# Conditioning, Updating and Lower Probability Zero

Jasper De Bock, Gert de Cooman

*Ghent University, SYSTeMS Research Group, Belgium.*

**Abstract**

We discuss the issue of conditioning on events with probability zero within an imprecise-probabilistic setting, where it may happen that the conditioning event has lower probability zero, but positive upper probability. In this situation, two different conditioning rules are commonly used: regular extension and natural extension. We explain the difference between them and discuss various technical and computational aspects. Both conditioning rules are often used to update an imprecise belief model after receiving the information that some event $O$ has occurred, simply by conditioning on $O$, but often little argumentation is given as to why such an approach would make sense. We help to address this problem by providing a firm foundational justification for the use of natural and regular extension as updating rules. Our results are presented in three different, closely related frameworks: sets of desirable gambles, lower previsions, and sets of probabilities. What makes our justification especially powerful is that it avoids making some of the unnecessary strong assumptions that are traditionally adopted. For example, we do not assume that lower and upper probabilities provide bounds on some 'true' probability mass function, on which we can then simply apply Bayes's rule. Instead a subject's lower probability for an event $O$ is taken to be the supremum betting rate at which he is willing to bet on $O$, and his upper probability is the infimum betting rate at which he is willing to take bets on $O$; we do not assume the existence of a fair betting rate that lies in between these bounds.

*Keywords:* conditioning, updating, probability zero, regular extension, natural extension, sets of desirable gambles

## 1. Introduction

Conditioning on events with probability zero is traditionally considered to be problematic because Bayes's rule cannot be applied. In those cases, depending on the approach that is taken, the conditional probability measure is either left undefined or chosen freely without enforcing any connection with the unconditional measure. The latter approach has the advantage of being more flexible, but it does not really solve the problem, because it provides no guidelines on how to come up with such a conditional model. In order to avoid all issues with countable additivity and measurability—which make the problem even more complicated—we restrict attention to finite state spaces, thereby allowing us to work with probability mass functions. In that case, the most common solution to this problem is to simply ignore it, because from a practical point of view, in a finite state space, events with probability zero are usually considered to be impossible anyway, which makes the task of conditioning on them rather irrelevant.

The situation becomes more complex when we consider a set of probability mass functions instead of a single one, because then, the lower and upper probability of an event need not coincide, and it may have lower probability zero, but positive upper probability. This happens frequently in practice. For example, many common methods for deriving sets of probabilities from data will assign lower probability zero to an event for which there is no data—because it might well be impossible—but not upper probability zero—because the fact that you have not seen it yet does not imply that it is impossible. Clearly, such events cannot be ignored, and we need to be able to condition on them.

Fortunately, working with sets of probability mass functions is already part of the solution, because it allows us to condition on an event $O$ that has probability zero in a trivial manner, simply by taking the conditional model to be the

*Email addresses:* Jasper.DeBock@UGent.be (Jasper De Bock), Gert.deCooman@UGent.be (Gert de Cooman)

set of all probability mass functions on $O$; this is called the vacuous model on $O$. Since Bayes's rule does not impose any constraints, this vacuous model represents all we know about the conditional probability mass function on $O$.

If we start from a set $\mathscr{M}$ of mass functions, then three cases can be distinguished. If every mass function in $\mathscr{M}$ assigns probability zero to $O$, we use the vacuous model on $O$ as our conditional model. If every element of $\mathscr{M}$ assigns positive probability to $O$, the conditional set of mass functions is obtained by applying Bayes's rule to every element of $\mathscr{M}$. The remaining possibility is the one that we have already mentioned: the case that $O$ has probability zero according to some elements of $\mathscr{M}$—lower probability zero—but positive probability according to others—positive upper probability. Here, imprecise-probabilistic frameworks—such as, but as we shall see not limited to, sets of probability mass functions—usually distinguish between two main conditioning rules: natural and regular extension. Natural extension again uses the vacuous model on $O$, consisting of all probability mass functions on $O$. Regular extension ignores the probability mass functions in $\mathscr{M}$ that assign probability zero to $O$ and simply applies Bayes's rule to the others.

**Example 1.** Consider a ternary state space $\Omega = \{a,b,c\}$ and let $f$ be the real-valued map on $\Omega$ that is defined by $f(a) := 1$, $f(b) := 2$ and $f(c) := 3$. Let $\mathscr{M}$ be the set of all mass functions $p$ on $\Omega$ for which the corresponding expectation $P_p(f) := \sum_{\omega\in\Omega} f(\omega)p(\omega)$ of $f$—also called the prevision of $f$—is higher than or equal to 2 or, equivalently, the largest set of mass functions on $\Omega$ such that the lower expectation

$$\underline{P}(f) := \inf_{p\in\mathscr{M}} P_p(f)$$

of $f$—also called the lower prevision of $f$—is equal to two. Since the unitary constraint for mass functions implies that $p(a)+2p(b)+3p(c) \geq 2 \Leftrightarrow p(c) \geq p(a)$, $\mathscr{M}$ is simply the set of all mass functions on $\Omega$ for which the probability of $c$ is at least as high as that of $a$.

Now let $O = \{a,c\}$ be the event that $b$ does not happen. According to $\mathscr{M}$, the lower and upper probability of $O$ are equal to 0 and 1, respectively. The lower probability 0 is attained by a single degenerate probability mass function $p_b$ that assigns all mass to $b$. If we use natural extension to condition the set $\mathscr{M}$ on the event $O$, then since $O$ has lower probability zero, we obtain the set of all mass functions on $O$—the vacuous model on $O$. Regular extension, on the other hand, corresponds to applying Bayes's rule to every mass function in $\mathscr{M}$ except $p_b$, and results in a conditional model that is significantly smaller, as it only contains those conditional probability mass functions on $O$ for which $p(c|O) \geq p(a|O)$. Hence, the conditional lower probability of $c$ that is provided by regular extension is equal to $1/2$, whereas it is 0 according to natural extension. $\lozenge$

As this example shows, regular extension can be a lot more informative than natural extension, in the sense that the resulting conditional set of mass functions is smaller—twice as small in this particular case. One can even come up with examples where regular extension results in a single conditional mass function and natural extension still returns the vacuous model [38, Appendix J2].

For this reason, from a practical point of view, regular extension is more useful than natural extension. Conditioning is not merely a technical concept, but also a popular updating tool. After being informed that an event $O$ has occurred, we wish to update our former model to obtain a new one, and the most popular approach for doing so is to condition the old model on the event $O$. For a single mass function, the use of Bayes's rule as an updating tool has been justified by many authors—see for example Refs. [29, 28, 30]—and this justification easily translates to sets of probabilities, by applying it elementwise. After updating, the updated set of mass functions takes the role of the original one, and is then used for inference, classification or decision making. Clearly, in order for this new model to be of practical use, it should be as informative as possible, and definitely not vacuous, which is why—as far as updating is concerned—regular extension is usually preferred over natural extension.

So far, so good. However, imprecision in probability theory can be represented by more models than just sets of probabilities. Consider for example the vacuous model, which is taken to model complete ignorance. In the framework of sets of probability mass functions, this corresponds to the use of the set of all possible mass functions. However, this is only a limited form of 'complete' ignorance, because it assumes the existence of such a mass function; we are only ignorant about which mass function it should be. Walley refers to this assumption as the dogma of ideal precision [38, Section 1.1.5]; we will simply call it ideal precision. It might seem like a trivial assumption, but it is in fact not. In order to understand why, we need to go back to the meaning of probability. Frequentists would argue that it is a limiting frequency. Indeed, in some cases, it is. But why should this be the case in general? And why should the frequencies converge? In any case, for a single experiment that is not repeated, this interpretation no longer works. People that favour the subjective approach will argue that probabilities are a subject's fair betting rates. However, from a subjective point of view, it seems hardly compelling that a subject should have such a fair betting rate for every event.

In order to avoid making such a strong assumption, some imprecise-probabilistic frameworks work directly with the bounds that are created by a set of probabilities, rather than the set itself. Walley's theory of coherent lower previsions is an important example [38]. In fact, it is not even necessary to refer to a set of probability mass functions at all. For example, the lower probability of an event $O$ can be taken to be the supremum betting rate for betting on $O$, and the upper probability can be regarded as the infimum betting rate for taking bets on $O$. These lower and upper probabilities need not coincide; there might be betting rates at which you are neither willing to bet on $O$ nor willing to take bets on $O$. It is important to realise that this interpretation does not imply the existence of some fair but unknown betting rate that lies in between these two bounds. More generally, we can consider supremum buying prices and infimum selling prices for arbitrary gambles—risky transactions whose payoff depends on the uncertain value of some variable—without assuming that the subject has a fair price for this gamble that lies in between these bounds. For example, the assessment $\underline{P}(f) = 2$ that is made in Example 1 need not be interpreted as a constraint on the expectation operator of some unknown probability mass function. Instead, the lower prevision $\underline{P}(f)$ can be interpreted directly as a subject's supremum buying price for the gamble $f$.

The nice thing about these alternative frameworks is that they are as powerful as their sets of probability mass functions counterpart, without the strong assumptions that come with the latter. For example, natural and regular extension can be defined directly in terms of lower previsions, using the consistency requirement of coherence to relate conditional lower previsions to unconditional ones [38, 19]. If $O$ has lower probability zero and positive upper probability, coherent conditioning allows for anything in between natural and regular extension. Natural extension is the most conservative conditional model, whereas regular extension is the most informative one. From a technical point of view, these conditioning rules can even be shown to be equivalent to their counterparts for sets of mass functions. It suffices to consider the set of compatible mass functions, apply natural or regular extension to that set, and then consider the corresponding lower and upper previsions or probabilities. However, without an assumption of ideal precision, these sets of compatible mass functions are just mathematical objects, and can no longer be interpreted as sets of candidates for the 'true' model. Therefore, in that case, the fact that Bayes's rule can be justified as an updating rule for mass functions no longer leads to a justification for updating by means of natural or regular extension.

*The main goal of this paper is to provide a firm foundational justification for the use of natural and regular extension as updating rules, directly in terms of lower previsions, and without any reference to ideal precision, Bayes's rule or sets of probability mass functions.* In order to achieve this goal, we study the problem of updating in terms of sets of desirable gambles, an imprecise-probabilistic framework that is closely related to—but slightly more general than—lower previsions. Instead of considering buying prices for gambles, this framework works directly with the gambles themselves, in particular those that are considered to be desirable—strictly preferred to the status quo. We consider two rules for updating these sets of desirable gambles, derived directly from basic principles. By translating them into the language of lower previsions, we eventually obtain a justification for the use of natural and regular extension as updating rules.

We start off in Section 2 by connecting the theory of sets of desirable gambles with that of lower previsions and sets of probability mass functions, restricting ourselves to the case where the uncertain variable of interest takes values in a finite space $\Omega$. We provide an overview of some of the most important results in the literature and fill in some gaps as we go along. For readers who are new to imprecise probabilities, this can be regarded as an introduction to the field. The more advanced reader will notice that our approach differs from some of the more conventional ones. We build the theory from scratch, using sets of desirable gambles as our starting point, in a way that closely resembles—but nevertheless differs slightly from—the approach advocated by Williams [43, 32]. In this way, we obtain a notion of coherence for conditional lower previsions that is completely structure-free, in the sense that the domain of the conditional lower prevision can be an arbitrary set of couples of the form $(f, B)$, with $B \subseteq \Omega$ an event and $f$ a gamble. Another subtle difference is that we take $f$ to be a gamble on $B$ rather than on $\Omega$. We prefer our choice because we think that a conditional model should be a model for—should be expressed in terms of gambles on—the elements of the conditioning event rather than those of the original state space, especially once we start using conditional models as updated ones.

After all this theory is in place, we start discussing the two conditioning rules that we are most interested in:

natural and regular extension. Section 3 introduces these rules in detail, defines them in terms of each of the different frameworks we consider, compares them with one another, provides pointers to the literature, and discusses various technical and computational issues. Every framework turns out to have its own merits: sets of probabilities provide the most intuitive—or ought we to say familiar—definition, sets of desirable gambles provide us with a conceptually simple interpretation, and lower previsions make the whole approach computationally tractable.

In Section 4, we shift the focus from conditioning to updating and study the problem of updating directly in terms of sets of desirable gambles. We introduce an asymmetric version of Walley's updating principle, discuss the conditions under which it makes sense to use it, and explain how it leads to a justification for updating by means of natural extension. It turns out that our approach leaves room for other updating rules as well, including more informative ones, such as regular extension. However, in order to justify them, our asymmetric version of Walley's updating principle is not sufficient, and needs to be combined with additional arguments.

This is exactly what we do in Section 5, for the particular case of updating by means of regular extension. The basic idea is that, since we are looking for an updated model that is meant to be used after some event $O$ has occurred, we are—in the process of coming up with such an updated model—making an implicit assumption that $O$ can occur. This assumption allows us to include an extra assessment that, when combined with our asymmetric version of Walley's updating principle, leads to conditional models that coincide with regular extension. We provide two versions of this justification for updating by means of regular extension. A simple version, which requires an assumption of ideal precision, and a more involved one, which does not.

We end the paper in Section 6 by summarising our main achievements and highlighting some aspects which we believe to be important. We also discuss remaining open problems and possible avenues for future work.

AppendixA gathers a number of additional technical results that we consider to be relevant, but which do not fit into the main storyline of the paper.

## 2. Modeling uncertainty within a finitary context

Consider a variable $X$ that takes values $\omega$ in a non-empty finite state space $\Omega$. A subject's uncertainty about the—possibly unknown—actual value of $X$ can then be represented in multiple ways, the most common of which is to use a single probability mass function on $\Omega$. In this section, we briefly introduce three alternative approaches: sets of desirable gambles, lower previsions, and (sets of) linear previsions.

Essential to each of these frameworks is the notion of a *gamble* on $\Omega$, which is a real-valued function on $\Omega$ that is interpreted as an uncertain payoff. If the actual value of $X$ turns out to be $\omega$, the owner of a gamble $f$ receives the—possibly negative—payoff $f(\omega)$, expressed in units of some predetermined linear utility scale.[1] We denote the set of all gambles on $\Omega$ as $\mathscr{G}(\Omega)$. This is a linear space under pointwise addition of gambles and pointwise multiplication of gambles with real numbers. For any two $f_1$ and $f_2$ in $\mathscr{G}(\Omega)$, we write '$f_1 \geq f_2$' if $(\forall \omega \in \Omega)\, f_1(\omega) \geq f_2(\omega)$ and '$f_1 > f_2$' if $f_1 \geq f_2$ and $f_1 \neq f_2$. Interesting subsets of $\mathscr{G}(\Omega)$ are denoted by using predicates as subscripts; for example: $\mathscr{G}(\Omega)_{>0} := \{f \in \mathscr{G}(\Omega) \colon f > 0\}$ is the set of all non-negative gambles on $\Omega$, excluding zero.

Events are identified with subsets of $\Omega$. Hence, the set of all events is the power set $\mathscr{P}(\Omega)$ of $\Omega$. We will often consider the set $\mathscr{P}_\emptyset(\Omega) := \mathscr{P}(\Omega) \setminus \{\emptyset\}$ of all non-empty events as well. Since $\Omega$ is finite, $\mathscr{P}(\Omega)$ and $\mathscr{P}_\emptyset(\Omega)$ are finite too. With every event $O \in \mathscr{P}(\Omega)$, we associate a special gamble $\mathbb{I}_O$ on $\Omega$, called its *indicator*, that assumes the value 1 on $O$ and 0 elsewhere.

### 2.1. Sets of desirable gambles

Sets of desirable gambles constitute the first framework we consider [40]. The basic idea here is to model a subject's uncertainty about the value of $X$ by looking at gambles on $\Omega$ that he finds desirable. A subject is said to find a gamble $f$ *desirable* if he strictly prefers it to the zero gamble. In other words, if he strictly prefers ownership of $f$ over ownership of the zero gamble—the status quo.

In order to reflect a rational subject's uncertainty, a *set of desirable gambles* $\mathscr{D}$ should be coherent, meaning that it satisfies the following consistency criteria.

[1] As long as the amounts of money remain limited, many people perceive the utility of a monetary reward to be a linear function of its monetary value; see Ref. [38, Sections 2.2.1 and 2.2.2] for additional discussion, including an example of a well-defined utility scale that is perfectly linear.

**Definition 1** (Coherence for sets of desirable gambles)**.** *A set $\mathscr{D}$ of desirable gambles on $\Omega$ is* coherent *if for all $\lambda \in \mathbb{R}_{>0}$ and all $f, f_1, f_2 \in \mathscr{G}(\Omega)$:*[2]

D1. $f = 0 \Rightarrow f \notin \mathscr{D}$; [avoiding null gain]

D2. $f > 0 \Rightarrow f \in \mathscr{D}$; [desiring partial gain]

D3. $f \in \mathscr{D} \Rightarrow \lambda f \in \mathscr{D}$; [positive scaling]

D4. $f_1, f_2 \in \mathscr{D} \Rightarrow f_1 + f_2 \in \mathscr{D}$. [combination]

Criteria D1 and D2 are rationality criteria; they follow directly from our interpretation of desirability. The zero gamble should not be desirable [D1]; gambles without negative payoffs and with the possibility of a positive payoff should always be desirable [D2]. Criteria D3 and D4 follow from the linearity of our utility scale.

Coherence has a number of useful consequences as well, which can be obtained by combining D1–D4. For example, for any coherent $\mathscr{D}$, and any $f, g \in \mathscr{G}(\Omega)$:

D5. $g \geq f$ and $f \in \mathscr{D} \Rightarrow g \in \mathscr{D}$; [monotonicity]

D6. $f \leq 0 \Rightarrow f \notin \mathscr{D}$. [avoiding non-positive gain]

*2.1.1. Natural extension*

In practice, we cannot expect a subject to specify for each gamble $f \in \mathscr{G}(\Omega)$ whether or not he finds it desirable. Instead, all that is usually obtained from an elicitation procedure is an *assessment* $\mathscr{A} \subseteq \mathscr{G}(\Omega)$, which may be only a subset of a subject's set of desirable gambles. Furthermore, such an assessment is often not coherent. However, by applying D2–D4, we can use $\mathscr{A}$ to infer the desirability of other gambles. The largest set of desirable gambles that can be constructed in this way is[3]

$$\mathscr{E}(\mathscr{A}) := \Big\{ \sum_{i=1}^{n} \lambda_i f_i : n \in \mathbb{N}, f_i \in \mathscr{A} \cup \mathscr{G}(\Omega)_{>0}, \lambda_i \in \mathbb{R}_{>0} \Big\}. \tag{1}$$

By construction, $\mathscr{E}(\mathscr{A})$ satisfies D2–D4. Consequently, $\mathscr{E}(\mathscr{A})$ is coherent if and only if it avoids null gain [D1]. Furthermore, if $\mathscr{E}(\mathscr{A})$ is coherent, then it is the smallest coherent set of desirable gambles that contains $\mathscr{A}$, and we then call $\mathscr{E}(\mathscr{A})$ the *natural extension* of $\mathscr{A}$. Since coherence is trivially preserved under taking intersections, this natural extension $\mathscr{E}(\mathscr{A})$ is then also equal to the intersection of all the coherent supersets of $\mathscr{A}$.

Even after enlarging an assessment by means of natural extension, the resulting set of desirable gambles is not guaranteed to be exhaustive. Further elicitation may result in additional desirable gambles. However, for various reasons, one may be unwilling or uncapable of performing such further elicitation; see Ref. [38, Section 2.10.3] for numerous examples. Hence, we will not require a set of desirable gambles $\mathscr{D}$ to be exhaustive, nor will we interpret it in this way.

*2.1.2. Conditional sets of desirable gambles*

One of the advantages of working with sets of desirable gambles is that conditioning is extremely elegant. Consider a set of desirable gambles $\mathscr{D}$ and an event $B \in \mathscr{P}_{\emptyset}(\Omega)$. Then the corresponding *conditional set of desirable gambles* is given by [10]

$$\mathscr{D}\rfloor B := \{ f \in \mathscr{G}(B) : \mathbb{I}_B f \in \mathscr{D} \}, \tag{2}$$

where, by introducing the convention that $0 \times \text{undefined} := 0$, we let $\mathbb{I}_B f$ be a gamble on $\Omega$ that coincides with $f$ on $B$ and is zero elsewhere. The intuition behind this definition is very simple: when $B$ occurs, the gambles $\mathbb{I}_B f$ and $f$ are indistinguishable in practice. Contingent on $B$ occurring, they yield the same payoff; if $B$ does not occur, then $\mathbb{I}_B f$ results in a zero payoff whereas $f$ is not defined. This definition of conditioning preserves coherence: if $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ is coherent, then $\mathscr{D}\rfloor B \subseteq \mathscr{G}(B)$ is clearly coherent as well.

[2]We use $\mathbb{R}_{>0}$ as a convenient shorthand notation for $\{\lambda \in \mathbb{R} : \lambda > 0\}$, and similarly for $\mathbb{R}_{\geq 0}$.

[3]We define the natural numbers $\mathbb{N}$ as the set of all *positive* integers. We use $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ to refer to the version that includes zero.

Alternative methods for conditioning a set of desirable gambles have also been proposed [38, 40, 22, 4], using the notation '$|$' rather than '$\rfloor$'; all of these alternative methods result in a set of gambles on $\Omega$ instead of $B$. We prefer the present version because we find it more intuitive that conditioning on an event $B$ produces a model for—a set of gambles on—that event. In any case, the choice between these definitions is mainly an aesthetical one, because they are all mathematically equivalent [10, Section 3.2].

### 2.2. *Lower previsions*

Instead of asking a subject to evaluate the desirability of a gamble directly, one can also ask him at which prices he would be willing to buy or sell that gamble. This is the approach that is taken in Walley's theory of lower previsions [38]. For any gamble $f$ on $\Omega$, the *lower prevision* $\underline{P}(f)$ of $f$ is a subject's supremum buying price for $f$. Similarly, the upper prevision $\overline{P}(f)$ of $f$ is his infimum selling price for $f$. Since selling $f$ for a price $\alpha$ is equivalent to buying $-f$ for a price $-\alpha$, lower and upper previsions are related by *conjugacy*: $\overline{P}(f) = -\underline{P}(-f)$. For this reason, it suffices to discuss only one of them. We follow Walley in concentrating on lower previsions.

Due to their interpretation as supremum buying prices, lower previsions can easily be related to sets of desirable gambles. In order to connect both approaches, it suffices to require that a subject considers the gamble $f-\alpha$ to be desirable if and only if he strictly prefers buying $f$ for the price $\alpha$ to the status quo—not buying any gamble at all. Using this connection, a coherent set of desirable gambles $\mathscr{D}$ trivially results in a lower prevision $\underline{P}_{\mathscr{D}}$, as defined by

$$\underline{P}_{\mathscr{D}}(f) := \sup\big\{\alpha \in \mathbb{R} \colon f-\alpha \in \mathscr{D}\big\} \text{ for every } f \in \mathscr{G}(\Omega). \tag{3}$$

Alternatively, lower previsions can be assessed directly as well. Any real valued function $\underline{P}$ with arbitrary domain $\mathscr{K} \subseteq \mathscr{G}(\Omega)$ can be interpreted as a lower prevision. We say that $\underline{P}$ is *coherent* if there is a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ such that $\underline{P}(f) = \underline{P}_{\mathscr{D}}(f)$ for all $f \in \mathscr{K}$.[4]

Lower previsions are less expressive than sets of desirable gambles. For a given coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$, there may be multiple coherent sets of desirable gambles $\mathscr{D}$ such that $\underline{P} = \underline{P}_{\mathscr{D}}$, the smallest of which is[5]

$$\mathscr{D}_{\underline{P}} := \{f \in \mathscr{G}(\Omega) \colon \underline{P}(f) > 0 \text{ or } f > 0\}.$$

All of these sets have the same associated set of *almost desirable gambles*

$$\mathrm{cl}(\mathscr{D}) := \big\{f \in \mathscr{G}(\Omega) \colon (\forall \delta \in \mathbb{R}_{>0})\ f+\delta \in \mathscr{D}\big\} = \{f \in \mathscr{G}(\Omega) \colon \underline{P}_{\mathscr{D}}(f) \geq 0\}. \tag{4}$$

We write $\mathrm{cl}(\mathscr{D})$ because, if $\mathscr{D}$ is coherent, then the set on the right-hand side of the defining equality in Eq. (4) is equal to the topological closure of $\mathscr{D}$, with respect to the topology that is induced by the Euclidean metric. Similarly, they will also have the same topological interior

$$\mathrm{int}(\mathscr{D}) := \big\{f \in \mathscr{G}(\Omega) \colon (\exists \delta \in \mathbb{R}_{>0})\ f-\delta \in \mathscr{D}\big\} = \{f \in \mathscr{G}(\Omega) \colon \underline{P}_{\mathscr{D}}(f) > 0\}. \tag{5}$$

It is furthermore easily proved that these conditions are equivalent: if $\mathscr{D}_1$ and $\mathscr{D}_2$ are two coherent sets of desirable gambles, then

$$\underline{P}_{\mathscr{D}_1} = \underline{P}_{\mathscr{D}_2} \Leftrightarrow \mathrm{cl}(\mathscr{D}_1) = \mathrm{cl}(\mathscr{D}_2) \Leftrightarrow \mathrm{int}(\mathscr{D}_1) = \mathrm{int}(\mathscr{D}_2). \tag{6}$$

Hence, coherent sets of desirable gambles with the same lower prevision $\underline{P}_{\mathscr{D}}$ differ only in their border $\mathrm{cl}(\mathscr{D}) \setminus \mathrm{int}(\mathscr{D})$. Nevertheless, which part of this border belongs to $\mathscr{D}$—the border structure of $\mathscr{D}$—may be important, for the following two reasons [40]. First of all, in a decision making context, it enables one to distinguish between strict and weak preference. For example, for two gambles $f, g \in \mathscr{G}(\Omega)$, we might say that $f$ is strictly preferred over $g$ if and only if $f-g \in \mathscr{D}$, whereas $f$ is weakly preferred over $g$ if and only if $f-g \in \mathrm{cl}(\mathscr{D}) \Leftrightarrow \underline{P}_{\mathscr{D}}(f-g) \geq 0$. The set of desirable gambles $\mathscr{D}$ is able to distinguish between these two notions, but the lower prevision $\underline{P}_{\mathscr{D}}$ is not.[6] Secondly, as we will illustrate further on, the border structure of a set of desirable gambles may have a significant impact on the conditional models it produces.

[4] The connection with other definitions of coherence for lower previsions is discussed in Section 2.2.3.
[5] In Ref. [38, Section 3.8.1], this set is denoted by $\mathscr{D}^+$ and is called the associated set of strictly desirable gambles.
[6] Similar observations can be made for other notions of strict and weak preference for sets of desirable gambles; see for example Refs. [38, 21, 40].

*2.2.1. Conditional lower previsions*

In the theory of lower previsions, conditional models are not merely regarded as derived concepts that are obtained through conditioning. Instead, they are primitive concepts, can be assessed directly, and are related to their unconditional counterparts by coherence. For any event $B \in \mathscr{P}_\emptyset(\Omega)$ and any gamble $f \in \mathscr{G}(B)$, we interpret the conditional lower prevision $\underline{P}(f\rfloor B)$ of $f$ given $B$ as a subject's supremum buying price for $f$, contingent on the occurrence of $B$.[7,8] When considered as an operator, a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ is a real-valued function whose domain can be any set $\mathscr{C}$ of couples $(f,B)$, with $B \in \mathscr{P}_\emptyset(\Omega)$ and $f \in \mathscr{G}(B)$. Hence, if we let $\mathscr{C}(\Omega) := \{(f,B)\colon B \in \mathscr{P}_\emptyset(\Omega), f \in \mathscr{G}(B)\}$ be the largest such set, then $\mathscr{C}$ can be any subset of $\mathscr{C}(\Omega)$. If $\mathscr{C}$ contains only couples of the form $(f,\Omega)$, with $f \in \mathscr{K} \subseteq \mathscr{G}(\Omega)$, then $\underline{P}(\cdot\rfloor\cdot)$ can be identified with an unconditional lower prevision $\underline{P}$ on $\mathscr{K}$, defined by $\underline{P}(f) := \underline{P}(f\rfloor\Omega)$ for all $f \in \mathscr{K}$. Furthermore, for any $B \in \mathscr{P}_\emptyset(\Omega)$, the operator $\underline{P}(\cdot\rfloor B)$ can be regarded as an unconditional lower prevision on $\mathscr{C}_B := \{f \in \mathscr{G}(B)\colon (f,B) \in \mathscr{C}\} \subseteq \mathscr{G}(B)$.

Due to the connection between desirability and buying prices, every coherent set of desirable gambles $\mathscr{D}$ has a unique corresponding conditional lower prevision $\underline{P}_\mathscr{D}(\cdot\rfloor\cdot)$, obtained by letting $\underline{P}_\mathscr{D}(\cdot\rfloor B)$ be the lower prevision that corresponds to $\mathscr{D}\rfloor B$:

$$\begin{aligned}\underline{P}_\mathscr{D}(f\rfloor B) &:= \sup\big\{\alpha \in \mathbb{R}\colon f-\alpha \in \mathscr{D}\rfloor B\big\} && (7)\\ &= \sup\big\{\alpha \in \mathbb{R}\colon \mathbb{I}_B[f-\alpha] \in \mathscr{D}\big\} \text{ for every } B \in \mathscr{P}_\emptyset(\Omega) \text{ and } f \in \mathscr{G}(B). && (8)\end{aligned}$$

However, sets of desirable gambles are still more expressive; different $\mathscr{D}$ can lead to the same $\underline{P}_\mathscr{D}(\cdot\rfloor\cdot)$ [21, Section 6].

A conditional lower prevision is said to be coherent if it can be derived from a coherent set of desirable gambles by means of Eq. (8).

**Definition 2** (Coherence for conditional lower previsions)**.** *A conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ is said to be* coherent *if there is some coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ such that*[9]

$$\underline{P}(f\rfloor B) = \underline{P}_\mathscr{D}(f\rfloor B) \text{ for every } (f,B) \in \mathscr{C}.$$

Since coherence for sets of desirable gambles is preserved under taking intersections, we find that coherence for conditional lower previsions is preserved under taking pointwise infima, or equivalently: the lower envelope of a set of coherent lower previsions is again a coherent lower prevision.

