Imprecise probability trees:
Bridging two theories of imprecise probability

Gert de Cooman [*], Filip Hermans

Ghent University, SYSTeMS Research Group, Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium

Abstract

We give an overview of two approaches to probability theory where lower and upper probabilities, rather than probabilities, are used: Walley's behavioural theory of imprecise probabilities, and Shafer and Vovk's game-theoretic account of probability. We show that the two theories are more closely related than would be suspected at first sight, and we establish a correspondence between them that (i) has an interesting interpretation, and (ii) allows us to freely import results from one theory into the other. Our approach leads to an account of probability trees and random processes in the framework of Walley's theory. We indicate how our results can be used to reduce the computational complexity of dealing with imprecision in probability trees, and we prove an interesting and quite general version of the weak law of large numbers.
© 2008 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, we have witnessed the growth of a number of theories of uncertainty, where imprecise (lower and upper) probabilities and previsions, rather than precise (or point-valued) probabilities and previsions, have a central part. Here we consider two of them, Glenn Shafer and Vladimir Vovk's game-theoretic account of probability [29], which is introduced in Section 2, and Peter Walley's behavioural theory [33], outlined in Section 3. These seem to have a rather different interpretation, and they certainly have been influenced by different schools of thought: Walley follows the tradition of Frank Ramsey [22], Bruno de Finetti [11] and Peter Williams [39] in trying to establish a rational model for a subject's beliefs in terms of her behaviour. Shafer and Vovk follow an approach that has many other influences as well, and is strongly coloured by ideas about gambling systems and martingales. They use Cournot's Principle to interpret lower and upper probabilities (see [28]; and [29, Chapter 2] for a nice historical overview), whereas on Walley's approach, lower and upper probabilities are defined in terms of a subject's betting rates.

What we set out to do here, [1] and in particular in Sections 4 and 5, is to show that in many practical situations, the two approaches are strongly connected.[2] This implies that quite a few results, valid in one theory, can automatically be converted and reinterpreted in terms of the other. Moreover, we shall see that we can develop an account of coherent immediate prediction in the context of Walley's behavioural theory, and prove, in Section 6, a weak law of large numbers with an intuitively appealing interpretation. We use this weak law in Section 7 to suggest a way of scoring a predictive model that satisfies A. Philip Dawid's *Prequential Principle* [5,6].

Why do we believe these results to be important, or even relevant, to AI? Probabilistic models are intended to represent an agent's beliefs about the world he is operating in, and which describe and even determine the actions he will take in a diversity of situations. Probability theory provides a normative system for reasoning and making decisions in the face of uncertainty. Bayesian, or precise, probability models have the property that they are completely decisive: a Bayesian agent always has an optimal choice when faced with a number of alternatives, whatever his state of information. While many may view this as an advantage, it is not always realistic. Imprecise probability models try to deal with this problem by explicitly allowing for indecision, while retaining the normative, or coherentist stance of the Bayesian approach. We refer to [8,33,34] for discussions about how this can be done.

Imprecise probability models appear in a number of AI-related fields. For instance in *probabilistic logic*: it was already known to George Boole [1] that the result of probabilistic inferences may be a set of probabilities (an imprecise probability model), rather than a single probability. This is also important for dealing with missing or incomplete data, leading to so-called partial identification of probabilities, see for instance [9,19]. There is also a growing literature on so-called *credal nets* [3,4]: these are essentially Bayesian nets with imprecise conditional probabilities.

We are convinced that it is mainly the mathematical and computational complexity often associated with imprecise probability models that is keeping them from becoming a more widely used tool for modelling uncertainty. But we believe that the results reported here can help make inroads in reducing this complexity. Indeed, the upshot of our being able to connect Walley's approach with Shafer and Vovk's, is twofold. First of all, we can develop a theory of *imprecise probability trees*: probability trees where the transition from a node to its children is described by an imprecise probability model in Walley's sense. Our results provide the necessary apparatus for making inferences in such trees. And because probability trees are so closely related to random processes, this effectively brings us into a position to start developing a theory of (event-driven) random processes where the uncertainty can be described using imprecise probability models. We illustrate this in Examples 1 and 3, and in Section 8.

Secondly, we are able to prove so-called Marginal Extension results (Theorems 3 and 7, Proposition 9), which lead to backwards recursion, and dynamic programming-like methods that allow for an exponential reduction in the computational complexity of making inferences in such imprecise probability trees. This is also illustrated in Examples 3 and Section 8. For (precise) probability trees, similar techniques were described in Shafer's book on causal reasoning [26]. They seem to go back to Christiaan Huygens, who drew the first probability tree, and showed how to reason with it, in his solution to Pascal and Fermat's Problem of Points.[3]

2. Shafer and Vovk's game-theoretic approach to probability

In their game-theoretic approach to probability [29], Shafer and Vovk consider games with two typical players, Reality and Sceptic, who play according to certain *protocols*. They obtain the most interesting results for what they call *coherent probability protocols*. This section is devoted to explaining what this means: what are the roles of Reality and Sceptic, how and by what rules do they play, and how can their game-play be related to probabilistic reasoning?

2.1. Reality's event tree

We begin with a first and basic assumption, dealing with how the first player, Reality, plays.

---

[1] An earlier and condensed version of this paper, with much less discussion and without proofs, was presented at the ISIPTA '07 Conference [7].

[2] Our line of reasoning here should be contrasted with the one in [28], where Shafer et al. use the game-theoretic framework developed in [29] to construct a theory of predictive upper and lower previsions whose interpretation is based on Cournot's Principle. See also the comments near the end of Section 5.

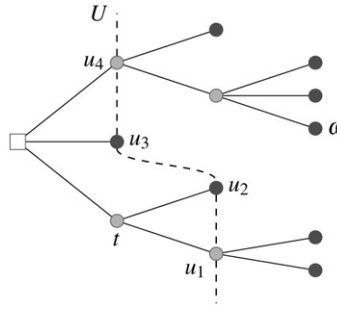[3] See Section 8 for more details and precise references.

Fig. 1. A simple event tree for Reality, displaying the initial situation $\Box$, other non-terminal situations (such as t) as grey circles, and paths, or terminal situations, (such as ω) as black circles. Also depicted is a cut $U = \{u_1, u_2, u_3, u_4\}$ of $\Box$. Observe that t (strictly) precedes u_1: $t \sqsubset u_1$, and that $C(t) = \{u_1, u_2\}$ is the children cut of t.

G1. Reality makes a number of moves, where the possible next moves may depend on the previous moves he has made, but do not in any way depend on the previous moves made by Sceptic.

This means that we can represent his game-play by an event tree (see also [26,27] for more information about event trees). We restrict ourselves here to the discussion of *bounded protocols*, where Reality makes only a finite and bounded number of moves from the beginning to the end of the game, whatever happens. But we don't exclude the possibility that at some point in the tree, Reality has the choice between an infinite number of next moves. We shall come back to these assumptions further on, once we have the appropriate notational tools to make them more explicit.[4]

Let us establish some terminology related to Reality's event tree.

2.1.1. Paths, situations and events

A *path* in the tree represents a possible sequence of moves for Reality from the beginning to the end of the game. We denote the set of all possible paths ω by Ω, the *sample space* of the game.

A *situation* t is some connected segment of a path that is *initial*, i.e., starts at the root of the tree. It identifies the moves Reality has made up to a certain point, and it can be identified with a node in the tree. We denote the set of all situations by $\Omega^\diamond$. It includes the set Ω of *terminal* situations, which can be identified with paths. All other situations are called *non-terminal*; among them is the *initial* situation $\Box$, which represents the empty initial segment. See Fig. 1 for a simple graphical example explaining these notions.

If for two situations s and t, s is a(n initial) segment of t, then we say that s *precedes* t or that t *follows* s, and write $s \sqsubseteq t$, or alternatively $t \sqsupseteq s$. If ω is a path and $t \sqsubseteq \omega$ then we say that the path ω *goes through* situation t. We write $s \sqsubset t$, and say that s *strictly precedes* t, if $s \sqsubseteq t$ and $s \neq t$.

An *event* A is a set of paths, or in other words, a subset of the sample space: $A \subseteq \Omega$. With an event A, we can associate its *indicator* I_A, which is the real-valued map on Ω that assumes the value 1 on A, and 0 elsewhere.

We denote by $\uparrow t := \{\omega \in \Omega : t \sqsubseteq \omega\}$ the set of all paths that go through t: $\uparrow t$ is the event that corresponds to Reality getting to a situation t. It is clear that not all events will be of the type $\uparrow t$. Shafer [26] calls events of this type *exact*. Further on, in Section 4, exact events will be the only events that can be legitimately conditioned on, because they are the only events that can be foreseen to occur as part of Reality's game-play.

2.1.2. Cuts of a situation

Call a *cut* U of a situation t any set of situations that follow t, and such that for all paths ω through t, there is a unique $u \in U$ that ω goes through. In other words:

(i) $(\forall u \in U)(u \sqsupseteq t)$; and
(ii) $(\forall \omega \sqsupseteq t)(\exists! u \in U)(\omega \sqsupseteq u)$;

---

[4] Essentially, the width of the tree may be infinite, but its depth should be finite.
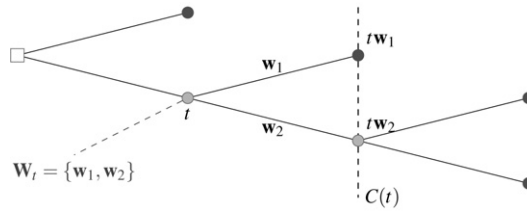
Fig. 2. An event tree for Reality, with the move space $\mathbf{W}_t$ and the corresponding children cut $C(t)$ of a non-terminal situation t.

see also Fig. 1. Alternatively, a set U of situations is a cut of t if and only if the corresponding set $\{\uparrow u: u \in U\}$ of exact events is a partition of the exact event $\uparrow t$. A cut can be interpreted as a (complete) stopping time.

If a situation $s \sqsupseteq t$ precedes (follows) some element of a cut U of t, then we say that s *precedes* (*follows*) U, and we write $s \sqsubseteq U$ ($s \sqsupseteq U$). Similarly for 'strictly precedes (follows)'. For two cuts U and V of t, we say that U *precedes* V if each element of U is followed by some element of V.

A *child* of a non-terminal situation t is a situation that immediately follows it. The set $C(t)$ of children of t constitutes a cut of t, called its *children cut*. Also, the set Ω of terminal situations is a cut of $\square$, called its *terminal cut*. The event $\uparrow t$ is the corresponding terminal cut of a situation t.

2.1.3. Reality's move spaces

We call a *move* $\mathbf{w}$ for Reality in a non-terminal situation t an arc that connects t with one of its children $s \in C(t)$, meaning that $s = t\mathbf{w}$ is the concatenation of the segment t and the arc $\mathbf{w}$. See Fig. 2.

Reality's *move space* in t is the set $\mathbf{W}_t$ of those moves $\mathbf{w}$ that Reality can make in t: $\mathbf{W}_t = \{\mathbf{w}: t\mathbf{w} \in C(t)\}$. We have already mentioned that $\mathbf{W}_t$ may be (countably or uncountably) infinite: there may be situations where reality has the choice between an infinity of next moves. But every $\mathbf{W}_t$ should contain at least two elements: otherwise there is no choice for Reality to make in situation t.

2.2. Processes and variables

We now have all the necessary tools to represent Reality's game-play. This game-play can be seen as a basis for an *event-driven*, rather than a time-driven, account of a theory of uncertain, or random, processes. The driving events are, of course, the moves that Reality makes.[5] In a theory of processes, we generally consider things that depend on (the succession of) these moves. This leads to the following definitions.

Any (partial) function on the set of situations $\Omega^\diamond$ is called a *process*, and any process whose domain includes all situations that follow a situation t is called a t-*process*. Of course, a t-process is also an s-process for all $s \sqsupseteq t$; when we call it an s-process, this means that we are restricting our attention to its values in all situations that follow s.

A special example of a t-process is the *distance* $d(t, \cdot)$ which for any situation $s \sqsupseteq t$ returns the number of steps $d(t, s)$ along the tree from t to s. When we said before that we are only considering *bounded protocols*, we meant that there is a natural number D such that $d(t, s) \leqslant D$ for all situations t and all $s \sqsupseteq t$.

Similarly, any (partial) function on the set of paths Ω is called a *variable*, and any variable on Ω whose domain includes all paths that go through a situation t is called a t-*variable*. If we restrict a t-process $\mathcal{F}$ to the set $\uparrow t$ of all terminal situations that follow t, we obtain a t-variable, which we denote by $\mathcal{F}_\Omega$.

If U is a cut of t, then we call a t-variable g U-*measurable* if for all u in U, g assumes the same value $g(u) := g(\omega)$ for all paths ω that go through u. In that case we can also consider g as a variable on U, which we denote as g_U.

If $\mathcal{F}$ is a t-process, then with any cut U of t we can associate a t-variable $\mathcal{F}_U$, which assumes the same value $\mathcal{F}_U(\omega) := \mathcal{F}(u)$ in all ω that follow $u \in U$. This t-variable is clearly U-measurable, and can be considered as a variable on U. This notation is consistent with the notation $\mathcal{F}_\Omega$ introduced earlier.

---

[5] These so-called *Humean* events shouldn't be confused with the *Moivrean* events we have considered before, and which are subsets of the sample space Ω. See Shafer [26, Chapter 1] for terminology and more explanation.
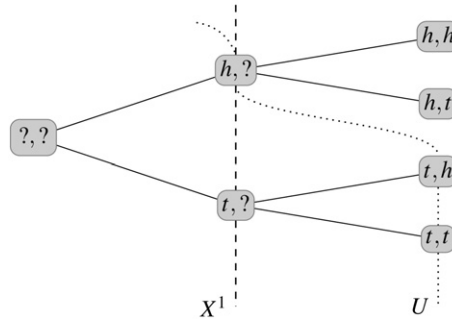
Fig. 3. The event tree associated with two successive coin flips. Also depicted are two cuts, X^1 and U, of the initial situation.

Similarly, we can associate with $\mathcal{F}$ a new, U-*stopped*, t-process $U(\mathcal{F})$, as follows:

$$U(\mathcal{F})(s) := \begin{cases} \mathcal{F}(s) & \text{if } t \sqsubseteq s \sqsubseteq U, \\ \mathcal{F}(u) & \text{if } u \in U \text{ and } u \sqsubseteq s. \end{cases}$$

The t-variable $U(\mathcal{F})_\Omega$ is U-measurable, and is actually equal to $\mathcal{F}_U$:

$$U(\mathcal{F})_\Omega = \mathcal{F}_U. \tag{1}$$

The following intuitive example will clarify these notions.

Example 1 *(Flipping coins).* Consider flipping two coins, one after the other. This leads to the event tree depicted in Fig. 3. The identifying labels for the situations should be intuitively clear: e.g., in the initial situation '$\square = ?, ?$' none of the coins have been flipped, in the non-terminal situation '$h, ?$' the first coin has landed 'heads' and the second coin hasn't been flipped yet, and in the terminal situation 't, t' both coins have been flipped and have landed 'tails'.

First, consider the real process $\mathcal{N}$, which in each situation s, returns the number $\mathcal{N}(s)$ of heads obtained so far, e.g., $\mathcal{N}(?, ?) = 0$ and $\mathcal{N}(h, ?) = 1$. If we restrict the process $\mathcal{N}$ to the set Ω of all terminal elements, we get a real variable $\mathcal{N}_\Omega$, whose values are: $\mathcal{N}_\Omega(h, h) = 2$, $\mathcal{N}_\Omega(h, t) = \mathcal{N}_\Omega(t, h) = 1$ and $\mathcal{N}_\Omega(t, t) = 0$.

Consider the cut U of the initial situation, which corresponds to the following stopping time: "stop after two flips, or as soon as an outcome is heads"; see Fig. 3. The values of the corresponding variable $\mathcal{N}_U$ are given by: $\mathcal{N}_U(h, h) = \mathcal{N}_U(h, t) = 1$, $\mathcal{N}_U(t, h) = 1$ and $\mathcal{N}_U(t, t) = 0$. So $\mathcal{N}_U$ is U-measurable, and can therefore be considered as a map on the elements $h, ?$ and t, h and t, t of U, with in particular $\mathcal{N}_U(h, ?) = 1$.

Next, consider the processes $\mathcal{F}, \mathcal{F}^1, \mathcal{F}^2 \colon \Omega^\diamond \to \{h, t, ?\}$, defined as follows:

s	$?, ?$	$h, ?$	$t, ?$	h, h	h, t	t, h	t, t
$\mathcal{F}(s)$	$?$	h	t	h	t	h	t
$\mathcal{F}^1(s)$	$?$	h	t	h	h	t	t
$\mathcal{F}^2(s)$	$?$	$?$	$?$	h	t	h	t

$\mathcal{F}$ returns the outcome of the latest, $\mathcal{F}^1$ the outcome of the first, and $\mathcal{F}^2$ that of the second coin flip. The associated variables $\mathcal{F}^1_\Omega$ and $\mathcal{F}^2_\Omega$ give, in each element of the sample space, the respective outcomes of the first and second coin flips.

