{\mathrm{PMM}}$ the family of pari-mutuel models, and we refer to [27, 30] 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 given a marginal and a conditional pari-mutuel model, there is *never* a PMM that is coherent with both of them. 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 [30, 41]:

$$\overline{E}(f) = f_\tau + (1+\delta)P((f - f_\tau)^+), \tag{4}$$

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

Proposition 14 *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 $\overline{P}(\bullet|B)$ 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 $\overline{P}_{\mathcal{B}}, \overline{P}(\bullet|\mathcal{B})$.*

Again, we should then look at the inner approximations of the marginal extension in order to find a suitable formulation of Jeffrey's rule for PMMs; we expect that the results in [26, Sect. 3.2] should be of interest, provided they are suitably extended from coherent lower probabilities to coherent lower previsions.

4.2.3. Total Variation Model

The third and final distortion model we consider in this paper is the one associated with the total variation distance.

Definition 15 *Given a probability measure P_0 on Ω and a distortion factor $\delta > 0$, the* total variation model *(TV model) they determine is given by the coherent lower probability* $\underline{P}(A) = \max\{P_0(A) - \delta, 0\}$.

We refer to [16] for a study of this model. The family of total variation models shall be denoted in this paper by $\mathcal{C}_{\mathrm{TV}}$.

As was the case with the other two families of distortion models, the class $\mathcal{C}_{\mathrm{TV}}$ is not closed under marginal extension. To see this, consider two finite spaces Ω_1, Ω_2, let P be a probability measure on Ω_1 and $P(\bullet|\omega_1)$ be a probability measure on Ω_2 for each $\omega_1 \in \Omega_1$. Let P_0 denote the probability measure on $\Omega_1 \times \Omega_2$ they determine. Assume that $P_0(\{(\omega_1, \omega_2)\}) > 0$ for every $(\omega_1, \omega_2) \in \Omega_1 \times \Omega_2$, and take $\delta_1 < \min_{A \subseteq \Omega_1} P(A)$ and $\delta_{\omega_1} < \min_{A \subseteq \Omega_2} P(A|\omega_1)$. Then we obtain

$$\begin{aligned}\widehat{\underline{E}}(\{(\omega_1, \omega_2)\}) &= \check{\underline{E}}(\underline{E}(\{\omega_2\}|\omega_1)) \\ &= \check{\underline{E}}(P(\{\omega_2\}|\omega_1) - \delta_{\omega_1}\mathbb{I}_{\{\omega_1\}}) \\ &= P_0(\{(\omega_1, \omega_2)\}) - \delta_{\omega_1}P(\{\omega_1\}) + \delta_{\omega_1}\delta_1 \\ &\qquad - P(\{\omega_2\}|\omega_1)\delta_1. \end{aligned} \tag{5}$$

With these considerations in mind, we can build the following example, which establishes that the class of TV models is not closed under marginal extension.

Example 4 *Consider $\Omega_1 = \{a, b\}, \Omega_2 = \{c, d\}$, $\delta_1 = 0.1 = \delta_a = \delta_b$ and the probability measure P_0 determined by*

$$P_0(\{(a,c)\}) = 0.15 = P_0(\{(a,d)\})$$
$$P_0(\{(b,c)\}) = 0.28 \quad P_0(\{(b,d)\}) = 0.42.$$

Let $P, P(\bullet|\Omega_1)$ be the marginal and conditional probability measures it determines, and let $\underline{P}, \underline{P}(\bullet|\Omega_1)$ be their associated TV models. Using Eq. (5), then the marginal extension $\widehat{\underline{E}} = \check{\underline{E}}(\underline{E}(\bullet|\Omega_1))$ satisfies

$$\widehat{\underline{E}}(\{(a,c)\}) = 0.08 \quad \widehat{\underline{E}}(\{(a,d)\}) = 0.08$$
$$\widehat{\underline{E}}(\{(b,c)\}) = 0.18 \quad \widehat{\underline{E}}(\{(b,d)\}) = 0.3.$$

Therefore, if this was a TV model, it should be associated with the distortion factor $\delta = 0.09$ and the probability mass function $(0.17, 0.17, 0.27, 0.39)$. *But that distortion model gives $\underline{P}(\{a\} \times \Omega_2) = 0.25, \underline{P}(\{b\} \times \Omega_2) = 0.57$, while $\widehat{\underline{E}}(\{a\} \times \Omega_2) = 0.2, \widehat{\underline{E}}(\{b\} \times \Omega_2) = 0.6$.* ◊

With respect to the inner approximations of a coherent lower probability by means of a TV model, it was showed in [26] that they correspond to the *incenter* of the associated credal set, as defined in [25]. An interesting open question arises in this way: how can we use this result in the particular case when our original model is the marginal extension of two TV models?

5. 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 Jeffrey's rule to sets of desirable gambles, and have generalised this rule to the even more versatile framework of sets of desirable gambles, and therefore to choice functions, too. We then 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 three classes of distortion models we considered do not allow for an expression of the marginal extension within its class.

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 Jeffrey's rule should provide the joint. While the marginal extension theorem has been generalised to a finite number of partitions in [11, 24] and the results on inner approximations from [26] 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 [14, 40].

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 in [24]; when we are interested in obtaining an optimal inner approximation, the case of minitive measures has been characterised in Sect. 4.1, while in the case of distortion models the connection with incenters of credal sets may allow us to use the results from [25].

Note also that in our compatibility study of Jeffrey's rule 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 in [42]. It would also be interesting to consider this problem from the point of view of belief revision, taking into account the discussions in [4, 5, 22] in the precise case. For this, the work in [8] and [21] would be particularly relevant.

Finally, even if the results above provide some analysis of the formulation of Jeffrey's rule within imprecise probability models, space and time limitations have prevented us from discussing a number of interesting side topics, such as (a) the connection with the approaches established in the context of belief functions under other conditioning rules; (b) 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 report on these problems in future work.

Acknowledgments

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. Arthur Van Camp was funded by Jason Konek's ERC Starting Grant "Epistemic Utility for Imprecise Probability" under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 852677). We also would like to thank Gert de Cooman, Kevin Blackwell and Catrin Campbell–Moore for helpful comments and discussion, and three anonymous reviewers for their careful reading, useful suggestions and interesting pointers to literature.

Author Contributions

The two authors contributed equally to this paper.

References

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

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

[3] 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.

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

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

[6] 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.

[7] 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.

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

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

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

[11] G. de Cooman and E. Miranda. Forward irrelevance. *Journal of Statistical Planning and Inference*, 139: 256–276, 2009. doi:10.1016/j.jspi.2008.01.012.

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

[13] 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.

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

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

[16] T. Herron, T. Seidenfeld, and L. Wasserman. Divisive conditioning: further results on dilation. *Philosophy of Science*, 64:411–444, 1997.

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

[18] R. Jeffrey. *The logic of decision*. McGraw, 1965.

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

[20] 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.

[21] 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.

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

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

[24] 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.

[25] 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.

[26] E. Miranda, I. Montes, and A. Presa. Inner approximations of credal sets by non-additive measures. In *Proceedings of IPMU'2022*, Communications in Computer and Information Science, pages 743–756, 2022.

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

[28] 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, 2020.

[29] 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, 2020.

[30] 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.

[31] 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.

[32] 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.

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

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

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

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

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

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

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

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

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

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

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

[44] 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.