**Proposition 1.** *Consider an arbitrary index set $I$ and, for every $i \in I$, a coherent conditional lower prevision $\underline{P}_i(\cdot\rfloor\cdot)$ on $\mathscr{C} \subseteq \mathscr{C}(\Omega)$. The conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, defined by*

$$\underline{P}(f\rfloor B) := \inf_{i\in I} \underline{P}_i(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}$$

*is then also coherent.*

*Proof of Proposition 1.* For every $i \in I$, the coherence of $\underline{P}_i(\cdot\rfloor\cdot)$ implies the existence of a coherent set of desirable gambles $\mathscr{D}_i \subseteq \mathscr{G}(\Omega)$ such that $\underline{P}_i(f\rfloor B) = \underline{P}_{\mathscr{D}_i}(f\rfloor B)$ for all $(f,B) \in \mathscr{C}$. Now let $\mathscr{D} := \cap_{i\in I}\mathscr{D}_i$. Since coherence for sets of desirable gambles is trivially preserved under taking intersections, we know that $\mathscr{D}$ is coherent. Consider now any $(f,B) \in \mathscr{C}$. In order to prove the result, it clearly suffices to show that $\underline{P}_\mathscr{D}(f\rfloor B) = \inf_{i\in I}\underline{P}_i(f\rfloor B)$.

For all $i \in I$, we have that $\mathscr{D} \subseteq \mathscr{D}_i$ and therefore also that $\underline{P}_\mathscr{D}(f\rfloor B) \leq \underline{P}_{\mathscr{D}_i}(f\rfloor B) = \underline{P}_i(f\rfloor B)$. This allows us to infer that $\underline{P}_\mathscr{D}(f\rfloor B) \leq \inf_{i\in I}\underline{P}_i(f\rfloor B)$. Consider now any $\varepsilon \in \mathbb{R}_{>0}$. Eq. (8) then implies that $\mathbb{I}_B[f-\underline{P}_\mathscr{D}(f\rfloor B)-\varepsilon] \notin \mathscr{D} = \cap_{i\in I}\mathscr{D}_i$. Hence, there is some $i^* \in I$ such that $\mathbb{I}_B[f-\underline{P}_\mathscr{D}(f\rfloor B)-\varepsilon] \notin \mathscr{D}_{i^*}$. Since $\mathscr{D}_{i^*}$ is coherent, this implies that $\mathbb{I}_B[f-\alpha] \notin \mathscr{D}_{i^*}$ for all $\alpha \geq \underline{P}_\mathscr{D}(f\rfloor B)+\varepsilon$, which in turn implies that $\underline{P}_{\mathscr{D}_{i^*}}(f\rfloor B) \leq \underline{P}_\mathscr{D}(f\rfloor B)+\varepsilon$. Since we also know that $\inf_{i\in I}\underline{P}_i(f\rfloor B) = \inf_{i\in I}\underline{P}_{\mathscr{D}_i}(f\rfloor B) \leq \underline{P}_{\mathscr{D}_{i^*}}(f\rfloor B)$, this allows us to infer that $\inf_{i\in I}\underline{P}_i(f\rfloor B) \leq \underline{P}_\mathscr{D}(f\rfloor B)+\varepsilon$. Since this is true for all $\varepsilon \in \mathbb{R}_{>0}$, we find that $\inf_{i\in I}\underline{P}_i(f\rfloor B) \leq \underline{P}_\mathscr{D}(f\rfloor B)$. We conclude that $\underline{P}_\mathscr{D}(f\rfloor B) = \inf_{i\in I}\underline{P}_i(f\rfloor B)$. □

[7] Other authors consider the conditional lower prevision $\underline{P}(f|B)$ of gambles $f \in \mathscr{G}(\Omega)$ instead; the connection with our approach will be established in Section 2.2.3.

[8] See Eq. (7) for an explicit definition.

[9] We will establish a connection with other definitions of coherence for conditional lower previsions in Section 2.2.3.

Coherence of $\underline{P}(\cdot\rfloor\cdot)$ implies that $\underline{P}(\cdot\rfloor\cdot)$ is *separately coherent* [38, Section 6.2.2], by which we mean that, for all $B\in\mathscr{P}_\emptyset(\Omega)$, $\underline{P}(\cdot\rfloor B)$ is a coherent lower prevision on $\mathscr{C}_B$.[10] However, separate coherence does not imply coherence: coherence of $\underline{P}(\cdot\rfloor B)$, for all $B\in\mathscr{P}_\emptyset(\Omega)$, is not sufficient for $\underline{P}(\cdot\rfloor\cdot)$ to be coherent. Whenever we want to clearly distinguish between separate coherence and coherence, we will refer to the latter as *joint coherence*.

If for every $B\in\mathscr{B}$, with $\mathscr{B}$ a subset of $\mathscr{P}_\emptyset(\Omega)$, we have a lower prevision $\underline{P}(\cdot\rfloor B)$ on $\mathscr{K}_B\subseteq\mathscr{G}(B)$, we use $\underline{P}(\cdot\rfloor\mathscr{B}):=\{\underline{P}(\cdot\rfloor B)\colon B\in\mathscr{B}\}$ to refer to this collection of lower previsions. It should be clear that such a collection can also be identified—trivially—with a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}:=\{(f,B)\colon B\in\mathscr{B},f\in\mathscr{K}_B\}$. We call a collection $\underline{P}(\cdot\rfloor\mathscr{B})$ separately coherent if each of its individual elements $\underline{P}(\cdot\rfloor B)$ is coherent; we call it (jointly) coherent if the corresponding conditional lower prevision is (jointly) coherent. If the collection we are referring to is clear from the context, we do not mention it explicitely. For example, we might say that a lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ is *coherent with* a lower prevision $\underline{P}(\cdot\rfloor B)$ on $\mathscr{G}(B)$; by this, we simply mean that the corresponding collection—consisting of the lower previsions $\underline{P}(\cdot\rfloor\Omega):=\underline{P}(\cdot)$ and $\underline{P}(\cdot\rfloor B)$—is jointly coherent.

### 2.2.2. Natural extension

Since $\mathscr{D}$ is not required to be exhaustive, $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ is not exhaustive either; the subject's actual supremum buying price for $f$ contingent on $B$ may be higher than $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$. If $\underline{P}(\cdot\rfloor\cdot)$ is assessed directly, then similarly, we do not require it to be exhaustive. A particularly useful advantage of this interpretation is that it allows us to turn a—possibly incoherent—lower prevision $\underline{P}(\cdot\rfloor\cdot)$ into a coherent one by correcting it upwards.

To understand how this comes about naturally, the first step is to realise that a coherent lower prevision is simply an assessment of desirable gambles: due to our interpretation for $\underline{P}(f\rfloor B)$, we know that for any $\varepsilon\in\mathbb{R}_{>0}$, there is some $\alpha\geq\underline{P}(f\rfloor B)-\varepsilon$ such that $\mathbb{I}_B[f-\alpha]$ is a desirable gamble. By combining this with D2 and D4, we find that the gambles in

$$\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}:=\left\{\mathbb{I}_B[f-\underline{P}(f\rfloor B)+\varepsilon]\colon (f,B)\in\mathscr{C}\text{ and }\varepsilon\in\mathbb{R}_{>0}\right\}.$$

are desirable and therefore also, by D2–D4, that the gambles in $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}:=\mathscr{E}(\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)})$ are desirable. Furthermore, since $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is the natural extension of the assessment $\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}$, we know from Section 2.1.1 that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is the smallest set of gambles whose desirability is implied by coherence (D2–D4) and (the assessment that corresponds to) $\underline{P}(\cdot\rfloor\cdot)$.

The next step is to consider the supremum buying prices that correspond to this set of desirable gambles $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, as given by

$$\underline{E}(f\rfloor B):=\underline{P}_{\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}}(f\rfloor B)\text{ for every }B\in\mathscr{P}_\emptyset(\Omega)\text{ and }f\in\mathscr{G}(B).\tag{9}$$

The resulting operator $\underline{E}(\cdot\rfloor\cdot)$ is defined on $\mathscr{C}(\Omega)$, and its restriction to $\mathscr{C}$ dominates $\underline{P}(\cdot\rfloor\cdot)$, in the sense that

$$\underline{E}(f\rfloor B)\geq\underline{P}(f\rfloor B)\text{ for all }(f,B)\in\mathscr{C}.\tag{10}$$

Let us begin by taking a look at what happens if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent. In that case, $\underline{E}(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision, and we will refer to it as the *natural extension* of $\underline{P}(\cdot\rfloor\cdot)$. Not only does it—if necessary—correct $\underline{P}(\cdot\rfloor\cdot)$ upwards on $\mathscr{C}$ to make it coherent, it also extends the domain of this correction to all of $\mathscr{C}(\Omega)$. Furthermore, out of all such coherent upwards corrections of $\underline{P}(\cdot\rfloor\cdot)$, $\underline{E}(\cdot\rfloor\cdot)$ is the most conservative—most imprecise—one that can always be inferred from $\underline{P}(\cdot\rfloor\cdot)$ without having to add extra assessments.

**Proposition 2.** *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ and let $\underline{P}'(\cdot\rfloor\cdot)$ be any coherent conditional lower prevision on $\mathscr{C}'\supseteq\mathscr{C}$ that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Then*

$$\underline{P}'(f\rfloor B)\geq\underline{E}(f\rfloor B)\textit{ for all }(f,B)\in\mathscr{C}'.$$

*Proof.* By Definition 2, there is some coherent set of desirable gambles $\mathscr{D}$ such that $\underline{P}_{\mathscr{D}}(f\rfloor B)=\underline{P}'(f\rfloor B)$ for all $(f,B)\in\mathscr{C}'$. Furthermore, by an argument similar to the one we provided for $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ in the main text, we know that $\mathscr{E}_{\underline{P}'(\cdot\rfloor\cdot)}$ consists of gambles whose desirability is implied by $\underline{P}'(\cdot\rfloor\cdot)$. Hence, we find that $\mathscr{E}_{\underline{P}'(\cdot\rfloor\cdot)}\subseteq\mathscr{D}$. Also, since $\underline{P}'(\cdot\rfloor\cdot)$ dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, $\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}$ is clearly a subset of $\mathscr{A}_{\underline{P}'(\cdot\rfloor\cdot)}$ and therefore $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is a subset of $\mathscr{E}_{\underline{P}'(\cdot\rfloor\cdot)}$. This implies that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}\subseteq\mathscr{D}$ and therefore also, by Eq. (7), that $\underline{P}'(f\rfloor B)=\underline{P}_{\mathscr{D}}(f\rfloor B)\geq\underline{P}_{\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}}(f\rfloor B)=\underline{E}(f\rfloor B)$ for all $(f,B)\in\mathscr{C}'$. $\square$

[10] It suffices to realise that there is a coherent set of desirable gambles on $\mathscr{G}(B)$—in particular: $\mathscr{D}\rfloor B$—from which $\underline{P}(\cdot\rfloor B)$ can be obtained by means of Eq. (3)—in particular: Eq. (7).

We conclude that if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent, the natural extension $\underline{E}(\cdot\rfloor\cdot)$ provides us with the most conservative—lowest—coherent supremum buying prices that are compatible with (the assessment that corresponds to) $\underline{P}(\cdot\rfloor\cdot)$.

So far, so good. But what if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent? As we know from Section 2.1.1, the only way for this to happen is if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ does not avoid null gain [D1], meaning that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ contains the zero gamble. Even worse, as we show in Proposition 3 below, there are $f \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ for which $f < 0$. In other words, there are gambles whose desirability is implied by $\underline{P}(\cdot\rfloor\cdot)$ and D2–D4, but which are guaranteed never to yield a positive payoff, and in some cases even yield a negative payoff. If this happens, then clearly, there is something wrong with $\underline{P}(\cdot\rfloor\cdot)$. Indeed, it turns out that if $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent, then $\underline{P}(\cdot\rfloor\cdot)$ is incoherent as well, and it cannot be made coherent by correcting it upwards. Hence, in that case, it is not possible to construct a coherent lower prevision that is consistent with (the non-exhaustive interpretation of) $\underline{P}(\cdot\rfloor\cdot)$, and the only option is to reassess $\underline{P}(\cdot\rfloor\cdot)$.

**Proposition 3.** *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$. Then the following statements are equivalent:*

(i) *$\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent;*

(ii) *$f < 0$ for some $f \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$;*

(iii) *$\underline{E}(f\rfloor B) = +\infty$ for some $(f,B) \in \mathscr{C}$;*

(iv) *Every conditional lower prevision that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$ is incoherent.*

*Proof.* It clearly suffices to show that (i)⇒(ii)⇒(i), (i)⇒(iii) and ¬(i)⇒ ¬(iv)⇒ ¬(iii).

We start by proving that (i)⇒(ii) and (i)⇒(iii). So assume that (i) holds: $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent. Then as explained in Section 2.1.1, $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ does not avoid null gain [D1], implying that there are $n \in \mathbb{N}$, $(\forall i \in \{1,\dots,n\})$ $(f_i,B_i) \in \mathscr{C}$, $\varepsilon_i \in \mathbb{R}_{>0}$, $\lambda_i \in \mathbb{R}_{>0}$ such that

$$g := \sum_{i=1}^{n} \lambda_i \mathbb{I}_{B_i}[f_i - \underline{P}(f_i\rfloor B_i) + \varepsilon_i] = 0.$$

Hence, for all $\lambda \in \mathbb{R}_{>0}$, we find that

$$-\lambda\mathbb{I}_{B_1} = -\frac{2\lambda}{\lambda_1\varepsilon_1}\lambda_1\mathbb{I}_{B_1}\frac{\varepsilon_1}{2} \geq \frac{2\lambda}{\lambda_1\varepsilon_1}\left(-\sum_{i=1}^{n}\lambda_i\mathbb{I}_{B_i}\frac{\varepsilon_i}{2} + g\right) = \frac{2\lambda}{\lambda_1\varepsilon_1}\left(\sum_{i=1}^{n}\lambda_i\mathbb{I}_{B_i}\left[f_i - \underline{P}(f_i\rfloor B_i) + \frac{\varepsilon_i}{2}\right]\right) =: h$$

and therefore, since $h \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, and because $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies D2 and D4 and therefore also D5, that $-\lambda\mathbb{I}_{B_1} \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$. This already implies that (ii) holds. Now assume *ex absurdo* that (iii) does not hold, implying in particular that $\underline{E}(f_1\rfloor B_1) < \infty$ and therefore, that there is some $\alpha \in \mathbb{R}$ such that $\mathbb{I}_{B_1}(f_1 - \alpha) \notin \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$. Consider any $\lambda \in \mathbb{R}_{>0}$ such that $\lambda > \alpha - \min f_1$. Then $\mathbb{I}_{B_1}(f_1 - \alpha) \geq -\lambda\mathbb{I}_{B_1}$ and hence, since (as we have just shown) $-\lambda\mathbb{I}_{B_1} \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, and because $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies D5, we find that $\mathbb{I}_{B_1}(f_1 - \alpha) \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$. This is a contradiction, allowing us to conclude that (iii) holds.

(ii)⇒(i) is trivial: by combining (ii) with the fact that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies D5, we immediately find that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ does not avoid null gain [D1].

For ¬(i)⇒ ¬(iv), it suffices to recall that, as explained earlier on in Section 2.2.2, coherence of $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ implies that $\underline{E}(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$.

Finally, we prove that ¬(iv)⇒ ¬(iii). So assume that ¬(iv): there is a coherent conditional lower prevision $\underline{P}'(\cdot\rfloor\cdot)$ that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Then by Proposition 2, for any $(f,B) \in \mathscr{C}$, $\underline{P}'(f\rfloor B) \geq \underline{E}(f\rfloor B)$. Hence, since $\underline{P}'(\cdot\rfloor\cdot)$ is by definition real-valued, we find that $\underline{E}(f\rfloor B) < \infty$ for all $(f,B) \in \mathscr{C}$. □

By combining this result with Proposition 2, we obtain the following alternative characterisation of coherence.

**Corollary 4.** *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$. Then $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if it coincides with $\underline{E}(\cdot\rfloor\cdot)$ on $\mathscr{C}$.*

*Proof.* First assume that $\underline{P}(\cdot\rfloor\cdot)$ is coherent. Then trivially, $\underline{P}(\cdot\rfloor\cdot)$ is a coherent lower prevision that dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Hence, by Proposition 2, $\underline{P}(\cdot\rfloor\cdot)$ dominates $\underline{E}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. However, by Eq. (10), the converse holds as well. Hence, $\underline{P}(\cdot\rfloor\cdot)$ and $\underline{E}(\cdot\rfloor\cdot)$ coincide on $\mathscr{C}$.

Next, assume that $\underline{P}(\cdot\rfloor\cdot)$ and $\underline{E}(\cdot\rfloor\cdot)$ coincide on $\mathscr{C}$. Now assume *ex absurdo* that $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is incoherent. Then by Proposition 3(i)&(iii), there is some $(f,B)\in\mathscr{C}$ such that $\underline{E}(f\rfloor B)=\infty$ and therefore, by assumption, also $\underline{P}(f\rfloor B)=\infty$. This is a contradiction because $\underline{P}(\cdot\rfloor\cdot)$ is a conditional lower prevision and therefore, by definition, real-valued. Hence, $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent. This implies that $\underline{E}(\cdot\rfloor\cdot)$ is coherent too, and therefore, since $\underline{P}(\cdot\rfloor\cdot)$ is by assumption the restriction of $\underline{E}(\cdot\rfloor\cdot)$ to $\mathscr{C}$, $\underline{P}(\cdot\rfloor\cdot)$ is coherent as well. □

If $\underline{P}(\cdot\rfloor\cdot)$ is coherent then, as one would intuitively expect, the most conservative coherent upwards correction of $\underline{P}(\cdot\rfloor\cdot)$ is $\underline{P}(\cdot\rfloor\cdot)$ itself. However, even in that case, natural extension is still an important tool, as it allows us to coherently extend $\underline{P}(\cdot\rfloor\cdot)$ to all of $\mathscr{C}(\Omega)$ in the most conservative way possible.

Consider for example the common case where $\mathscr{C}:=\{(f,\Omega)\colon f\in\mathscr{G}(\Omega)\}$, meaning that $\underline{P}(\cdot\rfloor\cdot)$ is essentially an unconditional lower prevision $\underline{P}(\cdot):=\underline{P}(\cdot\rfloor\Omega)$ on $\mathscr{G}(\Omega)$. In that case, natural extension can be regarded as a conditioning rule: for any $B\in\mathscr{P}_{\emptyset}(\Omega)$, it provides us with a corresponding coherent lower prevision $\underline{E}(\cdot\rfloor B)$ on $\mathscr{G}(B)$;[11] it is guaranteed to be coherent with $\underline{P}$, and out of all lower previsions on $\mathscr{G}(B)$ that are coherent with $\underline{P}$, it is the most conservative—most imprecise—one.

### 2.2.3. *Comparison with other approaches*

Besides Definition 2, many other definitions for coherence have been proposed as well; see Ref. [24] for an overview. As we are about to explain, within our finitary context—recall that $\Omega$ is finite—, many of them are mathematically equivalent to our approach.

First of all, in most of these other, more conventional definitions, the lower prevision $\underline{P}(\cdot|B)$ is defined for gambles on $\Omega$ rather than $B$. By analogy with conditioning for sets of desirable gambles, we use the notation '$\rfloor$' rather than the more conventional '$|$' as a reminder of this difference. On the more conventional approach, $\underline{P}_{\mathscr{D}}(f|B)$ is still defined by Eq. (8), but with $f$ an element of $\mathscr{G}(\Omega)$ instead of $\mathscr{G}(B)$, and Definition 2 remains identical: we just have to replace '$\rfloor$' by '$|$'. We prefer our present version because we find it more intuitive that conditioning on an event $B$ produces a model for—a lower prevision for gambles on—that event. Also, it allows us to use Eq. (7), which we think is particularly elegant because it illustrates that Eq. (8) follows directly from Eqs. (2) and (3). Mathematically, both approaches are equivalent. Indeed, if $\underline{P}(\cdot|\cdot)$ is coherent, then $\underline{P}(f|B)$ depends only on the restriction $f_B\in\mathscr{G}(B)$ of $f$ to $B$, allowing us to identify $\underline{P}(\cdot|\cdot)$ with a coherent lower prevision $\underline{P}(\cdot\rfloor\cdot)$. Conversely, if $\underline{P}(\cdot\rfloor\cdot)$ is coherent, then the lower prevision $\underline{P}(\cdot|\cdot)$ that is defined by

$$\underline{P}(f|B):=\underline{P}(f_B\rfloor B)\text{ for all }f\in\mathscr{G}(\Omega)\text{ and }B\in\mathscr{P}_{\emptyset}(\Omega)\text{ such that }(f_B,B)\in\mathscr{C}$$

is coherent. Even stronger: $\underline{P}(\cdot|\cdot)$ is coherent if and only if $\underline{P}(\cdot\rfloor\cdot)$ is coherent. Given this connection, we now compare Definition 2 with a number of alternative definitions for coherence.

Formally, the notion of coherence that resembles Definition 2 the most is that of Williams [44, 43, 35] (W-coherence). The main difference is that he considers so-called acceptable gambles rather than desirable ones. As was essentially pointed out by Williams himself—the cone $\mathscr{A}_0'$ in Ref. [44, Proposition 3] is identical to our $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$—, this leads to an equivalent definition; see Ref. [21] for some results in terms of desirable gambles. Two other differences are that Williams considers upper rather than lower previsions and that he imposes some structure on the domain of $\underline{P}(\cdot|\cdot)$.[12] Ref. [24, Section 3.1] explains that this does not make any difference either; structure-free generalisations of W-coherence for lower previsions can be found in Refs. [31, 24, 34] and [32, Chapter 13]. Another, more popular definition of coherence is that of Walley [38] (Walley-coherence); this definition is not structure-free. When it is applicable, Walley-coherence is known to be equivalent to W-coherence if $\Omega$ is finite [38, Appendix K]. We conclude that within our finitary context, Definition 2 is equivalent to both Walley- and W-coherence and that therefore, we can import all sorts of useful results that were developed for these other notions of coherence, ranging from theoretical properties [38, 32] to computational techniques [42, 34].

[11] We provide an explicit expression for this conditioning rule further on in Section 3.2.
[12] He requires the domain to be of the form in Corollary 6.

*2.2.4. Properties of coherent conditional lower previsions*

Coherence has many useful consequences; see for example Refs. [43, 38, 32]. We list only a few.

Consider a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with arbitrary domain $\mathscr{C}$ and let $\overline{P}(\cdot\rfloor\cdot)$ be the associated conditional upper prevision, as defined by

$$\overline{P}(f\rfloor B) := -\underline{P}(-f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}(\Omega) \text{ such that } (-f,B) \in \mathscr{C}.$$

Then for all $A,B \in \mathscr{P}_\emptyset(\Omega)$ such that $B \subseteq A$, all $\lambda \in \mathbb{R}_{\geq 0}$, all $\mu \in \mathbb{R}$ and all $f,g \in \mathscr{G}(B)$, the following properties hold whenever the expressions involved are well-defined:

C1. $\underline{P}(f\rfloor B) \geq \min f$;

C2. $\underline{P}(\lambda f\rfloor B) = \lambda \underline{P}(f\rfloor B)$; [non-negative homogeneity]

C3. $\underline{P}(f+g\rfloor B) \geq \underline{P}(f\rfloor B) + \underline{P}(g\rfloor B)$; [super-additivity]

C4. $\underline{P}(\mathbb{I}_B[f-\underline{P}(f\rfloor B)]\rfloor A) = 0$;

C5. $\overline{P}(f\rfloor B) + \overline{P}(g\rfloor B) \geq \overline{P}(f+g\rfloor B) \geq \overline{P}(f\rfloor B) + \underline{P}(g\rfloor B) \geq \underline{P}(f+g\rfloor B)$;

C6. $\underline{P}(f\rfloor B) \leq \overline{P}(f\rfloor B) \leq \max f$;

C7. $g \geq f \Rightarrow \underline{P}(g\rfloor B) \geq \underline{P}(f\rfloor B)$; [monotonicity]

C8. $\underline{P}(f+\mu\rfloor B) = \underline{P}(f\rfloor B) + \mu$. [constant additivity]

Conditions C1–C4 have a special status because—as the following result by Williams establishes—they can be used to characterise coherence.

**Proposition 5** ([43, 38, 32])**.** *Consider a non-empty subset $\mathscr{B}$ of $\mathscr{P}_\emptyset(\Omega)$ and, for all $B \in \mathscr{B}$, a linear subspace $\mathscr{K}_B$ of $\mathscr{G}(B)$. Let $\underline{P}(\cdot\rfloor\cdot)$ be any conditional lower prevision with domain $\mathscr{C} := \{(f,B)\colon B \in \mathscr{B} \text{ and } f \in \mathscr{K}_B\}$. Then $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if it satisfies* C1–C4. *Furthermore, for any $B \in \mathscr{B}$, $\underline{P}(\cdot\rfloor B)$ is coherent if and only if it satisfies* C1–C3. *Hence, in this particular case, $\underline{P}(\cdot\rfloor\cdot)$ is jointly coherent if and only if it is separately coherent and satisfies* C4.

*2.2.5. Betting rates*

Before we move on to the connection between lower previsions and probability measures, we would like to draw attention to a particular aspect of sets of desirable gambles and the lower previsions that are associated with them.

Clearly, the notion of a gamble is closely related to betting: similarly to what happens with betting, we either lose or gain utility (money), depending on the uncertain value of a variable $X$ (the outcome of some experiment). This connection is especially clear if we consider indicators of events. For any $B \in \mathscr{P}_\emptyset(\Omega)$ and $\lambda \in \mathbb{R}_{>0}$ and $\alpha \in \mathbb{R}$, the gamble $\lambda(\mathbb{I}_B - \alpha)$ corresponds to paying $\lambda\alpha$ in order to receive $\lambda$ if $B$ happens. In other words: betting on $B$, at a betting rate $\alpha$, and with stakes $\lambda$. Due to the linearity of our utility scale [C2], the desirability of such a bet does not depend on the stakes $\lambda$, but only on the betting rate $\alpha$: the gamble $\mathbb{I}_B - \alpha$ is desirable if and only if you are willing to bet on $B$ at a betting rate $\alpha$. Hence, we find that the supremum betting rate at which you are willing to bet on $B$, as defined by

$$\underline{P}_{\mathscr{D}}(B) := \sup\{\alpha \in \mathbb{R}\colon \mathbb{I}_B - \alpha \in \mathscr{D}\} = \underline{P}_{\mathscr{D}}(\mathbb{I}_B), \tag{11}$$

is equal to the lower prevision of $\mathbb{I}_B$. Similarly, since $\alpha - \mathbb{I}_B$ is desirable if and only if you are willing to *take* bets on $B$ at a betting rate $\alpha$—bet against $B$ at a betting rate $1-\alpha$—we find that the upper prevision of $\mathbb{I}_B$ is equal to the infimum betting rate

$$\overline{P}_{\mathscr{D}}(B) := \overline{P}_{\mathscr{D}}(\mathbb{I}_B) = \inf\{\alpha \in \mathbb{R}\colon \alpha - \mathbb{I}_B \in \mathscr{D}\} \tag{12}$$

at which you are willing to take bets on $B$. By coherence [C1 and C6], we find that

$$0 \leq \underline{P}_{\mathscr{D}}(B) \leq \overline{P}_{\mathscr{D}}(B) \leq 1, \tag{13}$$

as is to be expected for (supremum and infimum) betting rates. For reasons that should become clear shortly, $\underline{P}_{\mathscr{D}}(B)$ and $\overline{P}_{\mathscr{D}}(B)$ are often referred to as the lower and upper probability of $B$, respectively. However, this should not be taken to imply the existence of an unknown probability $P(B)$ of $B$, for which $\underline{P}_{\mathscr{D}}(B)$ and $\overline{P}_{\mathscr{D}}(B)$ provide lower and upper bounds; this may be the case in some situations, but in general, the interpretation in terms of betting rates is more fundamental. We discuss this point further in Section 2.4.3.

*2.3. Linear previsions*

If a subject's supremum rate $\underline{P}(B)$ for betting on $B$ coincides with the infimum rate $\overline{P}(B)$ at which he is willing to take bets on $B$, then $P(B) := \underline{P}(B) = \overline{P}(B)$ is his fair betting rate for the event $B$. Similarly, for a gamble $f \in \mathscr{G}(\Omega)$, if $\underline{P}(f) = \overline{P}(f)$, then $P(f) := \underline{P}(f) = \overline{P}(f)$ is the subject's fair price for $f$, called the prevision of $f$ by de Finetti [11]. If a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ assigns such fair prices to all gambles, in the sense that $(f,B) \in \mathscr{C}$ if and only if $(-f,B) \in \mathscr{C}$ and that

$$\underline{P}(f\rfloor B) = \overline{P}(f\rfloor B) =: -\underline{P}(-f\rfloor B) \text{ for all } (f,B) \in \mathscr{C},$$

then $\underline{P}(\cdot\rfloor\cdot)$ is said to be *self-conjugate*, is referred to as a *conditional prevision*, and we then simply write $P(\cdot\rfloor\cdot)$ instead of $\underline{P}(\cdot\rfloor\cdot)$ or $\overline{P}(\cdot\rfloor\cdot)$.

If a conditional prevision $P(\cdot\rfloor\cdot)$ is *coherent*, then by combining C1–C4 with self-conjugacy, we find that it satisfies the following properties. For all $A,B \in \mathscr{P}_\emptyset(\Omega)$ such that $B \subseteq A$, all $\lambda \in \mathbb{R}$, and all $f,g \in \mathscr{G}(B)$, and whenever the expressions involved are well-defined:

P1. $\min f \leq P(f\rfloor B) \leq \max f$;

P2. $P(\lambda f\rfloor B) = \lambda P(f\rfloor B)$; [homogeneity]

P3. $P(f+g\rfloor B) = P(f\rfloor B) + P(g\rfloor B)$; [additivity]

P4. $P(\mathbb{I}_B f\rfloor A) = P(f\rfloor B)P(B\rfloor A)$.

As we can see from conditions P2 and P3, for all $B \in \mathscr{P}_\emptyset(\Omega)$, $P(\cdot\rfloor B)$ is a linear operator, and for this reason, coherent conditional previsions are also referred to as *linear* conditional previsions. Proposition 5 leads to the following convenient characterisation.