The variable $\mathcal{F}^1_\Omega$ is X^1-measurable: as soon as we reach (any situation on) the cut X_1, its value is completely determined, i.e., we know the outcome of the first coin flip; see Fig. 3 for the definition of X^1.

We can associate with the process $\mathcal{F}$ the variable $\mathcal{F}_{X^1}$ that is also X^1-measurable: it returns, in any element of the sample space, the outcome of the first coin flip. Alternatively, we can stop the process $\mathcal{F}$ after one coin flip, which leads to the X^1-stopped process $X^1(\mathcal{F})$. This new process is of course equal to $\mathcal{F}^1$, and for the corresponding variable $\mathcal{F}^1_\Omega$, we have that $X^1(\mathcal{F})_\Omega = \mathcal{F}^1_\Omega = \mathcal{F}_{X^1}$; also see Eq. (1).

2.3. Sceptic's game-play

We now turn to the other player, Sceptic. His possible moves may well depend on the previous moves that Reality has made, in the following sense. In each non-terminal situation t, he has some set $\mathbf{S}_t$ of moves $\mathbf{s}$ available to him, called Sceptic's *move space* in t. We make the following assumption:

G2. In each non-terminal situation t, there is a (positive, negative, or zero) gain for Sceptic associated with each of the possible moves $\mathbf{s}$ in $\mathbf{S}_t$ that Sceptic can make. This gain depends only on the situation t and the next move $\mathbf{w}$ that Reality will make.

This means that for each non-terminal situation t there is a *gain function* $\lambda_t \colon \mathbf{S}_t \times \mathbf{W}_t \to \mathbb{R}$, such that $\lambda_t(\mathbf{s}, \mathbf{w})$ represents the change in Sceptic's capital in situation t when he makes move $\mathbf{s}$ and Reality makes move $\mathbf{w}$.

2.3.1. Strategies and capital processes

Let us introduce some further notions and terminology related to Sceptic's game-play. A *strategy* $\mathcal{P}$ for Sceptic is a partial process defined on the set $\Omega^\diamond \setminus \Omega$ of non-terminal situations, such that $\mathcal{P}(t) \in \mathbf{S}_t$ is the corresponding move that Sceptic will make in each non-terminal situation t. If Sceptic conceives of a strategy $\mathcal{P}$, this means he determines beforehand (before Reality leaves situation $\square$) what move $\mathcal{P}(t)$ to make in each non-terminal situation t.

With each such strategy $\mathcal{P}$ there corresponds a *capital process* $\mathcal{K}^\mathcal{P}$, whose value in each situation t gives us Sceptic's capital accumulated so far, when he starts out with zero capital in $\square$ and plays according to the strategy $\mathcal{P}$. It is given by the recursion relation

$$\mathcal{K}^\mathcal{P}(t\mathbf{w}) = \mathcal{K}^\mathcal{P}(t) + \lambda_t(\mathcal{P}(t), \mathbf{w}), \quad \mathbf{w} \in \mathbf{W}_t,$$

with initial condition $\mathcal{K}^\mathcal{P}(\square) = 0$. Of course, when Sceptic starts out (in $\square$) with capital α and uses strategy $\mathcal{P}$, his corresponding accumulated capital is given by the process $\alpha + \mathcal{K}^\mathcal{P}$. In the terminal situations, his accumulated capital is then given by the real variable $\alpha + \mathcal{K}^\mathcal{P}_\Omega$.

If we start in a non-terminal situation t, rather than in $\square$, then we can consider t-strategies $\mathcal{P}$ that tell Sceptic how to move starting from t onwards, and the corresponding capital process $\mathcal{K}^\mathcal{P}$ is then also a t-process, that tells us how much capital Sceptic has accumulated since starting with zero capital in situation t and using t-strategy $\mathcal{P}$.

2.3.2. Lower and upper prices

The assumptions G1 and G2 outlined above determine so-called *gambling protocols*. They are sufficient for us to be able to define lower and upper prices for real variables.

Consider a non-terminal situation t and a real t-variable f. The *upper price* $\overline{\mathbb{E}}_t(f)$ *for* f *in* t is defined as the infimum capital α that Sceptic has to start out with in t in order that there would be some t-strategy $\mathcal{P}$ such that his accumulated capital $\alpha + \mathcal{K}^\mathcal{P}$ allows him, at the end of the game, to hedge f, whatever moves Reality makes after t:

$$\overline{\mathbb{E}}_t(f) := \inf\{\alpha \colon \alpha + \mathcal{K}^\mathcal{P}_\Omega \geqslant f \text{ for some } t\text{-strategy } \mathcal{P}\}, \tag{2}$$

where $\alpha + \mathcal{K}^\mathcal{P}_\Omega \geqslant f$ is taken to mean that $\alpha + \mathcal{K}^\mathcal{P}(\omega) \geqslant f(\omega)$ for all terminal situations ω that go through t. Similarly, for the *lower price* $\underline{\mathbb{E}}_t(f)$ *for* f *in* t:

$$\underline{\mathbb{E}}_t(f) := \sup\{\alpha \colon \alpha - \mathcal{K}^\mathcal{P}_\Omega \leqslant f \text{ for some } t\text{-strategy } \mathcal{P}\}, \tag{3}$$

so $\underline{\mathbb{E}}_t(f) = -\overline{\mathbb{E}}_t(-f)$. If we start from the initial situation $t = \square$, we simply get the *upper and lower prices* for a real variable f, which we also denote by $\overline{\mathbb{E}}(f)$ and $\underline{\mathbb{E}}(f)$.

2.3.3. Coherent probability protocols

Requirements G1 and G2 for gambling protocols allow the moves, move spaces and gain functions for Sceptic to be just about anything. We now impose further conditions on Sceptic's move spaces.

A gambling protocol is called a *probability protocol* when besides G1 and G2, two more requirements are satisfied.

P1. For each non-terminal situation t, Sceptic's move space $\mathbf{S}_t$ is a convex cone in some linear space: $a_1\mathbf{s}_1 + a_2\mathbf{s}_2 \in \mathbf{S}_t$ for all non-negative real numbers a_1 and a_2 and all $\mathbf{s}_1$ and $\mathbf{s}_2$ in $\mathbf{S}_t$.

P2. For each non-terminal situation t, Sceptic's gain function λ_t has the following linearity property: $\lambda_t(a_1\mathbf{s}_1 + a_2\mathbf{s}_2, \mathbf{w}) = a_1\lambda_t(\mathbf{s}_1, \mathbf{w}) + a_2\lambda_t(\mathbf{s}_2, \mathbf{w})$ for all non-negative real numbers a_1 and a_2, all $\mathbf{s}_1$ and $\mathbf{s}_2$ in $\mathbf{S}_t$ and all $\mathbf{w}$ in $\mathbf{W}_t$.

Finally, a probability protocol is called *coherent*[6] when moreover:

C. For each non-terminal situation t, and for each $\mathbf{s}$ in $\mathbf{S}_t$ there is some $\mathbf{w}$ in $\mathbf{W}_t$ such that $\lambda_t(\mathbf{s}, \mathbf{w}) \leqslant 0$.

It is clear what this last requirement means: in each non-terminal situation, Reality has a strategy for playing from t onwards such that Sceptic can't (strictly) increase his capital from t onwards, whatever t-strategy he might use.

For such coherent probability protocols, Shafer and Vovk prove a number of interesting properties for the corresponding lower (and upper) prices. We list a number of them here. For any real t-variable f, we can associate with a cut U of t another special U-measurable t-variable $\underline{\mathbb{E}}_U$ by $\underline{\mathbb{E}}_U(f)(\omega) = \underline{\mathbb{E}}_u(f)$, for all paths ω through t, where u is the unique situation in U that ω goes through. For any two real t-variables f_1 and f_2, $f_1 \leqslant f_2$ is taken to mean that $f_1(\omega) \leqslant f_2(\omega)$ for all paths ω that go through t.

Proposition 1 (*Properties of lower and upper prices in a coherent probability protocol* [29]). *Consider a coherent probability protocol, let t be a non-terminal situation, f, f_1 and f_2 real t-variables, and U a cut of t. Then*

1. $\inf_{\omega \in \uparrow t} f(\omega) \leqslant \underline{\mathbb{E}}_t(f) \leqslant \overline{\mathbb{E}}_t(f) \leqslant \sup_{\omega \in \uparrow t} f(\omega)$ [*convexity*];
2. $\underline{\mathbb{E}}_t(f_1 + f_2) \geqslant \underline{\mathbb{E}}_t(f_1) + \underline{\mathbb{E}}_t(f_2)$ [*super-additivity*];
3. $\underline{\mathbb{E}}_t(\lambda f) = \lambda\underline{\mathbb{E}}_t(f)$ *for all real* $\lambda \geqslant 0$ [*non-negative homogeneity*];
4. $\underline{\mathbb{E}}_t(f + \alpha) = \underline{\mathbb{E}}_t(f) + \alpha$ *for all real* α [*constant additivity*];
5. $\underline{\mathbb{E}}_t(\alpha) = \alpha$ *for all real* α [*normalisation*];
6. $f_1 \leqslant f_2$ *implies that* $\underline{\mathbb{E}}_t(f_1) \leqslant \underline{\mathbb{E}}_t(f_2)$ [*monotonicity*];
7. $\underline{\mathbb{E}}_t(f) = \underline{\mathbb{E}}_t(\underline{\mathbb{E}}_U(f))$ [*law of iterated expectation*].

What is more, Shafer and Vovk use specific instances of such coherent probability protocols to prove various limit theorems (such as the law of large numbers, the central limit theorem, the law of the iterated logarithm), from which they can derive, as special cases, the well-known measure-theoretic versions. We shall come back to this in Section 6.

The game-theoretic account of probability we have described so far, is very general. But it seems to pay little or no attention to *beliefs* that Sceptic, or other, perhaps additional players in these games might entertain about how Reality will move through its event tree. This might seem strange, because at least according to the personalist and epistemicist school, probability is all about beliefs. In order to find out how we can incorporate beliefs into the game-theoretic framework, we now turn to Walley's imprecise probability models.

3. Walley's behavioural approach to probability

In his book on the behavioural theory of imprecise probabilities [33], Walley considers many different types of related uncertainty models. We shall restrict ourselves here to the most general and most powerful one, which also turns out to be the easiest to explain, namely coherent sets of really desirable gambles; see also [35].

Consider a non-empty set Ω of possible alternatives ω, only one of which actually obtains (or will obtain); we assume that it is possible, at least in principle, to determine which alternative does so. Also consider a subject who is uncertain about which possible alternative actually obtains (or will obtain). A *gamble* on Ω is a real-valued map on Ω, and it is interpreted as an uncertain reward, expressed in units of some predetermined linear utility scale: if ω actually obtains, then the reward is $f(\omega)$, which may be positive, negative, or zero. We use the notation $\mathcal{G}(\Omega)$ for the set of all gambles on Ω. Walley [33] assumes gambles to be bounded. We make no such boundedness assumption here.[7]

---

[6] For a discussion of the use of 'coherent' here, we refer to [28, Appendix C].

[7] The concept of a really desirable gamble (at least formally) allows for such a generalisation, because the coherence axioms for real desirability nowhere hinge on such a boundedness assumption, at least not from a technical mathematical point of view.

If a subject *accepts* a gamble f, this is taken to mean that she is willing to engage in the transaction where, (i) first it is determined which ω obtains, and (ii) then she receives the reward $f(\omega)$. We can try and model the subject's beliefs about Ω by considering which gambles she accepts.

3.1. Coherent sets of really desirable gambles

Suppose our subject specifies some set $\mathcal{R}$ of gambles she accepts, called a *set of really desirable gambles*. Such a set is called *coherent* if it satisfies the following *rationality requirements*:

D1. if $f < 0$ then $f \notin \mathcal{R}$ [avoiding partial loss];
D2. if $f \geqslant 0$ then $f \in \mathcal{R}$ [accepting partial gain];
D3. if f_1 and f_2 belong to $\mathcal{R}$ then their (point-wise) sum $f_1 + f_2$ also belongs to $\mathcal{R}$ [combination];
D4. if f belongs to $\mathcal{R}$ then its (point-wise) scalar product λf also belongs to $\mathcal{R}$ for all non-negative real numbers λ [scaling].

Here '$f < 0$' means '$f \leqslant 0$ and not $f = 0$'. Walley has also argued that, besides D1–D4, sets of really desirable gambles should satisfy an additional axiom:

D5. $\mathcal{R}$ is $\mathcal{B}$-conglomerable for any partition $\mathcal{B}$ of Ω: if $I_B f \in \mathcal{R}$ for all $B \in \mathcal{B}$, then also $f \in \mathcal{R}$ [full conglomerability].

When the set Ω is finite, all its partitions are finite too, and therefore full conglomerability becomes a direct consequence of the finitary combination axiom D3. But when Ω is infinite, its partitions may be infinite too, and then full conglomerability is a very strong additional requirement, that is not without controversy. If a model $\mathcal{R}$ is $\mathcal{B}$-conglomerable, this means that certain inconsistency problems when conditioning on elements B of $\mathcal{B}$ are avoided; see [33] for more details and examples. Conglomerability of belief models wasn't required by forerunners of Walley, such as Williams [39],[8] or de Finetti [11]. While we agree with Walley that conglomerability is a desirable property for sets of really desirable gambles, we do not believe that *full* conglomerability is always necessary: it seems that we only need to require conglomerability with respect to those partitions that we actually intend to condition our model on.[9] This is the path we shall follow in Section 4.

3.2. Conditional lower and upper previsions

Given a coherent set of really desirable gambles, we can define conditional lower and upper previsions as follows: for any gamble f and any non-empty subset B of Ω, with indicator I_B,

$$\overline{P}(f|B) := \inf\{\alpha: I_B(\alpha - f) \in \mathcal{R}\}, \tag{4}$$

$$\underline{P}(f|B) := \sup\{\alpha: I_B(f - \alpha) \in \mathcal{R}\}, \tag{5}$$

so $\underline{P}(f|B) = -\overline{P}(-f|B)$, and *the lower prevision $\underline{P}(f|B)$ of f, conditional on B* is the supremum price α for which the subject will buy the gamble f, i.e., accept the gamble $f - \alpha$, contingent on the occurrence of B. Similarly, *the upper prevision $\overline{P}(f|B)$ of f, conditional on B* is the infimum price α for which the subject will sell the gamble f, i.e., accept the gamble $\alpha - f$, contingent on the occurrence of B.

For any event A, we define the conditional lower probability $\underline{P}(A|B) := \underline{P}(I_A|B)$, i.e., the subject's supremum rate for betting on the event A, contingent on the occurrence of B, and similarly for $\overline{P}(A|B) := \overline{P}(I_A|B)$.

We want to stress here that by its definition [Eq. (5)], $\underline{P}(f|B)$ is a conditional lower prevision on what Walley [33, Section 6.1] has called the *contingent interpretation*: it is a supremum acceptable price for buying the gamble f *contingent* on the occurrence of B, meaning that the subject accepts the contingent gambles $I_B(f - \underline{P}(f|B) + \epsilon)$,

---

[8] Axioms related to (D1)–(D4), but not (D5), were actually suggested by Williams for bounded gambles. But it seems that we need at least some weaker form of (D5), namely the cut conglomerability (D5$'$) considered further on, to derive our main results: Theorems 3 and 6.

[9] The view expressed here seems related to Shafer's, as sketched near the end of [25, Appendix 1].

$\epsilon > 0$, which are called off unless B occurs. This should be contrasted with the *updating interpretation* for the conditional lower prevision $\underline{P}(f|B)$, which is a subject's *present* (before the occurrence of B) supremum acceptable price for buying f after receiving the information that B has occurred (and nothing else!). Walley's *Updating Principle* [33, Section 6.1.6], which we shall accept, and use further on in Section 4, (essentially) states that conditional lower previsions should be the same on both interpretations. There is also a third way of looking at a conditional lower prevision $\underline{P}(f|B)$, which we shall call the *dynamic interpretation*, and where $\underline{P}(f|B)$ stands for the subject's supremum acceptable buying price for f *after she gets to know B has occurred*. For precise conditional previsions, this last interpretation seems to be the one considered in [13,23,24]. It is far from obvious that there should be a relation between the first two and the third interpretations.[10] We shall briefly come back to this distinction in the following sections.