**Corollary 6** ([43, 38, 32])**.** *Consider a non-empty subset $\mathscr{B}$ of $\mathscr{P}_\emptyset(\Omega)$ and, for all $B \in \mathscr{B}$, a linear subspace $\mathscr{K}_B$ of $\mathscr{G}(B)$. Let $P(\cdot\rfloor\cdot)$ be any conditional prevision with domain $\mathscr{C} := \{(f,B)\colon B \in \mathscr{B}$ and $f \in \mathscr{K}_B\}$. Then $P(\cdot\rfloor\cdot)$ is coherent if and only if it satisfies* P1–P4. *Furthermore, for any $B \in \mathscr{B}$, $P(\cdot\rfloor B)$ is coherent if and only if it satisfies* P1–P3. *Hence, in this particular case, $P(\cdot\rfloor\cdot)$ is jointly coherent if and only if it is separately coherent and satisfies* P4.

We denote the set of all linear conditional previsions on $\mathscr{C}(\Omega)$ by $\mathbb{P}$. Furthermore, for any $B \in \mathscr{P}_\emptyset(\Omega)$, we let $\mathbb{P}_B$ be the set consisting of all linear previsions on $\mathscr{G}(B)$. In this way, for any $B \in \mathscr{P}_\emptyset(\Omega)$ and $P(\cdot\rfloor\cdot) \in \mathbb{P}$, we have that $P(\cdot\rfloor B) \in \mathbb{P}_B$. $\mathbb{P}_\Omega$ corresponds to an important special case.

*2.3.1. Full conditional probability measures*

One of the reasons why linear previsions are an important special case, is because they allow us to link the behavioural approach to modelling uncertainty, which we have just introduced, with the more conventional approach using probability measures.

Indeed, consider a linear prevision $P(\cdot\rfloor\cdot)$ on $\mathscr{C}(\Omega)$ and use $P(B\rfloor A)$ as a shorthand notation for $P(\mathbb{I}_B\rfloor A)$. Then by P1–P4, for all $A,B \in \mathscr{P}_\emptyset(\Omega)$ and $C,D \in \mathscr{P}(\Omega)$ such that $C \cap D = \emptyset$ and $C \subseteq B \subseteq A$:

F1. $P(B\rfloor B) = 1$;

F2. $P(C\rfloor B) \geq 0$;

F3. $P(C \cup D\rfloor B) = P(C\rfloor B) + P(D\rfloor B)$;

F4. $P(C\rfloor A) = P(C\rfloor B)P(B\rfloor A)$. [Bayes's rule]

Hence, formally, the restriction of $P(\cdot\rfloor\cdot)$ to indicators can be identified with a full conditional probability measure, because F1–F4 are the defining properties for such a measure [13].[13] Conditions F1–F3 assert that, for any $B \in \mathscr{P}_\emptyset(\Omega)$,

[13]see Ref. [6] for a recent overview of related literature.

$P(\cdot\rfloor B)$ satisfies the usual axioms of probability, and F4 corresponds to Bayes's rule. Furthermore, the original conditional prevision can be fully recovered from this full conditional probability measure. By linearity, for any $f \in \mathscr{G}(B)$, since $f = \sum_{\omega \in B} f(\omega)\mathbb{I}_{\{\omega\}}$, we find that

$$P(f\rfloor B) = \sum_{\omega \in B} f(\omega)p(\omega\rfloor B)$$

is the expected value of $f$ with respect to the probability mass function $p(\cdot\rfloor B)$ that corresponds to $P(\cdot\rfloor B)$, as defined by $p(\omega\rfloor B) := P(\{\omega\}\rfloor B) = P(\mathbb{I}_{\{\omega\}}\rfloor B)$ for all $\omega \in B$.

If every element of $\Omega$ has positive probability, Bayes's rule uniquely determines all conditional probabilities, and in that case, a full conditional probability measure and its associated linear conditional prevision is completely characterised by the unconditional mass function $p(\cdot) := p(\cdot\rfloor\Omega)$. However, if $P(B) = 0$, Bayes's rule imposes no non-trivial restrictions on $P(\cdot\rfloor B)$ and a full conditional measure then allows for $P(\cdot\rfloor B)$ to be specified separately.

### 2.4. *Sets of linear previsions*

The link between probability measures and lower previsions is not restricted to the special case of linear previsions. In general, lower previsions are related to *sets* of linear previsions and therefore, by the results in the previous section, to sets of probability measures.

#### 2.4.1. *The lower envelope theorem*

For any conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on an arbitrary domain $\mathscr{C}$, we can consider the corresponding set of dominating linear conditional previsions, as given by

$$K_{\underline{P}(\cdot\rfloor\cdot)} := \{P(\cdot\rfloor\cdot) \in \mathbb{P} \colon P(f\rfloor B) \geq \underline{P}(f\rfloor B) \text{ for all } (f,B) \in \mathscr{C}\}. \tag{14}$$

The following fundamental result by Williams [43] shows that $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if (a) there is at least one such dominating linear prevision and (b) the lower envelope of all these dominating linear previsions is equal to $\underline{P}(\cdot\rfloor\cdot)$.

**Theorem 7** (Lower envelope theorem [43])**.** *Consider a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$. Then $\underline{P}(\cdot\rfloor\cdot)$ is coherent if and only if $K_{\underline{P}(\cdot\rfloor\cdot)} \neq \emptyset$ and*

$$\underline{P}(f\rfloor B) = \min\{P(f\rfloor B) \colon P(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}\} \text{ for all } (f,B) \in \mathscr{C}.$$

Furthermore, in that case, by conjugacy, the corresponding conditional upper prevision is given by

$$\overline{P}(f\rfloor B) = \max\{P(f\rfloor B) \colon P(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}\} \text{ for all } (-f,B) \in \mathscr{C}.$$

By combining the lower envelope theorem with the results in Section 2.2.2, we obtain an alternative expression for the natural extension: $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent if and only $K_{\underline{P}(\cdot\rfloor\cdot)} \neq \emptyset$ and, in that case, we have that

$$\underline{E}(f\rfloor B) = \min\{P(f\rfloor B) \colon P(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}\} \text{ for all } (f,B) \in \mathscr{C}(\Omega). \tag{15}$$

#### 2.4.2. *Credal sets*

Although the set $K_{\underline{P}(\cdot\rfloor\cdot)}$ is extremely powerful from a theoretical point of view, it is often too complex to work with in practice. For this reason, it is sometimes convenient to restrict the domain of the conditional previsions in $K_{\underline{P}(\cdot\rfloor\cdot)}$. In particular, for any $B \in \mathscr{P}_{\emptyset}(\Omega)$, we may consider the set

$$K^B_{\underline{P}(\cdot\rfloor\cdot)} := \{P(\cdot\rfloor B) \colon P(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}\} \subseteq \mathbb{P}_B,$$

consisting of linear previsions on $\mathscr{G}(B)$. Alternatively, instead of restricting the domain of the conditional previsions in $K_{\underline{P}(\cdot\rfloor\cdot)}$, we may also regard $\underline{P}(\cdot\rfloor B)$ as an unconditional lower prevision on $\mathscr{C}_B := \{f \in \mathscr{G}(B) \colon (f,B) \in \mathscr{C}\}$—provided that $\mathscr{C}_B \neq \emptyset$—and consider the set of all linear previsions on $\mathscr{G}(B)$ that locally dominate $\underline{P}(\cdot\rfloor B)$:

$$K_{\underline{P}(\cdot\rfloor B)} := \{P \in \mathbb{P}_B \colon P(f) \geq \underline{P}(f\rfloor B) \text{ for all } f \in \mathscr{C}_B\}.$$

The following result establishes that it does not really matter which road we take. If the domain of $\underline{P}(\cdot\rfloor\cdot)$ is sufficiently large, $K^B_{\underline{P}(\cdot\rfloor\cdot)}$ and $K_{\underline{P}(\cdot\rfloor B)}$ coincide.

**Proposition 8.** *Consider an event $B \in \mathscr{P}_{\emptyset}(\Omega)$ and a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with domain $\mathscr{C}$ such that $\mathscr{C}_B = \mathscr{G}(B)$. Then $K_{\underline{P}(\cdot\rfloor B)} = K^B_{\underline{P}(\cdot\rfloor\cdot)}$.*[14]

*Proof.* We only prove that $K_{\underline{P}(\cdot\rfloor B)} \subseteq K^B_{\underline{P}(\cdot\rfloor\cdot)}$. The converse inclusion holds trivially. Consider therefore any $P \in K_{\underline{P}(\cdot\rfloor B)}$. We show that $P \in K^B_{\underline{P}(\cdot\rfloor\cdot)}$.

Let $\underline{P}^*(\cdot\rfloor\cdot)$ be the conditional lower prevision on $\mathscr{C}$ that is defined by

$$\underline{P}^*(f\rfloor A) := \begin{cases} \underline{P}(f\rfloor A) & \text{if } A \neq B \\ P(f) & \text{if } A = B \end{cases} \quad \text{for all } (f,A) \in \mathscr{C}.$$

and consider any $g \in \mathscr{E}_{\underline{P}^*(\cdot\rfloor\cdot)}$. We set out to prove that $g \not< 0$. If $g > 0$, this is trivial. If $g \not> 0$, we infer from $g \in \mathscr{E}_{\underline{P}^*(\cdot\rfloor\cdot)}$ that there are $n \in \mathbb{N}$ and, for all $i \in \{1,\ldots,n\}$, $\lambda_i \in \mathbb{R}_{>0}$, $(f_i,A_i) \in \mathscr{C}$ and $\varepsilon_i \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} g \geq \sum_{i=1}^{n} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}^*(f_i\rfloor A_i) + \varepsilon_i] &= \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}^*(f_i\rfloor A_i) + \varepsilon_i] + \sum_{i\in I'} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}^*(f_i\rfloor A_i) + \varepsilon_i] \\ &= \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \sum_{i\in I'} \lambda_i \mathbb{I}_{B}[f_i - P(f_i) + \varepsilon_i], \end{aligned}$$

with $I := \{i \in \{1\ldots n\} : A_i \neq B\}$ and $I' := \{i \in \{1\ldots n\} : A_i = B\}$. If $I' = \emptyset$—and therefore $I = \{1,\ldots,n\}$—we find that $g \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ and therefore, by Proposition 3 and the fact that $\underline{P}(\cdot\rfloor\cdot)$ is coherent, that $g \not< 0$. Hence, we may assume that $I \neq \emptyset$, allowing us to define $f := \sum_{i\in I'} \lambda_i f_i$ and $\varepsilon := \sum_{i\in I'} \lambda_i \varepsilon_i$. By the linearity of $P$, and the fact that $P \in K_{\underline{P}(\cdot\rfloor B)}$, we now have that

$$g \geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f + P(-f) + \varepsilon] \geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f + \underline{P}(-f\rfloor B) + \varepsilon],$$

where $\underline{P}(-f\rfloor B)$ is well-defined because $\mathscr{C}_B = \mathscr{G}(B)$ and therefore $(-f,B) \in \mathscr{C}$. Since $\underline{P}(\cdot\rfloor\cdot)$ is coherent, we infer from the lower envelope theorem [Theorem 7] that there is a linear conditional prevision $P'(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}$ such that $\underline{P}(-f\rfloor B) = P'(-f\rfloor B) = -P'(f\rfloor B)$ and therefore also

$$g \geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - \underline{P}(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f + \underline{P}(-f\rfloor B) + \varepsilon] \geq \sum_{i\in I} \lambda_i \mathbb{I}_{A_i}[f_i - P'(f_i\rfloor A_i) + \varepsilon_i] + \mathbb{I}_B[f - P'(f\rfloor B) + \varepsilon],$$

implying that $g \in \mathscr{E}_{P'(\cdot\rfloor\cdot)}$ and therefore, by Proposition 3 and the fact that $P'(\cdot\rfloor\cdot)$ is coherent, that $g \not< 0$. Since we have proved that this holds for any $g \in \mathscr{E}_{\underline{P}^*(\cdot\rfloor\cdot)}$, we infer from Proposition 3 that $K_{\underline{P}^*(\cdot\rfloor\cdot)}$ is non-empty.

Now let $P^*(\cdot\rfloor\cdot)$ be any element of this non-empty set $K_{\underline{P}^*(\cdot\rfloor\cdot)}$. We infer from $P \in K_{\underline{P}(\cdot\rfloor B)}$ that $P$ dominates $\underline{P}(\cdot\rfloor B)$ on $\mathscr{C}_B = \mathscr{G}(B)$ and therefore, that $\underline{P}^*(\cdot\rfloor\cdot)$ dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$. Furthermore, since $P^*(\cdot\rfloor\cdot) \in K_{\underline{P}^*(\cdot\rfloor\cdot)}$ implies that $P^*(\cdot\rfloor\cdot)$ dominates $\underline{P}^*(\cdot\rfloor\cdot)$ on $\mathscr{C}$, we find that $P^*(\cdot\rfloor\cdot)$ dominates $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, or equivalently, that $P^*(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}$. Also, by the linearity of $P$ and $P^*(\cdot\rfloor\cdot)$, we find that

$$P(f) = \underline{P}^*(f\rfloor B) \leq P^*(f\rfloor B) = -P^*(-f\rfloor B) \leq -\underline{P}^*(-f\rfloor B) = -P(-f) = P(f)$$

for all $f \in \mathscr{G}(B)$, implying that $P^*(\cdot\rfloor B) = P$ and therefore also, since $P^*(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)}$, that $P \in K^B_{\underline{P}(\cdot\rfloor\cdot)}$. □

Focussing on the 'local' sets $K_{\underline{P}(\cdot\rfloor B)}$ is not only useful from a practical point of view. They satisfy a fundamental theoretical property as well: under the weak*-topology—the topology of pointwise convergence—$K_{\underline{P}(\cdot\rfloor B)}$ is a closed and convex subset of $\mathbb{P}_B$ [38, Section 3.6.1]. Equivalently, and more intuitively, the corresponding set of probability mass functions on $B$ is convex and closed, when regarded as a subset of $\mathbb{R}^B$, and using the topology that is induced by the Euclidean metric [16, Section 10.2].[15] Any such closed and convex set of probability mass functions is called a *credal set* [18]. We will denote the credal set that corresponds to $K_{\underline{P}(\cdot\rfloor B)}$ by $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$.

[14] Many thanks to Enrique Miranda. The first author still remembers asking him if he knew whether this result was true. He said he did not know. Later that day, while strolling through town in search for a beer, he passed on a folded piece of paper. On it, he had written down the central idea of the proof we provide here.

[15] Other, arguably more intuitive, topologies can be used as well; see Refs. [37, 7] for more information.

One practical advantage of the fact that these local sets $K_{\underline{P}(\cdot\rfloor B)}$ are closed and convex is that it allows us to characterise them by means of their set of *extreme points* $\mathrm{ext}(K_{\underline{P}(\cdot\rfloor B)})$—those elements of $K_{\underline{P}(\cdot\rfloor B)}$ that cannot be written as a proper convex combination of two other elements. In particular: $K_{\underline{P}(\cdot\rfloor B)}$ is the convex hull of $\mathrm{ext}(K_{\underline{P}(\cdot\rfloor B)})$ [38, Section 3.6.2(b), note 5]. Equivalently, and perhaps more intuitively, the credal set $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$ is the convex hull of its vertices [27, Corollary 18.5.1]. The most important consequence of this result is that $\underline{P}(\cdot\rfloor B)$ is the lower envelope of $\mathrm{ext}(K_{\underline{P}(\cdot\rfloor B)})$ [38, Section 3.6.2(c)]:

$$\underline{P}(f\rfloor B) = \min\{P(f)\colon P \in \mathrm{ext}(K_{\underline{P}(\cdot\rfloor B)})\} \quad \text{for all } f \in \mathscr{G}(B). \tag{16}$$

This is especially useful if $K_{\underline{P}(\cdot\rfloor B)}$—or equivalently, $\mathscr{M}_{\underline{P}(\cdot\rfloor B)}$—is *finitely generated*, by which we mean that it has a finite number of extreme points. By Eq. (16), $\underline{P}(f\rfloor B)$ is then simply the minimum of a finite number of previsions.

*2.4.3. The sensitivity analysis interpretation*

The lower envelope theorem is not merely a theoretical result, it also suggest an alternative interpretation for coherent conditional lower previsions: a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ is simply a convenient representation for the set of linear conditional previsions $K_{\underline{P}(\cdot\rfloor\cdot)}$, or equivalently, the corresponding set of full conditonal probability measures. On this interpretation, one of these full conditional probability measures is believed to be the true, ideal model, but because of economic or time-constraints, or due to measurement errors, we are unable to specify it exactly, and can only provide a set of candidates. Walley [38, Section 1.1.5] refers to this belief in the existence of an ideal probability measure as the dogma of *ideal precision*;[16] the corresponding interpretation of conditional lower previsions is called the *sensitivity analysis interpretation*.

However, care should be taken in adopting this interpretation. As we are about to demonstrate by means of two examples, there are many situations in which it is not applicable. See Refs. [41, 38] for extensive additional discussion.

From a frequentist point of view, the existence of a fair betting rate—probability—for an event follows from an hypothesis that the experiment at hand is part of an infinite sequence of—independent or exchangeable—analogous experiments, where analogous is taken to mean that they have the same distribution. This notion of fair betting rate—probability—is problematic because it requires a predefined notion of probability; in that sense, the frequentist argument is circular. In any case, this frequentist hypothesis is only tenable if the limiting frequency of the event actually converges.[17] Nevertheless, it is sometimes applicable. For example, if each of the experiments corresponds to a flip of the same fair coin, it seems reasonable to regard $P(H) = 1/2$ as a fair betting rate for heads. Similarly, if we are told that the coin is not fair, but that its probability for heads lies in between $1/4$ and $3/4$, we are led to consider the set of all linear previsions for which $1/4 \leq P(H) \leq 3/4$, which can be conveniently represented by the—in this case unique—coherent lower prevision for which $\underline{P}(H) = 1/4$ and $\overline{P}(H) = 3/4$. However, the limiting frequency does not always converge. Consider for example a case where we do not know how the instances of heads and tails are generated—it might be a coin flip, but it might also be by some other generating process—and where we observe that the limiting frequency of heads does not converge, but rather cycles in between $1/4$ and $3/4$. In such a case, for a single experiment in this series, it seems again reasonable to bet on heads at any rate below $1/4$ and to take bets on heads at any rate above $3/4$, leading us to adopt $\underline{P}(\mathrm{H}) = 1/4$ and $\overline{P}(\mathrm{H}) = 3/4$ as our lower and upper prevision for (the indicator of) heads. However, in this case, there seems to be no reason to assume that there is such a thing as a fair (but unknown) betting rate $P(H)$ for betting on heads.[18]

A completely different, and rather extreme situation occurs when we want to model an experiment that is not repeated, and about which we know absolutely nothing. In that case, for any non-constant gamble $f \in \mathscr{G}(\Omega)$, it would be sensible to buy $f$ for any price below $\min f$, and sell it for any price higher than $\max f$, leading us to adopt the so-called *vacuous* lower prevision $\underline{P}_{\mathrm{v}}$ and the corresponding upper prevision $\overline{P}_{\mathrm{v}}$, as defined by

$$\underline{P}_{\mathrm{v}}(f) = \min f \text{ and } \overline{P}_{\mathrm{v}}(f) = \max f \text{ for all } f \in \mathscr{G}(\Omega).$$

However, here too, there seems to be no reason to assume the existence of some subject's fair (but unknown) price $P(f)$, in the sense that he should be willing to buy $f$ for any price below $P(f)$, and buy $f$ for any price above.

[16] We will henceforth refer to this idea as that of "ideal precision", to avoid the provocative use of the term 'dogma'.

[17] Under this hypothesis, convergence happens almost surely because of the law of large numbers.

[18] Such an assumption could be reasonable if there is some time-dependent fair betting rate, taking values in $[1/4, 3/4]$, of which the specific time evolution is unknown. However, we do not consider it reasonable to assume the existence of such a time-dependent fair betting rate based only on the observation that the limiting frequency cycles in between $1/4$ and $3/4$.

*2.4.4. Arbitrary sets of linear previsions*

If we choose to adopt the sensitivity analysis interpretation, and apply it to a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, we are led to model uncertainty by means of a set $K_{\underline{P}(\cdot\rfloor\cdot)}$ of linear conditional previsions. However, this is merely a special case, and as we have seen in Section 2.4.2, $K_{\underline{P}(\cdot\rfloor\cdot)}$ satisfies very specific properties. If we adopt ideal precision as a principle on its own—without the sensitivity analysis interpretation—we are not required to restrict attention to sets of linear previsions of this particular form. Any set of linear conditional previsions $K \subseteq \mathbb{P}$ can be used to model uncertainty. They do not need to be conditional either. Subsets $K$ of $\mathbb{P}_\Omega$ may also be considered.

With any set $K \subseteq \mathbb{P}$ of linear conditional previsions on $\mathscr{C}(\Omega)$, we can associate a coherent[19] conditional lower prevision $\underline{P}_K(\cdot\rfloor\cdot)$, as defined by

$$\underline{P}_K(f\rfloor B) = \inf\{P(f\rfloor B)\colon P(\cdot\rfloor\cdot) \in K\} \text{ for all } (f,B) \in \mathscr{C}(\Omega).$$

However, $\underline{P}_K(\cdot\rfloor\cdot)$ is not guaranteed to represent $K$, in the sense that $K_{\underline{P}_K(\cdot\rfloor\cdot)}$ might differ from $K$. In general, we only have that $K \subseteq K_{\underline{P}_K(\cdot\rfloor\cdot)}$. Similarly, the lower envelope $\underline{P}_K$ of a subset $K$ of $\mathbb{P}_\Omega$ may not represent $K$; the set of unconditional previsions that dominates $\underline{P}_K$ is guaranteed to include $K$, but the inclusion might be strict; equality is obtained if and only if $K$ is closed and convex.

Consider for example a situation in which you have a biased coin, but you do not know in which direction it is biased. You only know that it is three times more likely to fall on one of its sides than on the other. In that case, it seems reasonable to model this situation by means of a set $K$ consisting of two previsions $P_1$ and $P_2$, as defined by $P_1(H) = 1/4$ and $P_2(H) = 3/4$, respectively. The corresponding lower prevision is determined by $\underline{P}_K(H) = 1/4$ and $\overline{P}_K(H) = 3/4$—we obtain the same lower prevision as in Section 2.4.3. However, information has been lost, because $\underline{P}_K(\cdot)$ is dominated not only by $P_1$ and $P_2$, but also by any convex combination of these two, including for example the prevision that corresponds to a fair coin.

## 3. Conditioning by means of regular or natural extension

So far, we have come across two different imprecise-probabilistic methods for conditioning. For sets of desirable gambles, conditioning is fully determined by Eq. (2) and, for lower previsions, as explained at the end of Section 2.2.2, natural extension can be regarded as a conditioning rule. We have not stressed this yet, but both methods have a surprising property: they are always well-defined, regardless of whether or not the conditioning event has (lower or upper) probability zero. Bayes's rule on the other hand, the most famous probabilistic conditioning rule of all time, is ill-defined whenever the conditioning event $B$ has probability zero. Full conditional measures try to remedy this situation by allowing $P(\cdot\rfloor B)$ to be specified separately, but this does not resolve the issue, since the act of conditioning—deriving conditional models from unconditional ones—remains ill-defined: starting from an unconditional prevision $P$ on $\mathscr{G}(B)$, with $P(B) = 0$, Bayes's rule places no restrictions on $P(\cdot\rfloor B)$.

Should this lead us to conclude that imprecise-probabilistic approaches are more powerful when it comes to dealing with probability zero? Yes indeed! Does it mean that we should forget about Bayes's rule? Not at all! As we are about to show, Bayes's rule has a prominent place within imprecise-probabilistic conditioning as well. In many cases, it even leads to a unique conditioning rule, which will then coincide with natural extension. In the remaining cases, Bayes's rule also leads to another imprecise-probabilistic conditioning rule, called regular extension. Further on, we will argue that there are situations where this alternative conditioning rule is to be preferred over natural extension. But first, in this section, we will take a look at how this rule behaves, and to which extent it differs from natural extension.

Since regular extension is especially intuitive from a sensitivity analysis point of view, we start by introducing it in terms of sets of linear previsions. The connection with lower previsions and sets of desirable gambles will be established further on.

*3.1. In terms of sets of linear previsions*

If we adopt ideal precision,[20] and model uncertainty by means of a set of unconditional linear previsions $K \in \mathbb{P}_\Omega$, conditioning on an event $B \in \mathscr{P}_\emptyset(\Omega)$ is commonly performed by conditioning each of the elements of $K$ separately,

[19]This follows from the fact that coherence is preserved under taking pointwise infima; see Proposition 1.

[20]See Sections 2.4.3 and 2.4.4.

through Bayes's rule. For any $P \in K$ such that $P(B) > 0$, the resulting conditional prevision $P(\cdot\rfloor B)$ is defined by

$$P(f\rfloor B) := \frac{P(\mathbb{I}_B f)}{P(B)} \text{ for all } f \in \mathscr{G}(B). \tag{17}$$

If $P(B) > 0$ for all $P \in K$, this approach leads to a unique conditional model $K\rfloor B$, obtained by applying Eq. (17) to every element of $K$:

$$K\rfloor B := \{P(\cdot\rfloor B) \colon P \in K\} \subseteq \mathbb{P}_B. \tag{18}$$

If $P(B) = 0$ for all $P \in K$, Bayes's rule imposes no restrictions on the conditional prevision $P(\cdot\rfloor B)$, and Eq. (18) can no longer be applied. This leads us to consider the set of all previsions on $\mathscr{G}(B)$ as our conditional model $K\rfloor B := \mathbb{P}_B$.

The situation is less clear if $P(B) = 0$ for some $P \in K$, but $P(B) > 0$ for others. In such a case, we distinguish between two distinct approaches. By analogy with the corresponding notions for lower previsions—which will be discussed shortly—we call them natural and regular extension. Natural extension again considers the set $\mathbb{P}_B$ of all previsions on $\mathscr{G}(B)$, whereas regular extension ignores the previsions in $K$ for which $P(B) = 0$, and applies Bayes's rule to the others.

By including the aforementioned cases as well, we obtain two different conditioning rules. *Natural extension* leads us to consider the conditional models that are given by

$$K\rfloor^{\mathrm{n}} B := \begin{cases} \{P(\cdot\rfloor B) \colon P \in K\} & \text{if } P(B) > 0 \text{ for all } P \in K; \\ \mathbb{P}_B & \text{otherwise.} \end{cases} \tag{19}$$

*Regular extension* results in the use of the conditional sets that are defined by

$$K\rfloor^{\mathrm{r}} B := \begin{cases} \{P(\cdot\rfloor B) \colon P \in K \text{ and } P(B) > 0\} & \text{if } P(B) > 0 \text{ for at least one } P \in K; \\ \mathbb{P}_B & \text{otherwise.} \end{cases} \tag{20}$$

The first part of this formula—the case where $P(B) > 0$ for at least one $P \in K$—is called extended Bayesian conditioning in Ref. [36]. It has also been called generalised (Bayesian) conditioning [2]. However, care should be taken when using this terminology, because these names are used to refer to other concepts as well; see for example Refs. [17] and [1], respectively.

### *3.2. In terms of lower previsions*

For lower previsions, the only conditioning rule that we have discussed so far is natural extension: for a given unconditional coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$, and any event $B \in \mathscr{P}_\emptyset(\Omega)$, the natural extension $\underline{E}(\cdot\rfloor B)$ is the most conservative—most imprecise—conditional lower prevision on $\mathscr{G}(B)$ that is coherent with $\underline{P}$. Even stronger, as we will explain shortly, it is often the only one that is coherent with $\underline{P}$.

By Proposition 5, we know that a conditional lower prevision $\underline{P}(\cdot\rfloor B)$ on $\mathscr{G}(B)$ is coherent with $\underline{P}$ if and only if both of them are separately coherent and if they satisfy C4. In this particular context, with only these two lower previsions, C4 reduces to

$$\underline{P}(\mathbb{I}_B[f - \underline{P}(f\rfloor B)]) = 0 \quad \text{for all } f \in \mathscr{G}(B), \tag{21}$$

which is referred to as the generalised Bayes rule (GBR) [38, Section 6.4]. One of the reasons why it has this name, is because it reduces to Bayes's rule if $\underline{P}$ is a linear prevision; see P4 and F4. However, as we are about to show, the GBR has an even more fundamental connection with Bayes's rule.

If $\underline{P}(B) > 0$, the GBR is known to have a unique solution [38, Section 6.4.1]:[21] for any $f \in \mathscr{G}(B)$, there is a unique value of $\mu \in \mathbb{R}$ such that $\underline{P}(\mathbb{I}_B[f - \mu]) = 0$. Since $\underline{E}(\cdot\rfloor B)$ is coherent with $\underline{P}$, and therefore satisfies the GBR, this unique value coincides with $\underline{E}(f\rfloor B)$. In other words, if $\underline{P}(B) > 0$, $\underline{E}(\cdot\rfloor B)$ is the only coherent lower prevision on $\mathscr{G}(B)$ that satisfies the GBR, and therefore the only one that is jointly coherent with $\underline{P}$.

Let us now consider the set of linear previsions $K_{\underline{P}}$ that dominate $\underline{P}$. If we adopt the sensitivity analysis interpretation, we can condition this set by means of the methods in the previous section. If $\underline{P}(B) > 0$, then $P(B) > 0$ for all $P \in K_{\underline{P}}$,

[21] This is only true if the domain of $\underline{P}$ is large enough, which is clearly the case here because it is—assumed to be—equal to $\mathscr{G}(\Omega)$.

and we can simply apply Bayes's rule to each such $P$ to obtain a linear prevision $P(\cdot\rfloor B)$ on $\mathscr{G}(B)$, leading us to adopt the set $K_{\underline{P}}\rfloor B$ as our conditional model. Now let $\underline{P}(\cdot\rfloor B)$ be the lower envelope of this set. Then $\underline{P}(\cdot\rfloor B)$ is jointly coherent with $\underline{P}$, because for every $P\in K_{\underline{P}}$, $P$ and $P(\cdot\rfloor B)$ are jointly coherent, and therefore their lower envelopes—$\underline{P}$ and $\underline{P}(\cdot\rfloor B)$, respectively—are jointly coherent as well—see Proposition 1. Hence, by the results in the previous paragraph, $\underline{P}(\cdot\rfloor B)$ coincides with $\underline{E}(\cdot\rfloor B)$; see Ref. [38, Section 6.4.2] as well. Even stronger, as the following result establishes, $\underline{E}(\cdot\rfloor B)$ is not only the lower envelope of $K_{\underline{P}}\rfloor B$, it even represents it exactly.

**Corollary 9.** *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{E}(\cdot\rfloor\cdot)$ be its natural extension. Then for all $B\in\mathscr{P}_\emptyset(\Omega)$ such that $\underline{P}(B)>0$, we have that $K_{\underline{E}(\cdot\rfloor B)}=K_{\underline{P}}\rfloor B$.*

*Proof.* We only prove that $K_{\underline{E}(\cdot\rfloor B)}\subseteq K_{\underline{P}}\rfloor B$, because the converse inclusion follows trivially from the fact that $\underline{E}(\cdot\rfloor B)$ is the lower envelope of $K_{\underline{P}}\rfloor B$. So consider any $P^*\in K_{\underline{E}(\cdot\rfloor B)}$. Then by Proposition 8, there is some $P(\cdot\rfloor\cdot)\in K_{\underline{E}(\cdot\rfloor\cdot)}$ such that $P^*(\cdot)=P(\cdot\rfloor B)$. Since $P(B)\geq\underline{E}(B)=\underline{P}(B)>0$, we infer from the coherence of $P(\cdot\rfloor\cdot)$ [P4] that $P(\cdot\rfloor B)$ is related to $P$ by means of Eq. (17). Since $P\in K_{\underline{P}}$, this implies that $P^*=P(\cdot\rfloor B)\in K_{\underline{P}}\rfloor B$. □

Hence, in summary: if $\underline{P}(B)>0$, then by applying Bayes's rule to the linear previsions that dominate $\underline{P}$, we obtain a unique conditional model $K_{\underline{P}}\rfloor B$ that is fully characterised by its lower envelope. This lower envelope is equal to the natural extension $\underline{E}(\cdot\rfloor B)$, and is the unique solution to Eq. (21), which is the ultimate reason why this equation is called the generalised Bayes rule.

Unfortunately, these nice results no longer hold if $\underline{P}(B)=0$. In that case, the GBR might have multiple solutions and coherence alone is not guaranteed to lead to a unique value of $\underline{P}(f\rfloor B)$. The most conservative option is then to resort to the vacuous lower prevision, as defined by $\underline{P}_{\mathrm{v}}(f\rfloor B):=\min f$ for all $f\in\mathscr{G}(B)$; it is the most conservative—most imprecise—coherent lower prevision on $\mathscr{G}(B)$, and it satisfies the GBR whenever $\underline{P}(B)=0$. Hence, in that case, $\underline{P}_{\mathrm{v}}(\cdot\rfloor B)$ is the most conservative lower prevision on $\mathscr{G}(B)$ that is coherent with $\underline{P}$, and it is therefore equal to the natural extension $\underline{E}(\cdot\rfloor B)$.

By combining this with the results for $\underline{P}(B)>0$, and also using the fact that $\underline{P}(\mathbb{I}_B[f-\mu])$ is non-increasing in $\mu$ because of (C7), it can be shown that conditioning by means of natural extension results in the use of the following expressions:

$$\underline{E}(f\rfloor B)=\begin{cases}\max\{\mu\in\mathbb{R}\colon\underline{P}(\mathbb{I}_B[f-\mu])\geq 0\} & \text{if } \underline{P}(B)>0\\ \min f & \text{otherwise}\end{cases}\quad\text{for all } f\in\mathscr{G}(B). \tag{22}$$

If we adopt the sensitivity analysis interpretation, this conditioning rule can be regarded as a special case of the notion of natural extension that was introduced in the previous section: if we apply Eq. (19) to the set of dominating linear previsions $K_{\underline{P}}$, then the resulting conditional model $K_{\underline{P}}\rfloor^{\mathrm{n}}B$ has $\underline{E}(\cdot\rfloor B)$ as its lower envelope, and is even fully characterised by this lower envelope.

**Corollary 10.** *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{E}(\cdot\rfloor\cdot)$ be its natural extension. Then for all $B\in\mathscr{P}_\emptyset(\Omega)$, we have that $K_{\underline{E}(\cdot\rfloor B)}=K_{\underline{P}}\rfloor^{\mathrm{n}}B$.*

*Proof.* If $\underline{P}(B)>0$, this is exactly what is stated in Corollary 9. If $\underline{P}(B)=0$, it follows trivially from P1. □

A similar, but slightly weaker result can be obtained for regular extension as well. If we apply Eq. (20) to $K_{\underline{P}}$, then the lower envelope of the resulting conditional model $K_{\underline{P}}\rfloor^{\mathrm{r}}B$ is given by

$$\underline{R}(f\rfloor B):=\begin{cases}\max\{\mu\in\mathbb{R}\colon\underline{P}(\mathbb{I}_B[f-\mu])\geq 0\} & \text{if } \overline{P}(B)>0\\ \min f & \text{otherwise}\end{cases}\quad\text{for all } f\in\mathscr{G}(B). \tag{23}$$

and the resulting conditioning rule is called regular extension [38, Appendix J]. It coincides with natural extension whenever $\underline{P}(B)>0$ or $\overline{P}(B)=0$, but may differ from it when $0=\underline{P}(B)<\overline{P}(B)$. In the latter case, in contradistinction with what we found for natural extension, $\underline{R}(\cdot\rfloor B)$ is not guaranteed to fully characterise the conditional model $K_{\underline{P}}\rfloor^{\mathrm{r}}B$ it is the lower envelope of. The set $K_{\underline{R}(\cdot\rfloor B)}$ of linear previsions that dominate $\underline{R}(\cdot\rfloor B)$ is convex and closed—see Section 2.4.2—but $K_{\underline{P}}\rfloor^{\mathrm{r}}B$ is only guaranteed to be convex [18, 5], and may not be closed [5, Example 1]. However, only very little information is lost; by convexity, and since $\underline{R}(f\rfloor B)$ is the lower envelope of $K_{\underline{P}}\rfloor^{\mathrm{r}}B$, the latter lies in between $K_{\underline{R}(\cdot\rfloor B)}$ and its relative interior, and therefore approximates it very closely. At the expense of this minimal loss of information, $K_{\underline{P}}\rfloor^{\mathrm{r}} B$ can be conveniently represented by $\underline{R}(\cdot\rfloor B)$. If $K_{\underline{P}}$ is finitely generated, this representation is even exact [5, Section 2].

Regular extension can also be introduced without any reference to the sensitivity analysis interpretation or the set $K_{\underline{P}}$ of dominating linear previsions. Again, as with the natural extension, the resulting lower prevision $\underline{R}(\cdot\rfloor B)$ is coherent with the original model $\underline{P}$ [20, Section 3.3.4]. It only differs from the natural extension if $0 = \underline{P}(B) < \overline{P}(B)$ and is then the largest solution to the GBR, and therefore the *least* conservative—most precise—model that is coherent with $\underline{P}$ [19], whereas natural extension provides the *most* conservative—most imprecise—such model. In order to turn the regular extension into a most conservative model, coherence needs to be combined with additional axioms. Ref. [38, Appendix J3, Eq. (C16)] provides an abstract condition[22] that does the job, but does not justify why a subject should want to impose this condition as an axiom.

### 3.3. Computational aspects

From a computational point of view, calculating $\underline{E}(f\rfloor B)$ or $\underline{R}(f\rfloor B)$ requires two things: evaluating the sign of $\underline{P}(B)$ or $\overline{P}(B)$, respectively, and—in case it is positive—computing the value of $\max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\}$. We consider two distinctly different approaches.

The first approach is to use the extreme points of $K_{\underline{P}}$. By Eq. (16) and conjugacy, we know that $\underline{P}(B)$ is positive if and only if $P(B)$ is positive for all $P \in \operatorname{ext}(K_{\underline{P}})$, and that $\overline{P}(B)$ is positive whenever there is at least one $P \in \operatorname{ext}(K_{\underline{P}})$ for which $P(B) > 0$. The following result establishes that $\max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\}$ can be evaluated by applying Bayes's rule to every element of $\operatorname{ext}(K_{\underline{P}})$, whenever possible, and then taking the lower envelope of the resulting models.

**Proposition 11.** *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$. Then for any $B \in \mathscr{P}_{\emptyset}(\Omega)$ and any $f \in \mathscr{G}(B)$:*[23]

$$\max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\} = \inf\left\{\frac{P(f\mathbb{I}_B)}{P(B)} \colon P \in \operatorname{ext}(K_{\underline{P}}), P(B) > 0\right\}. \tag{24}$$

*Proof.* By Eq. (16), we know that $\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0$ if and only if $P(\mathbb{I}_B[f-\mu]) \geq 0$ for all $P \in \operatorname{ext}(K_{\underline{P}})$. Since each of these $P$ is coherent, we also know that $P(\mathbb{I}_B[f-\mu]) = P(\mathbb{I}_B f) - \mu P(B)$, and that $P(\mathbb{I}_B f) = 0$ whenever $P(B) = 0$. Hence, we find that

$$\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0 \Leftrightarrow (\forall P \in \operatorname{ext}(K_{\underline{P}}) \colon P(B) > 0)\ P(\mathbb{I}_B f) - \mu P(B) \geq 0.$$

This completes the proof, because it implies that $\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0$ if and only if $\mu$ is lower than or equal to the right-hand side of Eq. (24). $\square$

Theoretically, this approach always works. In practice, it usually only works if the number of extreme points is finite and reasonably small. It also requires that these extreme points are given, or that they can be computed efficiently from $\underline{P}$.

Alternatively, we can work directly with the lower prevision $\underline{P}$, and in particular, with the corresponding real-valued function $\rho_f$, defined by

$$\rho_f(\mu) := \underline{P}(\mathbb{I}_B[f-\mu]) \quad \text{for all } \mu \in \mathbb{R}.$$

By Eq. (16), we know that $\rho_f$ is the pointwise minimum of a set of linear, non-increasing functions $P(\mathbb{I}_B f) - \mu P(B)$, with $P \in \operatorname{ext}(K_{\underline{P}})$. It is therefore (Lipschitz[24]) continuous, concave, and non-increasing, and its first derivative, whenever it exists, lies between $-\overline{P}(B)$ and $-\underline{P}(B)$. The left and right derivatives always exist, and are guaranteed to lie between the same bounds.

Let $\mu_0 < \min f$ and $\mu_1 > \max f$. It then follows from the coherence of $\underline{P}$ that $\rho_f(\mu_0)$ is positive if and only if $\underline{P}(B)$ is positive too, and similarly for $\rho_f(\mu_1)$ and $\overline{P}(B)$. Alternatively, the signs of $\underline{P}(B)$ and $\overline{P}(B)$ can be evaluated directly as well. Evaluating $\mu^* := \max\{\mu \in \mathbb{R} \colon \underline{P}(\mathbb{I}_B[f-\mu]) \geq 0\}$ is done iteratively. If $\underline{P}(B) > 0$, then $\rho_f$ is a strictly decreasing function of $\mu$, and $\mu^*$ is its only root. By coherence, this root is guaranteed to lie between $\min f$ and $\max f$, and it can

[22] If $\overline{P}(B) > 0$ and $\underline{P}(\mathbb{I}_B f) \geq 0$, then $\underline{P}(f\rfloor B) \geq 0$.

[23] This result is essentially well-known; we provide its proof for the sake of completeness.

[24] See Ref. [9, Section 6.1].

therefore be found easily by means of the bisection method, or any other root-finding procedure; see Refs. [9, Section 6.3] and [39, p. 18] for methods that have been specifically designed to exploit the properties of the function $\rho_f$. If we are able to evaluate $\rho_f(\mu)$ up to some numerical error $\varepsilon$, and iterate sufficiently many times, then by the bounds on the first (left and right) derivative of $\rho_f$, the error that is made by the bisection method will not exceed $\varepsilon/\underline{P}(B)$. If $\overline{P}(B) > \underline{P}(B) = 0$, then $\rho_f$ is identically zero in $]-\infty, \mu^*]$ and strictly decreasing in $]\mu^*, +\infty[$, and in this decreasing part, the first derivative, whenever it exists, is bounded above by $-\underline{P}^{>0}(B)$, with $\underline{P}^{>0}(B) := \inf\{P(B)\colon P \in \operatorname{ext}(K_{\underline{P}}), P(B) > 0\}$, and similarly for the left and right derivative, which always exist. Finding $\mu^*$ is now a bit more tricky. Since coherence again implies that $\mu^*$ lies in between $\min f$ and $\max f$, we could in principle directly apply the bisection algorithm here as well. However, if during this procedure, numerical errors lead us to mistakenly conclude that $\rho_f(\mu)$ is negative for some $\mu < \mu^*$, the obtained solution could greatly underestimate the actual value $\mu^*$. The simplest way to fix this is to look for the unique root of $\rho_f + \delta$, for some sufficiently small $\delta > \varepsilon$. If $\underline{P}^{>0}(B) > 0$—for example, if $K_{\underline{P}}$ is finitely generated—then the obtained solution will overestimate the actual value of $\mu^*$, but by no more than $(\delta+\varepsilon)/\underline{P}^{>0}(B)$.

### 3.4. *What about sets of desirable gambles?*

After all this elaboration about conditioning with lower previsions, and in particular with natural and regular extension, one could wonder why we even bother to deal with these notions. Why do we not simply work with sets of desirable gambles? The answer is a practical one: from a computational point of view, conditioning by means of natural or regular extension—Eqs. (22) and (23), respectively—is more tractable. Although Eq. (2) provides a conceptually very simple conditioning rule, it is difficult to use in practice. If $\mathscr{D}$ has a complex border structure, it can be very difficult—if not impossible—to check whether a gamble $f \in \mathscr{G}(B)$ belongs to $\mathscr{D}\rfloor B$ or, equivalently, whether $\mathbb{I}_B f \in \mathscr{D}$, especially if $\{\mathbb{I}_B f\colon f \in \mathscr{G}(B)\}$ is a subset of the border of $\mathscr{D}$. The two main reasons are numerical errors and the fact that it is difficult—and sometimes even impossible due to memory limitations—to provide a computer representation for the exact border structure of $\mathscr{D}$.[25]

Nevertheless, even if the actual calculations are performed in terms of lower previsions, sets of desirable gambles remain important, both theoretically and philosophically. One of the key reasons for their importance, is that they provide conditional lower previsions with an interpretation, without any reference to Bayes's rule or the sensitivity analysis interpretation: conceptually, for every coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$, there is a—possibly not given—coherent set of desirable gambles $\mathscr{D}$ such that $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot)$ coincides with $\underline{P}(\cdot\rfloor\cdot)$ on $\mathscr{C}$.

If $\mathscr{D}$ is given, then in theory, conditioning should be done by applying Eq. (2) to $\mathscr{D}$, resulting in a conditional set of desirable gambles $\mathscr{D}\rfloor B$. In practice however, we are usually only interested in the corresponding lower prevision $\underline{P}_{\mathscr{D}\rfloor B}(f) = \underline{P}_{\mathscr{D}}(f\rfloor B)$, for some gamble $f \in \mathscr{G}(B)$. Hence, rather than constructing $\mathscr{D}\rfloor B$, which is often intractable, we will instead try and calculate $\underline{P}_{\mathscr{D}}(f\rfloor B)$ directly. However, even this may be very hard, because—as is the case for $\mathscr{D}\rfloor B$—$\underline{P}_{\mathscr{D}}(f\rfloor B)$ may crucially depend on the exact border structure of $\mathscr{D}$. In order to avoid this dependency on the border structure, the trick is to focus on the unconditional prevision $\underline{P}_{\mathscr{D}}$. Unlike $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot)$, $\underline{P}_{\mathscr{D}}(\cdot) := \underline{P}_{\mathscr{D}}(\cdot\rfloor\Omega)$ does not depend on the exact border structure of $\mathscr{D}$, and can therefore be evaluated in a more reliable manner. If we now use the techniques in Section 3.3 to obtain the natural extension $\underline{E}(f\rfloor B)$ of $\underline{P}_{\mathscr{D}}$, then by the results of Section 3.2, $\underline{E}(f\rfloor B)$ is guaranteed to provide a lower—conservative—bound on $\underline{P}_{\mathscr{D}}(f\rfloor B)$, and when $\underline{P}_{\mathscr{D}}(B) > 0$, this bound will even be tight. For some sets $\mathscr{D}$, it is even possible to prove, on theoretical grounds, that $\underline{P}_{\mathscr{D}}(f\rfloor B)$ is bounded from below by the regular extension $\underline{R}(f\rfloor B)$ of $\underline{P}_{\mathscr{D}}$,[26] thereby providing a lower bound on $\underline{P}_{\mathscr{D}}(f\rfloor B)$ that is guaranteed to be exact whenever $\overline{P}_{\mathscr{D}}(B) > 0$.[27] We will construct such sets in Section 5 and Appendix AppendixA and prove that they indeed satisfy this property. In any case, for now, the main message is that even if the underlying model is a set of desirable gambles $\mathscr{D}$, we can still perform all the calculations in terms of lower previsions, using natural or regular extension. This approach is usually more feasible from a computational point of view, and is guaranteed to provide lower—conservative—bounds on $\underline{P}_{\mathscr{D}}(f\rfloor B)$, that are often even tight.

[25] See Refs. [4, 25] for some ingenious but complex methods that are able to deal with the special case of so-called finitely generated sets of desirable gambles.

[26] This holds trivialy if $\underline{P}_{\mathscr{D}}(B) > 0$; if $\overline{P}_{\mathscr{D}}(B) > 0$, this can either be proved directly—as we do in Corrolary 21 and Proposition 28—or, alternatively, by verifying the necessary and sufficient conditions in Ref. [21, Appendix A.1] or the sufficient conditions in Ref. [4, Theorem 10].

[27] Because $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ is coherent with $\underline{P}_{\mathscr{D}}$, and because, as we have seen in Section 3.2, $\underline{R}(\cdot\rfloor B)$ is the least conservative lower prevision for which the case.

In many cases, $\mathscr{D}$ is not given, and is simply an underlying theoretical concept. For example, in the case of conditioning, all we may have to start from is the unconditional lower prevision $\underline{P}(\cdot) := \underline{P}_{\mathscr{D}}(\cdot\rfloor\Omega)$ on $\mathscr{G}(\Omega)$, with $\mathscr{C} = \{(f,\Omega)\colon f \in \mathscr{G}(\Omega)\}$. In that case, by specifying the set $\mathscr{D}$ ourselves, we are in fact specifying a conditioning rule: for every $B \in \mathscr{P}_\emptyset(\Omega)$, the set $\mathscr{D}$ will provide us with a corresponding coherent lower prevision $\underline{P}_{\mathscr{D}}(\cdot\rfloor B)$ on $\mathscr{G}(B)$. Natural and regular extension correspond to particular choices of $\mathscr{D}$. By definition—see Section 2.2.2—natural extension can be obtained by using the set $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, which in this case is equal to $\mathscr{D}_{\underline{P}}$.

**Proposition 12.** *Consider a coherent 'conditional' lower prevision $\underline{P}(\cdot\rfloor\cdot)$ whose domain $\mathscr{C}$ is equal to $\{(f,\Omega)\colon f \in \mathscr{G}(\Omega)\}$, and let $\underline{P}(\cdot) := \underline{P}(\cdot\rfloor\Omega)$ be the corresponding unconditional lower prevision on $\mathscr{G}(\Omega)$. Then $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)} = \mathscr{D}_{\underline{P}}$.*

*Proof.* In this particular case, we find that

$$\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)} = \{f - \underline{P}(f) + \varepsilon\colon f \in \mathscr{G}(\Omega), \varepsilon \in \mathbb{R}_{>0}\} = \{g \in \mathscr{G}(\Omega)\colon \underline{P}(g) > 0\}, \tag{25}$$

using C8 to obtain the second equality. Hence, $\mathscr{D}_{\underline{P}}$ is clearly a subset of $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)} := \mathscr{E}(\mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)})$. In order to establish the converse inclusion, let us consider any $f \in \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ and prove that $f \in \mathscr{D}_{\underline{P}}$. If $f > 0$, this is trivial. If $f \not> 0$, then by Eq. (1), $f \geq \sum_{i=1}^n \lambda_i f_i$, with $n \in \mathbb{N}$ and, for all $i \in \{1,\dots,n\}$, $f_i \in \mathscr{A}_{\underline{P}(\cdot\rfloor\cdot)}$ and $\lambda_i \in \mathbb{R}_{>0}$. Hence, by coherence [C2, C3 and C7] and Eq. (25), we find that $\underline{P}(f) \geq \underline{P}(\sum_{i=1}^n \lambda_i f_i) \geq \sum_{i=1}^n \lambda_i \underline{P}(f_i) > 0$ and therefore, that $f \in \mathscr{D}_{\underline{P}}$. □

Regular extension can be obtained by using the set

$$\mathscr{D}^{\mathrm{r}}_{\underline{P}} := \{f \in \mathscr{G}(\Omega)\colon (\underline{P}(f) \geq 0 \text{ and } \overline{P}(f) > 0) \text{ or } f > 0\} = \mathscr{D}_{\underline{P}} \cup \{f \in \mathscr{G}(\Omega)\colon \underline{P}(f) = 0 \text{ and } \overline{P}(f) > 0\}.$$

**Proposition 13.**[28] *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{R}(\cdot\rfloor\cdot)$ be the conditional lower prevision on $\mathscr{C}(\Omega)$ that is defined by Eq. (23). Then $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ and $\underline{R}(\cdot\rfloor\cdot)$ are both coherent, $\underline{P} = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}$ and*

$$\underline{R}(f\rfloor B) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) \quad \textit{for all } (f,B) \in \mathscr{C}(\Omega).$$

*Proof.* Coherence of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ follows by straightforward verification of D1–D4. D2 holds trivially. D1 and D3 follow directly from the coherence of $\underline{P}$ [C8 and C2, respectively]. In order to prove D4, we consider any $f,g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$, and show that $f+g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. If $f,g \in \mathscr{D}_{\underline{P}}$ this follows from the coherence of $\mathscr{D}_{\underline{P}}$. Otherwise, we may assume without loss of generality that $\underline{P}(f) = 0$ and $\overline{P}(f) > 0$. Since $g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$ clearly implies that $\underline{P}(g) \geq 0$, we infer that $\underline{P}(f+g) \geq \underline{P}(f) + \underline{P}(g) \geq 0$ [C3] and $\overline{P}(f+g) \geq \overline{P}(f) + \underline{P}(g) > 0$ [C5], which in turn implies that $f+g \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$.

Next, we prove that $\underline{P} = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}$. Consider any $f \in \mathscr{G}(\Omega)$ and any $\alpha \in \mathbb{R}$. If $\alpha < \underline{P}(f)$, then $\underline{P}(f-\alpha) > 0$ [C8] and therefore $f - \alpha \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. If $\alpha > \underline{P}(f)$, then $\underline{P}(f-\alpha) < 0$ [C8] and therefore also $f - \alpha \not> 0$ [C1], which implies that $f - \alpha \notin \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. Hence, by Eq. (3), $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f) = \underline{P}(f)$.

Since $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ is coherent, $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor\cdot)$ is coherent by construction, and therefore, the only thing that we still need to prove is that $\underline{R}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor\cdot)$. So consider any $(f,B) \in \mathscr{C}(\Omega)$. We will show that $\underline{R}(f\rfloor B) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$. Since $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor\cdot)$ is coherent, $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor B)$ is clearly coherent with $\underline{P} = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}$. First, assume that $\underline{P}(B) > 0$. Then $\underline{P}$ uniquely determines $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$ through the GBR [Eq. (21)], and therefore $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$ coincides with $\underline{R}(f\rfloor B)$. Next, assume that $\overline{P}(B) > \underline{P}(B) = 0$. Then coherence of $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(\cdot\rfloor B)$ and $\underline{P}$ implies that $\underline{R}(f\rfloor B) \geq \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$. Consider now any $\mu < \underline{R}(f\rfloor B)$. Since $\underline{P}(\mathbb{I}_B[f - \underline{R}(f\rfloor B)]) = 0$, we know that $\underline{P}(\mathbb{I}_B[f-\mu]) \geq 0$ [C7] and

$$\overline{P}(\mathbb{I}_B[f-\mu]) \geq \underline{P}(\mathbb{I}_B[f - \underline{R}(f\rfloor B)]) + \overline{P}(\mathbb{I}_B\lambda) \geq \lambda\overline{P}(B) > 0$$

[C5 and C2], with $\lambda := \underline{R}(f\rfloor B) - \mu > 0$, which implies that $\mathbb{I}_B[f-\mu] \in \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. Since this holds for all $\mu < \underline{R}(f\rfloor B)$, we infer from Eq. (8) that $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) \geq \underline{R}(f\rfloor B)$ and therefore that $\underline{R}(f\rfloor B) = \underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B)$. Finally, assume that $\overline{P}(B) = 0$. Consider any $\mu > \min f$ and choose $\alpha \in \mathbb{R}$ such that $\alpha > \max f$ and $\alpha > 0$. Then $\overline{P}(\mathbb{I}_B[f-\mu]) \leq \overline{P}(\mathbb{I}_B\alpha) = \alpha\overline{P}(B) = 0$ [C7 and C2] and $\mathbb{I}_B[f-\mu] \not> 0$, and therefore $\mathbb{I}_B[f-\mu] \notin \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. Since this holds for all $\mu > \min f$, we infer from Eq. (8) that $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) \leq \min f$ and therefore, by coherence [C1], that $\underline{P}_{\mathscr{D}^{\mathrm{r}}_{\underline{P}}}(f\rfloor B) = \min f = \underline{R}(f\rfloor B)$. □

[28] This result was stated without proof in Ref. [26, Section 2.6.6]; Refs. [38, Appendix F4] and [4] provide earlier, but less direct statements of the same result.

Since sets of desirable gambles are more expressive than conditional lower previsions, the sets $\mathscr{D}_{\underline{P}}$ and $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ are not necessarily the only ones that result in the use of natural or regular extension as a conditioning rule, respectively. For natural extension, the set $\mathscr{D}_{\underline{P}}$ is however fundamental, because it is the smallest—most conservative, most imprecise—set of gambles whose desirability is implied by $\underline{P}$ and coherence. For regular extension, $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ does not have such a special status. We will construct another example of a set of desirable gambles that results in regular extension in Appendix AppendixA, and will show that it is a—sometimes strict—subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$; see Corollary 32 and Example 2.

## 4. There is more to updating than just conditioning

Conditioning is commonly—and successfully—used for a multitude of practical purposes, the most important of which is to solve the following updating problem: starting from an initial belief model for an uncertain variable $X$ that takes values in $\Omega$, and given the additional information that some event $O \in \mathscr{P}_{\emptyset}(\Omega)$ has occurred, how then should we update this belief model to reflect this new information? If the belief model is taken to be a probability mass function, then the traditional solution to this problem is to condition the original model by means of Bayes's rule. In fact, in that context, the act of "conditioning on an event $O$" is often even identified with solving this updating problem. Similarly, if the belief model is an imprecise probability model, the conditioning rules that we have discussed before are commonly used as updating rules.

However, it should not be forgotten that conditioning is just a mathematical concept. Before using this concept to solve a practical problem such as the updating task, one should at least try and justify why it is indeed reasonable to use it for this purpose. The goal of this section is to perform this exercise for the two imprecise-probabilistic conditioning rules that we discussed earlier on: natural and regular extension. We intend to argue that, in many cases, it does indeed make sense to use these rules to update a belief model in the light of the occurrence of an event. We shall start by arguing for the use of Eq. (2) as a conservative updating rule, which will then lead to a rather straightforward justification for updating by means of natural extension. It will follow from our analysis that there is room for less conservative updating strategies as well, such as regular extension. The actual justification for updating by means of regular extension is left to Section 5.

### 4.1. Narrowing down the problem

Imprecise-probabilistic belief updating, in a general sense, is highly complicated. Many different, and often ill-defined aspects might come into play: additional background information other than the occurrence of $O$, complex dynamical aspects, changes of mind, biased information, etcetera; see for example Refs. [46, 33, 30]. For our present purposes, we will restrict ourselves to a specific, less general setting, which is nevertheless still applicable in a wide variety of situations.

#### 4.1.1. Updating by means of a rule

Our most important restriction is that we only consider situations where the updated models are provided beforehand. In other words: we are looking for an *updating rule*, which—by definition of a rule—is stated in advance. For any event $O \in \mathscr{P}_{\emptyset}(\Omega)$ that is considered, this rule provides an associated updated belief model, that is obtained simply by applying the rule to the original belief model. If $O$ occurs, this updated model is then simply adopted. This situation for example reflects the typical practical two-phased approach to modelling, where the model is first built by experts, based on expert knowledge and/or data, only to be queried afterwards by the user. The user is then usually not an expert in the field, and is therefore provided not only with a model, but also with an updating rule to go with it.

An important consequence of restricting ourselves to this particular case, is that it rules out the possibility of taking into account any additional information other than the occurrence of some event $O$, because such additional information cannot be anticipated when designing—or choosing—the rule. In other words, all that is learned is that $O$ has occurred, and nothing else. However, our setting does not rule out the possibility of changing your mind, it only requires you to do so in a way that is specified in advance, and that is based on no information other than that $O$ has occurred.

*4.1.2. What does it mean to learn that an event has occurred?*

Since our updating rule is based only on the initial belief model and the information that $O$ has occurred, we should clearly define what 'to be informed that $O$ has occurred' means. The '$O$ has occurred' part is easy: it simply means that $X$ has taken a value in $O$. The 'to be informed that' part is a bit more tricky. In our set-up, we will only deal with cases where the information that $O$ has occurred is reported to us honestly, correctly and whenever it applies, and where we do not receive any additional information about which particular value $X$ takes in $O$. We discuss each of these separate requirements below, but basically, they reduce to the following single requirement: *it should be agreed upon beforehand that we will be informed about the occurrence of O—and nothing more—if and only if it actually occurs*.

The requirement that we should not receive any additional information about which particular value $X$ takes in $O$ may seem straightforward, but it is in fact not that trivial. For example, among other things, it implies that the point in time when we are informed that $O$ has occurred should not provide us with any additional information about the value of $X$, other than that it belongs to $O$. For example, if $X$ is the outcome of two consecutive coin flips, and $O$ is the event that the outcome of at least one of them is heads, it could be that after the first flip, this event is already known to be true. However, we should not be told so until after the second flip, because otherwise, we would not only learn that the outcome of at least one of the coin flips was heads—that $O$ has occurred—but we would also receive the additional information that the outcome of the first flip was heads.