For any partition $\mathcal{B}$ of Ω, we let $\underline{P}(f|\mathcal{B}) := \sum_{B\in\mathcal{B}} I_B \underline{P}(f|B)$ be the gamble on Ω that in any element ω of B assumes the value $\underline{P}(f|B)$, where B is any element of $\mathcal{B}$.

The following properties of conditional lower and upper previsions associated with a coherent set of really desirable bounded gambles were (essentially) proved by Walley [33], and by Williams [39]. We give the extension to potentially unbounded gambles:

Proposition 2 *(Properties of conditional lower and upper previsions [33]). Consider a coherent set of really desirable gambles $\mathcal{R}$, let B be any non-empty subset of Ω, and let f, f_1 and f_2 be gambles on Ω. Then*[11]

1. $\inf_{\omega\in B} f(\omega) \leqslant \underline{P}(f|B) \leqslant \overline{P}(f|B) \leqslant \sup_{\omega\in B} f(\omega)$ *[convexity]*;
2. $\underline{P}(f_1 + f_2|B) \geqslant \underline{P}(f_1|B) + \underline{P}(f_2|B)$ *[super-additivity]*;
3. $\underline{P}(\lambda f|B) = \lambda \underline{P}(f|B)$ *for all real $\lambda \geqslant 0$ [non-negative homogeneity]*;
4. $\underline{P}(f + \alpha|B) = \underline{P}(f|B) + \alpha$ *for all real α [constant additivity]*;
5. $\underline{P}(\alpha|B) = \alpha$ *for all real α [normalisation]*;
6. $f_1 \leqslant f_2$ *implies that $\underline{P}(f_1|B) \leqslant \underline{P}(f_2|B)$ [monotonicity]*;
7. *if $\mathcal{B}$ is any partition of Ω that refines the partition $\{B, B^c\}$ and $\mathcal{R}$ is $\mathcal{B}$-conglomerable, then $\underline{P}(f|B) \geqslant \underline{P}(\underline{P}(f|\mathcal{B})|B)$ [conglomerative property]*.

The analogy between Propositions 1 and 2 is striking, even if there is an equality in Proposition 1.7, where we have only an inequality in Proposition 2.7.[12] In the next section, we set out to identify the exact correspondence between the two models. We shall find a specific situation where applying Walley's theory leads to equalities rather than the more general inequalities of Proposition 2.7.[13]

We now show that there can indeed be a strict inequality in Proposition 2.7.

Example 2. Consider an urn with red, green and blue balls, from which a ball will be drawn at random. Our subject is uncertain about the colour of this ball, so $\Omega = \{r, g, b\}$. Assume that she assesses that she is willing to bet on this colour being red at rates up to (and including) $1/4$, i.e., that she accepts the gamble $I_{\{r\}} - 1/4$. Similarly for the other two colours, so she also accepts the gambles $I_{\{g\}} - 1/4$ and $I_{\{b\}} - 1/4$. It is not difficult to prove using the coherence requirements D1–D4 and Eq. (5) that the smallest coherent set of really desirable gambles $\mathcal{R}$ that includes these assessments satisfies $f \in \mathcal{R} \Leftrightarrow \underline{P}(f) \geqslant 0$, where

$$\underline{P}(f) = \frac{3}{4}\frac{f(r) + f(g) + f(b)}{3} + \frac{1}{4}\min\{f(r), f(g), f(b)\}.$$

---

[10] In [28], the authors seem to confuse the updating interpretation with the dynamic interpretation when they claim that their new understanding of lower and upper previsions "justifies Peter Walley's updating principle".

[11] Here, as in Proposition 1, we implicitly assume that whatever we write down is well-defined, meaning that for instance no sums of $-\infty$ and $+\infty$ appear, and that the function $\underline{P}(f|\mathcal{B})$ is real-valued, and nowhere infinite. Shafer and Vovk don't seem to mention the need for this.

[12] Concatenation inequalities for lower prices do appear in the more general context described in [28].

[13] This seems to happen generally for what is called *marginal extension* in a situation of immediate prediction, meaning that we start out with, and extend, an initial model where we condition on increasingly finer partitions, and where the initial conditional model for any partition deals with gambles that are measurable with respect to the finer partitions; see [33, Theorem 6.7.2] and [20].

For the partition $\mathcal{B} = \{b, \{r, g\}\}$ (a Daltonist has observed the colour of the ball and tells the subject about it), it follows from Eq. (5) after some manipulations that

$$\underline{P}\big(f|\{b\}\big) = f(b) \quad \text{and} \quad \underline{P}\big(f|\{r, g\}\big) = \frac{2}{3}\frac{f(r) + f(g)}{2} + \frac{1}{3}\min\big\{f(r), f(g)\big\}.$$

If we consider $f = I_{\{g\}}$, then in particular $\underline{P}(\{g\}|\{b\}) = 0$ and $\underline{P}(\{g\}|\{r, g\}) = 1/3$, so $\underline{P}(\{g\}|\mathcal{B}) = 1/3I_{\{r,g\}}$ and therefore

$$\underline{P}\big(\underline{P}(\{g\}|\mathcal{B})\big) = \frac{3}{4}\frac{1/3 + 1/3}{3} + \frac{1}{4}0 = \frac{1}{6},$$

whereas $\underline{P}(\{g\}) = 1/4$, and therefore $\underline{P}(\{g\}) > \underline{P}(\underline{P}(\{g\}|\mathcal{B}))$.

The difference $\overline{P}(f|B) - \underline{P}(f|B)$ between infimum selling and supremum buying prices for gambles f represents imprecision present in our subject's belief model. If we look at the inequalities in Proposition 2.1, we are led to consider two extreme cases. One extreme maximises the 'degrees of imprecision' $\overline{P}(f|B) - \underline{P}(f|B)$ by letting $\underline{P}(f|B) = \inf_{\omega \in B} f(\omega)$ and $\overline{P}(f|B) = \sup_{\omega \in B} f(\omega)$. This leads to the so-called *vacuous model*, corresponding to $\mathcal{R} = \{f: f \geqslant 0\}$, and intended to represent complete ignorance on the subject's part.

The other extreme minimises the degrees of imprecision $\overline{P}(f|B) - \underline{P}(f|B)$ by letting $\underline{P}(f|B) = \overline{P}(f|B)$ everywhere. The common value $P(f|B)$ is then called the *prevision*, or *fair price*, for f conditional on B. We call the corresponding functional $P(\cdot|B)$ a (conditional) *linear prevision*. Linear previsions are the precise probability models considered by de Finetti [11]. They of course have all properties of lower and upper previsions listed in Proposition 2, with equality rather than inequality for statements 2 and 7. The restriction of a linear prevision to (indicators of) events is a finitely additive probability measure.

4. Connecting the two approaches

In order to lay bare the connections between the game-theoretic and the behavioural approach, we enter Shafer and Vovk's world, and consider another player, called Forecaster, who, *in situation* $\square$, has certain *piece-wise* beliefs about what moves Reality will make.

4.1. Forecaster's local beliefs

More specifically, for each non-terminal situation $t \in \Omega^\diamond \setminus \Omega$, she has beliefs (in situation $\square$) about which move **w** Reality will choose from the set $\mathbf{W}_t$ of moves available to him if he gets to t. We suppose she represents those beliefs in the form of a *coherent*[14] set $\mathcal{R}_t$ of really desirable gambles on $\mathbf{W}_t$. These beliefs are conditional on the updating interpretation, in the sense that they represent Forecaster's beliefs in situation $\square$ about what Reality will do *immediately after he gets to situation t*. We call any specification of such coherent $\mathcal{R}_t$, $t \in \Omega^\diamond \setminus \Omega$, an *immediate prediction model* for Forecaster. We want to stress here that $\mathcal{R}_t$ should *not* be interpreted dynamically, i.e., as a set of gambles on $\mathbf{W}_t$ that Forecaster accepts in situation t.

We shall generally call an event tree, provided with local predictive belief models in each of the non-terminal situations t, an *imprecise probability tree*. These local belief models may be coherent sets of really desirable gambles $\mathcal{R}_t$. But they can also be lower previsions $\underline{P}_t$ (perhaps derived from such sets $\mathcal{R}_t$). When all such local belief models are precise previsions, or equivalently (finitely additive) probability measures, we simply get a *probability tree* in Shafer's [26, Chapter 3] sense.

4.2. From local to global beliefs

We can now ask ourselves what the behavioural implications of these conditional assessments $\mathcal{R}_t$ in the immediate prediction model are. For instance, what do they tell us about whether or not Forecaster should accept certain gambles[15] on Ω, the set of possible paths for Reality? In other words, how can these beliefs (in $\square$) about which next

---

[14] Since we don't immediately envisage conditioning this local model on subsets of $\mathbf{W}_t$, we impose no extra conglomerability requirements here, only the coherence conditions D1–D4.

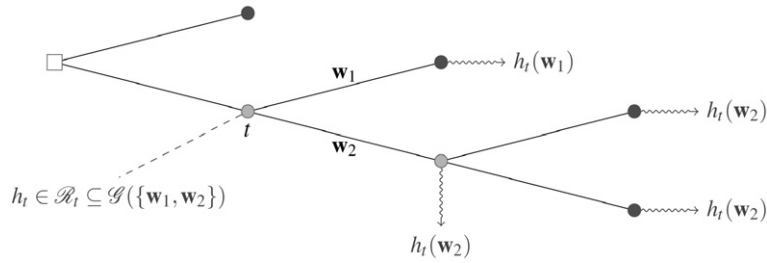[15] In Shafer and Vovk's language, gambles are real variables.

Fig. 4. In a non-terminal situation t, we consider a gamble h_t on Reality's move space $\mathbf{W}_t$ that Forecaster accepts, and turn it into a process, also denoted by h_t. The values $h_t(s)$ in situations $s \sqsupseteq t$ are indicated by curly arrows.

move Reality will make in each non-terminal situation t be combined coherently into beliefs (in $\square$) about Reality's complete sequence of moves?

In order to investigate this, we use Walley's very general and powerful method of *natural extension*, which is just *conservative coherent reasoning*: we construct, using the local pieces of information $\mathcal{R}_t$, a set of really desirable gambles on Ω for Forecaster in situation $\square$ that is (i) coherent, and (ii) as small as possible, meaning that no more gambles should be accepted than is actually required by coherence.

4.2.1. Collecting the pieces

Consider any non-terminal situation $t \in \Omega^\Diamond \setminus \Omega$ and any gamble h_t in $\mathcal{R}_t$. With h_t we can associate a t-gamble,[16] also denoted by h_t, and defined by

$$h_t(\omega) := h_t\big(\omega(t)\big)$$

for all $\omega \sqsupseteq t$, where we denote by $\omega(t)$ the unique element of $\mathbf{W}_t$ such that $t\omega(t) \sqsubseteq \omega$. The t-gamble h_t is U-measurable for any cut U of t that is non-trivial, i.e., such that $U \neq \{t\}$. This implies that we can interpret h_t as a map on U. In fact, we shall even go further, and associate with the gamble h_t on $\mathbf{W}_t$ a t-process, also denoted by h_t, by letting $h_t(s) := h_t(\omega(t))$ for any $s \sqsupseteq t$, where ω is any terminal situation that follows s; see also Fig. 4.

$I_{\uparrow t} h_t$ represents the gamble on Ω that is called off unless Reality ends up in situation t, and which, when it isn't called off, depends only on Reality's move immediately after t, and gives the same value $h_t(\mathbf{w})$ to all paths ω that go through $t\mathbf{w}$. The fact that Forecaster, in situation $\square$, accepts h_t on $\mathbf{W}_t$ conditional on Reality's getting to t, translates immediately to the fact that Forecaster accepts the contingent gamble $I_{\uparrow t} h_t$ on Ω, by Walley's Updating Principle. We thus end up with a set

$$\mathcal{R} := \bigcup_{t \in \Omega^\Diamond \setminus \Omega} \{I_{\uparrow t} h_t \colon h_t \in \mathcal{R}_t\}$$

of gambles on Ω that Forecaster accepts in situation $\square$.

The only thing left to do now, is to find the smallest coherent set $\mathcal{E}_\mathcal{R}$ of really desirable gambles that includes $\mathcal{R}$ (if indeed there is any such coherent set). Here we take coherence to refer to conditions D1–D4, together with D5$'$, a variation on D5 which refers to conglomerability with respect to those partitions that we actually intend to condition on, as suggested in Section 3.

4.2.2. Cut conglomerability

These partitions are what we call *cut partitions*. Consider any cut U of the initial situation $\square$. The set of events $\mathcal{B}_U := \{\uparrow u \colon u \in U\}$ is a partition of Ω, called the *U-partition*. D5$'$ requires that our set of really desirable gambles should be *cut conglomerable*, i.e., conglomerable with respect to every cut partition $\mathcal{B}_U$:[17]

---

[16] Just as for variables, we can define a t-gamble as a partial gamble whose domain includes $\uparrow t$.

[17] Again, when all of Reality's move spaces $\mathbf{W}_t$ are finite, cut conglomerability (D5$'$) is a consequence of D3, and therefore needs no extra attention. But when some or all move spaces are infinite, then a cut U may contain an infinite number of elements, and the corresponding cut partition $\mathcal{B}_U$ will then be infinite too, making cut conglomerability a non-trivial additional requirement.
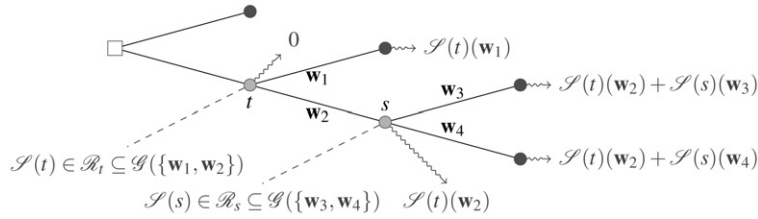
Fig. 5. The t-selection $\mathcal{S}$ in this event tree is a process defined in the two non-terminal situations t and s; it selects, in each of these situations, a really desirable gamble for Forecaster. The values of the corresponding gamble process $\mathcal{G}^{\mathcal{S}}$ are indicated by curly arrows.

D5′. $\mathcal{R}$ is $\mathcal{B}_U$-conglomerable for any cut-partition $\mathcal{B}_U$ of Ω: if $I_{\uparrow u} f \in \mathcal{R}$ for all u in some cut U of $\square$, then also $f \in \mathcal{R}$ [cut conglomerability].

Why do we only require conglomerability for cut partitions? Simply because we are interested in *predictive infer-ence*: we eventually will want to find out about the gambles on Ω that Forecaster accepts in situation $\square$, conditional (contingent) on Reality getting to a situation t. This is related to finding lower previsions for Forecaster conditional on the corresponding events $\uparrow t$. A collection $\{\uparrow t: t \in T\}$ of such events constitutes a partition of the sample space Ω if and only if T is a cut of $\square$.

Because we require cut conglomerability, it follows in particular that $\mathcal{E}_{\mathcal{R}}$ will contain the sums of gambles $g := \sum_{u \in U} I_{\uparrow u} h_u$ for all *non-terminal* cuts U of $\square$ and all choices of $h_u \in \mathcal{R}_u$, $u \in U$. This is because $I_{\uparrow u} g = I_{\uparrow u} h_u \in \mathcal{R}$ for all $u \in U$. Because moreover $\mathcal{E}_{\mathcal{R}}$ should be a convex cone [by D3 and D4], any sum of such sums $\sum_{u \in U} I_{\uparrow u} h_u$ over a finite number of non-terminal cuts U should also belong to $\mathcal{E}_{\mathcal{R}}$. But, since in the case of bounded protocols we are discussing here, Reality can only make a bounded and finite number of moves, $\Omega^{\Diamond} \setminus \Omega$ is a finite union of such non-terminal cuts, and therefore the sums $\sum_{u \in \Omega^{\Diamond} \setminus \Omega} I_{\uparrow u} h_u$ should belong to $\mathcal{E}_{\mathcal{R}}$ for all choices $h_u \in \mathcal{R}_u$, $u \in \Omega^{\Diamond} \setminus \Omega$.

4.2.3. Selections and gamble processes

Consider any non-terminal situation t, and call t-*selection* any partial process $\mathcal{S}$ defined on the non-terminal $s \sqsupseteq t$ such that $\mathcal{S}(s) \in \mathcal{R}_s$. With a t-selection $\mathcal{S}$, we associate a t-process $\mathcal{G}^{\mathcal{S}}$, called a *gamble process*, where

$$\mathcal{G}^{\mathcal{S}}(s) = \sum_{t \sqsubseteq u \sqsubset s} \mathcal{S}(u)(s) \tag{6}$$

in all situations $s \sqsupseteq t$; see also Fig. 5. Alternatively, $\mathcal{G}^{\mathcal{S}}$ is given by the recursion relation

$$\mathcal{G}^{\mathcal{S}}(s\mathbf{w}) = \mathcal{G}^{\mathcal{S}}(s) + \mathcal{S}(s)(\mathbf{w}), \quad \mathbf{w} \in \mathbf{W}_s$$

for all non-terminal $s \sqsupseteq t$, with initial value $\mathcal{G}^{\mathcal{S}}(t) = 0$. In particular, this leads to the t-gamble $\mathcal{G}^{\mathcal{S}}_{\Omega}$ defined on all terminal situations ω that follow t, by letting

$$\mathcal{G}^{\mathcal{S}}_{\Omega} = \sum_{t \sqsubseteq u, u \in \Omega^{\Diamond} \setminus \Omega} I_{\uparrow u} \mathcal{S}(u). \tag{7}$$

Then we have just argued that the gambles $\mathcal{G}^{\mathcal{S}}_{\Omega}$ should belong to $\mathcal{E}_{\mathcal{R}}$ for all non-terminal situations t and all t-selections $\mathcal{S}$. As before for strategy and capital processes, we call a $\square$-selection $\mathcal{S}$ simply a *selection*, and a $\square$-gamble process simply a *gamble process*.

4.2.4. The Marginal Extension Theorem

It is now but a technical step to prove Theorem 3 below. It is a significant generalisation, in terms of sets of really desirable gambles rather than coherent lower previsions,[18] of the *Marginal Extension Theorem* first proved by Walley [33, Theorem 6.7.2], and subsequently extended by De Cooman and Miranda [20]. It essentially tells us that a t-gamble is (globally) desirable if and only if it dominates some t-gamble obtained by (additively) piecing together locally desirable gambles in the manner we have described above.

---

[18] The difference in language may obscure that this is indeed a generalisation. But see Theorem 7 for expressions in terms of predictive lower previsions that should make the connection much clearer.

Theorem 3 *(Marginal Extension Theorem). There is a smallest set of gambles that satisfies D1–D4 and D5′ and includes $\mathcal{R}$. This natural extension of $\mathcal{R}$ is given by*

$$\mathcal{E}_{\mathcal{R}} := \big\{ g\colon g \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}} \text{ for some selection } \mathcal{S} \big\}.$$

Moreover, for any non-terminal situation t and any t-gamble g, it holds that $I_{\uparrow t}g \in \mathcal{E}_{\mathcal{R}}$ if and only if there is some t-selection $\mathcal{S}_t$ such that $g \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}_t}$, where as before, $g \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}_t}$ is taken to mean that $g(\omega) \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}_t}(\omega)$ for all terminal situations ω that follow t.

4.3. Predictive lower and upper previsions

We now use the coherent set of really desirable gambles $\mathcal{E}_{\mathcal{R}}$ to define special lower previsions $\underline{P}(\cdot|t) := \underline{P}(\cdot|\uparrow t)$ for Forecaster in situation $\square$, conditional on an event $\uparrow t$, i.e., on Reality getting to situation t, as explained in Section 3.[19] We shall call such conditional lower previsions *predictive* lower previsions. We then get, using Eq. (5) and Theorem 3, that for any non-terminal situation t,

$$\underline{P}(f|t) := \sup\big\{ \alpha\colon I_{\uparrow t}(f - \alpha) \in \mathcal{E}_{\mathcal{R}} \big\} \tag{8}$$

$$= \sup\big\{ \alpha\colon f - \alpha \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}} \text{ for some } t\text{-selection } \mathcal{S} \big\}. \tag{9}$$

We also use the notation $\underline{P}(f) := \underline{P}(f|\square) = \sup\{\alpha\colon f - \alpha \in \mathcal{E}_{\mathcal{R}}\}$. It should be stressed that Eq. (8) is also valid in terminal situations t, whereas Eq. (9) clearly isn't.

Besides the properties in Proposition 2, which hold in general for conditional lower and upper previsions, the predictive lower (and upper) previsions we consider here also satisfy a number of additional properties, listed in Propositions 4 and 5.

Proposition 4 *(Additional properties of predictive lower and upper previsions). Let t be any situation, and let f, f_1 and f_2 be gambles on Ω.*

1. *If t is a terminal situation ω, then $\underline{P}(f|\omega) = \overline{P}(f|\omega) = f(\omega)$;*
2. *$\underline{P}(f|t) = \underline{P}(f I_{\uparrow t}|t)$ and $\overline{P}(f|t) = \overline{P}(f I_{\uparrow t}|t)$;*
3. *$f_1 \leqslant f_2$ (on $\uparrow t$) implies that $\underline{P}(f_1|t) \leqslant \underline{P}(f_2|t)$ [monotonicity].*

Before we go on, there is an important point that must be stressed and clarified. It is an immediate consequence of Proposition 4.2 that when f and g are any two gambles that coincide on $\uparrow t$, then $\underline{P}(f|t) = \underline{P}(g|t)$. This means that $\underline{P}(f|t)$ is completely determined by the values that f assumes on $\uparrow t$, and it allows us to define $\underline{P}(\cdot|t)$ on gambles that are only necessarily defined on $\uparrow t$, i.e., on t-gambles. We shall do so freely in what follows.

For any cut U of a situation t, we may define the t-gamble $\underline{P}(f|U)$ as the gamble that assumes the value $\underline{P}(f|u)$ in any $\omega \sqsupseteq u$, where $u \in U$. This t-gamble is U-measurable by construction, and it can be considered as a gamble on U.

Proposition 5 *(Separate coherence). Let t be any situation, let U be any cut of t, and let f and g be t-gambles, where g is U-measurable.*

1. *$\underline{P}(\uparrow t|t) = 1$;*
2. *$\underline{P}(g|U) = g_U$;*
3. *$\underline{P}(f + g|U) = g_U + \underline{P}(f|U)$;*
4. *if g is moreover non-negative, then $\underline{P}(gf|U) = g_U \underline{P}(f|U)$.*

4.4. Correspondence between immediate prediction models and coherent probability protocols

There appears to be a close correspondence between the expressions [such as (3)] for lower prices $\underline{\mathbb{E}}_t(f)$ associated with coherent probability protocols and those [such as (9)] for the predictive lower previsions $\underline{P}(f|t)$ based on an

---

[19] We stress again that these are conditional lower previsions on the contingent/updating interpretation.

immediate prediction model. Say that a given coherent probability protocol and given immediate prediction model *match* whenever they lead to identical corresponding lower prices $\underline{\mathbb{E}}_t$ and predictive lower previsions $\underline{P}(\cdot|t)$ for all non-terminal $t \in \Omega^{\Diamond} \setminus \Omega$.

The following theorem marks the culmination of our search for the correspondence between Walley's, and Shafer and Vovk's approaches to probability theory.

Theorem 6 *(Matching Theorem). For every coherent probability protocol there is an immediate prediction model such that the two match, and conversely, for every immediate prediction model there is a coherent probability protocol such that the two match.*

The ideas underlying the proof of this theorem should be clear. If we have a coherent probability protocol with move spaces $\mathbf{S}_t$ and gain functions λ_t for Sceptic, define the immediate prediction model for Forecaster to be (essentially) $\mathcal{R}_t := \{-\lambda(\mathbf{s}, \cdot) \colon \mathbf{s} \in \mathbf{S}_t\}$. If, conversely, we have an immediate prediction model for Forecaster consisting of the sets $\mathcal{R}_t$, define the move spaces for Sceptic by $\mathbf{S}_t := \mathcal{R}_t$, and his gain functions by $\lambda_t(h, \cdot) := -h$ for all h in $\mathcal{R}_t$. We discuss the interpretation of this correspondence in more detail in Section 5.

4.5. Calculating predictive lower prevision using backwards recursion

The Marginal Extension Theorem allows us to calculate the most conservative global belief model $\mathcal{E}_{\mathcal{R}}$ that corresponds to the local immediate prediction models $\mathcal{R}_t$. Here beliefs are expressed in terms of sets of really desirable gambles. Can we derive a result that allows us to do something similar for the corresponding lower previsions?

To see what this question entails, first consider a local model $\mathcal{R}_s$: a set of really desirable gambles on $\mathbf{W}_s$, where $s \in \Omega^{\Diamond} \setminus \Omega$. Using Eq. (5), we can associate with $\mathcal{R}_s$ a lower prevision $\underline{P}_s$ on $\mathcal{G}(\mathbf{W}_s)$. Each gamble g_s on $\mathbf{W}_s$ can be seen as an uncertain reward, whose outcome $g_s(\mathbf{w})$ depends on the (unknown) move $\mathbf{w} \in \mathbf{W}_s$ that Reality will make if it gets to situation s. And Forecaster's *local* (predictive) lower prevision

$$\underline{P}_s(g_s) := \sup\{\alpha \colon g_s - \alpha \in \mathcal{R}_s\} \tag{10}$$

for g_s is her supremum acceptable price (in $\square$) for buying g_s when Reality gets to s.

But as we have seen in Section 4.3, we can also, in each situation t, derive *global* predictive lower previsions $\underline{P}(\cdot|t)$ for Forecaster from the global model $\mathcal{E}_{\mathcal{R}}$, using Eq. (8). For each t-gamble f, $\underline{P}(f|t)$ is Forecaster inferred supremum acceptable price (in $\square$) for buying f, contingent on Reality getting to t.

Is there a way to construct the global predictive lower previsions $\underline{P}(\cdot|t)$ directly from the local predictive lower previsions $\underline{P}_s$? We can infer that there is from the following theorem, together with Propositions 8 and 9 below.

Theorem 7 *(Concatenation Formula). Consider any two cuts U and V of a situation t such that U precedes V. For all t-gambles f on Ω,*[20]

1. $\underline{P}(f|t) = \underline{P}(\underline{P}(f|U)|t)$;
2. $\underline{P}(f|U) = \underline{P}(\underline{P}(f|V)|U)$.

To make clear what the following Proposition 8 implies, consider any t-selection $\mathcal{S}$, and define the *U-called off t-selection* $\mathcal{S}^U$ as the selection that mimics $\mathcal{S}$ until we get to U, where we begin to select the zero gambles: for any non-terminal situation $s \sqsupseteq t$, let $\mathcal{S}^U(s) := \mathcal{S}(s)$ if s strictly precedes (some element of) U, and let $\mathcal{S}^U(s) := 0 \in \mathcal{R}_s$ otherwise. If we stop the gamble process $\mathcal{G}^{\mathcal{S}}$ at the cut U, we readily infer from Eq. (6) that for the U-stopped process $U(\mathcal{G}^{\mathcal{S}})$

$$U(\mathcal{G}^{\mathcal{S}}) = \mathcal{G}^{\mathcal{S}^U} \quad \text{and therefore, also using Eq. (1),} \quad \mathcal{G}_U^{\mathcal{S}} = \mathcal{G}_{\Omega}^{\mathcal{S}^U}. \tag{11}$$

---

[20] Here too, it is implicitly assumed that all expressions are well-defined, e.g., that in the second statement, $\underline{P}(f|v)$ is a real number for all $v \in V$, making sure that $\underline{P}(f|V)$ is indeed a gamble.
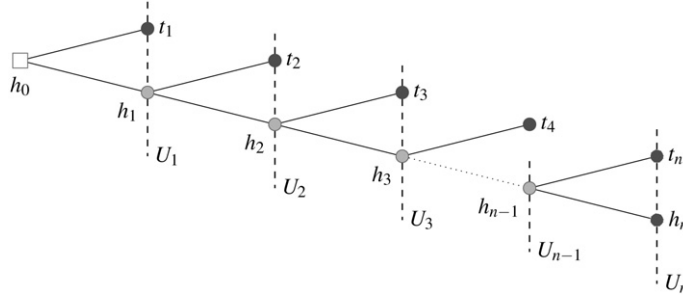
Fig. 6. The event tree for the uncertain process involving n successive coin flips described in Example 3.

We see that stopped gamble processes are gamble processes themselves, that correspond to selections being 'called off' as soon as Reality reaches a cut. This also means that we can actually restrict ourselves to selections $\mathcal{S}$ that are U-called off in Proposition 8.

Proposition 8. *Let t be a non-terminal situation, and let U be a cut of t. Then for any U-measurable t-gamble f, $I_{\uparrow t} f \in \mathcal{E}_{\mathcal{R}}$ if and only is there is some t-selection $\mathcal{S}$ such that $I_{\uparrow t} f \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}^U}$, or equivalently, $f_U \geqslant \mathcal{G}_U^{\mathcal{S}}$. Consequently,*

$$\underline{P}(f|t) = \sup\{\alpha \colon f - \alpha \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}^U} \text{ for some } t\text{-selection } \mathcal{S}\} = \sup\{\alpha \colon f_U - \alpha \geqslant \mathcal{G}_U^{\mathcal{S}} \text{ for some } t\text{-selection } \mathcal{S}\}.$$

If a t-gamble h is measurable with respect to the children cut $C(t)$ of a non-terminal situation t, then we can interpret it as gamble on $\mathbf{W}_t$. For such gambles, the following immediate corollary of Proposition 8 tells us that the predictive lower previsions $\underline{P}(h|t)$ are completely determined by the local model $\mathcal{R}_t$.

Proposition 9. *Let t be a non-terminal situation, and consider a $C(t)$-measurable gamble h. Then $\underline{P}(h|t) = \underline{P}_t(h)$.*

These results tells us that all predictive lower (and upper) previsions can be calculated using backwards recursion, by starting with the trivial predictive previsions $\overline{P}(f|\Omega) = \underline{P}(f|\Omega) = f$ for the terminal cut Ω, and using only the local models $\underline{P}_t$. This is illustrated in the following simple example. We shall come back to this idea in Section 8.

Example 3. Suppose we have $n > 0$ coins. We begin by flipping the first coin: if we get tails, we stop, and otherwise we flip the second coin. Again, we stop if we get tails, and otherwise we flip the third coin, ... In other words, we continue flipping new coins until we get one tails, or until all n coins have been flipped. This leads to the event tree depicted in Fig. 6. Its sample space is $\Omega = \{t_1, t_2, \ldots, t_n, h_n\}$. We will also consider the cuts $U_1 = \{t_1, h_1\}$ of $\square$, $U_2 = \{t_2, h_2\}$ of h_1, $U_3 = \{t_3, h_3\}$ of h_2, ..., and $U_n = \{t_n, h_n\}$ of h_{n-1}. It will be convenient to also introduce the notation h_0 for the initial situation $\square$.

For each of the non-terminal situations h_k, $k = 0, 1, \ldots, n - 1$, Forecaster has beliefs (in $\square$) about what move Reality will make in that situation, i.e., about the outcome of the $k + 1$th coin flip. These beliefs are expressed in terms of a set of really desirable gambles $\mathcal{R}_{h_k}$ on Reality's move space $\mathbf{W}_{h_k}$ in h_k. Each such move space $\mathbf{W}_{h_k}$ can clearly be identified with the children cut U_{k+1} of h_k.

For the purpose of this example, it will be enough to consider the local predictive lower previsions $\underline{P}_{h_k}$ on $\mathcal{G}(U_{k+1})$, associated with $\mathcal{R}_{h_k}$ through Eq. (10). Forecaster assumes all coins to be approximately fair, in the sense that she assesses that the probability of heads for each flip lies between $\frac{1}{2} - \delta$ and $\frac{1}{2} + \delta$, for some $0 < \delta < \frac{1}{2}$. This assessment leads to the following local predictive lower previsions:[21]

$$\underline{P}_{h_k}(g) = (1 - 2\delta)\left[\frac{1}{2}g(h_{k+1}) + \frac{1}{2}g(t_{k+1})\right] + 2\delta \min\{g(h_{k+1}), g(t_{k+1})\}, \tag{12}$$

where g is any gamble on U_{k+1}.

---

[21] These so-called linear-vacuous mixtures, or contamination models, are the natural extensions of the probability assessments $\underline{P}_{h_k}(\{h_{k+1}\}) = \frac{1}{2} - \delta$ and $\overline{P}_{h_k}(\{h_{k+1}\}) = \frac{1}{2} + \delta$; see [33, Chapters 3–4] for more details.