The requirement of honesty is typically relevant if the information is the answer to some sensitive question, such as whether or not someone is a smoker. The answer may be biased; patients who smoke will often say that they do not. These kind of biases occur frequently, and should be taken into account. However, we do not consider them to be part of the updating problem, but rather of the modelling problem. Whenever such a bias is suspected, the model should be extended in such a way as to differentiate between the answer that is given and the actual truth, and this extended model should try and capture the relationship between both. The event $O$ is then simply the—possibly dishonest—answer that is given by the patient, which is *not* required to coincide with the truthful answer to the question. In this way the event $O$ is trivially guaranteed to be reported honestly, and hence our set-up applies to the extended model.

The requirement of correctness is similar to that of honesty, but refers to unintended errors rather than intentional ones, for example due to measurement errors, miscommunication, etcetera. Again, we do not consider this to be part of the updating problem, but rather of the modelling problem. The event $O$ should be taken to refer to a—possibly wrong—actual observation; the observation process itself, as well as the errors it produces, are taken to be part of the model. This guarantees that the observation that $O$ has occurred is trivially correct, thereby making our set-up apply.

Finally, the information that the event $O$ has occurred should be reported whenever it applies. More exactly: it should be agreed upon beforehand that if $O$ occurs, we are guaranteed to receive the information that it indeed does. At first sight, this might seem to follow from the requirements of honesty and correctness, but it does not: honesty and correctness only guarantee that whenever we are informed that $O$ has occurred, it has indeed occurred, but not the other way around. Consider for example a situation where we come to know that the outcome of the roll of a die is even. Then, under the assumptions of correctness and honesty alone, this only allows us to infer that the event $O$ on which we should update is a subset of $O^* \coloneqq \{2,4,6\}$. For example, it might be that it was decided beforehand—without our knowing—that the information that the outcome is even would only be provided to us if the actual outcome is 2 or 4, and kept from us whenever the outcome is 6. In that case, the information that the outcome is even is correct, as well as honest, but it is too weak, since it also implies that the outcome is not 6. Hence, the event on which we should be updating is $O \coloneqq \{2,4\}$ rather than $O^*$.

*4.1.3. Which events should be considered?*

So far, we have restricted attention to a generic single event $O \in \mathscr{P}_\emptyset(\Omega)$. However, that is not the end of the story. Can it be any event? Can we consider multiple events? No and yes, respectively.

Since updating is by its very definition concerned with providing a new belief model after getting to know that $O$ has occurred, it clearly only applies to observable events. We do not necessarily have to be able to observe it ourselves, but someone has to be, and needs to be able to communicate it to us. In contrast, conditioning can be applied to any event, regardless of whether or not it is observable. It is for this reason that we use $O$—from 'observation'—to refer to a generic event on which we update, rather than $B$, which we use for events on which we condition.

Many authors do not single out just one event $O$, but consider it to be part of a structured collection of events that reflects the actual process of gathering information. For example, Walley [38] focuses on partitions of $\Omega$, the elements of which typically correspond to the outcome of some experiment, or the answer to some question. Shafer [29] considers more complex structures, which he calls protocols; they allow for multi-phased set-ups as well, where additional experiments and/or questions can be used to further refine the sample space. These situations are covered by our setting as well; it suffices for the requirements of Section 4.1.2 to hold for every event in the considered collection. However, our approach is more general. For example, if $O$ is part of some partition, then we do not require that there is some predetermined point in time where we are guaranteed to know which event in this partition has occurred. In any case, it is not necessary to consider a collection of events. Furthermore, if $O$ does belong to some collection, then most of our analysis will not depend on it. Therefore, we will usually refrain from mentioning any collection, and will simply consider a single event $O \in \mathscr{P}_{\emptyset}(\Omega)$. Whenever the specific collection to which $O$ belongs does become relevant, it will be mentioned explicitly; see for example Section 4.3.2.

### *4.2. Conditioning as an updating strategy*

Now that we know exactly what we are dealing with, let us start solving the problem. First of all, by our assumptions, we are guaranteed that whenever we are informed that $O$ has occurred, it has indeed occurred, meaning that $X$ takes a value in $O$. Hence, after learning that $O$ has occurred, we no longer need to consider the elements in $\Omega \setminus O$. This implies that the updated belief model should capture the uncertainty about which value $X$ takes in the remaining set of possibilities $O$. Depending on the framework we adopt, this updated belief model can be a set of desirable gambles on $O$, a lower prevision on $\mathscr{G}(O)$, or any of the other uncertainty models that we discussed in Section 2.

#### *4.2.1. An asymmetric version of Walley's updating principle*

We first consider the framework of sets of desirable gambles. The initial belief model is then a coherent set of desirable gambles on $\Omega$, which we denote by $\mathscr{D}$, and the updated belief model is a coherent set of desirable gambles on $O$, which we denote by $\mathscr{D}_O$. By definition of an updated model, $\mathscr{D}_O$ consists of gambles that are desirable *after O* occurs, whereas $\mathscr{D}$ consists of gambles that are desirable *now—before O* occurs. The central question of the updating problem is whether or not, and if yes, how, $\mathscr{D}_O$ should be related to $\mathscr{D}$.

In order to be able to answer this question, we distinguish $\mathscr{D}_O$ from two other sets of gambles on $O$. The set $\mathscr{D}\rfloor O$ has already been discussed at length; it is fully determined by $\mathscr{D}$, as it consists of those gambles $f \in \mathscr{G}(O)$ for which the gamble $\mathbb{I}_O f$ is desirable *now*—is an element of $\mathscr{D}$. The set $\mathscr{D}^O$ is new. It consists of those gambles on $O$ of which our subject *now* thinks that they *should* become desirable *after O* occurs; following Walley's terminology [38, Section 6.1.5], we call these gambles $O$-desirable. So how is $\mathscr{D}_O$ related to $\mathscr{D}^O$ and $\mathscr{D}\rfloor O$?

The connection between $\mathscr{D}_O$ and $\mathscr{D}^O$ follows trivially from our assumptions. As explained in Section 4.1.1, we are restricting attention to situations in which the updated model is provided in advance by an expert and then simply adopted by the user once $O$ actually occurs. Within this setting, $\mathscr{D}_O$ and $\mathscr{D}^O$ coincide because the user—by assumption—will consider a gamble $f \in \mathscr{G}(O)$ to be desirable after $O$ occurs if and only if it was stated beforehand by the expert that $f$ should be desirable after $O$ occurs, or equivalently, if the expert considered it to be $O$-desirable.

The remaining task is to establish a relationship between $\mathscr{D}^O$ and $\mathscr{D}\rfloor O$. The best-known solution is that of Walley [38, Section 6.1.6], who says that $\mathscr{D}^O$ should be equal to $\mathscr{D}\rfloor O$. We do not agree. We will argue that this requirement is too strong, and that the only thing that can be reasonably imposed is for $\mathscr{D}^O$ to be a superset of $\mathscr{D}\rfloor O$. We start by repeating Walley's argument. The central idea is very elegant: under the assumptions of Section 4.1.2,[29] the effect of owning a gamble $f \in \mathscr{G}(O)$ after being informed that $O$ has occurred is indistinguishable from the effect of owning the gamble $\mathbb{I}_O f \in \mathscr{G}(\Omega)$ now; as explained in Section 2.1.2, they result in the same payoff if $O$ occurs and have no effect otherwise. So far, we agree. However, Walley does not stop here. He uses this fact to infer—without any actual argumentation—that $f$ should be $O$-desirable if and only if $\mathbb{I}_O f$ is desirable: $f \in \mathscr{D}^O \Leftrightarrow \mathbb{I}_O f \in \mathscr{D} \Leftrightarrow f \in \mathscr{D}\rfloor O$; he calls this the updating principle.

We think that in order for the updating principle to be compelling, considering a gamble to be desirable should mean being willing to accept it, as Walley seems to assume. If this is not the case, then the fact that two transactions have the same effect does not imply that their desirability should be equivalent. In our framework, where desirability of $f$ means *strict* preference of $f$ over the status quo—a notion that is stronger than just being willing to accept $f$—such an equivalence would require the status quo to remain identical as well, and this is not the case, because the frame of

[29] Actually, Walley seems to impose even stronger assumptions, as he requires $O$ to be part of a partition of $\Omega$. We do not consider this to be necessary.

reference changes: after $O$ has occurred, the status quo is no longer the zero gamble on $\Omega$, but rather the zero gamble on $O$.

It seems to us that there are two situations in which this change of status quo may influence the desirability of a gamble: if—*before* the occurrence of $O$—the subject believes that $O$ cannot occur or if he is undecided about whether or not $O$ can occur. In those two cases, even if the subject does not prefer $\mathbb{I}_O f$ strictly over the status quo *before* the occurrence of $O$, it may still make perfect sense for him to think that $f$ should be strictly preferred over the status quo *after* the occurrence of $O$, because it is then clear that $O$ can occur. It is only if the subject believes that $O$ can occur——again, and obviously, *before* the occurrence of $O$—that it seems compelling that $O$-desirability of $f$ should imply the desirability of $\mathbb{I}_O f$. In that case, it can be argued that $\mathscr{D}^O$ should be a subset of $\mathscr{D}\rfloor O$. However, in general, $O$-desirability of a gamble $f$ does not need to imply the desirability of $\mathbb{I}_O f$.

The converse relation does hold in general: $\mathscr{D}\rfloor O$ should be a subset of $\mathscr{D}^O$. Indeed, if $\mathbb{I}_O f$ is desirable now, then it must be that the restriction of $\mathbb{I}_O f$ to $O$—the gamble $f$—is strictly preferred over the restriction of the status quo to $O$—the zero gamble on $O$—because outside of $O$, $\mathbb{I}_O f$ is identical to the status quo and therefore clearly not strictly preferred to it. Since, under the conditions imposed in Section 4.1.2, these restrictions are exactly the payoffs that become relevant after $O$ has occurred, we conclude that $f$ is $O$-desirable.

By combining the arguments above, we find that under the conditions described in Section 4.1:

$$\mathscr{D}_O = \mathscr{D}^O \supseteq \mathscr{D}\rfloor O. \tag{26}$$

Since we also want $\mathscr{D}_O$ to be coherent, we find that the updated set of desirable gambles $\mathscr{D}_O$ should be a coherent superset of the conditional set of desirable gambles $\mathscr{D}\rfloor O$. Furthermore, as we have seen, if the subject believes that $O$ can occur, $\mathscr{D}^O$ and therefore also $\mathscr{D}_O$ should be equal to $\mathscr{D}\rfloor O$. Of course, this requires a clear definition of what it means for a subject to believe that $O$ can occur. We discuss this further in Section 4.3.2; until then, as we explain in the next section, $\mathscr{D}_O$ will automatically be equal to $\mathscr{D}\rfloor O$.

*4.2.2. Conditioning as a conservative updating strategy*

Without any further assumptions, the only reasonable updating strategy that follows from Eq. (26) is to use the conditional model $\mathscr{D}\rfloor O$ as our updated model $\mathscr{D}_O$, simply because it is the most conservative—most imprecise—choice of $\mathscr{D}_O$ that is compatible with Eq. (26), and therefore the only one that is truly implied by it. Furthermore, provided that $\mathscr{D}$ is coherent, $\mathscr{D}\rfloor O$ will be coherent as well. Hence, after all this effort to distinguish between conditioning and updating, it turns out—rather amusingly—that indeed, as is commonly done, conditioning can be regarded as an updating strategy—when the conditions of Section 4.1 are satisfied.

*4.2.3. Justifying natural extension as an updating rule*

The result above easily translates to the framework of lower previsions. If the original belief model is a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, it suffices to apply the result of Section 4.2.2 to the smallest—most conservative—associated set of desirable gambles $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ to find that the most conservative updated lower prevision on $\mathscr{G}(O)$ that is compatible with—the only one that is implied by—Eq. (26) is given by the natural extension $\underline{E}(\cdot\rfloor O)$ of $\underline{P}(\cdot\rfloor\cdot)$.

If $\underline{P}(\cdot\rfloor\cdot)$ is effectively an unconditional lower prevision $\underline{P}(\cdot) = \underline{P}(\cdot\rfloor\Omega)$ on $\mathscr{G}(\Omega)$, this natural extension can be calculated using the computational techniques of Section 3.3. Otherwise, $\mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is a (possibly strict) superset of $\mathscr{D}_{\underline{E}(\cdot\rfloor\Omega)}$ and therefore, the natural extension $\underline{E}(\cdot\rfloor O)$ of $\underline{P}(\cdot\rfloor\cdot)$ (possibly strictly) dominates the natural extension $\underline{E}^*(\cdot\rfloor O)$ of $\underline{E}(\cdot\rfloor\Omega)$. If the domain of $\underline{P}(\cdot\rfloor\cdot)$ is large enough, such that $\underline{P}(\cdot\rfloor\Omega)$ is defined on $\mathscr{G}(\Omega)$, $\underline{E}(\cdot\rfloor\Omega)$ can be replaced by $\underline{P}(\cdot\rfloor\Omega)$ because they are then equal. If $\underline{P}(\cdot\rfloor\Omega)$ is known—or if $\underline{E}(\cdot\rfloor\Omega)$ can be calculated—$\underline{E}^*(\cdot\rfloor O)$ can be obtained by applying the computational techniques of Section 3.3. The result is guaranteed to coincide with $\underline{E}(\cdot\rfloor O)$ whenever $\underline{P}(O\rfloor\Omega) > 0$—$\underline{E}(O\rfloor\Omega) > 0$—but will be vacuous and might therefore only be a (safe) lower approximation of $\underline{E}(\cdot\rfloor O)$ if $\underline{P}(O\rfloor\Omega) = 0$—$\underline{E}(O\rfloor\Omega) = 0$.

*4.2.4. Justifying Bayes's rule as an updating rule*

For the special case of linear previsions—and hence also probability mass functions—natural extension coincides with Bayes's rule whenever $O$ has positive probability, and therefore, in that case, we obtain a justification for using Bayes's rule as an updating strategy. Since it is impossible for a coherent lower prevision to dominate a linear one without coinciding with it, we can even conclude that, whenever $P(O) > 0$, Bayes's rule is the *unique* updating strategy that is compatible with Eq. (26). However, this special case is not our main point of interest here, as there are many other justifications available for updating by means of Bayes's rule; see for example Refs. [29, 28, 30].

### 4.3. What about other updating strategies?

As we have just shown, the use of conditioning as an updating rule is justified by Eq. (26), and this expression applies whenever the conditions of Section 4.1 are met. However, this is not the end of the story. Other updating strategies can be justified as well, both more and less conservative.

#### 4.3.1. More conservative strategies

More conservative updating rules—smaller updated sets of desirable gambles—do not require any actual justification. Since we do not adopt an exhaustive interpretation—see Section 2.1.1—they are fully consistent with the commitments that are implied by the use of conditioning as an updating rule. However, it is rather silly to use these more conservative updating strategies because, as we have just shown, they are unnecessarily weak. The reason why it may nevertheless be reasonable to apply them is a practical one: conditional models may be intractable to compute, and in those cases, tractable more conservative updating strategies can serve as a useful safe approximation. For example, as explained in Section 3.4, the natural extension $\underline{E}(\cdot\rfloor O)$ of the unconditional prevision $\underline{P}_{\mathscr{D}}$ serves as a tractable outer approximation of $\underline{P}_{\mathscr{D}}(\cdot\rfloor O)$, and $\underline{P}_{\mathscr{D}}(\cdot\rfloor O)$ itself is already more tractable to compute, as well as more conservative, than the actual conditional set of desirable gambles $\mathscr{D}\rfloor O$. In much the same way, as explained in Section 4.2.3, the natural extension $\underline{E}^*(\cdot\rfloor O)$ of $\underline{E}(\cdot\rfloor\Omega)$—or $\underline{P}(\cdot\rfloor\Omega)$—can be used as a tractable lower bound for the natural extension $\underline{E}(\cdot\rfloor O)$ that we are actually after, which is the one that corresponds to $\underline{P}(\cdot\rfloor\cdot)$.

#### 4.3.2. Less conservative strategies

Far more attractive are updating strategies that are less conservative, because they result in belief models that are more informative and therefore, ultimately, more powerful. One of the nice aspects of our argumentation in Section 4.2.1—and Eq. (26) in particular—is that it is compatible with such strategies: $\mathscr{D}_O$ may be strictly larger than $\mathscr{D}\rfloor O$. However, the fact that these updating strategies are not ruled out by Eq. (26) does not suffice to justify them. In order to truly justify adding extra gambles to $\mathscr{D}_O$, we need to come up with a compelling principle that implies their addition. We will introduce such a principle in Section 5, and show that it implies the use of regular extension as an updating rule.

It is also important to realise that there are limits to how much larger $\mathscr{D}_O$ can be made. We have already mentioned two of these upper constraints in Section 4.2.1. First of all: $\mathscr{D}_O$ should be kept small enough to keep it coherent. Secondly: if the subject believes that $O$ can occur, then $\mathscr{D}_O$ should be equal to $\mathscr{D}\rfloor O$. In order to make this constraint exact, we need to define what "believing that $O$ can occur" means. In our finitary context, and within the framework of sets of desirable gambles, we consider it reasonable to use the following definition: a subject believes that $O$ can occur if and only if he is willing to bet on its occurrence at some strictly positive (but possibly very small) betting rate—if $\underline{P}_{\mathscr{D}}(O) > 0$; see Section 2.2.5.

A third upper constraint is that, $\mathscr{D}_O$ and $\mathscr{D}$ should avoid partial loss, in the sense that it should not be possible to combine a gamble $f \in \mathscr{D}$ with a gamble $f_O \in \mathscr{D}_O$ such that the combined transaction $f + \mathbb{I}_O f_O$ results in a payoff that is never positive and sometimes negative.[30] This is important because $\mathscr{D}$ and $\mathscr{D}_O$ are both announced before the occurrence of $O$, and therefore, a subject who has $\mathscr{D}$ and $\mathscr{D}_O$ as its belief models can be forced to accept such a combination of transactions. However, fortunately, as long as $\mathscr{D}_O$ is a coherent superset of $\mathscr{D}\rfloor O$, this will never happen.

**Proposition 14.** *Let $\mathscr{D}$ be a coherent set of desirable gambles on $\Omega$, let $O$ be an event in $\mathscr{P}_\emptyset(\Omega)$, and let $\mathscr{D}_O$ be a coherent set of desirable gambles on $O$ such that $\mathscr{D}_O \supseteq \mathscr{D}\rfloor O$. Then*

$$(f \in \mathscr{D} \text{ and } f_O \in \mathscr{D}_O) \Rightarrow f + \mathbb{I}_O f_O \not\leq 0.$$

[30] It suffices to consider just a single gamble from each of the sets involved because, by coherence of the individual sets, any finite combination of gambles of the same set is again a member of that set [D4].

*Proof.* Consider any $f \in \mathscr{D}$ and $f_O \in \mathscr{D}_O$ and let $f'_O$ be the restriction of $f$ to $O$. If there is some $\omega \in \Omega \setminus O$ such that $f(\omega) > 0$, then $(f + \mathbb{I}_O f_O)(\omega) = f(\omega) > 0$ and therefore $f + \mathbb{I}_O f_O \not\leq 0$. Hence, without loss of generality, we may assume that $\mathbb{I}_O f \geq f$. By coherence of $\mathscr{D}$ [D5], this implies that $\mathbb{I}_O f'_O = \mathbb{I}_O f \in \mathscr{D}$ and therefore also that $f'_O \in \mathscr{D}\rfloor O \subseteq \mathscr{D}_O$. Invoking the coherence of $\mathscr{D}_O$, we find that $f'_O + f_O \in \mathscr{D}_O$ [D4] and therefore also that $f'_O + f_O \not\leq 0$ [D6]. Hence, there is some $\omega \in O$ such that $f'_O(\omega) + f_O(\omega) > 0$. Since $f'_O(\omega) = f(\omega)$ and $f_O(\omega) = \mathbb{I}_O(\omega) f_O(\omega)$, this implies that $f(\omega) + \mathbb{I}_O(\omega) f_O(\omega) > 0$ and therefore also that $f + \mathbb{I}_O f_O \not\leq 0$. □

The situation becomes more tricky if a subject announces, besides $\mathscr{D}$, updated models $\mathscr{D}_O$ for multiple events $O$, for example for every element of some partition $\mathscr{O}$ of $\Omega$, or for the set of events $\mathscr{O}$ that corresponds to a protocol—see Section 4.1.3. In those cases, even if $\mathscr{D}_O \supseteq \mathscr{D}\rfloor O$ for all $O \in \mathscr{O}$, it is often very easy to combine gambles in $\mathscr{D}$ with gambles from these different updated models $\mathscr{D}_O$, $O \in \mathscr{O}$, in such a way that the combined transaction makes the subject who announced these models suffer a partial loss, or sometimes even a sure loss (a negative payoff regardless of the outcome). Dempster's rule of conditioning is for example known to suffer from this problem [38, Sections 5.13.9–11]. We will not discuss the exact conditions under which a subject can be made to suffer from such a partial or sure loss any further; see Ref. [46] for detailed technical discussions of these and other related consistency criteria between initial and updated belief models. For our present purposes, it suffices to realise that updating by means of conditioning will always avoid partial loss, simply because $\mathscr{D}$ is coherent and therefore satisfies D6. As an immediate consequence, we find that whenever there is a coherent set of desirable gambles $\mathscr{D}^*$ such that $\mathscr{D} \subseteq \mathscr{D}^*$ and such that, for all $O \in \mathscr{O}$, $\mathscr{D}_O \subseteq \mathscr{D}^*\rfloor O$, then updating by means of these models $\mathscr{D}_O$, $O \in \mathscr{O}$, is guaranteed to avoid partial loss. Regular extension provides a nice example: by Proposition 13, it can be seen to correspond to the use of updated models that are obtained by conditioning the set of desirable gambles $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ rather than the actual model $\mathscr{D}_{\underline{P}}$; since $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ is a coherent superset of $\mathscr{D}_{\underline{P}}$, we find that updating $\mathscr{D}_{\underline{P}}$ in this way is guaranteed to avoid partial loss, even if multiple updated models are announced at the same time.

*4.3.3. Different settings and interpretations*

It should not be forgotten that our justification for updating by means of natural extension only applies if the conditions that were discussed in Section 4.1 are met, and that it furthermore crucially depends on our subjective interpretation in terms of desirable gambles. If these conditions are not met, or if this interpretation is not adopted, our argumentation is no longer compelling, and other updating strategies could be considered, provided of course that one can find a way to justify them. Ref. [38, Section 6.11] and Refs. [46, 30, 33] provide some ideas on how to deal with situations where the conditions in Section 4.1 are relaxed. Ref. [12, Section 6.3.2] compares our approach with updating rules that are not based on interpretations in terms of gambles. Interestingly, many of the alternative updating rules that are provided in the literature, regardless of the setting they consider or the interpretation they adopt, tend to be at least as informative as natural extension.[31] Since we do not adopt an exhaustive interpretation, this implies that, rather surprisingly, and despite the fact that they come from a completely different direction, these other updating rules nevertheless turn out to be compatible with our approach. This being said, we will now refocus on our setting, and our interpretation, and we will use it to explain that it is possible to justify the use of regular extension as an updating rule by combining the results of this section with additional arguments.

## 5. Justifying the use of regular extension as an updating rule

Regular extension comes across as an intuitive updating rule because of its clear interpretation in terms of sets of probability mass functions or sets of linear previsions. Because of the popularity of Bayes's rule as an updating tool, it seems natural to simply apply it whenever possible, and to ignore the models to which it cannot be applied—those that assign probability zero to the event of interest. The goal of this section is to justify this approach, in two different ways. Our first justification is based on an assumption of ideal precision. It is expressed in terms of sets of linear previsions and closely resembles the intuitive idea sketched above. Our second justification for the use of regular extension as an updating rule starts from less restrictive assumptions; it does not require an assumption of ideal precision and is expressed directly in terms of sets of desirable gambles and/or lower previsions.

[31] We invite the reader to check; maximum likelihood updating (a special case of Dempster's rule of conditioning) [15] and $\alpha$-cut updating [3] are two examples.

### 5.1. *Using an assumption of ideal precision*

As soon as ideal precision is adopted, justifying the use of regular extension as an updating rule for sets of linear previsions is fairly straightforward. This justification is often taken for granted, but for the sake of completeness, let us make the argument explicit.

Let $K$ be a set of linear previsions. Due to the assumption of ideal precision, each of these linear previsions is considered to be a candidate for the 'correct' linear prevision, but we do not know which one of them it is. However, so the argument goes, after an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ occurs, some of these candidates can be ruled out, in particular those that assign probability zero to $O$. From a behavioural point of view, adopting a prevision that assigns $P(O) = 0$ implies that you are willing to bet against $O$ at betting rates that are arbitrarily close to 1—at all odds. With hindsight, after observing $O$, and given that $\Omega$ is finite, this seems like an unreasonable commitment, which is the reason why these previsions are no longer judged to be a candidate for the 'correct' model. Hence, with hindsight, after observing $O$, the initial set of candidates should have been $K' := \{P \in K \colon P(O) > 0\}$ rather than $K$. If $K' \neq \emptyset$, then since every prevision in $K'$ is a separate candidate model, each of these is to be updated individually. If the conditions of Section 4.1 are met, then as explained in Section 4.2.2, the fact that $O$ has positive probability implies that this should be done by means of Bayes's rule, leading us to use $K' \rfloor O = K \rfloor^{\mathrm{r}} O$ as our updated set of candidate models. If $K' = \emptyset$, then none of the candidate models are reasonable, and this procedure cannot be applied. In that case, we are led to consider the set $\mathbb{P}_O = K \rfloor^{\mathrm{r}} O$ of all possible previsions on $O$ as our updated set of candidates, simply because by the assumption of ideal precision, one of these previsions is guaranteed to be the correct model. Either way, the approach above leads us to adopt $K \rfloor^{\mathrm{r}} O$ as our updated set of candidate models, and thereby seems to provide a justification for the use of regular extension as an updating rule.

However, there is still a slight issue with this justification, at least in the way we have presented it so far. Indeed, one could argue that after $O$ has occurred, the initial candidate models are no longer relevant, and that therefore, it makes no sense to go back and remove some of them. This problem is solved by the fact that we are adopting the setting of Section 4.1.1. Since this setting requires that the updated model is provided in advance, the act of constructing this updated model is necessarily a thought experiment that is conducted in advance, before the occurrence of any event. Within this thought experiment, it is implicitly assumed that $O$ can occur, because otherwise, it makes no sense to provide an updated model for when $O$ actually occurs, and it is this implicit assumption that leads us to adopt $K'$ instead of $K$. It is important to realise that $K'$ exists within this thought experiment only. Our initial belief model, before any event has occurred, is still the set $K$.

Similar arguments can also be used to try and justify other updating rules. For example, taking $K'$ to be the set of all previsions that assign maximal probability to $O$ results in maximum likelihood updating, of which Dempster's rule of conditioning can be regarded as a special case [15]. $\alpha$-cut updating corresponds to the removal of all previsions $P$ in $K$ for which $P(O) < \alpha \overline{P}_K(O)$, for some $\alpha \in (0,1)$ [3]. However, we consider the case $P(O) = 0$ to be the more fundamental. The removal of extra previsions seems hard to justify on theoretical grounds. For example, the value of $\alpha$ is bound to be arbitrary. Furthermore, the resulting rules are no longer guaranteed to avoid partial or sure loss. Nevertheless, these rules have proved useful in practice, and the argumentation above can be used to motivate their use on theoretical grounds.

These arguments can also be used to justify regular extension as an updating rule for lower previsions. It suffices to adopt the sensitivity analysis interpretation, and to apply the reasoning above to the set of dominating linear previsions. However, this is unnecessarily restrictive, because, as we are about to show, a similar justification can also be obtained in a more direct manner, without any reference to the sensitivity analysis interpretation or ideal precision.

### 5.2. *A direct justification, without ideal precision*

One of the crucial points in the previous section was that, since the updated model is (assumed to be) specified in advance, the act of constructing it is necessarily a thought experiment that makes the implicit assumption that the event $O$ can occur. Under this assumption, some of the previsions in $K$ can be removed, and in this way, we obtain regular extension.

The very same idea can be applied to sets of desirable gambles. Our current set of desirable gambles $\mathscr{D}$ need not be the same as the set of desirable gambles $\mathscr{D}'$ that we would adopt under the extra assumption that $O$ can occur—not to be confused with the set of gambles $\mathscr{D} \rfloor O$ that are desirable contingent on the actual occurrence of $O$. But what should $\mathscr{D}'$ be? Is it related to $\mathscr{D}$? Can we construct it in an automated way?

*5.2.1. Adding an assessment*

We think that an assumption that $O$ can occur should lead us to add the following assessment:[32]

$$\text{For } \varepsilon \in (0,1) \text{ sufficiently small, } \mathbb{I}_O - \varepsilon \text{ should be desirable.} \tag{27}$$

In other words, there should be some positive—but possibly very small—betting rate at which you are willing to bet on $O$. Although we prefer not to stress this because it might easily be associated with an assumption of ideal precision—which do not want to make—it might be useful to realise that in terms of probabilities, this assessment simply means that $O$ has some positive—but possibly very small—probability $\varepsilon$. We consider our assumption that $\Omega$ is finite to be crucial here, because it guarantees that $O$ is 'sufficiently large' with respect to $\Omega$. If $\Omega$ were to be infinite, we would not be inclined to adopt Assessment (27) for singleton events. Nevertheless, even for finite $\Omega$, one might think that Assessment (27) is still not compelling; we leave it to the reader to decide for himself.

We want to stress that we are not assuming that $O$ can occur. Our suggestion here is simply that *if such an assumption is made*, then Assessment (27) should be adopted. Our further analysis is based on this principle, and our conclusions therefore only apply to events $O$ for which it is considered reasonable. The main idea is that, while constructing an updated model that is to be used after the occurrence of $O$, we are conducting a thought experiment in which $O$ can obviously occur, thereby allowing us to invoke Assessment (27). However, outside of this thought experiment, we do not assume that $O$ can occur.

In any case, if we choose to adopt Assessment (27), the first problem we are confronted with is the meaning of "sufficiently small": how small should $\varepsilon$ be? An obvious suggestion is to return to the subject whose beliefs are modelled by means of $\mathscr{D}$—often an expert—and ask him to provide us with an $\varepsilon \in (0,1)$ such that, under the assumption that $O$ can occur, $\mathbb{I}_O - \varepsilon$ would be desirable to him. Alternatively, the choice of $\varepsilon$ can be based on someone else's opinion—possibly your own. For now, let us assume that $\varepsilon$ is known; we will come back to this shortly.