Let us see how we can for instance calculate, from the local predictive models $\underline{P}_{h_k}$, the predictive lower probabilities $\underline{P}(\{h_n\}|s)$ for a gamble f on Ω and any situation s in the tree. First of all, for the terminal situations it is clear from Proposition 4.1 that

$$\underline{P}(\{h_n\}|t_k) = 0 \quad \text{and} \quad \underline{P}(\{h_n\}|h_n) = 1. \tag{13}$$

We now turn to the calculation of $\underline{P}(\{h_n\}|h_{n-1})$. It follows at once from Proposition 9 that $\underline{P}(\{h_n\}|h_{n-1}) = \underline{P}_{h_{n-1}}(\{h_n\})$, and therefore, substituting $g = I_{\{h_n\}}$ in Eq. (12) for $k = n - 1$,

$$\underline{P}(\{h_n\}|h_{n-1}) = \frac{1}{2} - \delta. \tag{14}$$

To calculate $\underline{P}(\{h_n\}|h_{n-2})$, consider that, since $h_{n-1} \sqsubseteq U_{n-1}$,

$$\underline{P}(\{h_n\}|h_{n-2}) = \underline{P}(\underline{P}(\{h_n\}|U_{n-1})|h_{n-2}) = \underline{P}_{h_{n-2}}(\underline{P}(\{h_n\}|U_{n-1}))$$

where the first equality follows from Theorem 7, and the second from Proposition 9, taking into account that $g_{n-1} := \underline{P}(\{h_n\}|U_{n-1})$ is a gamble on the children cut U_{n-1} of h_{n-2}. It follows from Eq. (13) that $g_{n-1}(t_{n-1}) = \underline{P}(\{h_n\}|t_{n-1}) = 0$ and from Eq. (14) that $g_{n-1}(h_{n-1}) = \underline{P}(\{h_n\}|h_{n-1}) = \frac{1}{2} - \delta$. Substituting $g = g_{n-1}$ in Eq. (12) for $k = n - 2$, we then find that

$$\underline{P}(\{h_n\}|h_{n-2}) = \left(\frac{1}{2} - \delta\right)^2. \tag{15}$$

Repeating this course of reasoning, we find that more generally

$$\underline{P}(\{h_n\}|h_k) = \left(\frac{1}{2} - \delta\right)^{n-k}, \quad k = 0, \ldots, n - 1. \tag{16}$$

This illustrates how we can use a backwards recursion procedure to calculate global from local predictive lower previsions.[22]

5. Interpretation of the Matching Theorem

In Shafer and Vovk's approach, there sometimes also appears, besides Reality and Sceptic, a third player, called *Forecaster*. Her role consists in determining what Sceptic's move space $\mathbf{S}_t$ and gain function λ_t are, in each non-terminal situation t. Shafer and Vovk leave largely unspecified just how Forecaster should do that, which makes their approach quite general and abstract.

But the Matching Theorem now tells us that we can connect their approach with Walley's, and therefore inject a notion of belief modelling into their game-theoretic framework. We can do that by being more specific about how Forecaster should determine Sceptic's move spaces $\mathbf{S}_t$ and gain functions λ_t: they should be determined by Forecaster's beliefs (in $\square$) about what Reality will do immediately after getting to non-terminal situations t.[23] Let us explain this more carefully.

Suppose that Forecaster has certain beliefs, *in situation* $\square$, about what move Reality will make next in each non-terminal situation t, and suppose she models those beliefs by specifying a coherent set $\mathcal{R}_t$ of really desirable gambles on $\mathbf{W}_t$. This brings us to the situation described in the previous section.

When Forecaster specifies such a set, she is making certain behavioural commitments: she is committing herself to accepting, in situation $\square$, any gamble in $\mathcal{R}_t$, contingent on Reality getting to situation t, and to accepting any combination of such gambles according to the combination axioms D3, D4 and D5$'$. This implies that we can derive predictive lower previsions $\underline{P}(\cdot|t)$, with the following interpretation: in situation $\square$, $\underline{P}(f|t)$ is the supremum price

---

[22] It also indicates why we need to work in the more general language of lower previsions and gambles, rather than the perhaps more familiar one of lower probabilities and events: even if we only want to calculate a global predictive lower probability, already after one recursion step we need to start working with lower previsions of gambles. More discussion on the prevision/gamble versus probability/event issue can be found in [33, Chapter 4].

[23] The germ for this idea, in the case that Forecaster's beliefs can be expressed using precise probability models on the $\mathcal{G}(\mathbf{W}_t)$, is already present in Shafer's work, see for instance [29, Chapter 8] and [25, Appendix 1]. We extend this idea here to Walley's imprecise probability models.

Forecaster can be made to buy the t-gamble f for, conditional on Reality's getting to t, and on the basis of the commitments she has made in the initial situation $\square$.

What Sceptic can now do, is take Forecaster up on her commitments. This means that in situation $\square$, he can use a selection $\mathcal{S}$, which for each non-terminal situation t, selects a gamble (or equivalently, any non-negative linear combination of gambles) $\mathcal{S}(t) = h_t$ in $\mathcal{R}_t$ and offer the corresponding gamble $\mathcal{G}_\Omega^{\mathcal{S}}$ on Ω to Forecaster, who is bound to accept it. If Reality's next move in situation t is $\mathbf{w} \in \mathbf{W}_t$, this changes Sceptic's capital by (the positive, negative, or zero amount) $-h_t(\mathbf{w})$. In other words, his move space $\mathbf{S}_t$ can then be identified with the convex set of gambles $\mathcal{R}_t$ and his gain function λ_t is then given by $\lambda_t(h_t, \cdot) = -h_t$. But then the *selection* $\mathcal{S}$ can be identified with a *strategy* $\mathcal{P}$ for Sceptic, and $\mathcal{K}_\Omega^{\mathcal{P}} = -\mathcal{G}_\Omega^{\mathcal{S}}$ (this is the essence of the proof of Theorem 6), which tells us that we are led to a coherent probability protocol, and that the corresponding lower prices $\underline{\mathbb{E}}_t$ for Sceptic coincide with Forecaster's predictive lower previsions $\underline{P}(\cdot|t)$.

In a very nice paper [28], Shafer, Gillett and Scherl discuss ways of introducing and interpreting lower previsions in a game-theoretic framework, not in terms of prices that a subject is willing to pay for a gamble, but in terms of whether a subject believes she can make a lot of money (utility) at those prices. They consider such conditional lower previsions both on a contingent and on a dynamic interpretation, and argue that there is equality between them in certain cases. Here, we have decided to stick to the more usual interpretation of lower and upper previsions, and concentrated on the contingent/updating interpretation. We see that on our approach, the game-theoretic framework is useful too.

This is of particular relevance to the laws of large numbers that Shafer and Vovk derive in their game-theoretic framework, because such laws can now be given a behavioural interpretation in terms of Forecaster's predictive lower and upper previsions. To give an example, we now turn to deriving a very general weak law of large numbers.

6. A more general weak law of large numbers

Consider a non-terminal situation t and a cut U of t. Define the t-variable n_U such that $n_U(\omega)$ is the distance $d(t, u)$, measured in moves along the tree, from t to the unique situation u in U that ω goes through. n_U is clearly U-measurable, and $n_U(u)$ is simply the distance $d(t, u)$ from t to u. We assume that $n_U(u) > 0$ for all $u \in U$, or in other words that $U \neq \{t\}$. Of course, in the bounded protocols we are considering here, n_U is bounded, and we denote its minimum by N_U.

Now consider for each s between t and U a *bounded* gamble h_s and a real number m_s such that $h_s - m_s \in \mathcal{R}_s$, meaning that Forecaster in situation $\square$ accepts to buy h_s for m_s, contingent on Reality getting to situation s. Let $B > 0$ be any common upper bound for $\sup h_s - \inf h_s$, for all $t \sqsubseteq s \sqsubset U$. It follows from the coherence of $\mathcal{R}_s$ [D1] that $m_s \leqslant \sup h_s$. To make things interesting, we shall also assume that $\inf h_s \leqslant m_s$, because otherwise $h_s - m_s \geqslant 0$ and accepting this gamble represents no real commitment on Forecaster's part. As a result, we see that $|h_s - m_s| \leqslant \sup h_s - \inf h_s \leqslant B$.

We are interested in the following t-gamble G_U, given by

$$G_U = \frac{1}{n_U} \sum_{t \sqsubseteq s \sqsubset U} I_{\uparrow s}[h_s - m_s],$$

which provides a measure for how much, on average, the gambles h_s yield an outcome above Forecaster's accepted buying prices m_s, along segments of the tree starting in t and ending right before U. In other words, G_U measures the average gain for Forecaster along segments from t to U, associated with commitments she has made and is taken up on, because Reality has to move along these segments. This gamble G_U is U-measurable too. We may therefore interpret G_U as a gamble on U. Also, for any h_s and any $u \in U$, we know that because $s \sqsubset u$, h_s has the same value $h_s(u) := h_s(\omega(s))$ in all ω that go through u. This allows us to write

$$G_U(u) = \frac{1}{n_U(u)} \sum_{t \sqsubseteq s \sqsubset u} [h_s(u) - m_s].$$

We would like to study Forecaster's beliefs (in the initial situation $\square$ and contingent on Reality getting to t) in the occurrence of the event

$$\{G_U \geqslant -\epsilon\} := \{\omega \in \uparrow t: G_U(\omega) \geqslant -\epsilon\},$$

where $\epsilon > 0$. In other words, we want to know $\underline{P}(\{G_U \geqslant -\epsilon\}|t)$, which is Forecaster's supremum rate for betting on the event that his average gain from t to U will be at least $-\epsilon$, contingent on Reality's getting to t.

Theorem 10 *(Weak Law of Large Numbers). For all $\epsilon > 0$,*

$$\underline{P}(\{G_U \geqslant -\epsilon\}|t) \geqslant 1 - \exp\left(-\frac{N_U \epsilon^2}{4B^2}\right).$$

We see that as N_U increases this lower bound increases to one, so the theorem can be very loosely formulated as follows: *As the horizon recedes, Forecaster, if she is coherent, should believe increasingly more strongly that her average gain along any path from the present to the horizon won't be negative.*

This is a very general version of the weak law of large numbers. It can be seen as a generalisation of Hoeffding's inequality for martingale differences [14] (see also [37, Chapter 4] and [30, Appendix A.7]) to coherent lower previsions on event trees.

7. Scoring a predictive model

We now look at an interesting consequence of Theorem 10: we shall see that it can be used to score a predictive model in a manner that satisfies Dawid's *Prequential Principle* [5,6]. We consider the special case of Theorem 10 where $t = \square$.

Suppose Reality follows a path up to some situation u_o in U, which leads to an average gain $G_U(u_o)$ for Forecaster. Suppose this average gain is negative: $G_U(u_o) < 0$. We see that $\uparrow u_o \subseteq \{G_U < -\epsilon\}$ for all $0 < \epsilon < -G_U(u_o)$, and therefore all these events $\{G_U < -\epsilon\}$ have actually occurred (because $\uparrow u_o$ has). On the other hand, Forecaster's upper probability (in $\square$) for their occurrence satisfies $\overline{P}(\{G_U < -\epsilon\}) \leqslant \exp(-\frac{N_U \epsilon^2}{4B^2})$, by Theorem 10. Coherence then tells us that Forecaster's upper probability (in $\square$) for the event $\uparrow u_o$, which has actually occurred, is then at most $S_{N_U}(\gamma_U(u_o))$, where

$$S_N(x) = \exp\left(-\frac{N}{4}x^2\right) \quad \text{and} \quad \gamma_U(u) := \frac{G_U(u_o)}{B}.$$

Observe that $\gamma_U(u_o)$ is a number in $[-1, 0)$, by assumption. Coherence requires that Forecaster, because of her local predictive commitments, can be forced (by Sceptic, if he chooses his strategy well) to bet against the occurrence of the event $\uparrow u_o$ at a rate that is at least $1 - S_{N_U}(\gamma_U(u_o))$. So we see that Forecaster is losing utility because of her local predictive commitments. Just how much depends on how close $\gamma_U(u_o)$ lies to -1, and on how large N_U is; see Fig. 7.

The upper bound $S_{N_U}(\gamma_U(u_o))$ we have constructed for the upper probability of $\uparrow u_o$ has a very interesting property, which we now try to make more explicit. Indeed, if we were to calculate Forecaster's upper probability $\overline{P}(\uparrow u_o)$ for $\uparrow u_o$ directly using Eq. (9), this value would generally depend on Forecaster's predictive assessments $\mathcal{R}_s$ for situations s that don't precede u_o, and that Reality therefore never got to. We shall see that such is not the case for the upper bound $S_{N_U}(\gamma_U(u_o))$ constructed using Theorem 10.
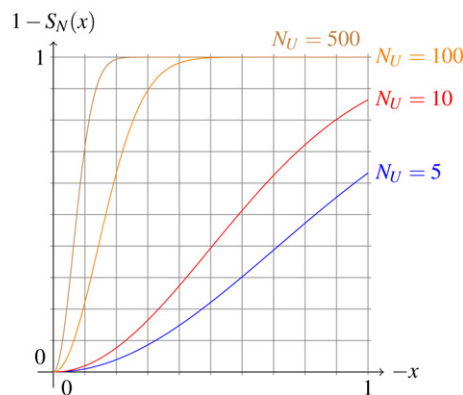

Fig. 7. What Forecaster can be made to pay, $1 - S_N(x)$, as a function of $x = \gamma_U(u)$, for different values of $N = N_U$.

Consider any situation s before U but not on the path through u_o, meaning that Reality never got to this situation s. Therefore the corresponding gamble $h_s - m_s$ in the expression for G_U isn't used in calculating the value of $G_U(u_o)$, so we can change it to anything else, and still obtain the same value of $G_U(u_o)$.

Indeed, consider any other predictive model, where the only thing we ask is that the $\mathcal{R}'_s$ coincide with the $\mathcal{R}_s$ for all s that precede u_o. For other s, the $\mathcal{R}'_s$ can be chosen arbitrarily, but still coherently. Now construct a new average gain gamble G'_U for this alternative predictive model, where the only restriction is that we let $h'_s = h_s$ and $m'_s = m_s$ if s precedes u_o. We know from the reasoning above that $G'_U(u_o) = G_U(u_o)$, so the new upper probability that the event $\uparrow u_o$ will be observed is at most

$$S_{N_U}\left(\frac{G'_U(u_o)}{B}\right) = S_{N_U}\left(\frac{G_U(u_o)}{B}\right) = S_{N_U}\big(\gamma_U(u_o)\big).$$

In other words, the upper bound $S_N(\gamma_U(u))$ we found for Forecaster's upper probability of Reality getting to a situation *u_o depends only on Forecaster's local predictive assessments $\mathcal{R}_s$ for situations s that Reality has actually got to, and not on her assessments for other situations*. This means that this method for scoring a predictive model satisfies Dawid's *Prequential Principle*; see for instance [5,6].

8. Concatenation and backwards recursion

As we have discovered in Section 4.5, Theorem 7 and Proposition 9 enable us to calculate the global predictive lower previsions $\underline{P}(\cdot|t)$ in imprecise probability trees from local predictive lower previsions $\underline{P}_s$, $s \sqsupseteq t$, using a backwards recursion method. That this is possible in probability trees, where the probability models are precise (previsions), is well-known,[24] and was arguably discovered by Christiaan Huygens in the middle of the 17th century.[25] It allows for an exponential, dynamic programming-like reduction in the complexity of calculating previsions (or expectations); it seems to be essentially this phenomenon that leads to the computational efficiency of such machine learning tools as, for instance, Needleman and Wunsch's [21] sequence alignment algorithm.

In this section, we want to give an illustration of such exponential reduction in complexity, by looking at a problem involving Markov chains. Assume that the state $X(n)$ of a system at consecutive times $n = 1, 2, \ldots, N$ can assume any value in a finite set $\mathcal{X}$. Forecaster has some beliefs about the state $X(1)$ at time 1, leading to a coherent lower prevision $\underline{P}_1$ on $\mathcal{G}(\mathcal{X})$. She also assesses that when the system jumps from state $X(n) = x_n$ to a new state $X(n+1)$, where the system goes to will only depend on the state $X(n)$ the system was in at time n, and not on the states $X(k)$ of the system at previous times $k = 1, 2, \ldots, n - 1$. Her beliefs about where the system in $X(n) = x_n$ will go to at time $n + 1$ are represented by a lower prevision $\underline{P}_{x_n}$ on $\mathcal{G}(\mathcal{X})$.

The time evolution of this system can be modelled as Reality traversing an event tree. An example of such a tree for $\mathcal{X} = \{a, b\}$ and $N = 3$ is given in Fig. 8. The situations of the tree have the form $(x_1, \ldots, x_k) \in \mathcal{X}^k$, $k = 0, 1, \ldots, N$; for $k = 0$ this gives some abuse of notation as we let $\mathcal{X}^0 := \{\square\}$. In each cut $X^k := \mathcal{X}^k$ of $\square$, the value $X(k)$ of the state at time k is revealed.

This leads to an imprecise probability tree with local predictive models $\underline{P}_\square := \underline{P}_1$ and

$$\underline{P}_{(x_1,\ldots,x_k)} = \underline{P}_{x_k} \tag{17}$$

expressing the usual *Markov conditional independence condition*, but here in terms of lower previsions. For notational convenience, we now introduce a (generally non-linear) *lower transition operator* $\underline{T}$ on the linear space $\mathcal{G}(\mathcal{X})$ as follows:

$$\underline{T}: \mathcal{G}(\mathcal{X}) \to \mathcal{G}(\mathcal{X}): x \mapsto \underline{P}_x(f),$$

---

[24] See Chapter 3 of Shafer's book [26] on causal reasoning in probability trees. This chapter contains a number of propositions about calculating probabilities and expectations in probability trees that find their generalisations in Sections 4.3 and 4.5. For instance, Theorem 7 generalises Proposition 3.11 in [26] to imprecise probability trees.

[25] See Appendix A of Shafer's book [26]. Shafer discusses Huygens's treatment of a special case of the so-called *Problem of Points*, where Huygens draws what is probably the first recorded probability tree, and solves the problem by backwards calculation of expectations in the tree. Huygens's treatment can be found in Appendix VI of [15].
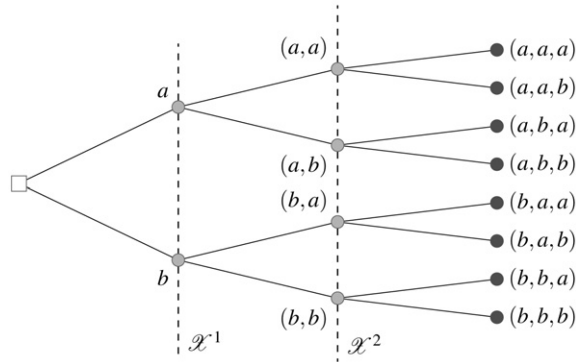
Fig. 8. The event tree for the time evolution of system that can be in two states, a and b, and can change state at each time instant $n = 1, 2, 3$. Also depicted are the respective cuts $\mathcal{X}^1$ and $\mathcal{X}^2$ of $\square$ where the state at times 1 and 2 are revealed.

or in other words, $\underline{\mathrm{T}}f$ is a gamble on $\mathcal{X}$ whose value $\underline{\mathrm{T}}f(x)$ in the state $x \in \mathcal{X}$ is given by $\underline{P}_x(f)$. The lower transition operator $\underline{\mathrm{T}}$ completely describes Forecaster's beliefs about how the system changes its state from one instant to the next.

We now want to find the corresponding model for Forecaster's beliefs (in $\square$) about the state the system will be in at time n. So let us consider a gamble f_n on $\mathcal{X}^N$ that actually only depends on the value $X(n)$ of X at this time n. We then want to calculate its lower prevision $\underline{P}(f_n) := \underline{P}(f_n|\square)$.

Consider a time instant $k \in \{0, 1, \ldots, n-1\}$, and a situation $(x_1, \ldots, x_k) \in \mathcal{X}^k$. For the children cut $C(x_1, \ldots, x_k) := \{(x_1, \ldots, x_k, x_{k+1}): x_{k+1} \in \mathcal{X}\}$ of $(x_1, \ldots, x_k)$, we see that $\underline{P}(f_n|C(x_1, \ldots, x_k))$ is a gamble that only depends on the value of $X(k+1)$ in $\mathcal{X}$, and whose value in x_{k+1} is given by $\underline{P}(f_n|x_1, \ldots, x_{k+1})$. We then find that

$$\underline{P}(f_n|x_1, \ldots, x_k) = \underline{P}\big(\underline{P}(f_n|C(x_1, \ldots, x_k))|x_1, \ldots, x_k\big) = \underline{P}_{x_k}\big(\underline{P}(f_n|C(x_1, \ldots, x_k))\big), \qquad (18)$$

where the first equality follows from Theorem 7, and the second from Proposition 9 and Eq. (17). We first apply Eq. (18) for $k = n - 1$. By Proposition 5.2, $\underline{P}(f_n|C(x_1, \ldots, x_{n-1})) = f_n$, so we are led to $\underline{P}(f_n|x_1, \ldots, x_{n-1}) = \underline{P}_{x_{n-1}}(f_n) = \underline{\mathrm{T}}f_n(x_{n-1})$, and therefore

$$\underline{P}\big(f_n|C(x_1, \ldots, x_{n-2})\big) = \underline{\mathrm{T}}f_n.$$

Substituting this in Eq. (18) for $k = n - 2$, yields $\underline{P}(f_n|x_1, \ldots, x_{n-2}) = \underline{P}_{x_{n-2}}(\underline{\mathrm{T}}f_n)$, and therefore

$$\underline{P}\big(f_n|C(x_1, \ldots, x_{n-3})\big) = \underline{\mathrm{T}}^2 f_n.$$

Proceeding in this fashion until we get to $k = 1$, we get $\underline{P}(f_n|C(\square)) = \underline{\mathrm{T}}^{n-1} f_n$, and going one step further to $k = 0$, Eq. (18) yields $\underline{P}(f_n|\square) = \underline{P}_\square(\underline{P}(f_n|C(\square)))$ and therefore

$$\underline{P}(f_n) = \underline{P}_1\big(\underline{\mathrm{T}}^{n-1} f_n\big). \qquad (19)$$

We see that the complexity of calculating $\underline{P}(f_n)$ in this way is essentially *linear* in the number of time steps n.

In the literature on imprecise probability models for Markov chains [2,17,31,32], another so-called *credal set*, or *set of probabilities*, approach is generally used to calculate $\underline{P}(f_n)$. The point we want to make here is that such an approach typically has a worse (exponential) complexity in the number of time steps. To see this, recall [33] that a lower prevision $\underline{P}$ on $\mathcal{G}(\mathcal{X})$ that is derived from a coherent set of really desirable gambles, corresponds to a convex closed set $\mathcal{M}(\underline{P})$ of probability mass functions p on $\mathcal{X}$, called a *credal set*, and given by

$$\mathcal{M}(\underline{P}) := \big\{p: \; (\forall g \in \mathcal{G}(\mathcal{X}))\underline{P}(g) \leqslant E_p(g)\big\}$$

where we let $E_p(g) := \sum_{x \in \mathcal{X}} p(x)g(x)$ be the expectation of the gamble g associated with the mass function p; E_p is a linear prevision in the language of Section 3.2. It then also holds that for all gambles g on $\mathcal{X}$,

$$\underline{P}(g) = \min\big\{E_p(g): \; p \in \mathcal{M}(\underline{P})\big\} = \min\big\{E_p(g): \; p \in \mathrm{ext}\,\mathcal{M}(\underline{P})\big\}$$

where $\mathrm{ext}\,\mathcal{M}(\underline{P})$ is the set of extreme points of the convex closed set $\mathcal{M}(\underline{P})$. Typically on this approach, $\mathrm{ext}(\mathcal{M}(\underline{P}))$ is assumed to be finite, and then $\mathcal{M}(\underline{P})$ is called a *finitely generated credal set*. See for instance [3,4] for a discussion of credal sets with applications to Bayesian networks.

Then $\underline{P}(f_n)$ can also be calculated as follows:[26] Choose for each non-terminal situation $t = (x_1, \ldots, x_k) \in \mathcal{X}^k$, $k = 0, 1, \ldots, n - 1$, a mass function p_t in the set $\mathcal{M}(\underline{P}_t)$ given by Eq. (17), or equivalently, in its set of extreme points $\text{ext}\,\mathcal{M}(\underline{P}_t)$. This leads to a (precise) probability tree for which we can calculate the corresponding expectation of f_n. Then $\underline{P}(f_n)$ is the minimum of all such expectations, calculated for all possible assignments of mass functions to the nodes. We see that, roughly speaking, when all $\mathcal{M}(\underline{P}_t)$ have a typical number of extreme points M, then the complexity of calculating $\underline{P}(f_n)$ will be essentially M^n, i.e., exponential in the number of time steps.

This shows that the 'lower prevision' approach can for some problems lead to more efficient algorithms than the 'credal set' approach. This may be especially relevant for probabilistic inferences involving graphical models, such as credal networks [3,4]. Another nice example of this phenomenon, concerned with checking coherence for precise and imprecise probability models, is due to Walley et al. [36].

9. Additional remarks

We have proved the correspondence between the two approaches only for event trees with a bounded horizon. For games with infinite horizon, the correspondence becomes less immediate, because Shafer and Vovk implicitly make use of coherence axioms that are stronger than D1–D4 and D5′, leading to lower prices that dominate the corresponding predictive lower previsions. Exact matching would be restored of course, provided we could argue that these additional requirements are rational for any subject to comply with. This could be an interesting topic for further research.

We haven't paid much attention to the special case that the coherent lower previsions and their conjugate upper previsions coincide, and are therefore (precise) *previsions* or *fair prices* in de Finetti's [11] sense. When all the local predictive models $\underline{P}_t$ (see Proposition 9) happen to be precise, meaning that $\underline{P}_t(f) = \overline{P}_t(f) = -\underline{P}_t(-f)$ for all gambles f on $\mathbf{W}_t$, then the immediate prediction model we have described in Section 4 becomes very closely related, and arguably identical to, the probability trees introduced and studied by Shafer in [26]. Indeed, we then get predictive previsions $P(\cdot|s)$ that can be obtained through concatenation of the local models P_t, as guaranteed by Theorem 7.[27]

Moreover, as indicated in Section 8, it is possible to prove lower envelope theorems to the effect that (i) the local lower previsions $\underline{P}_t$ correspond to lower envelopes of sets $\mathcal{M}_t$ of local previsions P_t; (ii) each possible choice of previsions P_t in $\mathcal{M}_t$ over all non-terminal situations t, leads to a *compatible* probability tree in Shafer's [26] sense, with corresponding predictive previsions $P(\cdot|s)$; and (iii) the predictive lower previsions $\underline{P}(\cdot|s)$ are the lower envelopes of the predictive previsions $P(\cdot|s)$ for the compatible probability trees. Of course, the law of large numbers of Section 6 remains valid for probability trees.

Finally, we want to recall that Theorem 7 and Proposition 9 allow for a calculation of the predictive models $\underline{P}(\cdot|s)$ using only the local models and *backwards recursion*, in a manner that is strongly reminiscent of dynamic programming techniques. This should allow for a much more efficient computation of such predictive models than, say, an approach that exploits lower envelope theorems and sets of probabilities/previsions. We think that there may be lessons to be learnt from this for dealing with other types of graphical models, such as credal networks [3,4], as well.

What makes this more efficient approach possible is, ultimately, the Marginal Extension Theorem (Theorem 3), which leads to the Concatenation Formula (Theorem 7), i.e., to the specific equality, rather than the general inequalities, in Proposition 2.7. Generally speaking (see for instance [33, Section 6.7] and [20]), such marginal extension results can be proved because the models that Forecaster specifies are *local*, or *immediate* prediction models: they relate to her beliefs, in each non-terminal situation t, about what move Reality is going to make *immediately after* getting to t.

Acknowledgements

---

[26] An explicit proof of this statement would take us to far, but it is an immediate application of Theorems 3 and 4 in [20].

[27] This should for instance be compared with Proposition 3.11 in [26].

Erik Quaeghebeur have read and commented on earlier drafts. We are also grateful for the insightful and generous comments of three reviewers, which led us to better discuss the significance and potential applications of our results, and helped us improve the readability of this paper.

Appendix A. Proofs of main results

In this appendix, we have gathered proofs for the most important results in the paper.

A.1. Proofs of results in Section 3

We begin with a proof of Proposition 2. Although similar results were proved for bounded gambles by Walley [33], and by Williams [39] before him, our proof also works for the extension to possibly unbounded gambles we are considering in this paper.

Proof of Proposition 2. For the first statement, we only give a proof for the first two inequalities. The proof for the remaining inequality is similar. For the first inequality, we may assume without loss of generality that $\inf\{\omega \in B \colon f(\omega)\} > -\infty$ and is therefore a real number, which we denote by β. So we know that $I_B(f - \beta) \geqslant 0$ and therefore $I_B(f - \beta) \in \mathcal{R}$, by D2. It then follows from Eq. (5) that $\beta \leqslant \underline{P}(f|B)$. To prove the second inequality, assume *ex absurdo* that $\overline{P}(f|B) < \underline{P}(f|B)$, then it follows from Eqs. (4) and (5) that there are real α and β such that $\beta < \alpha$, $I_B(f - \alpha) \in \mathcal{R}$ and $I_B(\beta - f) \in \mathcal{R}$. By D3, $I_B(\beta - \alpha) = I_B(f - \alpha) + I_B(\beta - f) \in \mathcal{R}$, but this contradicts D1, since $I_B(\beta - \alpha) < 0$.

We now turn to the second statement. As announced in Footnote 11, we may assume that the sum of the terms $\underline{P}(f_1|B)$ and $\underline{P}(f_2|B)$ is well-defined. If either of these terms is equal to $-\infty$, the resulting inequality then holds trivially, so we may assume without loss of generality that both terms are strictly greater than $-\infty$. Consider any real $\alpha < \underline{P}(f_1|B)$ and $\beta < \underline{P}(f_2|B)$, then by Eq. (5) we see that both $I_B(f_1 - \alpha) \in \mathcal{R}$ and $I_B(f_2 - \beta) \in \mathcal{R}$. Hence $I_B[(f_1 + f_2) - (\alpha + \beta)] \in \mathcal{R}$, by D3, and therefore $\underline{P}(f_1 + f_2|B) \geqslant \alpha + \beta$, using Eq. (5) again. Taking the supremum over all real $\alpha < \underline{P}(f_1|B)$ and $\beta < \underline{P}(f_2|B)$ leads to the desired inequality.

To prove the third statement, first consider $\lambda > 0$. Since by D4, $I_B(\lambda f - \alpha) \in \mathcal{R}$ if and only if $I_B(f - \alpha/\lambda) \in \mathcal{R}$, we get, using Eq. (5)

$$\underline{P}(\lambda f|B) = \sup\{\alpha \colon I_B(\lambda f - \alpha) \in \mathcal{R}\} = \sup\{\lambda \beta \colon I_B(f - \beta) \in \mathcal{R}\} = \lambda \underline{P}(f|B).$$

For $\lambda = 0$, consider that $\underline{P}(0|B) = \sup\{\alpha \colon -I_B\alpha \in \mathcal{R}\} = 0$, where the last equality follows from D1 and D2.

For the fourth statement, use Eq. (5) to find that

$$\underline{P}(f + \alpha|B) = \sup\{\beta \colon I_B(f + \alpha - \beta) \in \mathcal{R}\} = \sup\{\alpha + \gamma \colon I_B(f - \gamma) \in \mathcal{R}\} = \alpha + \underline{P}(f|B).$$

The fifth statement is an immediate consequence of the first.

To prove the sixth statement, observe that $f_1 \leqslant f_2$ implies that $I_B(f_2 - f_1) \geqslant 0$ and therefore $I_B(f_2 - f_1) \in \mathcal{R}$, by D2. Now consider any real α such that $I_B(f_1 - \alpha) \in \mathcal{R}$, then by D3, $I_B(f_2 - \alpha) = I_B(f_1 - \alpha) + I_B(f_2 - f_1) \in \mathcal{R}$. Hence

$$\{\alpha \colon I_B(f_1 - \alpha) \in \mathcal{R}\} \subseteq \{\alpha \colon I_B(f_2 - \alpha) \in \mathcal{R}\}$$

and by taking suprema and considering Eq. (5), we deduce that indeed $\underline{P}(f_1|B) \leqslant \underline{P}(f_2|B)$.

For the final statement, assume that $\underline{P}(f|C)$ is a real number for all $C \in \mathcal{B}$. Also observe that $\underline{P}(f|D) = \underline{P}(f I_D|D)$ for all non-empty D. Define the gamble g as follows: $g(\omega) := \underline{P}(f|C)$ for all $\omega \in C$, where $C \in \mathcal{B}$. We have to prove that $\underline{P}(g|B) \leqslant \underline{P}(f|B)$. We may assume without loss of generality that $\underline{P}(g|B) > -\infty$ [because otherwise the inequality holds trivially]. Fix $\epsilon > 0$, and consider the gamble $I_B(f - g + \epsilon)$. Also consider any $C \in \mathcal{B}$. If $C \subseteq B$ then $I_C I_B(f - g + \epsilon) = I_C(f - \underline{P}(f|C) + \epsilon) \in \mathcal{R}$, using Eq. (5). If $C \cap B = \emptyset$ then again $I_C I_B(f - g + \epsilon) = 0 \in \mathcal{R}$, by D2. Since $\mathcal{R}$ is $\mathcal{B}$-conglomerable, it follows that $I_B(f - g + \epsilon) \in \mathcal{R}$, whence $\underline{P}(f - g|B) \geqslant -\epsilon$, again using Eq. (5). Hence $\underline{P}(h|B) \geqslant 0$, where $h := f - g$. Consequently,

$$\underline{P}(f|B) = \underline{P}(h + g|B) \geqslant \underline{P}(h|B) + \underline{P}(g|B) \geqslant \underline{P}(g|B),$$

where we use the second statement, and the fact that $\underline{P}(g|B) > -\infty$ and $\underline{P}(h|B) \geqslant 0$ implies that the sum on the right-hand side of the inequality is well-defined as an extended real number. □

A.2. Proofs of results in Section 4

Proof of Theorem 3. We have already argued that any coherent set of really desirable gambles that includes $\mathcal{R}$, must contain all gambles $\mathcal{G}^\mathcal{S}$ [by D3 and D5′]. By D2 and D3, it must therefore include the set $\mathcal{E}_\mathcal{R}$. If we can show that $\mathcal{E}_\mathcal{R}$ is coherent, i.e., satisfies D1–D4 and D5′, then we have proved that $\mathcal{E}_\mathcal{R}$ is the natural extension of $\mathcal{R}$. This is what we now set out to do.