We are now faced with a classical *belief expansion* problem [14, 18]: we have an initial belief model $\mathscr{D}$—a coherent set of desirable gambles—and want to incorporate the additional assessment that $\mathbb{I}_O - \varepsilon$ is desirable. Similarly to what is done in propositional logic, this can be achieved by considering the deductive closure of the union of these assessments, where in the language of sets of desirable gambles, the deductive closure is obtained by applying the operator $\mathscr{E}$; see Refs. [45, 23, 8] for more information. Applying this procedure, we obtain the following set of desirable gambles:

$$\mathscr{E}_O^{\varepsilon}(\mathscr{D}) := \mathscr{E}(\mathscr{D} \cup \{\mathbb{I}_O - \varepsilon\}). \tag{28}$$

This set is not guaranteed to be coherent; the assessment $\mathbb{I}_O - \varepsilon$ can be inconsistent with $\mathscr{D}$. It is easy to see that $\mathscr{E}_O^{\varepsilon}(\mathscr{D})$ will be coherent if and only if $\varepsilon - \mathbb{I}_O \notin \mathscr{D}$. It is useful to compare this with what happens in propositional logic: a belief base can be expanded with a proposition $a$ if and only if this belief base does not contain the negation of $a$.

Let us now come back to the problem of choosing $\varepsilon$. In practice, it is often very difficult to do so. The fact that we think that there should be some $\varepsilon \in (0,1)$ for which $\mathbb{I}_O - \varepsilon$ is desirable does not imply that we can actually provide such an $\varepsilon$. This typically occurs if $\mathscr{D}$ was provided by an expert that is no longer available for extra questions. Therefore, instead of fixing $\varepsilon$ in some arbitrary way, we propose to restrict attention to the set of desirable gambles

$$\mathscr{E}_O(\mathscr{D}) := \bigcap_{\varepsilon \in (0,1)} \mathscr{E}_O^{\varepsilon}(\mathscr{D}), \tag{29}$$

which consists exactly of those gambles whose desirability can *always* be inferred by expanding $\mathscr{D}$ with $\mathbb{I}_O - \varepsilon$, *regardless* of the value of $\varepsilon \in (0,1)$. Assessment (27) should clearly lead us to consider—at the very least—the gambles in $\mathscr{E}_O(\mathscr{D})$ as desirable. Other gambles might be desirable as well, but in order to find out which ones, some kind of domain expertise seems necessary. If this kind of expertise is not available, then $\mathscr{E}_O(\mathscr{D})$ seems to be a reasonable, conservative choice of model. Basically, we are then no longer adopting Assessment (27), but merely some of its consequences.

[32] This follows from our definition of "believing that $O$ can occur" in Section 4.3.2.

*5.2.2. Investigating the consequences of the assessment*

In its current form, our expression for $\mathscr{E}_O(\mathscr{D})$ is rather indirect, making it difficult to get a feeling for which gambles it contains. Therefore, before drawing any conclusions with respect to updating, we start with a theoretical study of the set $\mathscr{E}_O(\mathscr{D})$. We restrict ourselves to results that are directly related to the updating problem we are trying to solve; see Appendix AppendixA for additional properties that are—although they are relevant—not directly related to the present discussion.

The following proposition establishes that in order for $\mathscr{E}_O(\mathscr{D})$ to be coherent, it is sufficient as well as necessary for $\overline{P}_{\mathscr{D}}(O)$ to be strictly positive.

**Proposition 15.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$. Then $\mathscr{E}_O(\mathscr{D})$ is incoherent if and only if $\overline{P}_{\mathscr{D}}(O) = 0$. Furthermore, if $\mathscr{E}_O(\mathscr{D})$ is incoherent, then $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$.*

*Proof.* First, assume that $\mathscr{E}_O(\mathscr{D})$ is incoherent. Since, by construction, $\mathscr{E}_O(\mathscr{D})$ satisfies D2–D4, this implies that $\mathscr{E}_O(\mathscr{D})$ fails to satisfy D1 and therefore, that $0 \in \mathscr{E}_O(\mathscr{D})$.

Fix $\varepsilon \in (0,1)$. By Eq. (29), we know that $0 \in \mathscr{E}_O^{\varepsilon}(\mathscr{D})$. By Eq. (28), and since $\mathscr{D}$ is coherent, this implies that either $\mathbb{I}_O - \varepsilon = 0$ or $g + \lambda(\mathbb{I}_O - \varepsilon) = 0$, with $\lambda \in \mathbb{R}_{>0}$ and $g \in \mathscr{D}$. The first option is impossible because $\varepsilon < 1$. Hence, we have that $\varepsilon - \mathbb{I}_O = {}^{1}\!/\!_{\lambda} g$ and therefore, by the coherence of $\mathscr{D}$ and because ${}^{1}\!/\!_{\lambda} > 0$ and $g \in \mathscr{D}$, that $\varepsilon - \mathbb{I}_O \in \mathscr{D}$. Since this holds for all $\varepsilon \in (0,1)$, we infer from Eq. (12) that $\overline{P}_{\mathscr{D}}(O) \leq 0$ and therefore, by Eq. (13), that $\overline{P}_{\mathscr{D}}(O) = 0$.

Next, still assuming that $\mathscr{E}_O(\mathscr{D})$ is incoherent, we prove that $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$. Choose any $f \in \mathscr{G}(\Omega)$. Now choose any $\varepsilon \in (0,1)$. From the first part of this proof, we know that ${}^{\varepsilon}\!/\!_{2} - \mathbb{I}_O \in \mathscr{D}$. Now choose $\alpha \in \mathbb{R}_{>0}$ high enough such that $f + \alpha > 0$. Then, by Eq. (28),

$$f = (f+\alpha) + {}^{2\alpha}\!/\!_{\varepsilon}({}^{\varepsilon}\!/\!_{2} - \mathbb{I}_O) + {}^{2\alpha}\!/\!_{\varepsilon}(\mathbb{I}_O - \varepsilon) \in \mathscr{E}_O^{\varepsilon}(\mathscr{D}).$$

Since this holds for all $\varepsilon \in \mathbb{R}_{>0}$, we infer from Eq. (29) that $f \in \mathscr{E}_O(\mathscr{D})$. Since this holds for all $f \in \mathscr{G}(\Omega)$, we find that $\mathscr{E}_O(\mathscr{D}) = \mathscr{G}(\Omega)$.

Finally, assume that $\overline{P}_{\mathscr{D}}(O) = 0$. Consider any $\varepsilon \in (0,1)$. Then by Eq. (12), we know that there is some $0 \leq \alpha < \varepsilon$ such that $\alpha - \mathbb{I}_O \in \mathscr{D}$. Hence, by Eq. (28) and the coherence of $\mathscr{D}$, we find that $0 = (\varepsilon - \alpha) + (\alpha - \mathbb{I}_O) + (\mathbb{I}_O - \varepsilon) \in \mathscr{E}_O^{\varepsilon}(\mathscr{D})$. Since this holds for all $\varepsilon \in (0,1)$, we infer from Eq. (29) that $0 \in \mathscr{E}_O(\mathscr{D})$, implying that $\mathscr{E}_O(\mathscr{D})$ is incoherent. □

Therefore, if $\overline{P}_{\mathscr{D}}(O) = 0$, the set $\mathscr{E}_O(\mathscr{D})$ is clearly not very useful. For now, let us assume that $\overline{P}_{\mathscr{D}}(O)$ is strictly positive. In that case, perhaps surprisingly, none of the gambles $\mathbb{I}_O - \varepsilon$ in Assessment (27) is actually added to $\mathscr{D}$.

**Proposition 16.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > 0$. Then for all $\varepsilon \in (0,1)$:*

$$\mathbb{I}_O - \varepsilon \in \mathscr{E}_O(\mathscr{D}) \Leftrightarrow \mathbb{I}_O - \varepsilon \in \mathscr{D}.$$

*Proof.* Since $\mathscr{D}$ is clearly a subset of $\mathscr{E}_O(\mathscr{D})$, we only need to prove the direct implication. So consider any $\varepsilon \in (0,1)$ and assume that $\mathbb{I}_O - \varepsilon \in \mathscr{E}_O(\mathscr{D})$. We will prove that then $\mathbb{I}_O - \varepsilon \in \mathscr{D}$. If we choose $0 < \varepsilon' < \varepsilon$, then by Eq. (29), we know that $\mathbb{I}_O - \varepsilon \in \mathscr{E}_O^{\varepsilon'}(\mathscr{D})$. By Eq. (28), and because $\mathscr{D}$ is coherent, we now have that, without loss of generality, either (a) $\mathbb{I}_O - \varepsilon \in \mathscr{D}$ or (b) $\mathbb{I}_O - \varepsilon = g + \lambda(\mathbb{I}_O - \varepsilon')$, with $g \in \mathscr{D} \cup \{0\}$ and $\lambda > 0$. In case of (a), the proof is concluded. In case of (b), we find that, with $\lambda' := 1 - \lambda$,

$$\lambda'(\mathbb{I}_O - \varepsilon) = \mathbb{I}_O - \varepsilon + \lambda(\varepsilon' - \mathbb{I}_O) + \lambda(\varepsilon - \varepsilon') = g + \lambda(\varepsilon - \varepsilon') \in \mathscr{D},$$

where the last inclusion is a consequence of the coherence of $\mathscr{D}$. If $\lambda' = 0$, then $0 \in \mathscr{D}$, contradicting the coherence of $\mathscr{D}$. If $\lambda' < 0$, then, again by the coherence of $\mathscr{D}$, $\varepsilon - \mathbb{I}_O \in \mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D})$. By combining this with our assumption, and using the coherence of $\mathscr{E}_O(\mathscr{D})$ [which is a consequence of Proposition 15 and our assumption that $\overline{P}_{\mathscr{D}}(O) > 0$], we find that $0 = (\varepsilon - \mathbb{I}_O) + (\mathbb{I}_O - \varepsilon) \in \mathscr{E}_O(\mathscr{D})$, contradicting the coherence of $\mathscr{E}_O(\mathscr{D})$. The only remaining possibility is that $\lambda' > 0$. In this case, by the coherence of $\mathscr{D}$, we find that, indeed, $\mathbb{I}_O - \varepsilon \in \mathscr{D}$. □

If $\underline{P}_{\mathscr{D}}(O)$ is strictly positive, we even find that $\mathscr{E}_O(\mathscr{D})$ is equal to $\mathscr{D}$.

**Proposition 17.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and an event $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\underline{P}_{\mathscr{D}}(O) > 0$. Then $\mathscr{E}_O(\mathscr{D}) = \mathscr{D}$.*

*Proof.* By Eq. (11), and since $\underline{P}_{\mathscr{D}}(O) > 0$, we know that there is an $\varepsilon > 0$ such that $\mathbb{I}_O - \varepsilon \in \mathscr{D}$. Since $\mathscr{D}$ is coherent, we also know that $\varepsilon < 1$. From Eq. (28), we now infer that $\mathscr{E}_O^{\varepsilon}(\mathscr{D}) = \mathscr{E}(\mathscr{D})$ and therefore, since $\mathscr{D}$ is coherent, that $\mathscr{E}_O^{\varepsilon}(\mathscr{D}) = \mathscr{D}$. Applying Eq. (29), we find that $\mathscr{E}_O(\mathscr{D}) \subseteq \mathscr{D}$. Since clearly also $\mathscr{D} \subseteq \mathscr{E}_O(\mathscr{D})$, we conclude that $\mathscr{E}_O(\mathscr{D}) = \mathscr{D}$. □

By Eq. (13), the only remaining option is that $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$. In that case, $\mathscr{E}_O(\mathscr{D})$ is completely characterised by the following theorem.

**Theorem 18.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and let $O \in \mathscr{P}_{\emptyset}(\Omega)$ be any event such that $\overline{P}_{\mathscr{D}}(O) > 0$. Then*

$$f \in \mathscr{E}_O(\mathscr{D}) \Leftrightarrow f \in \mathscr{D} \text{ or } (\forall \varepsilon \in (0,1))(\exists \lambda \in \mathbb{R}_{>0})\ f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathscr{D} \tag{30}$$

$$\Leftrightarrow f \in \mathscr{D} \text{ or } (\forall \varepsilon \in (0,1))(\exists \overline{\lambda} \in \mathbb{R}_{>0})(\forall \lambda \in (0,\overline{\lambda}])\ f + \lambda(\varepsilon - \mathbb{I}_O) \in \operatorname{int}(\mathscr{D}). \tag{31}$$

*Proof.* First, assume that the right-hand side of Eq. (30) or Eq. (31) holds. Then clearly, for all $\varepsilon \in (0,1)$, by Eq. (28), $f \in \mathscr{E}_O^{\varepsilon}(\mathscr{D})$. Hence, by Eq. (29), $f \in \mathscr{E}_O(\mathscr{D})$.

Next, assume that $f \in \mathscr{E}_O(\mathscr{D})$. We will prove that this implies the right-hand side of Eqs. (30) and (31). Clearly, without loss of generality, we may assume that $f \in \mathscr{E}_O(\mathscr{D}) \setminus \mathscr{D}$. Now fix $\varepsilon \in (0,1)$. By Eq. (28), and because $\mathscr{D}$ is coherent, we find that there are $g \in \mathscr{D} \cup \{0\}$ and $\overline{\lambda}_\varepsilon \in \mathbb{R}_{>0}$ such that $f = g + \overline{\lambda}_\varepsilon(\mathbb{I}_O - \varepsilon)$. If $g = 0$, then by the coherence of $\mathscr{E}_O(\mathscr{D})$ [because of Proposition 15 and because $\overline{P}_{\mathscr{D}}(O) > 0$], $\mathbb{I}_O - \varepsilon \in \mathscr{E}_O(\mathscr{D})$. Due to Proposition 16 and the coherence of $\mathscr{D}$, this implies that $f = \overline{\lambda}_\varepsilon(\mathbb{I}_O - \varepsilon) \in \mathscr{D}$, a contradiction. Hence, $g \in \mathscr{D}$ and therefore $f + \overline{\lambda}_\varepsilon(\varepsilon - \mathbb{I}_O) \in \mathscr{D}$. By repeating this argument, we obtain a set of coefficients $\overline{\lambda}_\varepsilon \in \mathbb{R}_{>0}$, one for every $\varepsilon \in (0,1)$, that satisfies the following condition:

$$(\forall \varepsilon \in (0,1))\ f + \overline{\lambda}_\varepsilon(\varepsilon - \mathbb{I}_O) \in \mathscr{D}. \tag{32}$$

This already proves Eq. (30) [simply denote $\overline{\lambda}_\varepsilon$ as $\lambda$].

Now fix $\varepsilon \in (0,1)$ and $\lambda \in (0, \overline{\lambda}_{\varepsilon^*}]$, with $\varepsilon^* := \varepsilon/2$. We will show that $f + \lambda(\varepsilon - \mathbb{I}_O) \in \operatorname{int}(\mathscr{D})$ and thereby finish the proof for Eq. (31) [simply denote $\overline{\lambda}_{\varepsilon^*}$ as $\overline{\lambda}$]. We consider two possibilities. The first possibility is that, for all $\varepsilon' \in (0, \varepsilon^*]$, $\overline{\lambda}_{\varepsilon'}\varepsilon' \geq \lambda\varepsilon^*$. Then for all $\varepsilon'' \in (0,1)$, we can choose $\alpha > \max\{0, \max f\}$ and $\varepsilon' \in (0, \varepsilon^*]$ small enough such that $\varepsilon'(1 + \alpha/\lambda\varepsilon^*) \leq \varepsilon''$ and therefore also

$$\overline{\lambda}_{\varepsilon'}(\varepsilon'' - \mathbb{I}_O) \geq \overline{\lambda}_{\varepsilon'}(\varepsilon'(1 + \frac{\alpha}{\lambda\varepsilon^*}) - \mathbb{I}_O) \geq \overline{\lambda}_{\varepsilon'}(\varepsilon'(1 + \frac{\alpha}{\overline{\lambda}_{\varepsilon'}\varepsilon'}) - \mathbb{I}_O) = \alpha + \overline{\lambda}_{\varepsilon'}(\varepsilon' - \mathbb{I}_O) > f + \overline{\lambda}_{\varepsilon'}(\varepsilon' - \mathbb{I}_O).$$

By combining this with Eq. (32) and the coherence of $\mathscr{D}$, we find that $\varepsilon'' - \mathbb{I}_O \in \mathscr{D}$. Since this holds for all $\varepsilon'' \in (0,1)$, we infer from Eq. (12) that $\overline{P}_{\mathscr{D}}(O) \leq 0$, which contradicts our assumption. Hence, we only have to consider the second, remaining possibility, namely that there is some $\varepsilon' \in (0, \varepsilon^*]$ for which $\overline{\lambda}_{\varepsilon'}\varepsilon' < \lambda\varepsilon^*$. Since also $\lambda\varepsilon^* \leq \overline{\lambda}_{\varepsilon^*}\varepsilon^*$, we can use this particular $\varepsilon'$ to define

$$\delta_1 := \frac{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \lambda\varepsilon^*}{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'} \geq 0,\ \delta_2 := 1 - \delta_1 = \frac{\lambda\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'}{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'} > 0 \text{ and } \beta := \delta_1\overline{\lambda}_{\varepsilon'} + \delta_2\overline{\lambda}_\varepsilon \geq \lambda,$$

where the last inequality holds because

$$\beta = \frac{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \lambda\varepsilon^*}{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'}\overline{\lambda}_{\varepsilon'} + \frac{\lambda\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'}{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'}\overline{\lambda}_{\varepsilon^*} = \frac{\overline{\lambda}_{\varepsilon^*}\varepsilon^*\overline{\lambda}_{\varepsilon'} - \lambda\varepsilon^*\overline{\lambda}_{\varepsilon'} + \lambda\varepsilon^*\overline{\lambda}_{\varepsilon^*} - \overline{\lambda}_{\varepsilon'}\varepsilon'\overline{\lambda}_{\varepsilon^*}}{\overline{\lambda}_{\varepsilon^*}\varepsilon^* - \overline{\lambda}_{\varepsilon'}\varepsilon'}$$

and therefore

$$\begin{aligned}
\beta \geq \lambda &\Leftrightarrow \overline{\lambda}_{\varepsilon^*}\varepsilon^*\overline{\lambda}_{\varepsilon'} - \lambda\varepsilon^*\overline{\lambda}_{\varepsilon'} + \lambda\varepsilon^*\overline{\lambda}_{\varepsilon^*} - \overline{\lambda}_{\varepsilon'}\varepsilon'\overline{\lambda}_{\varepsilon^*} \geq \overline{\lambda}_{\varepsilon^*}\varepsilon^*\lambda - \overline{\lambda}_{\varepsilon'}\varepsilon'\lambda \\
&\Leftrightarrow \overline{\lambda}_{\varepsilon^*}\varepsilon^*\overline{\lambda}_{\varepsilon'} - \lambda\varepsilon^*\overline{\lambda}_{\varepsilon'} - \overline{\lambda}_{\varepsilon'}\varepsilon'\overline{\lambda}_{\varepsilon^*} + \overline{\lambda}_{\varepsilon'}\varepsilon'\lambda \geq 0 \\
&\Leftrightarrow \overline{\lambda}_{\varepsilon^*}\varepsilon^* - \lambda\varepsilon^* - \varepsilon'\overline{\lambda}_{\varepsilon^*} + \varepsilon'\lambda \geq 0 \Leftrightarrow (\overline{\lambda}_{\varepsilon^*} - \lambda)(\varepsilon^* - \varepsilon') \geq 0.
\end{aligned}$$

Since

$$\delta_1\overline{\lambda}_{\varepsilon'}\varepsilon' + \delta_2\overline{\lambda}_{\varepsilon^*}\varepsilon^* = \delta_1\overline{\lambda}_{\varepsilon'}\varepsilon' + (1 - \delta_1)\overline{\lambda}_{\varepsilon^*}\varepsilon^* = \overline{\lambda}_{\varepsilon^*}\varepsilon^* + \delta_1(\overline{\lambda}_{\varepsilon'}\varepsilon' - \overline{\lambda}_{\varepsilon^*}\varepsilon^*) = \lambda\varepsilon^*,$$

we find that

$$f+\lambda(\varepsilon^*-\mathbb{I}_O)=f+\lambda\varepsilon^*-\lambda\mathbb{I}_O\geq f+\lambda\varepsilon^*-\beta\mathbb{I}_O=(\delta_1+\delta_2)f+\delta_1\overline{\lambda}_{\varepsilon'}\varepsilon'+\delta_2\overline{\lambda}_{\varepsilon^*}\varepsilon^*-(\delta_1\overline{\lambda}_{\varepsilon'}+\delta_2\overline{\lambda}_{\varepsilon})\mathbb{I}_O$$
$$=\delta_1[f+\overline{\lambda}_{\varepsilon'}(\varepsilon'-\mathbb{I}_O)]+\delta_2[f+\overline{\lambda}_{\varepsilon^*}(\varepsilon^*-\mathbb{I}_O)].$$

By Eq. (32) and the coherence of $\mathscr{D}$, this implies that $f+\lambda(\varepsilon^*-\mathbb{I}_O)\in\mathscr{D}$. Hence, if we define $\delta:=\lambda\varepsilon^*>0$, then since we know that $\lambda\varepsilon^*=\lambda\varepsilon-\delta$ [because $\varepsilon^*=\varepsilon/2$], we find that $f+\lambda(\varepsilon-\mathbb{I}_O)-\delta\in\mathscr{D}$ and therefore also, by Eq. (5), that $f+\lambda(\varepsilon-\mathbb{I}_O)\in\operatorname{int}(\mathscr{D})$. □

Although this result applies whenever $\overline{P}_{\mathscr{D}}(O)>0$, we are of course mainly interested in the case $\underline{P}_{\mathscr{D}}(O)=0$. If $\underline{P}_{\mathscr{D}}(O)$ is positive, we already know from Proposition 17 that $\mathscr{E}_O(\mathscr{D})=\mathscr{D}$. If $\underline{P}_{\mathscr{D}}(O)=0$, $\mathscr{E}_O(\mathscr{D})$ might be strictly larger than $\mathscr{D}$; however, as the following result shows, only slightly—or should we say, marginally—so.

**Corollary 19.** *Consider a coherent set of desirable gambles $\mathscr{D}\subseteq\mathscr{G}(\Omega)$ and let $O\in\mathscr{P}_\emptyset(\Omega)$ be any event such that $\overline{P}_{\mathscr{D}}(O)>0$. Then $\mathscr{D}\subseteq\mathscr{E}_O(\mathscr{D})\subseteq\operatorname{cl}(\mathscr{D})=\operatorname{cl}(\mathscr{E}_O(\mathscr{D}))$ and*

$$\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f)=\underline{P}_{\mathscr{D}}(f)\ \text{ for all } f\in\mathscr{G}(\Omega).$$

*Proof.* We first prove that $\mathscr{D}\subseteq\mathscr{E}_O(\mathscr{D})\subseteq\operatorname{cl}(\mathscr{D})$. Since $\mathscr{D}$ is clearly a subset of $\mathscr{E}_O(\mathscr{D})$, it suffices to show that $\mathscr{E}_O(\mathscr{D})\subseteq\operatorname{cl}(\mathscr{D})$. So fix any $f\in\mathscr{E}_O(\mathscr{D})$ and $\delta\in\mathbb{R}_{>0}$. By Eq. (4), we need to prove that $f+\delta\in\mathscr{D}$. If $f\in\mathscr{D}$, this follows trivially from the coherence of $\mathscr{D}$. Hence, by Theorem 18, we can assume without loss of generality that

$$(\forall\varepsilon\in(0,1))(\exists\overline{\lambda}\in\mathbb{R}_{>0})(\forall\lambda\in(0,\overline{\lambda}])\ f+\lambda(\varepsilon-\mathbb{I}_O)\in\operatorname{int}(\mathscr{D})\subseteq\mathscr{D}.$$

If we choose $\varepsilon\in(0,1)$ and $\lambda\in\mathbb{R}_{>0}$ small enough, then $f+\delta\geq f+\lambda(\varepsilon-\mathbb{I}_O)\in\mathscr{D}$ and therefore, by the coherence of $\mathscr{D}$, $f+\delta\in\mathscr{D}$.

We now know that $\mathscr{D}\subseteq\mathscr{E}_O(\mathscr{D})$ and $\mathscr{E}_O(\mathscr{D})\subseteq\operatorname{cl}(\mathscr{D})$. By applying the operator cl to both sides of each of these inclusions, we find that

$$\operatorname{cl}(\mathscr{D})\subseteq\operatorname{cl}(\mathscr{E}_O(\mathscr{D}))\subseteq\operatorname{cl}(\operatorname{cl}(\mathscr{D}))=\operatorname{cl}(\mathscr{D}),$$

where the last equality follows trivially from the fact that for coherent $\mathscr{D}$, cl coincides with the topological closure operator. Hence, we find that $\operatorname{cl}(\mathscr{D})=\operatorname{cl}(\mathscr{E}_O(\mathscr{D}))$. By Eq. (6), and since $\mathscr{E}_O(\mathscr{D})$ is coherent because of Proposition 15, this implies that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f)=\underline{P}_{\mathscr{D}}(f)$ for all $f\in\mathscr{G}(\Omega)$. □

Hence, if $\overline{P}_{\mathscr{D}}(O)>\underline{P}_{\mathscr{D}}(O)=0$, the difference between $\mathscr{D}$ and $\mathscr{E}_O(\mathscr{D})$ is situated on their border. Nevertheless, this difference could be important, especially if we start to condition these models.

One particular conditional model that will be especially useful to us is $\mathscr{E}_O(\mathscr{D})\rfloor O$. It has the nice property that the associated set of almost desirable gambles does not depend on the border structure of $\mathscr{E}_O(\mathscr{D})$ or $\mathscr{D}$.

**Proposition 20.** *Consider a coherent set of desirable gambles $\mathscr{D}\subseteq\mathscr{G}(\Omega)$ and an event $O\in\mathscr{P}_\emptyset(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O)>0$. Then for all $f\in\mathscr{G}(O)$:*

$$f\in\operatorname{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)\Leftrightarrow\mathbb{I}_Of\in\operatorname{cl}(\mathscr{E}_O(\mathscr{D}))\Leftrightarrow\mathbb{I}_Of\in\operatorname{cl}(\mathscr{D}).$$

*Proof.* We only prove that $f\in\operatorname{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)\Leftrightarrow\mathbb{I}_Of\in\operatorname{cl}(\mathscr{D})$; this implies the other equivalences because, as we have shown in the proof of Corollary 19, $\operatorname{cl}(\mathscr{E}_O(\mathscr{D}))=\operatorname{cl}(\mathscr{D})$.

First assume that $\mathbb{I}_Of\in\operatorname{cl}(\mathscr{D})$. Consider any $\lambda\in\mathbb{R}_{>0}$ and $\varepsilon\in(0,1)$. Then by Eq. (4), $\mathbb{I}_O(f+\lambda)+\lambda(\varepsilon-\mathbb{I}_O)=\mathbb{I}_Of+\lambda\varepsilon\in\mathscr{D}$ [because $\delta:=\lambda\varepsilon\in\mathbb{R}_{>0}$]. Since this holds for all $\varepsilon\in(0,1)$, we infer from Theorem 18 that $\mathbb{I}_O(f+\lambda)\in\mathscr{E}_O(\mathscr{D})$ or, equivalently, that $f+\lambda\in\mathscr{E}_O(\mathscr{D})\rfloor O$. Since this holds for all $\lambda\in\mathbb{R}_{>0}$, we find that $f\in\operatorname{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)$.

Conversely, assume that $f\in\operatorname{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)$. Consider any $\delta\in\mathbb{R}_{>0}$ and choose $\alpha\in(0,1)$ such that $\alpha<\delta$. Then, by Eq. (4), $f+\alpha/2\in\mathscr{E}_O(\mathscr{D})\rfloor O$ and therefore $\mathbb{I}_O(f+\alpha/2)\in\mathscr{E}_O(\mathscr{D})$. If $\mathbb{I}_O(f+\alpha/2)\in\mathscr{D}$, then by coherence of $\mathscr{D}$ also $\mathbb{I}_O(f+\delta)\in\mathscr{D}$. If $\mathbb{I}_O(f+\alpha/2)\notin\mathscr{D}$, then by applying Theorem 18 with $\varepsilon:=\alpha/2$, we find some $\overline{\lambda}\in\mathbb{R}_{>0}$ such that $\mathbb{I}_O(f+\alpha/2)+\lambda(\alpha/2-\mathbb{I}_O)\in\operatorname{int}(\mathscr{D})\subseteq\mathscr{D}$, with $\lambda:=\min\{1,\overline{\lambda}\}\in(0,\overline{\lambda}]$. Since $\lambda\leq1$ and $0<\alpha<\delta$, we know that $\delta>\alpha\geq\alpha/2(1+\mathbb{I}_O)\geq\alpha/2(\lambda+\mathbb{I}_O)$, which implies that

$$\mathbb{I}_Of+\delta>\mathbb{I}_Of+\frac{\alpha}{2}(\lambda+\mathbb{I}_O)\geq\mathbb{I}_Of+\frac{\alpha}{2}(\lambda+\mathbb{I}_O)-\mathbb{I}_O\lambda=\mathbb{I}_O(f+\frac{\alpha}{2})+\lambda(\frac{\alpha}{2}-\mathbb{I}_O).$$

Therefore, because $\mathbb{I}_O(f+\alpha/2)+\lambda(\alpha/2-\mathbb{I}_O)\in\mathscr{D}$, it follows from the coherence of $\mathscr{D}$ that, again, $\mathbb{I}_O f+\delta\in\mathscr{D}$. Hence, in all cases, $\mathbb{I}_O f+\delta\in\mathscr{D}$. Since this holds for all $\delta\in\mathbb{R}_{>0}$, we infer from Eq. (4) that $\mathbb{I}_O f\in\mathrm{cl}(\mathscr{D})$. □

Therefore, if we are only interested in $\mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)$, or equivalently, in $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$, all we need to know is $\mathrm{cl}(\mathscr{D})$, or equivalently, $\underline{P}_{\mathscr{D}}$. The connection between $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ and $\underline{P}_{\mathscr{D}}$ is provided by regular extension.