We first show that D1 is satisfied. It clearly suffices to show that for no selection $\mathcal{S}$, it holds that $\mathcal{G}_\Omega^\mathcal{S} < 0$. This follows at once from Lemma 2 below.

To prove that D2 holds, consider the selection $\mathcal{S}_0 := 0$, then $\mathcal{G}^{\mathcal{S}_0} = 0$, and if $f \geqslant 0$ it follows that $f \geqslant \mathcal{G}^{\mathcal{S}_0}$ whence indeed $f \in \mathcal{E}_\mathcal{R}$.

To prove that D3 and D4 hold, consider any f_1 and f_2 in $\mathcal{E}_\mathcal{R}$, and any non-negative real numbers a_1 and a_2. We know there are selections $\mathcal{S}_1$ and $\mathcal{S}_2$ such that $f_1 \geqslant \mathcal{G}^{\mathcal{S}_1}$ and $f_2 \geqslant \mathcal{G}^{\mathcal{S}_2}$. But $a_1\mathcal{S}_1 + a_2\mathcal{S}_2$ is a selection as well [because the $\mathcal{R}_t$ satisfy D3 and D4], and $\mathcal{G}^{a_1\mathcal{S}_1+a_2\mathcal{S}_2} = a_1\mathcal{G}^{\mathcal{S}_1} + a_2\mathcal{G}^{\mathcal{S}_2} \leqslant a_1 f_1 + a_2 f_2$, whence indeed $a_1 f_1 + a_2 f_2 \in \mathcal{E}_\mathcal{R}$.

To conclude, we show that D5′ is satisfied. Consider any cut U of $\square$. Consider a gamble f and assume that $I_{\uparrow u} f \in \mathcal{E}_\mathcal{R}$ for all $u \in U$. We must prove that $f \in \mathcal{E}_\mathcal{R}$. Let $U_t := U \cap \Omega$ and $U_{nt} := U \setminus \Omega$, so U is the disjoint union of U_t and U_{nt}. For $\omega \in U_t$, $I_{\uparrow\omega} f = I_{\uparrow\omega} f(\omega) \in \mathcal{E}_\mathcal{R}$ implies that $f(\omega) \geqslant 0$, by D1. For $u \in U_{nt}$, we invoke Lemma 3 to find that there is some u-selection $\mathcal{S}_u$ such that $I_{\uparrow u} f \geqslant \mathcal{G}_\Omega^{\mathcal{S}_u}$. Now construct a selection $\mathcal{S}$ as follows. Consider any s in $\Omega^\lozenge \setminus \Omega$. If $u \sqsubseteq s$ for some [unique, because U is a cut] $u \in U_{nt}$, let $\mathcal{S}(s) := \mathcal{S}_u(s)$. Otherwise let $\mathcal{S}(s) := 0$. Then

$$\mathcal{G}^\mathcal{S} = \sum_{u \in U_{nt}} I_{\uparrow u} \mathcal{G}^{\mathcal{S}_u} \leqslant \sum_{u \in U_{nt}} I_{\uparrow u} f \leqslant \sum_{u \in U} I_{\uparrow u} f = f,$$

so indeed $f \in \mathcal{E}_\mathcal{R}$; the first equality can be seen as immediate, or as a consequence of Lemma 1, and the second inequality holds because we have just shown that $f(\omega) \geqslant 0$ for all $\omega \in U_t$. The rest of the proof now follows from Lemma 3. $\square$

Lemma 1. *Let t be any non-terminal situation, and let U be any cut of t. Consider a t-selection $\mathcal{S}$, and let, for any $u \in U \setminus \Omega$, $\mathcal{S}_u$ be the u-selection given by $\mathcal{S}_u(s) = \mathcal{S}(s)$ if the non-terminal situation s follows u, and $\mathcal{S}_u(s) := 0$ otherwise. Moreover, let $\mathcal{S}^U$ be the U-called off t-selection for $\mathcal{S}$ (as defined after Theorem 7). Then*

$$\mathcal{G}_\Omega^\mathcal{S} = \sum_{u \in U \cap \Omega} I_{\uparrow u} \mathcal{G}^\mathcal{S}(u) + \sum_{u \in U \setminus \Omega} I_{\uparrow u}\big[\mathcal{G}^\mathcal{S}(u) + \mathcal{G}_\Omega^{\mathcal{S}_u}\big]$$

$$= \mathcal{G}_U^\mathcal{S} + \sum_{u \in U \setminus \Omega} I_{\uparrow u} \mathcal{G}_\Omega^{\mathcal{S}_u} = \mathcal{G}_\Omega^{\mathcal{S}^U} + \sum_{u \in U \setminus \Omega} I_{\uparrow u} \mathcal{G}_\Omega^{\mathcal{S}_u}.$$

Proof. It is immediate that the second equality holds; see Eq. (11) for the third. For the first equality, it obviously suffices to consider the values of the left- and right-hand sides in any $\omega \in {\uparrow}u$ for $u \in U \setminus \Omega$. The value of the right-hand side is then, using Eqs. (6) and (7),

$$\mathcal{G}^\mathcal{S}(u) + \mathcal{G}_\Omega^{\mathcal{S}_u}(\omega) = \sum_{t \sqsubseteq s \sqsubset u} \mathcal{S}(s)(u) + \sum_{u \sqsubseteq s \sqsubset \omega} \mathcal{S}(s)(\omega) = \sum_{t \sqsubseteq s \sqsubset \omega} \mathcal{S}(s)(\omega) = \mathcal{G}_\Omega^\mathcal{S}(\omega). \qquad \square$$

Lemma 2. *Consider any non-terminal situation t and any t-selection $\mathcal{S}$. Then it doesn't hold that $\mathcal{G}_\Omega^\mathcal{S} < 0$ (on ${\uparrow}t$). As a corollary, consider any cut U of t, and the gamble $\mathcal{G}_U^\mathcal{S}$ on U defined by $\mathcal{G}_U^\mathcal{S}(u) = \mathcal{G}^\mathcal{S}(u)$. Then it doesn't hold that $\mathcal{G}_U^\mathcal{S} < 0$ (on U).*

Proof. Define the set $P_\mathcal{S} := \{s \in \Omega^\lozenge \setminus \Omega : t \sqsubseteq s \text{ and } \mathcal{S}(s) \geqslant 0\}$, and its (relative) complement $N_\mathcal{S} := \{s \in \Omega^\lozenge \setminus \Omega : t \sqsubseteq s \text{ and } \mathcal{S}(s) \not\geqslant 0\}$. If $N_\mathcal{S} = \emptyset$ then $\mathcal{G}_\Omega^\mathcal{S} \geqslant 0$, by Eq. (7), so we can assume without loss of generality that $N_\mathcal{S}$ is non-empty. Consider any minimal element t_1 of $N_\mathcal{S}$, meaning that there is no s in $N_\mathcal{S}$ such that $s \sqsubset t_1$ [there is such a minimal element in $N_\mathcal{S}$ because of the bounded horizon assumption]. So for all $t \sqsubseteq s \sqsubset t_1$ we have that $\mathcal{S}(s) \geqslant 0$. Choose $\mathbf{w}_1$ in $\mathbf{W}_{t_1}$ such that $\mathcal{S}(t_1)(\mathbf{w}_1) > 0$ [this is possible because $\mathcal{R}_{t_1}$ satisfies D1]. This brings us to the situation $t_2 := t_1\mathbf{w}_1$. If $t_2 \in N_\mathcal{S}$, then choose $\mathbf{w}_2$ in $\mathbf{W}_{t_2}$ such that $\mathcal{S}(t_2)(\mathbf{w}_2) > 0$ [again possible by D1]. If $t_2 \in P_\mathcal{S}$ then we

know that $\mathcal{S}(t_2)(\mathbf{w}_2) \geqslant 0$ for any choice of $\mathbf{w}_2$ in $\mathbf{W}_{t_2}$. We can continue in this way until we reach a terminal situation $\omega = t_1 \mathbf{w}_1 \mathbf{w}_2 \ldots$ after a finite number of steps [because of the bounded horizon assumption]. Moreover

$$\mathcal{G}_{\Omega}^{\mathcal{S}}(\omega) = \sum_{t \sqsubset t_1} \mathcal{S}(t)\big(\omega(t)\big) + \sum_k \mathcal{S}(t_k)(\mathbf{w}_k) \geqslant 0 + \mathcal{S}(t_1)(\mathbf{w}_1) + 0 > 0.$$

It therefore can't hold that $\mathcal{G}_{\Omega}^{\mathcal{S}} < 0$ (on $\uparrow t$).

To prove the second statement, consider the U-called off t-selection $\mathcal{S}^U$ derived from $\mathcal{S}$ by letting $\mathcal{S}^U(s) := \mathcal{S}(s)$ if s (follows t and) strictly precedes some u in U, and zero otherwise. Then $\mathcal{G}^{\mathcal{S}}(u) = \sum_{t \sqsubseteq s \sqsubset u} \mathcal{S}(s)(u) = \mathcal{G}_{\Omega}^{\mathcal{S}^U}(\omega)$ for all ω that go through u, where $u \in U$ [see also Eq. (11)]. Now apply the above result for the t-selection $\mathcal{S}^U$. □

Lemma 3. *Consider any non-terminal situation t and any gamble f. Then $I_{\uparrow t} f \in \mathcal{E}_{\mathcal{R}}$ if and only if there is some t-selection $\mathcal{S}_t$ such that $I_{\uparrow t} f \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}_t}$ (on $\uparrow t$).*

Proof. It clearly suffices to prove the necessity part. Assume therefore that $I_{\uparrow t} f \in \mathcal{E}_{\mathcal{R}}$, meaning [definition of the set $\mathcal{E}_{\mathcal{R}}$] that there is some selection $\mathcal{S}$ such that $I_{\uparrow t} f \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}}$. Let $\mathcal{S}_t$ be the t-selection defined by letting $\mathcal{S}_t(s) := \mathcal{S}(s)$ if $t \sqsubseteq s$, and zero otherwise. It follows from Lemma 1 [use the cut of $\square$ made up of t and the terminal situations that do not follow t] that

$$I_{\uparrow t} f \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}} = I_{\uparrow t}[\mathcal{G}^{\mathcal{S}}(t) + \mathcal{G}_{\Omega}^{\mathcal{S}_t}] + \sum_{\omega' \notin \uparrow t} I_{\uparrow \omega'} \mathcal{G}_{\Omega}^{\mathcal{S}}(\omega'),$$

whence, for all $\omega \in \Omega$,

$$\mathcal{G}_{\Omega}^{\mathcal{S}}(\omega) \leqslant 0, \quad \omega \not\sqsupseteq t, \tag{A.1}$$

$$\mathcal{G}^{\mathcal{S}}(t) + \mathcal{G}_{\Omega}^{\mathcal{S}_t}(\omega) \leqslant f(\omega), \quad \omega \sqsupseteq t. \tag{A.2}$$

Then, by (A.2), the proof is complete if we can prove that $\mathcal{G}^{\mathcal{S}}(t) \geqslant 0$. Assume *ex absurdo* that $\mathcal{G}^{\mathcal{S}}(t) < 0$. Consider the cut of $\square$ made up of t and the terminal situations that don't follow t. Applying Lemma 2 for this cut and for the initial situation $\square$, we see that there must be some $\omega \in \Omega \setminus \uparrow t$ such that $\mathcal{G}_{\Omega}^{\mathcal{S}}(\omega) > 0$. But this contradicts (A.1). □

Proof of Proposition 4. For the first statement, consider a terminal situation ω and a gamble f on Ω. Then $\uparrow \omega = \{\omega\}$ and therefore $I_{\uparrow \omega}(f - \alpha) = I_{\{\omega\}}(f(\omega) - \alpha) \in \mathcal{E}_{\mathcal{R}}$ if and only if $\alpha \leqslant f(\omega)$, by D1 and D2. Using Eq. (8), we find that indeed $\underline{P}(f|\omega) = f(\omega)$. By conjugacy, $\overline{P}(f|\omega) = -\underline{P}(-f|\omega) = -(-f(\omega)) = f(\omega)$ as well.

For the second statement, use Eq. (8) and observe that $I_{\uparrow t}(f - \alpha) = I_{\uparrow t}(f I_{\uparrow t} - \alpha)$. The last statement is an immediate consequence of the second and Proposition 2.6. □

Proof of Proposition 5. The first statement follows from Eq. (8) if we observe that $I_{\uparrow t}(I_{\uparrow t} - \alpha) = I_{\uparrow t}(1 - \alpha) \in \mathcal{E}_{\mathcal{R}}$ if and only if $\alpha \leqslant 1$, by D1 and D2.

For the second statement, consider any $u \in U$, then we must show that $\underline{P}(g|u) = g_U(u)$. But the U-measurability of g tells us that $I_{\uparrow u}(g - \alpha) = I_{\uparrow u}(g_U(u) - \alpha)$, and this gamble belongs to $\mathcal{E}_{\mathcal{R}}$ if and only if $\alpha \leqslant g_U(u)$, by D1 and D2. Now use Eq. (8).

The proofs of the third and fourth statements are similar, and based on the observation that $I_{\uparrow u}(f + g - \alpha) = I_{\uparrow u}(f + g_U(u) - \alpha)$ and $I_{\uparrow u}(gf - \alpha) = I_{\uparrow u}(g_U(u)f - \alpha)$. □

Proof of Theorem 6. First, consider an immediate prediction model $\mathcal{R}_t$, $t \in \Omega^{\diamond} \setminus \Omega$. Define Sceptic's move spaces to be $\mathbf{S}_t := \mathcal{R}_t$ and his gain functions $\lambda_t \colon \mathbf{S}_t \times \mathbf{W}_t$ by $\lambda_t(h, \mathbf{w}) := -h(\mathbf{w})$ for all $h \in \mathcal{R}_t$ and $\mathbf{w} \in \mathbf{W}_t$. Clearly P1 and P2 are satisfied, because each $\mathcal{R}_t$ is a convex cone by D3 and D4. But so is the coherence requirement C. Indeed, if it weren't satisfied there would be some non-terminal situation t and some gamble h in $\mathcal{R}_t$ such that $h(\mathbf{w}) < 0$ for all $\mathbf{w}$ in $\mathbf{W}_t$, contradicting the coherence requirement D1 for $\mathcal{R}_t$. We are thus led to a coherent probability protocol. We show there is matching. Consider any non-terminal situation t, and any t-selection $\mathcal{S}$. For all terminal situations $\omega \sqsupseteq t$,

$$\mathcal{G}_{\Omega}^{\mathcal{S}}(\omega) = \sum_{t \sqsubseteq u \sqsubset \omega} \mathcal{S}(u)\big(\omega(u)\big) = \sum_{t \sqsubseteq u \sqsubset \omega} -\lambda_u\big(\mathcal{S}(u), \omega(u)\big) = -\mathcal{K}_{\Omega}^{\mathcal{S}}(\omega),$$

or in other words, selections and strategies are in a one-to-one correspondence (are actually the same things), and the corresponding gamble and capital processes are each other's additive inverses. It is therefore immediate from Eqs. (3) and (9) that $\mathbb{E}_t = \underline{P}(\cdot|t)$.