**Corollary 21.** *Consider a coherent set of desirable gambles $\mathscr{D}\subseteq\mathscr{G}(\Omega)$ and an event $O$ in $\mathscr{P}_\emptyset(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O)>0$. Let $\underline{R}(\cdot\rfloor O)$ be the regular extension of $\underline{P}_{\mathscr{D}}$, as given by Eq.* (23)*. Then*

$$\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O)=\underline{R}(f\rfloor O)\quad\text{for all } f\in\mathscr{G}(O).$$

*Proof.* Fix any $f\in\mathscr{G}(O)$ and any $\mu\in\mathbb{R}$. By Proposition 20 and Eq. (4), we have that

$$\underline{P}_{\mathscr{D}}(\mathbb{I}_O[f-\mu])\geq 0\Leftrightarrow\mathbb{I}_O[f-\mu]\in\mathrm{cl}(\mathscr{D})\Leftrightarrow f-\mu\in\mathrm{cl}(\mathscr{E}_O(\mathscr{D})\rfloor O)\Leftrightarrow\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f-\mu)\geq 0.$$

Since $\overline{P}_{\mathscr{D}}(O)>0$, we know from Proposition 15 that $\mathscr{E}_O(\mathscr{D})$ is coherent. This implies that $\mathscr{E}_O(\mathscr{D})\rfloor O$ and therefore also $\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}$ is coherent, which allows us to infer that $\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f-\mu)=\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f)-\mu$ [C8]. Furthermore, due to Eqs. (3) and (7), we know that $\underline{P}_{\mathscr{E}_O(\mathscr{D})\rfloor O}(f)=\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O)$. Hence, putting it all together, we find that

$$\underline{P}_{\mathscr{D}}(\mathbb{I}_O[f-\mu])\geq 0\Leftrightarrow\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O)\geq\mu.$$

Since this holds for all $\mu\in\mathbb{R}$, we infer from Eq. (23) and the fact that $\overline{P}_{\mathscr{D}}(O)>0$ that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O)\geq\underline{R}(f\rfloor O)$.

Next, since $\mathscr{E}_O(\mathscr{D})$ is coherent, we know that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ is coherent with $\underline{P}_{\mathscr{E}_O(\mathscr{D})}$ and therefore also, since $\underline{P}_{\mathscr{E}_O(\mathscr{D})}=\underline{P}_{\mathscr{D}}$ because of Corollary 19, that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(\cdot\rfloor O)$ is coherent with $\underline{P}_{\mathscr{D}}$. Since $\overline{P}_{\mathscr{D}}(O)>0$ guarantees that $\underline{R}(\cdot\rfloor O)$ is the largest lower prevision on $\mathscr{G}(O)$ that is coherent with $\underline{P}_{\mathscr{D}}$, this implies that $\underline{P}_{\mathscr{E}_O(\mathscr{D})}(f\rfloor O)\leq\underline{R}(f\rfloor O)$. □

*5.2.3. Turning the assessment into an updating rule*

Plenty of mathematics so far, but still no updating rule. So let us get back to the beginning: a coherent set of desirable gambles $\mathscr{D}\subseteq\mathscr{G}(\Omega)$ and an event $O\in\mathscr{P}_\emptyset(\Omega)$. We are looking for an updated set of desirable gambles $\mathscr{D}_O$ that we intend to adopt after the occurrence of $O$, but which is specified in advance.

If $O$ cannot occur, it makes no sense to provide an updated model $\mathscr{D}_O$, nor does it matter if we get it wrong. Hence, *while constructing the updated model $\mathscr{D}_O$*, we might as well assume that $O$ can occur. Within this thought experiment, if we are willing to adopt Assessment (27)—and we will assume that we are—then as explained in Section 5.2.1, we can combine this assessment with our set of desirable gambles $\mathscr{D}$ to obtain a new set of gambles that is guaranteed to include $\mathscr{E}_O(\mathscr{D})$. Given that our interpretation for sets of desirable gambles is non-exhaustive, using $\mathscr{E}_O(\mathscr{D})$ itself puts us on the safe, conservative side. The rest of the argument now depends on whether or not $\mathscr{E}_O(\mathscr{D})$ is coherent.

If it is, it means that Assessment (27) is compatible with $\mathscr{D}$, and we are led to consider $\mathscr{D}'=\mathscr{E}_O(\mathscr{D})$ as our set of desirable gambles. It is important to realise that $\mathscr{D}'$ is only adopted within the thought experiment of—in advance—constructing an updated model for after the occurrence of $O$; the belief model of our subject is still $\mathscr{D}$. Within this thought experiment, we can now apply the arguments of Section 4, and this results in the use of $\mathscr{D}_O=\mathscr{D}'\rfloor O=\mathscr{E}_O(\mathscr{D})\rfloor O$ as a conservative choice of updated model. If $\underline{P}_{\mathscr{D}}(O)>0$, Proposition 17 tells us that $\mathscr{E}_O(\mathscr{D})=\mathscr{D}$, which implies that in that case, the updated model $\mathscr{D}_O=\mathscr{D}\rfloor O$ is exactly the same as the one we would have used had we not included Assessment (27).

If $\mathscr{E}_O(\mathscr{D})$ is incoherent, or equivalently, by Proposition 15, if $\overline{P}_{\mathscr{D}}(O)=0$, it means that Assessment (27) is not compatible with $\mathscr{D}$. In fact, it even implies that $\mathscr{E}_O(\mathscr{D})=\mathscr{G}(\Omega)$. Hence, in that case, it clearly makes no sense to use $\mathscr{E}_O(\mathscr{D})$ as the set $\mathscr{D}'$ that is adopted within the thought experiment of construction $\mathscr{D}_O$. A possible solution to this problem is to drop Assessment (27) and take $\mathscr{D}'$ to be equal to $\mathscr{D}$. By applying the arguments of Section 4 to $\mathscr{D}'$, this results in the use of $\mathscr{D}_O=\mathscr{D}'\rfloor O=\mathscr{D}\rfloor O$ as our updated model. However, the same conclusion can also be reached without dropping Assessment (27). The fact that Assessment (27) is incompatible with $\mathscr{D}$ does not imply that this is also the case for subsets of $\mathscr{D}$. Indeed, in particular, the gambles that are of the form $\mathbb{I}_O f$, with $f\in\mathscr{D}\rfloor O$, will always be compatible with Assessment (27).[33] Since these gambles are the only ones that are relevant for Eq. (26), the

[33] Because coherence of $\mathscr{D}$ implies that $\mathbb{I}_O f\not\lesssim 0$, which in turn implies that $f\not\lesssim 0$. Hence, for any $\varepsilon\in(0,1)$ and any $\lambda\in\mathbb{R}_{>0}$, it holds that $f+\lambda(1-\varepsilon)\not\lesssim 0$ and therefore also that $\mathbb{I}_O f+\lambda(\mathbb{I}_O-\varepsilon)\not\lesssim 0$.

argument of Section 4 can still be applied, and it leads us to conclude that using $\mathscr{D}_O = \mathscr{D}\rfloor O$ as our updated model is a conservative approach.

By combining the case in which $\mathscr{E}_O(\mathscr{D})$ is coherent with the one in which it is not, we end up with the following updating rule:

$$\mathscr{D}^{\mathrm{r}}_O := \begin{cases} \mathscr{E}_O(\mathscr{D})\rfloor O = \mathscr{D}\rfloor O & \text{if } \underline{P}_{\mathscr{D}}(O) > 0 \\ \mathscr{E}_O(\mathscr{D})\rfloor O & \text{if } \overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0 \\ \mathscr{D}\rfloor O & \text{if } \overline{P}_{\mathscr{D}}(O) = 0 \end{cases} \quad \text{for all } O \in \mathscr{P}_{\emptyset}(\Omega). \tag{33}$$

It is identical to simply conditioning, except if $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$, in which case it is guaranteed to be at least as informative—since $\mathscr{E}_O(\mathscr{D}) \supseteq \mathscr{D}$. Nevertheless, despite it being more informative, this strategy still avoids partial loss, even if we announce updated models for multiple events $O$ at the same time.[34] Furthermore, it clearly also satisfies the second upper constraint that was discussed in Section 4.3.2.

If we let $\underline{R}(\cdot\rfloor\cdot)$ be the regular extension of $\underline{P}_{\mathscr{D}}$, then by Theorem 18, for all $O \in \mathscr{P}_{\emptyset}(\Omega)$, the lower prevision that is associated with the updated set $\mathscr{D}^{\mathrm{r}}_O$ is given by

$$\underline{P}_{\mathscr{D}^{\mathrm{r}}_O}(f) = \begin{cases} \underline{R}(f\rfloor O) = \underline{P}_{\mathscr{D}}(f\rfloor O) & \text{if } \underline{P}_{\mathscr{D}}(O) > 0 \\ \underline{R}(f\rfloor O) & \text{if } \overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0 \\ \underline{P}_{\mathscr{D}}(f\rfloor O) & \text{if } \overline{P}_{\mathscr{D}}(O) = 0 \end{cases} \quad \text{for all } f \in \mathscr{G}(O). \tag{34}$$

*5.2.4. Rephrasing the result in terms of lower previsions*

The updating rule that was derived in the previous section can be translated to the framework of lower previsions in a straightforward manner. For any coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ and any event $O \in \mathscr{P}_{\emptyset}(\Omega)$, we simply apply Eq. (33) to the smallest associated set of desirable gambles $\mathscr{D} = \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ and consider the corresponding lower prevision $\underline{P}_{\mathscr{D}^{\mathrm{r}}_O}$, as given by Eq. (34). In order to reflect this particular choice of $\mathscr{D}$ in our notation, we denote this lower prevision as $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor O)$. The following result establishes that $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ is a coherent conditional lower prevision on $\mathscr{C}(\Omega)$ that is furthermore completely characterised by the natural extension of $\underline{P}(\cdot\rfloor\cdot)$.

**Corollary 22.** *Consider a coherent conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with arbitrary domain $\mathscr{C}$. Let $\underline{E}(\cdot\rfloor\cdot)$ be its natural extension and let $\underline{R}(\cdot\rfloor\cdot)$ be the regular extension of $\underline{E}(\cdot\rfloor\Omega)$. Then $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ is a coherent lower prevision on $\mathscr{C}(\Omega)$ and, for all $(f,O) \in \mathscr{C}(\Omega)$:*

$$\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(f\rfloor O) = \begin{cases} \underline{R}(f\rfloor O) = \underline{E}(f\rfloor O) & \text{if } \underline{E}(O\rfloor\Omega) > 0 \\ \underline{R}(f\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega) > \underline{E}(O\rfloor\Omega) = 0 \\ \underline{E}(f\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega) = 0 \end{cases}$$

*Proof.* The equality follows from Eq. (34) and the fact that, with $\mathscr{D} = \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$, by definition of natural extension, $\underline{P}_{\mathscr{D}}(\cdot\rfloor\cdot) = \underline{E}(\cdot\rfloor\cdot)$ and therefore also $\underline{P}_{\mathscr{D}}(\cdot) = \underline{E}(\cdot\rfloor\Omega)$. Since $\underline{P}(\cdot\rfloor\cdot)$ is coherent, $\mathscr{D} = \mathscr{E}_{\underline{P}(\cdot\rfloor\cdot)}$ is coherent as well, and therefore, Corollary 31—see Appendix AppendixA—implies that $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot) = \underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})}(\cdot\rfloor\cdot)$ is a coherent lower prevision on $\mathscr{C}(\Omega)$. □

Interestingly, if we apply this procedure to an unconditional lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$, the resulting conditional lower prevision $\underline{R}_{\underline{P}}(\cdot\rfloor\cdot)$ is equal to the regular extension of $\underline{P}$.

**Corollary 23.** *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{R}(\cdot\rfloor\cdot)$ be its regular extension. Then*

$$\underline{R}_{\underline{P}}(f\rfloor O) = \underline{R}(f\rfloor O) \quad \textit{for all } (f,O) \in \mathscr{C}(\Omega).$$

*Proof.* As explained in the proof of Corollary 22, $\underline{R}_{\underline{P}}(\cdot\rfloor\cdot)$ is equal to $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})}(\cdot\rfloor\cdot)$, with, in this particular case, $\mathscr{D} = \mathscr{E}_{\underline{P}} = \mathscr{D}_{\underline{P}}$. The proof now follows immediately from Corollary 32—see Appendix AppendixA. □

[34] There is a coherent set of desirable gambles $\mathscr{D}^* = \mathscr{E}^{\mathrm{r}}(\mathscr{D})$—see Appendix AppendixA, Corollary 32—such that $\mathscr{D} \subseteq \mathscr{D}^*$ and $\mathscr{E}_O(\mathscr{D}) \subseteq \mathscr{D}^*$ for all $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) > \underline{P}_{\mathscr{D}}(O) = 0$. Hence, by Eq. (33), for all $O \in \mathscr{P}_{\emptyset}(\Omega)$, $\mathscr{D}^{\mathrm{r}}_O \subseteq \mathscr{D}^*\rfloor O$. As explained in Section 4.3.2, this implies that updating by means of the updated sets $\mathscr{D}^{\mathrm{r}}_O$ is guaranteed to avoid partial loss, even if multiple updated sets are announced at the same time.

This result implies that $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ can be regarded as a generalisation of the notion of regular extension to coherent *conditional* lower previsions. Therefore, from now on, for any conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, we will refer to $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$ as the *regular extension* of $\underline{P}(\cdot\rfloor\cdot)$ and, whenever it is clear from the context which conditional lower prevision it is derived from, we drop the index and write $\underline{R}(\cdot\rfloor\cdot)$ instead of $\underline{R}_{\underline{P}(\cdot\rfloor\cdot)}(\cdot\rfloor\cdot)$.

If we are considering the regular extension $\underline{R}(\cdot\rfloor\cdot)$ of a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$, some notational confusion might arise, because in Corollary 22, we used $\underline{R}(\cdot\rfloor\cdot)$ to refer to the regular extension of $\underline{E}(\cdot\rfloor\Omega)$, with $\underline{E}(\cdot\rfloor\cdot)$ the natural extension of $\underline{P}(\cdot\rfloor\cdot)$. In order to avoid this confusion, we will denote the regular extension of $\underline{E}(\cdot\rfloor\Omega)$ by $\underline{R}^*(\cdot\rfloor\cdot)$. Similarly, as we did in Section 4.2.3, we denote the natural extension of $\underline{E}(\cdot\rfloor\Omega)$ by $\underline{E}^*(\cdot\rfloor\cdot)$. Using these conventions, the regular extension $\underline{R}(\cdot\rfloor\cdot)$ of $\underline{P}(\cdot\rfloor\cdot)$ is given by

$$\underline{R}(\cdot\rfloor O) = \begin{cases} \underline{R}^*(\cdot\rfloor O) = \underline{E}^*(\cdot\rfloor O) & \text{if } \underline{E}(O\rfloor\Omega) > 0 \\ \underline{R}^*(\cdot\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega) > \underline{E}(O\rfloor\Omega) = 0 \\ \underline{E}(\cdot\rfloor O) & \text{if } \overline{E}(O\rfloor\Omega) = 0 \end{cases} \quad \text{for all } O \in \mathscr{P}_\emptyset(\Omega).$$

If $\underline{P}(f\rfloor\Omega)$ is defined for all $f \in \mathscr{G}(\Omega)$, then $\underline{E}(\cdot\rfloor\Omega)$ will be equal to $\underline{P}(\cdot\rfloor\Omega)$ and $\underline{E}^*(\cdot\rfloor\cdot)$ and $\underline{R}^*(\cdot\rfloor\cdot)$ can be taken to be the natural and regular extension of $\underline{P}(\cdot\rfloor\Omega)$, respectively. If $\underline{P}(\cdot\rfloor\Omega)$ is known—or if $\underline{E}(\cdot\rfloor\Omega)$ can be calculated—$\underline{R}^*(\cdot\rfloor O)$ can be obtained by applying the computational techniques of Section 3.3. The result is guaranteed to coincide with $\underline{R}(\cdot\rfloor O)$ whenever $\overline{P}(O\rfloor\Omega) > 0$—$\overline{E}(O\rfloor\Omega) > 0$—but will be vacuous and might therefore only be a (safe) lower approximation of $\underline{R}(\cdot\rfloor O)$ if $\overline{P}(O\rfloor\Omega) = 0$—$\overline{E}(O\rfloor\Omega) = 0$.

Natural and regular extension coincide whenever $\underline{E}(O\rfloor\Omega) > 0$ or $\overline{E}(O\rfloor\Omega) = 0$. If $\overline{E}(O\rfloor\Omega) > \underline{E}(O\rfloor\Omega) = 0$, regular extension may differ from natural extension and is guaranteed to dominate it. Eq. (15) provides the natural extension of a conditional lower prevision $\underline{P}(\cdot\rfloor\cdot)$ with an interpretation in terms of linear previsions: it is the lower envelop of the set $K_{\underline{P}(\cdot\rfloor\cdot)}$ consisting of the conditional lower previsions that dominate $\underline{P}(\cdot\rfloor\cdot)$. A similar interpretation can be given to the regular extension of $\underline{P}(\cdot\rfloor\cdot)$ as well: in case it differs from the natural extension—if $\overline{E}(O\rfloor\Omega) > \underline{E}(O\rfloor\Omega) = 0$—then the regular extension $\underline{R}(\cdot\rfloor O)$ of $\underline{P}(\cdot\rfloor\cdot)$ is equal to the regular extension $\underline{R}^*(\cdot\rfloor O)$ of $\underline{E}(\cdot\rfloor\Omega)$, and therefore given by

$$\begin{aligned} \underline{R}(f\rfloor O) &= \inf\{P(f\rfloor O) \colon P \in K_{\underline{E}(\cdot\rfloor\Omega)} \text{ and } P(O) > 0\} \\ &= \inf\{P(f\rfloor O) \colon P(\cdot\rfloor\cdot) \in K_{\underline{P}(\cdot\rfloor\cdot)} \text{ and } P(O\rfloor\Omega) > 0\} \quad \text{for all } f \in \mathscr{G}(O). \end{aligned}$$

Given that the only assumptions that were made to obtain $\underline{R}(\cdot\rfloor\cdot)$ are (i) that we are working within the setting that was described in Section 4.1 and (ii) that within the though experiment of constructing the updated model, the subject whose beliefs are being modelled is willing to adopt Assessment (27), we have finally found our justification for updating by means of regular extension. Whenever these two conditions are met, regular extension serves as a conservative updating strategy, for both conditional and unconditional coherent lower previsions.

## 6. Conclusions

Updating and conditioning are two very different matters. Conditioning is only a mathematical concept, expressed solely in terms of current beliefs. Updating, on the other hand, is concerned with how to change these beliefs, after being informed that some event has occurred. A claim that these two concepts should somehow be related to one another, let alone that they should coincide, is by no means trivial.

This paper has addressed this issue within an imprecise-probabilistic setting, in terms of sets of desirable gambles, lower previsions and sets of linear previsions/mass functions. If we are looking for an updating rule that is specified in advance, and furthermore assume that the occurrence of the events on which we want to update—and nothing more—will be reported to us if and only if they actually occur, then conditioning can be justified as a conservative updating strategy. If we condition by means of natural extension, no further assumptions are required. For regular extension, these conditions need to be supplemented by an additional one: provided that an event can occur, we should be willing to bet on the occurrence of that event at some arbitrarily small but nevertheless positive betting rate.

All of our main results have been obtained directly in terms of sets of desirable gambles. It is only afterwards that we have translated them to the framework of lower previsions. Nevertheless, from a more practical point of view, as the reader will no doubt have noticed, our results in terms of lower previsions seem to be the easiest to apply. The reason for this is that, whenever the event has positive lower probability, the corresponding conditional and/or updated lower prevision can be obtained directly from the unconditional one, using relatively simple computational techniques. For regular extension, it even suffices for the upper probability to be positive, as will be the case for most events of interest. This might seem trivial and well-known, but it is in fact not, because it also holds if we apply regular extension to a conditional lower prevision, or to a set of desirable gambles: if the upper probability of an event is positive and we are interested in the corresponding updated lower prevision—the lower prevision that corresponds to the updated set of desirable gambles—then all that is needed in order to compute it is the unconditional part of (the natural extension of) the lower prevision we started from—the closure of the original set of desirable gambles. Hence, for the purpose of updating by means of regular extension, and unlike what is the case for natural extension, the border structure of a set of desirable gambles—the very thing that makes it more general than other imprecise-probabilistic models—seems to be of little relevance, unless we want to update on an event with upper probability zero.

As far as future work is concerned, the main open problem is whether our results can be extended to infinite state spaces, and if yes, to which extent. At first sight, it seems like our justifications for updating by means of natural or regular extension should still hold. However, the conditions that need to be fulfilled in order for them to be applicable become hard to satisfy, and are perhaps even unrealistic.

Most importantly, the requirement that we should be informed about the occurrence of the event of interest—and nothing more—if and only if it actually occurs becomes problematic. If the state space is the euclidean plane, it is common to condition on and/or update after the observation of one of its coordinates, resulting in a model for the remaining coordinate. However, observing the exact value of a single coordinate is an idealisation of what we have really observed: some larger event that contains the line that corresponds to the coordinate. Unfortunately, we are never informed about the actual occurrence of this larger event, thereby making it impossible to apply our justifications for updating. Instead, this larger event is usually assumed to have some mathematically convenient shape, and a limit argument is employed to be able to condition on the idealised observation, ultimately leading to a conditioning rule for densities rather than probability measures. Along the way, an implicit assumption of countable additivity is usually added to the mix as well. This leads to mathematically convenient expressions, but it becomes hard to pinpoint what they mean exactly, let alone to justify such procedures as an updating strategy. All of this is already highly non-trivial in the precise-probabilistic case, and it is bound to become even harder in the imprecise-probabilistic case. See for example Ref. [38, Section 6.10] for some interesting examples and initial work on these topics.

If we were able to deal with these issues, the notion of regular extension could be extended to infinite state spaces in a way that differs from the approach that is taken by Walley [38, Appendix J]. The requirement that, under the assumption that an event can occur, the subject should be willing to bet on it at some possibly small but nevertheless positive betting rate, would only have to be imposed on the series of events of which the idealised event is the limit, rather than on the limit event itself. By taking the limit of the resulting updated models, we obtain a notion of regular extension that is more informative than natural extension, even if the limit event has upper probability zero, as long as the elements of the series of events of which it is the limit each have positive upper probability. In contrast, in order for Walley's notion of regular extension to be informative, the upper probability of the limit event needs to be positive.

In addition to these perhaps rather philosophical challenges, dealing with infinite spaces also leads to a host of technical issues that have to be dealt with. The theory of conditional lower previsions introduced in Section 2.2.1 is no longer equivalent with the alternatives that were discussed in Section 2.2.3, mostly due to issues with countable additivity and conglomerability. Depending on the theory that is adopted, coherence tends to be either too weak or too restrictive. Some of the proofs in this paper no longer work, and extending them to the infinite case seems hard, if not impossible. Nevertheless, being able to deal with the infinite case would be useful, and we consider it be an important as well as challenging line of future research.

**Acknowledgements**

Jasper De Bock is a PhD Fellow of the Research Foundation – Flanders (FWO) and wishes to acknowledge its financial support. Gert de Cooman's research was partially funded through project number 3G012512 of the FWO. The authors would like to thank the two anonymous referees for their constructive criticism and their many insightful comments and suggestions.

## AppendixA. Additional technical results

Besides the ones that were already discussed in Section 5.2.2, the operator $\mathscr{E}_O$ that was introduced in Section 5.2.1 has some additional nice properties as well. Since they do not fit nicely into the main discussion of the paper, we gather them in this appendix.

We start by investigating what happens if we apply $\mathscr{E}_O$ multiple, say $n \in \mathbb{N}$, times. Consider a sequence of events $O_i \in \mathscr{P}_\emptyset(\Omega)$, $i \in \{1,\dots,n\}$, and let us apply the corresponding sequence of operators $\mathscr{E}_{O_i}$, $i \in \{1,\dots,n\}$, to a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$, in that order, resulting in a set of gambles

$$\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) \coloneqq \mathscr{E}_{O_n}(\mathscr{E}_{O_{n-1}}(\cdots \mathscr{E}_{O_2}(\mathscr{E}_{O_1}(\mathscr{D})) \cdots)). \tag{A.1}$$

What does this set look like? And does it depend on the order in which the operators are applied? The following results provide an answer to these questions.

**Proposition 24.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a sequence of $n \in \mathbb{N}$ events $O_i \in \mathscr{P}_\emptyset(\Omega)$, $i \in \{1,\dots,n\}$. Then $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is coherent if and only if $\overline{P}_\mathscr{D}(O_i) > 0$ for all $i \in \{1,\dots,n\}$. If it is coherent, then $\mathrm{cl}(\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})) = \mathrm{cl}(\mathscr{D})$. If it is incoherent, then $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{G}(\Omega)$.*

*Proof.* We provide a proof by induction. For $n = 1$, the result follows trivially from Proposition 15 and Corollary 19. Consider now the case $n > 1$. Then by the induction hypothesis, we may assume that the result is true for $n-1$.

First, assume that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is coherent. Assume *ex absurdo* that $\mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})$ is incoherent. Then, by the induction hypothesis, $\mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D}) = \mathscr{G}(\Omega)$, and therefore

$$\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})) = \mathscr{E}_{O_n}(\mathscr{G}(\Omega)) = \mathscr{G}(\Omega)$$

is incoherent, a contradiction. Hence, we find that $\mathscr{D}' \coloneqq \mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})$ must be coherent. By the induction hypothesis, this implies that $\mathrm{cl}(\mathscr{D}') = \mathrm{cl}(\mathscr{D})$ and, for all $i \in \{1,\dots,n-1\}$, that $\overline{P}_\mathscr{D}(O_i) > 0$. Since $\mathscr{D}'$ and $\mathscr{D}$ are both coherent, we infer from $\mathrm{cl}(\mathscr{D}') = \mathrm{cl}(\mathscr{D})$ and Eq. (6) that $\underline{P}_{\mathscr{D}'} = \underline{P}_\mathscr{D}$ and therefore, by conjugacy, that $\overline{P}_{\mathscr{D}'}(O_n) = \overline{P}_\mathscr{D}(O_n)$. Since $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{D}')$, we infer from Proposition 15 and the coherence of $\mathscr{D}'$ and $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ that $\overline{P}_{\mathscr{D}'}(O_n) > 0$, which in turn implies that $\mathrm{cl}(\mathscr{E}_{O_n}(\mathscr{D}')) = \mathrm{cl}(\mathscr{D}')$ [Corollary 19]. In conclusion, we have found that $\mathrm{cl}(\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})) = \mathrm{cl}(\mathscr{E}_{O_n}(\mathscr{D}')) = \mathrm{cl}(\mathscr{D}') = \mathrm{cl}(\mathscr{D})$ and that, for all $i \in \{1,\dots,n\}$, $\overline{P}_\mathscr{D}(O_i) > 0$.

Next, assume that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is incoherent. We will prove that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{G}(\Omega)$ and that there is some $i \in \{1,\dots,n\}$ such that $\overline{P}_\mathscr{D}(O_i) = 0$. If $\mathscr{D}' \coloneqq \mathscr{E}_{O_1,\dots,O_{n-1}}(\mathscr{D})$ is incoherent, it follows from the induction hypothesis that (a) there is some $i \in \{1,\dots,n-1\}$ such that $\overline{P}_\mathscr{D}(O_i) = 0$ and (b) $\mathscr{D}' = \mathscr{G}(\Omega)$ and therefore also $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{D}') = \mathscr{G}(\Omega)$. Hence, without loss of generality, we can assume that $\mathscr{D}'$ is coherent. As shown in the first part of this proof, this implies that $\overline{P}_{\mathscr{D}'}(O_n) = \overline{P}_\mathscr{D}(O_n)$. Since $\mathscr{D}'$ is coherent and $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{E}_{O_n}(\mathscr{D}')$, we can now combine Proposition 15 with the fact that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$ is incoherent to infer that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{G}(\Omega)$ and $\overline{P}_{\mathscr{D}'}(O_n) = \overline{P}_\mathscr{D}(O_n) = 0$. □

**Proposition 25.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a sequence of $n \in \mathbb{N}$ events $O_i \in \mathscr{P}_\emptyset(\Omega)$, $i \in \{1,\dots,n\}$. Then*

$$\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \bigcup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D}).$$

*Proof.* Let us define $\mathscr{D}_0 \coloneqq \mathscr{D}$ and, for all $i \in \{1,\dots,n\}$, $\mathscr{D}_i \coloneqq \mathscr{E}_{O_i}(\mathscr{D}_{i-1}) = \mathscr{E}_{O_1\dots,O_i}(\mathscr{D})$. Since $\mathscr{D}_0 \subseteq \mathscr{D}_1 \subseteq \cdots \subseteq \mathscr{D}_{n-1} \subseteq \mathscr{D}_n$, we find that, for all $i \in \{1,\dots,n\}$,

$$\mathscr{E}_{O_i}(\mathscr{D}) = \mathscr{E}_{O_i}(\mathscr{D}_0) \subseteq \mathscr{E}_{O_i}(\mathscr{D}_{i-1}) = \mathscr{D}_i \subseteq \mathscr{D}_n = \mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}).$$

Hence, we are left to prove that $\mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) \subseteq \cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$. Fix any $f \in \mathscr{E}_{O_1,\dots,O_n}(\mathscr{D})$. We will prove that $f \in \cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$. By Proposition 15, we may assume without loss of generality that, for all $i \in \{1,\dots,n\}$, $\overline{P}_\mathscr{D}(O_i) > 0$, because otherwise, $\cup_{i=1}^{n} \mathscr{E}_{O_i}(\mathscr{D})$ would be equal to $\mathscr{G}(\Omega)$, which would make the proof trivial. Now let $i^*$ be the smallest $i \in \{0,\dots,n\}$ for which it holds that $f \in \mathscr{D}_i$ [since $f \in \mathscr{E}_{O_1,\dots,O_n}(\mathscr{D}) = \mathscr{D}_n$, $i^*$ always exists]. If $i^* = 0$, then $f \in \mathscr{D}$, which makes the proof trivial. Hence, without loss of generality, we may assume that $i^* \geq 1$. This allows us to consider the set $\mathscr{D}_{i^*-1}$, of which we know, by definition of $i^*$, that $f \notin \mathscr{D}_{i^*-1}$. Since $\overline{P}_{\mathscr{D}}(O_i) > 0$ for all $i \in \{1, \ldots, i^*-1\}$, we know that $\mathscr{D}_{i^*-1}$ is coherent and that $\mathrm{cl}(\mathscr{D}_{i^*-1}) = \mathrm{cl}(\mathscr{D})$ [this is trivial if $i^*-1 = 0$ and otherwise follows from Proposition 24 with $n = i^*-1$]. Since $\mathscr{D}_{i^*-1}$ is coherent and $\overline{P}_{\mathscr{D}}(O_{i^*}) > 0$, we can combine Theorem 18 with the fact that $f \in \mathscr{D}_{i^*} \setminus \mathscr{D}_{i^*-1} = \mathscr{E}_{O_{i^*}}(\mathscr{D}_{i^*-1}) \setminus \mathscr{D}_{i^*-1}$ to infer that

$$(\forall \varepsilon \in (0,1))(\exists \overline{\lambda} \in \mathbb{R}_{>0})(\forall \lambda \in (0, \overline{\lambda}])\ f + \lambda(\varepsilon - \mathbb{I}_{O_{i^*}}) \in \mathrm{int}(\mathscr{D}_{i^*-1}). \tag{A.2}$$

Since $\mathscr{D}$ and $\mathscr{D}_{i^*-1}$ are both coherent, $\mathrm{cl}(\mathscr{D}_{i^*-1}) = \mathrm{cl}(\mathscr{D})$ implies that $\mathrm{int}(\mathscr{D}_{i^*-1}) = \mathrm{int}(\mathscr{D})$ [Eq. (6)]. Therefore, since $\mathscr{D}$ is coherent and $\overline{P}_{\mathscr{D}}(O_{i^*}) > 0$, we can infer from Eq. (A.2) and Theorem 18 that $f \in \mathscr{E}_{O_{i^*}}(\mathscr{D})$ and therefore also, that $f \subseteq \cup_{i=1}^n \mathscr{E}_{O_i}(\mathscr{D})$. □

It follows from this last result that $\mathscr{E}_{O_1,\ldots,O_n}(\mathscr{D})$ is fully determined by the set of events $\mathscr{O} = \{O_i \colon i \in \{1, \ldots, n\}\}$, and we will therefore simply denote it by $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$. The order of the events $O_1, \ldots, O_n$ does not matter, neither does the fact that some of the events might appear multiple times. For this reason, from now on, we no longer consider sequences of events, but non-empty subsets $\mathscr{O}$ of $\mathscr{P}_{\emptyset}(\Omega)$.

For any non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$, and any coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$, we define

$$\mathscr{E}_{\mathscr{O}}(\mathscr{D}) \coloneqq \bigcup_{O \in \mathscr{O}} \mathscr{E}_O(\mathscr{D}).$$

By the results above, and since $\mathscr{P}_{\emptyset}(\Omega)$ and therefore also $\mathscr{O}$ is finite, the following properties are immediate.

**Corollary 26.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$. Then $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ is coherent if and only if $\overline{P}_{\mathscr{D}}(O) > 0$ for all $O \in \mathscr{O}$. If it is coherent, then $\mathrm{cl}(\mathscr{E}_{\mathscr{O}}(\mathscr{D})) = \mathrm{cl}(\mathscr{D})$. If it is incoherent, then $\mathscr{E}_{\mathscr{O}}(\mathscr{D}) = \mathscr{G}(\Omega)$.*

*Proof.* Trivial consequence of Propositions 24 and 25. □

The necessary and sufficient condition for $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ to be coherent—$\overline{P}_{\mathscr{D}}(O) > 0$ for all $O \in \mathscr{O}$—that is provided in the result above simplifies if $\mathscr{O} = \mathscr{P}_{\emptyset}(\Omega)$. In that case, because $\mathscr{D}$ and therefore also $\underline{P}_{\mathscr{D}}$ is coherent, this condition is satisfied if and only if $\overline{P}_{\mathscr{D}}(\omega) > 0$ for all $\omega \in \Omega$.

It is also possible to characterise $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ differently, in a way that closely resembles our definition for $\mathscr{E}_O(\mathscr{D})$.

**Theorem 27.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$. For any $\varepsilon \in (0,1)^{\mathscr{O}}$,[35] let $\mathscr{E}_{\mathscr{O}}^{\varepsilon}(\mathscr{D}) \coloneqq \mathscr{E}(\mathscr{D} \cup \{\mathbb{I}_O - \varepsilon_O \colon O \in \mathscr{O}\})$. Then*

$$\mathscr{E}_{\mathscr{O}}(\mathscr{D}) = \bigcap_{\varepsilon \in (0,1)^{\mathscr{O}}} \mathscr{E}_{\mathscr{O}}^{\varepsilon}(\mathscr{D}).$$

*Proof of Theorem 27.* For the sake of notational convenience, let us denote the right-hand side of the desired equality by $\mathscr{E}_{\mathscr{O}}^*(\mathscr{D})$. We only show that $\mathscr{E}_{\mathscr{O}}^*(\mathscr{D}) \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. The converse inclusion holds trivially because, for all $O \in \mathscr{O}$, $\mathscr{E}_O(\mathscr{D})$ is clearly a subset of $\mathscr{E}_{\mathscr{O}}^*(\mathscr{D})$.

Consider any $f \in \mathscr{E}_{\mathscr{O}}^*(\mathscr{D})$ and assume *ex absurdo* that $f \notin \mathscr{E}_{\mathscr{O}}(\mathscr{D}) = \bigcup_{O \in \mathscr{O}} \mathscr{E}_O(\mathscr{D})$, which already implies that $f \notin \mathscr{D}$. Then for all $O \in \mathscr{O}$, $f \notin \mathscr{E}_O(\mathscr{D})$ and therefore, there is some $\varepsilon_O \in (0,1)$ such that $f \notin \mathscr{E}_O^{\varepsilon_O}(\mathscr{D})$. Let $\varepsilon_{\min} \coloneqq \min_{O \in \mathscr{O}} \varepsilon_O$ [this minimum is well-defined because $\Omega$ and therefore also $\mathscr{P}_{\emptyset}(\Omega)$ and $\mathscr{O}$ are finite] and let $\varepsilon' \in (0,1)^{\mathscr{O}}$ be defined, for all $O \in \mathscr{O}$, by $\varepsilon'_O \coloneqq \varepsilon_{\min} |\mathscr{O}|^{-1}$. Since we know that $f \in \mathscr{E}_{\mathscr{O}}^*(\mathscr{D}) \setminus \mathscr{D} \subseteq \mathscr{E}_{\mathscr{O}}^{\varepsilon'}(\mathscr{D}) \setminus \mathscr{D}$, we can infer from the definition of $\mathscr{E}_{\mathscr{O}}^{\varepsilon'}(\mathscr{D})$ and the coherence of $\mathscr{D}$ that there are $n \in \mathbb{N}$ such that $n \leq |\mathscr{O}|$, $g \in \mathscr{D} \cup \{0\}$, $(\forall i \in \{1, \ldots, n\})\ \lambda_i \in \mathbb{R}_{>0}$ and $O_i \in \mathscr{O}$ such that

$$f = g + \sum_{i=1}^n \lambda_i (\mathbb{I}_{O_i} - \varepsilon'_{O_i}) = g + \sum_{i=1}^n \lambda_i \mathbb{I}_{O_i} - \frac{\varepsilon_{\min}}{|\mathscr{O}|} \sum_{i=1}^n \lambda_i. \tag{A.3}$$

[35] An element $\varepsilon$ of $(0,1)^{\mathscr{O}}$ is a map from $\mathscr{O}$ to $(0,1)$. For any $O \in \mathscr{O}$, the corresponding value is an element of $(0,1)$ and will be denoted by $\varepsilon_O$.

If we now define $\lambda_{\max} := \max\{\lambda_i \colon 1 \le i \le n\}$ and let $i_{\max} \in \arg\max\{\lambda_i \colon 1 \le i \le n\}$, then

$$\sum_{i=1}^{n} \lambda_i \mathbb{I}_{O_i} - \frac{\varepsilon_{\min}}{|\mathscr{O}|} \sum_{i=1}^{n} \lambda_i \ge \lambda_{\max} \mathbb{I}_{O_{i_{\max}}} - \frac{\varepsilon_{\min}}{|\mathscr{O}|} n \lambda_{\max} \ge \lambda_{\max} \mathbb{I}_{O_{i_{\max}}} - \varepsilon_{\min} \lambda_{\max} \ge \lambda_{\max} (\mathbb{I}_{O_{i_{\max}}} - \varepsilon_{O_{i_{\max}}})$$

and therefore, by Eq. (A.3), $f \ge g + \lambda_{\max}(\mathbb{I}_{O_{i_{\max}}} - \varepsilon_{O_{i_{\max}}})$, which implies that $f \in \mathscr{E}_{O_{i_{\max}}}^{\varepsilon_{O_{i_{\max}}}}(\mathscr{D})$. This is a contradiction, allowing us to conclude that $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. Since this holds for all $f \in \mathscr{E}_{\mathscr{O}}^{*}(\mathscr{D})$, we find that $\mathscr{E}_{\mathscr{O}}^{*}(\mathscr{D}) \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. □

Again, as was the case for $\mathscr{E}_{O}(\mathscr{D})$, $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ is closely connected with regular extension. For every event $O \in \mathscr{O}$, $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O)$ is equal to the regular extension $\underline{R}(\cdot \rfloor O)$ of $\underline{P}_{\mathscr{D}}$.

**Proposition 28.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$ such that, for all $O \in \mathscr{O}$, $\overline{P}_{\mathscr{D}}(O) > 0$. Let $\underline{R}(\cdot \rfloor \cdot)$ be the regular extension of $\underline{P}_{\mathscr{D}}$, as defined by Eq.* (23). *Then $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})} = \underline{P}_{\mathscr{D}}$,*

$$\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O) = \underline{R}(\cdot \rfloor O) \text{ for all } O \in \mathscr{O}$$

*and*

$$\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O) = \underline{P}_{\mathscr{D}}(\cdot \rfloor O) \text{ for all } O \in \mathscr{P}_{\emptyset}(\Omega) \text{ such that } \overline{P}_{\mathscr{D}}(O) = 0.$$

*Proof.* We know from Corollary 26 that $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ is coherent, and that $\operatorname{cl}(\mathscr{E}_{\mathscr{O}}(\mathscr{D})) = \operatorname{cl}(\mathscr{D})$. Therefore, by Eq. (6), $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})} = \underline{P}_{\mathscr{D}}$.

Fix $O \in \mathscr{O}$. The coherence of $\mathscr{E}_{\mathscr{O}}(\mathscr{D})$ implies that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O)$ is coherent with $\underline{P}_{\mathscr{E}_{O}(\mathscr{D})}$ and therefore also with $\underline{P}_{\mathscr{D}}$. Since $\overline{P}_{\mathscr{D}}(O) > 0$, this implies that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O) \le \underline{R}(\cdot \rfloor O)$ [see Section 3.2]. By Corollary 21, we know that $\underline{P}_{\mathscr{E}_{O}(\mathscr{D})}(\cdot \rfloor O) = \underline{R}(\cdot \rfloor O)$. Furthermore, $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O) \ge \underline{P}_{\mathscr{E}_{O}(\mathscr{D})}(\cdot \rfloor O)$ because $\mathscr{E}_{O}(\mathscr{D}) \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$. Hence, putting it all together, we find that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O) = \underline{R}(\cdot \rfloor O)$.

Consider any $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) = 0$. By Lemma 29, we know that $\mathscr{E}_{\mathscr{O}}(\mathscr{D}) \rfloor O = \mathscr{D} \rfloor O$, which implies that $\underline{P}_{\mathscr{E}_{\mathscr{O}}(\mathscr{D})}(\cdot \rfloor O) = \underline{P}_{\mathscr{D}}(\cdot \rfloor O)$. □

**Lemma 29.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$ such that, for all $O \in \mathscr{O}$, $\overline{P}_{\mathscr{D}}(O) > 0$. Then*

$$\mathscr{E}_{\mathscr{O}}(\mathscr{D}) \rfloor O = \mathscr{D} \rfloor O \text{ for all } O \in \mathscr{P}_{\emptyset}(\Omega) \text{ such that } \overline{P}_{\mathscr{D}}(O) = 0.$$

*Proof.* Fix $O \in \mathscr{P}_{\emptyset}(\Omega)$ such that $\overline{P}_{\mathscr{D}}(O) = 0$. We only prove that $\mathscr{E}_{\mathscr{O}}(\mathscr{D}) \rfloor O \subseteq \mathscr{D} \rfloor O$, because the other inclusion follows trivially from the fact that $\mathscr{D} \subseteq \mathscr{E}_{\mathscr{O}}(\mathscr{D})$.

Consider any $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \rfloor O$ and assume *ex absurdo* that $f \notin \mathscr{D} \rfloor O$. Then $\mathbb{I}_O f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$, which implies that there is some $O^* \in \mathscr{O}$ such that $\mathbb{I}_O f \in \mathscr{E}_{O^*}(\mathscr{D}) \setminus \mathscr{D}$. Using Lemma 30, we find that $\overline{P}_{\mathscr{D}}(\mathbb{I}_O f) > 0$. Choose $\lambda \in \mathbb{R}_{>0}$ such that $\lambda \ge \max f$. Then by the coherence of $\underline{P}_{\mathscr{D}}$, we have that $\overline{P}_{\mathscr{D}}(\mathbb{I}_O f) \le \overline{P}_{\mathscr{D}}(\mathbb{I}_O \lambda) = \lambda \overline{P}_{\mathscr{D}}(O) = 0$. This is a contradiction, allowing us to infer that $f \in \mathscr{D} \rfloor O$. Since this holds for all $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \rfloor O$, we find that $\mathscr{E}_{\mathscr{O}}(\mathscr{D}) \rfloor O \subseteq \mathscr{D} \rfloor O$. □

**Lemma 30.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and a non-empty set of events $\mathscr{O} \subseteq \mathscr{P}_{\emptyset}(\Omega)$ such that, for all $O \in \mathscr{O}$, $\overline{P}_{\mathscr{D}}(O) > 0$. Then for all $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$: $0 = \underline{P}_{\mathscr{D}}(f) < \overline{P}_{\mathscr{D}}(f)$.*

*Proof.* Consider any $f \in \mathscr{E}_{\mathscr{O}}(\mathscr{D}) \setminus \mathscr{D}$, which implies that there is some $O \in \mathscr{O}$ such that $f \in \mathscr{E}_{O}(\mathscr{D}) \setminus \mathscr{D}$. Choose $\varepsilon \in (0,1)$ such that $\varepsilon < \overline{P}_{\mathscr{D}}(O)$. Since $\overline{P}_{\mathscr{D}}(O) > 0$, we know from Theorem 18 that there is some $\lambda \in \mathbb{R}_{>0}$ such that $f + \lambda(\varepsilon - \mathbb{I}_O) \in \mathscr{D}$, which implies that $\underline{P}_{\mathscr{D}}(f + \lambda[\varepsilon - \mathbb{I}_O]) \ge 0$. Using the coherence of $\underline{P}_{\mathscr{D}}$ [C2, C5 and C8], this allows us to infer that

$$\begin{aligned}\overline{P}_{\mathscr{D}}(f) = \overline{P}_{\mathscr{D}}(\lambda[\mathbb{I}_O - \varepsilon] + f + \lambda[\varepsilon - \mathbb{I}_O]) &\ge \overline{P}_{\mathscr{D}}(\lambda[\mathbb{I}_O - \varepsilon]) + \underline{P}_{\mathscr{D}}(f + \lambda[\varepsilon - \mathbb{I}_O]) \\ &\ge \overline{P}_{\mathscr{D}}(\lambda[\mathbb{I}_O - \varepsilon]) = \lambda(\overline{P}_{\mathscr{D}}(O) - \varepsilon) > 0.\end{aligned}$$

Since $f \in \mathscr{E}_{O}(\mathscr{D})$, we know that $\underline{P}_{\mathscr{E}_{O}(\mathscr{D})}(f) \ge 0$ and therefore, by Corollary 19, that $\underline{P}_{\mathscr{D}}(f) \ge 0$. Finally, the fact that $f \notin \mathscr{D}$ implies that, for all $\alpha \ge 0$, $f - \alpha \notin \mathscr{D}$, which in turn implies that $\underline{P}_{\mathscr{D}}(f) \le 0$. □

If the set $\mathscr{O}$ is taken to be equal to $\mathscr{O}_1(\mathscr{D}) \coloneqq \{O \in \mathscr{P}_\emptyset(\Omega) \colon \overline{P}_\mathscr{D}(O) > \underline{P}_\mathscr{D}(O) = 0\}$ or $\mathscr{O}_2(\mathscr{D}) \coloneqq \{O \in \mathscr{P}_\emptyset(\Omega) \colon \overline{P}_\mathscr{D}(O) > 0\}$, or any set in between these two, we obtain an important special instance of the operator $\mathscr{E}_\mathscr{O}$, which we denote by $\mathscr{E}^{\mathrm{r}}(\mathscr{D})$. For any coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$, it is defined by

$$\mathscr{E}^{\mathrm{r}}(\mathscr{D}) \coloneqq \bigcup_{\substack{O \in \mathscr{P}_\emptyset(\Omega) \\ \overline{P}_\mathscr{D}(O) > \underline{P}_\mathscr{D}(O) = 0}} \mathscr{E}_O(\mathscr{D}) = \bigcup_{\substack{O \in \mathscr{P}_\emptyset(\Omega) \\ \overline{P}_\mathscr{D}(O) > 0}} \mathscr{E}_O(\mathscr{D}).$$

The last equality is a consequence of Proposition 17, which is also the reason why $\mathscr{E}^{\mathrm{r}}(\mathscr{D}) = \mathscr{E}_\mathscr{O}(\mathscr{D})$ for all $\mathscr{O}_1(\mathscr{D}) \subseteq \mathscr{O} \subseteq \mathscr{O}_2(\mathscr{D})$.

**Corollary 31.** *Consider a coherent set of desirable gambles $\mathscr{D} \subseteq \mathscr{G}(\Omega)$ and let $\underline{R}(\cdot \rfloor \cdot)$ be the regular extension of $\underline{P}_\mathscr{D}$. Then $\mathscr{E}^{\mathrm{r}}(\mathscr{D})$ is coherent, $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})} = \underline{P}_\mathscr{D}$ and*

$$\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D})}(f \rfloor O) = \begin{cases} \underline{R}(f \rfloor O) & \text{if } \overline{P}_\mathscr{D}(O) > 0 \\ \underline{P}_\mathscr{D}(f \rfloor O) & \text{if } \overline{P}_\mathscr{D}(O) = 0 \end{cases} \quad \text{for all } (f, O) \in \mathscr{C}(\Omega).$$

*Proof.* Since $\mathscr{E}^{\mathrm{r}}(\mathscr{D}) = \mathscr{E}_{\mathscr{O}_2(\mathscr{D})}(\mathscr{D})$, this follows immediately from Corollary 26 and Proposition 28. □

If $\mathscr{D}$ is the set of desirable gambles $\mathscr{D}_{\underline{P}}$ that corresponds to a coherent lower prevision $\underline{P}$, then $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})$ is a subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$ and—as is the case for $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$—its associated conditional lower prevision $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})}(\cdot \rfloor \cdot)$ is equal to the regular extension of $\underline{P}$.

**Corollary 32.** *Consider a coherent lower prevision $\underline{P}$ on $\mathscr{G}(\Omega)$ and let $\underline{R}(\cdot \rfloor \cdot)$ be its regular extension. Then $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})$ is a coherent subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$, $\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})} = \underline{P}$ and*

$$\underline{P}_{\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})}(f \rfloor O) = \underline{R}(f \rfloor O) \quad \text{for all } (f, O) \in \mathscr{C}(\Omega).$$

*Proof.* Since $\mathscr{D}_{\underline{P}}$ is coherent and $\underline{P}_{\mathscr{D}_{\underline{P}}} = \underline{P}$, we infer from Corollary 31 that it suffices to prove that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) \subseteq \mathscr{D}^{\mathrm{r}}_{\underline{P}}$ and that, for all $(f, O) \in \mathscr{C}(\Omega)$ such that $\overline{P}(O) = 0$, $\underline{P}_{\mathscr{D}_{\underline{P}}}(f \rfloor O) = \underline{R}(f \rfloor O)$. The inclusion follows from Lemma 30 and the fact that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) = \mathscr{E}_{\mathscr{O}_2(\mathscr{D}_{\underline{P}})}(\mathscr{D}_{\underline{P}})$. So consider any $(f, O) \in \mathscr{C}(\Omega)$ such that $\overline{P}(O) = 0$. Let $\underline{E}(\cdot \rfloor \cdot)$ be the natural extension of $\underline{P}$. Then as explained in Section 3.4, $\underline{P}_{\mathscr{D}_{\underline{P}}}(f \rfloor O) = \underline{E}(f \rfloor O)$. Hence, since natural and regular extension coincide if the conditioning event has upper probability zero, $\underline{P}_{\mathscr{D}_{\underline{P}}}(f \rfloor O) = \underline{R}(f \rfloor O)$. □

The following example illustrates that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}})$ can be a strict subset of $\mathscr{D}^{\mathrm{r}}_{\underline{P}}$.

**Example 2.** Let $\Omega = \{a, b\}$ and consider the lower prevision $\underline{P}$ that is defined by

$$\underline{P}(f) \coloneqq \min\{f(a), \frac{f(a) + f(b)}{2}\} \text{ for all } f \in \mathscr{G}(\Omega).$$

Then

$$\begin{aligned} \mathscr{D}_{\underline{P}} &= \{f \in \mathscr{G}(\Omega) \colon f(a) > 0 \text{ and } f(a) + f(b) > 0\} \cup \mathscr{G}(\Omega)_{>0} \\ &= \{f \in \mathscr{G}(\Omega) \colon f(a) \geq 0 \text{ and } f(a) + f(b) > 0\} \end{aligned} \tag{A.4}$$

and

$$\mathscr{D}^{\mathrm{r}}_{\underline{P}} = \mathscr{D}_{\underline{P}} \cup \{f \in \mathscr{G}(\Omega) \colon -f(b) = f(a) > 0\} = \{f \in \mathscr{G}(\Omega) \colon f(a) \geq 0 \text{ and } f(a) + f(b) \geq 0\} \setminus \{0\}.$$

Since $\underline{P}(\Omega) = \overline{P}(\Omega) = 1$, $1/2 = \underline{P}(\{a\}) < \overline{P}(\{a\}) = 1$, $0 = \underline{P}(\{b\}) < \overline{P}(\{b\}) = 1/2$ and $\underline{P}_{\mathscr{D}_{\underline{P}}} = \underline{P}$, we know that $\mathscr{O}_1(\mathscr{D}_{\underline{P}})$ consists of the singleton $\{b\}$, which implies that $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) = \mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}})$. Assume *ex absurdo* that $\mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) \setminus \mathscr{D}_{\underline{P}} \neq \emptyset$. Consider any $f \in \mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) \setminus \mathscr{D}_{\underline{P}}$. Then by Theorem 18, and since $\overline{P}_{\mathscr{D}_{\underline{P}}}(\{b\}) > 0$, there is some $\overline{\lambda} \in \mathbb{R}_{>0}$ such that, for all $\lambda \in (0, \overline{\lambda}]$, $f + \lambda(1/2 - \mathbb{I}_{\{b\}}) \in \mathscr{D}_{\underline{P}}$, which implies, by Eq. (A.4), that $f(a) + \lambda 1/2 \geq 0$ and $f(a) + f(b) > 0$. Since this holds for all $\lambda \in (0, \overline{\lambda}]$, we infer that $f(a) \geq 0$ and $f(a) + f(b) > 0$, or equivalently, that $f \in \mathscr{D}_{\underline{P}}$. This is a contradiction, allowing us to infer that $\mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) \setminus \mathscr{D}_{\underline{P}} = \emptyset$ and therefore, by Theorem 18, and since $\overline{P}_{\mathscr{D}_{\underline{P}}}(\{b\}) > 0$, that $\mathscr{E}_{\{b\}}(\mathscr{D}_{\underline{P}}) = \mathscr{D}_{\underline{P}}$ and therefore also $\mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) = \mathscr{D}_{\underline{P}}$. Hence, in conclusion: $\mathscr{D}_{\underline{P}} = \mathscr{E}^{\mathrm{r}}(\mathscr{D}_{\underline{P}}) \subset \mathscr{D}^{\mathrm{r}}_{\underline{P}}$. ◊

[1] Richard Bradley. The kinematics of belief and desire. *Synthese*, 156(3):513–535, 2007.

[2] Seamus Bradley and Katie Steele. Uncertainty, learning, and the “problem” of dilation. *Erkenntnis*, pages 1–17, 2013.

[3] Marco Cattaneo. A continuous updating rule for imprecise probabilities. In Anne Laurent, Oliver Strauss, Bernadette Bouchon-Meunier, and Ronald R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, volume 444 of *Communications in Computer and Information Science*, pages 426–435. Springer International Publishing, 2014.

[4] Inés Couso and Serafín Moral. Sets of desirable gambles: conditioning, representation, and precise probabilities. *International Journal of Approximate Reasoning*, 52(7):1034–1055, 2011.

[5] Fabio G. Cozman and Cassio Polpo de Campos. Kuznetsov independence for interval-valued expectations and sets of probability distributions: Properties and algorithms. *International Journal of Approximate Reasoning*, 55(2):666–682, 2014.

[6] Fabio G. Cozman and Teddy Seidenfeld. Independence for full conditional measures and their graphoid properties. In *Foundations of the Formal Sciences VI, Reasoning about Probabilities and Probabilistic Reasoning*, pages 1–29. College Publications, London, 2009.

[7] Jasper De Bock and Gert de Cooman. Extreme lower previsions. *Journal of Mathematical Analysis and Applications*, 421(2):1042–1080, January 2015.

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

[9] Gert de Cooman, Filip Hermans, Alessandro Antonucci, and Marco Zaffalon. Epistemic irrelevance in credal nets: the case of imprecise Markov trees. *International Journal of Approximate Reasoning*, 51(9):1029–1052, 2010.

[10] Gert de Cooman and Erik 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.

[11] Bruno de Finetti. *Teoria delle Probabilità*. Einaudi, Turin, 1970.

[12] Sébastien Destercke and Didier Dubois. Other uncertainty theories based on capacities. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

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

[14] Peter Gärdenfors. *Knowledge in Flux – Modeling the Dynamics of Epistemic States*. The MIT Press, Cambridge, MA, 1988.

[15] Itzhak Gilboa and David Schmeidler. Updating ambiguous beliefs. In Y. Moses, editor, *Theoretical Aspects of Reasoning About Knowledge*, pages 143–162. Morgan and Kaufmann, San Mateo, CA, 1992.

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

[17] Elmar Kriegler. Updating under unknown unknowns: An extension of Bayes' rule. *International Journal of Approximate Reasoning*, 50(4):583 – 596, 2009.

[18] Isaac Levi. *The Enterprise of Knowledge*. MIT Press, London, 1980.

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

[20] Enrique Miranda and Gert de Cooman. Lower previsions. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[21] Enrique Miranda and Marco Zaffalon. Notes on desirability and conditional lower previsions. *Annals of Mathematics and Artificial Intelligence*, 60:251–309, 2010.

[22] Serafín Moral. Epistemic irrelevance on sets of desirable gambles. *Annals of Mathematics and Artificial Intelligence*, 45:197–214, 2005.

[23] Serafín Moral and Nic Wilson. Revision rules for convex sets of probabilities. In Giulianella Coletti, Didier Dubois, and Romano Scozzafava, editors, *Mathematical Models for Handling Partial Knowledge in Artificial Intelligence*, pages 113–128. Plenum Press, New York, 1995.

[24] Renato Pelessoni and Paulo Vicig. Williams coherence and beyond. *International Journal of Approximate Reasoning*, 50:612–6276, 2009.

[25] Erik Quaeghebeur. The CONEstrip algorithm. In *Proceedings of SMPS 2012*, 2012.

[26] Erik Quaeghebeur. Desirability. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[27] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[28] Glenn Shafer. Bayes's two arguments for the Rule of Conditioning. *The Annals of Statistics*, 10:1075–1089, 1982.

[29] Glenn Shafer. Conditional probability. *International Statistical Review*, 53:261–277, 1985.

[30] Glenn Shafer, Peter R. Gillett, and Richard B. Scherl. A new understanding of subjective probability and its generalization to lower and upper prevision. *International Journal of Approximate Reasoning*, 33:1–49, 2003.

[31] Matthias C. M. Troffaes. Conditional lower previsions for unbounded random quantities. In Jonathan Lawry, Enrique Miranda, Alberto Bugarin, Shoumei Li, Mariá Ángeles Gil, Przemyslaw Grzegorzewski, and Olgierd Hryniewicz, editors, *Soft Methods in Probability for Integrated Uncertainty Modelling, Advances in Soft Computing*, pages 201–209. Springer, 2006.

[32] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley, 2014.

[33] Matthias C. M. Troffaes and Michael Goldstein. A note on the temporal sure preference principle and the updating of lower previsions. In F. Cozman, T. Denœux, S. Destercke, and T. Seidenfeld, editors, *ISIPTA'13: Proceedings of the Eighth International Symposium on Imprecise Probability: Theories and Applications*, pages 319–328, Compiègne, 2013. SIPTA.

[34] Matthias C. M. Troffaes and Robert Hable. Computation. In Thomas Augustin, Frank P. A. Coolen, Gert de Cooman, and Matthias C. M. Troffaes, editors, *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[35] Paolo Vicig, Marco Zaffalon, and Fabio G. Cozman. Notes on “Notes on conditional previsions”. *International Journal of Approximate Reasoning*, 44(3):358 – 365, 2007. Reasoning with Imprecise Probabilities.

[36] Frans Voorbraak. Probabilistic belief expansion and conditioning. Technical report, 1996.

[37] Damjan Škulj and Robert Hable. Coefficients of ergodicity for Markov chains with uncertain parameters. *Metrika*, 76:107–133, 2013.

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

[39] Peter Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83(1):1–58, 1996.

[40] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24:125–148, 2000.

[41] Peter Walley and Terrence L. Fine. Towards a frequentist theory of upper and lower probability. *Annals of Statistics*, 10:741–761, 1982.

[42] Peter Walley, Renato Pelessoni, and Paolo Vicig. Direct algorithms for checking consistency and making inferences from conditional probability assessments. *Journal of Statistical Planning and Inference*, 126:119–151, 2004.

[43] Peter M. Williams. Notes on conditional previsions. Technical report, School of Mathematical and Physical Science, University of Sussex, UK, 1975. Revised journal version: [44].

[44] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007.

[45] Nic Wilson and Serafín Moral. A logical view of probability. In *Proceedings of ECAI 1994*, pages 386–390, 1994.

[46] Marco Zaffalon and Enrique Miranda. Probability and time. *Artificial Intelligence*, 198(0):1–51, 2013.