Conversely, consider a coherent probability protocol with move spaces $\mathbf{S}_t$ and gain functions $\lambda_t \colon \mathbf{S}_t \times \mathbf{W}_t$ for all non-terminal t. Define $\mathcal{R}'_t := \{-\lambda_t(\mathbf{s}, \cdot) \colon \mathbf{s} \in \mathbf{S}_t\}$. By a similar argument to the one above, we see that $\underline{P}'(\cdot|t) = \mathbb{E}_t$, where the $\underline{P}'(\cdot|t)$ are the predictive lower previsions associated with the sets $\mathcal{R}'_t$. But each $\mathcal{R}'_t$ is a convex cone of gambles by P1 and P2, and by C we know that for all non-terminal situations t and all gambles h in $\mathcal{R}'_t$ there is some $\mathbf{w}$ in $\mathbf{W}_t$ such that $h(\mathbf{w}) \geqslant 0$. This means that the conditions for Lemma 4 are satisfied, and therefore also $\underline{P}'(\cdot|t) = \underline{P}(\cdot|t)$, where the $\underline{P}(\cdot|t)$ are the predictive lower previsions associated with the immediate prediction model $\mathcal{R}_t$ that is the smallest convex cone containing all non-negative gambles and including $\{-\lambda_t(\mathbf{s}, \cdot) + \delta \colon \mathbf{s} \in \mathbf{S}_t, \delta > 0\}$. □

Lemma 4. *Consider, for each non-terminal situation $t \in \Omega^\Diamond \setminus \Omega$, a set of gambles $\mathcal{R}'_t$ on $\mathbf{W}_t$ such that* (i) *$\mathcal{R}'_t$ is a convex cone, and* (ii) *for all $h \in \mathcal{R}'_t$ there is some $\mathbf{w}$ in $\mathbf{W}_t$ such that $h(\mathbf{w}) \geqslant 0$. Then each set $\mathcal{R}_t := \{\alpha(h + \delta) + f \colon h \in \mathcal{R}'_t, \delta > 0, f \geqslant 0, \alpha \geqslant 0\}$ is a coherent set of really desirable gambles on $\mathbf{W}_t$. Moreover, all predictive lower previsions obtained using the sets $\mathcal{R}_t$ coincide with the ones obtained using the $\mathcal{R}'_t$.*

Proof. Fix a non-terminal situation t. We first show that $\mathcal{R}_t$ is a coherent set of really desirable gambles, i.e., that D1–D4 are satisfied. Observe that $\mathcal{R}_t$ is the smallest convex cone of gambles including the set $\{h + \delta \colon h \in \mathcal{R}'_t, \delta > 0\}$ and containing all non-negative gambles. So D2–D4 are satisfied. To prove that D1 holds, consider any $g < 0$ and assume *ex absurdo* that $g \in \mathcal{R}_t$. Then there are h in $\mathcal{R}'_t$, $\delta > 0$, $f \geqslant 0$ and $\alpha \geqslant 0$ such that $0 > g = \alpha(h + \delta) + f$, whence $\alpha(h + \delta) < 0$ and therefore $\alpha > 0$ and $h + \delta < 0$. But by (ii), there is some $\mathbf{w}$ in $\mathbf{W}_t$ such that $h(\mathbf{w}) \geqslant 0$, whence $h(\mathbf{w}) + \delta > 0$. This contradicts $h + \delta < 0$.

We now move to the second part. Consider any gamble f on Ω. Fix t in $\Omega^\Diamond \setminus \Omega$ and $\epsilon > 0$. First consider any t-selection $\mathcal{S}'$ associated with the $\mathcal{R}'_s$, i.e., such that $\mathcal{S}'(s) \in \mathcal{R}'_s$ for all $s \sqsupseteq t$. Since Reality can only make a finite and *bounded* number of moves, whatever happens, it is possible to choose $\delta_s > 0$ for each non-terminal $s \sqsupseteq t$ such that $\sum_{t \sqsubseteq s \sqsubset \omega} \delta_s < \epsilon$ for all ω in Ω that follow t. Define the t-selection $\mathcal{S}$ associated with the $\mathcal{R}_s$ by $\mathcal{S}(s) := \mathcal{S}'(s) + \delta_s \in \mathcal{R}_s$ for all non-terminal s that follow t. Clearly $\mathcal{G}^\mathcal{S}_\Omega \leqslant \epsilon + \mathcal{G}^{\mathcal{S}'}_\Omega$, and therefore

$$\underline{P}'(f|t) = \sup_{\mathcal{S}'} \sup \{\alpha \colon f - \alpha \geqslant \mathcal{G}^{\mathcal{S}'}_\Omega\} \leqslant \sup_{\mathcal{S}'} \sup \{\alpha \colon f - \alpha + \epsilon \geqslant \mathcal{G}^\mathcal{S}_\Omega\}$$

$$= \sup_{\mathcal{S}'} \sup \{\alpha \colon f - \alpha \geqslant \mathcal{G}^\mathcal{S}_\Omega\} + \epsilon \leqslant \underline{P}(f|t) + \epsilon.$$

Since this inequality holds for all $\epsilon > 0$, we find that $\underline{P}'(f|t) \leqslant \underline{P}(f|t)$.

Conversely, consider any t-selection $\mathcal{S}$ associated with the $\mathcal{R}_s$. For all $s \sqsupseteq t$, we have that there are h_s in $\mathcal{R}'_s$, $\delta_s > 0$, $f_s \geqslant 0$ and $\alpha_s \geqslant 0$ such that $\mathcal{S}(s) = \alpha_s(h_s + \delta_s) + f_s$. Define the t-selection $\mathcal{S}'$ associated with the $\mathcal{R}'_s$ by $\mathcal{S}'(s) := \alpha_s h_s = \mathcal{S}(s) - \alpha_s \delta_s - f_s \leqslant \mathcal{S}(s)$. Clearly then also $\mathcal{G}^{\mathcal{S}'}_\Omega \leqslant \mathcal{G}^\mathcal{S}_\Omega$, and therefore

$$\underline{P}(f|t) = \sup_{\mathcal{S}} \sup \{\alpha \colon f - \alpha \geqslant \mathcal{G}^\mathcal{S}_\Omega\} \leqslant \sup_{\mathcal{S}} \sup \{\alpha \colon f - \alpha \geqslant \mathcal{G}^{\mathcal{S}'}_\Omega\} \leqslant \underline{P}'(f|t).$$

This proves that indeed $\underline{P}'(f|t) = \underline{P}(f|t)$. □

Proof of Theorem 7. It isn't difficult to see that the second statement is a consequence of the first, so we only prove the first statement.

Consider any t-gamble f on Ω. Recall that it is implicitly assumed that $\underline{P}(f|U)$ is again a t-gamble. Then we have to prove that $\underline{P}(f|t) = \underline{P}(\underline{P}(f|U)|t)$. Let, for ease of notation, $g := \underline{P}(f|U)$, so the t-gamble g is U-measurable, and we have to prove that $\underline{P}(f|t) = \underline{P}(g|t)$. Now, there are two possibilities.

First, if t is a terminal situation ω, then, on the one hand, $\underline{P}(f|t) = f(\omega)$ by Proposition 4.1. On the other hand, again by Proposition 4.1,

$$\underline{P}(g|t) = g(\omega) = \underline{P}(f|U)(\omega).$$

Now, since U is a cut of $t = \omega$, the unique element u of U that $t = \omega$ goes through, is $u = \omega$, and therefore $\underline{P}(f|U)(\omega) = \underline{P}(f|\omega) = f(\omega)$, again by Proposition 4.1. This tells us that in this case indeed $\underline{P}(f|t) = \underline{P}(g|t)$.

Secondly, suppose that t is not a terminal situation. Then it follows from Proposition 2.7 and the cut conglomerability of $\mathcal{E}_{\mathcal{R}}$ that $\underline{P}(f|t) \geqslant \underline{P}(\underline{P}(f|U)|t) = \underline{P}(g|t)$ [recall that $\underline{P}(\cdot|t) = \underline{P}(\cdot|\uparrow t)$ and that $\underline{P}(\cdot|U) = \underline{P}(\cdot|\mathcal{B}_U)$]. It therefore remains to prove the converse inequality $\underline{P}(f|t) \leqslant \underline{P}(g|t)$. Choose $\epsilon > 0$, then using Eq. (9) we see that there is some t-selection $\mathcal{S}$ such that $f - \underline{P}(f|t) + \epsilon \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}}$ on all paths that go through t. Invoke Lemma 1, using the notations introduced there, to find that

$$f - \underline{P}(f|t) + \epsilon \geqslant \mathcal{G}_U^{\mathcal{S}} + \sum_{u \in U \setminus \Omega} I_{\uparrow u} \mathcal{G}_{\Omega}^{\mathcal{S}_u} \quad \text{(on } \uparrow t). \tag{A.3}$$

Now consider any $u \in U$. If u is a terminal situation ω, then by Proposition 4.1, $g(u) = \underline{P}(f|\omega) = f(\omega)$, and therefore Eq. (A.3) yields

$$g(\omega) - \underline{P}(f|t) + \epsilon \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}^U}(\omega), \tag{A.4}$$

also taking into account that $\mathcal{G}_U^{\mathcal{S}} = \mathcal{G}_{\Omega}^{\mathcal{S}^U}$ [see Eq. (11)]. If u is not a terminal situation then for all $\omega \in \uparrow u$, Eq. (A.3) yields

$$f(\omega) - \underline{P}(f|t) + \epsilon \geqslant \mathcal{G}_U^{\mathcal{S}}(u) + \mathcal{G}_{\Omega}^{\mathcal{S}_u}(\omega),$$

and since $\mathcal{S}_u$ is a u-selection, this inequality together with Eq. (9) tells us that $\underline{P}(f|u) \geqslant \underline{P}(f|t) - \epsilon + \mathcal{G}_U^{\mathcal{S}}(u)$, and therefore, for all $\omega \in \uparrow u$,

$$g(\omega) - \underline{P}(f|t) + \epsilon \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}^U}(\omega). \tag{A.5}$$

If we combine the inequalities (A.4) and (A.5), and recall Eq. (9), we get that $\underline{P}(g|t) \geqslant \underline{P}(f|t) - \epsilon$. Since this holds for all $\epsilon > 0$, we may indeed conclude that $\underline{P}(g|t) \geqslant \underline{P}(f|t)$. $\quad\square$

Proof of Proposition 8. The condition is clearly sufficient, so let us show that it is also necessary. Suppose that $I_{\uparrow t} f \in \mathcal{E}_{\mathcal{R}}$, then there is some t-selection $\mathcal{S}$ such that $f \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}}$, by Theorem 3 [or Lemma 3]. Define, for any $u \in U \setminus \Omega$, the selection $\mathcal{S}_u$ as follows: $\mathcal{S}_u(s) := \mathcal{S}(s)$ if $s \sqsupseteq u$ and $\mathcal{S}_u(s) := 0$ elsewhere. Then, by Lemma 1,

$$\mathcal{G}_{\Omega}^{\mathcal{S}} = \mathcal{G}_U^{\mathcal{S}} + \sum_{u \in U \setminus \Omega} I_{\uparrow u} \mathcal{G}_{\Omega}^{\mathcal{S}_u}.$$

Now fix any u in U. If u is a terminal situation ω, then it follows from the equality above that

$$f_U(u) = f(\omega) \geqslant \mathcal{G}_U^{\mathcal{S}}(u).$$

If u is not a terminal situation, we get for all $\omega \in \uparrow u$:

$$f_U(u) = f(\omega) \geqslant \mathcal{G}_U^{\mathcal{S}}(u) + \mathcal{G}_{\Omega}^{\mathcal{S}_u}(\omega),$$

whence, by taking the supremum of all $\omega \in \uparrow u$,

$$f_U(u) \geqslant \mathcal{G}_U^{\mathcal{S}}(u) + \sup_{\omega \in \uparrow u} \mathcal{G}_{\Omega}^{\mathcal{S}_u}(\omega) \geqslant \mathcal{G}_U^{\mathcal{S}}(u),$$

where the last inequality follows since $\sup_{\omega \in \uparrow u} \mathcal{G}_{\Omega}^{\mathcal{S}_u}(\omega) \geqslant 0$ by Lemma 2 [with $t = u$ and $\mathcal{S} = \mathcal{S}_u$]. Now recall that $f_U \geqslant \mathcal{G}_U^{\mathcal{S}}(u)$ is equivalent to $I_{\uparrow t} f \geqslant \mathcal{G}_{\Omega}^{\mathcal{S}^U}$ [see Eq. (11)]. $\quad\square$

A.3. Proofs of results in Section 6

Proof of Theorem 10. This proof builds on an intriguing idea, used by Shafer and Vovk in a different situation and form; see [29, Lemma 3.3].

Because $|h_s - m_s| \leqslant B$ for all $t \sqsubseteq s \sqsubset u$, it follows that $G_U(u) \geqslant -B$, and it therefore suffices to prove the inequality for $\epsilon < B$. We work with the upper probability $\overline{P}(\Delta_{t,\epsilon}^c|t)$ of the complementary event $\Delta_{t,\epsilon}^c := \{G_U < -\epsilon\}$. It is given by

$$\inf\{\alpha \colon \alpha - \mathcal{G}_{\Omega}^{\mathcal{S}} \geqslant I_{\Delta_{t,\epsilon}^c} \text{ for some } t\text{-selection } \mathcal{S}\}. \tag{A.6}$$

Because G_U is U-measurable, we can (and will) consider $\Delta_{t,\epsilon}^c$ as an event on U. In the expression (A.6), we may assume that $\alpha \geqslant 0$. Indeed, if we had $\alpha < 0$ and $\alpha - \mathcal{G}_\Omega^\mathcal{S} \geqslant I_{\Delta_{t,\epsilon}^c}$ for some t-selection $\mathcal{S}$, then it would follow that $\mathcal{G}_\Omega^\mathcal{S} \leqslant \alpha < 0$, contradicting Lemma 2. Fix therefore $\alpha > 0$ and $\delta > 0$ and consider the selection $\mathcal{S}$ such that $\mathcal{S}(s) := \lambda_s (h_s - m_s) \in \mathcal{R}_s$ for all $t \sqsubseteq s \sqsubset U$ and let $\mathcal{S}(s)$ be zero elsewhere. Here

$$\lambda_s := \alpha\delta \prod_{t \sqsubseteq v \sqsubset s} \left[1 + \delta\big(m_v - h_v(s)\big)\right] = \alpha\delta \prod_{t \sqsubseteq v \sqsubset s} \left[1 + \delta\big(m_v - h_v(u)\big)\right], \tag{A.7}$$

where u is any element of U that follows s. Recall again that $-B \leqslant h_s - m_s \leqslant B$, so if we choose $\delta < \frac{1}{2B}$, we are certainly guaranteed that $\lambda_s > 0$ and therefore indeed $\lambda_s(h_s - m_s) \in \mathcal{R}_s$. After some elementary manipulations we get for any $u \in U$ and any $\omega \in \uparrow u$:

$$\mathcal{G}_\Omega^\mathcal{S}(\omega) = \sum_{t \sqsubseteq s \sqsubset u} \big(h_s(u) - m_s\big)\lambda_s = \sum_{t \sqsubseteq s \sqsubset u} \big(h_s(u) - m_s\big)\alpha\delta \prod_{t \sqsubseteq v \sqsubset s} \left[1 + \delta\big(m_v - h_v(u)\big)\right]$$

where the second equality follows from Eq. (A.7). [The gamble $\mathcal{G}_\Omega^\mathcal{S}$ is U-measurable.] If we let $\xi_s := m_s - h_s(u)$ for ease of notation, then we get

$$\mathcal{G}_\Omega^\mathcal{S}(u) = -\alpha \sum_{t \sqsubseteq s \sqsubset u} \delta\xi_s \prod_{t \sqsubseteq v \sqsubset s} [1 + \delta\xi_v] = \alpha \sum_{t \sqsubseteq s \sqsubset u} \prod_{t \sqsubseteq v \sqsubseteq s} [1 + \delta\xi_v] - \alpha \sum_{t \sqsubseteq s \sqsubset u} \prod_{t \sqsubseteq v \sqsubset s} [1 + \delta\xi_v]$$

$$= \alpha - \alpha \prod_{t \sqsubseteq v \sqsubset u} [1 + \delta\xi_v] = \alpha - \alpha \prod_{t \sqsubseteq v \sqsubset u} \left[1 + \delta\big(m_v - h_v(u)\big)\right]$$

for all u in U. Then it follows from (A.6) that if we can find an $\alpha \geqslant 0$ such that

$$\alpha \prod_{t \sqsubseteq v \sqsubset u} \left[1 + \delta\big(m_v - h_v(u)\big)\right] \geqslant 1$$

whenever u belongs to $\Delta_{t,\epsilon}^c$, then this α is an upper bound for $\overline{P}(\Delta_{t,\epsilon}^c|t)$. By taking logarithms on both sides of the inequality above, we get the equivalent condition

$$\ln\alpha + \sum_{t \sqsubseteq s \sqsubset u} \ln\left[1 + \delta\big(m_s - h_s(u)\big)\right] \geqslant 0. \tag{A.8}$$

Since $\ln(1 + x) \geqslant x - x^2$ for $x > -\frac{1}{2}$, and $\delta(m_s - h_s(u)) \geqslant -\delta B > -\frac{1}{2}$ by our previous restrictions on δ, we find

$$\sum_{t \sqsubseteq s \sqsubset u} \ln\left[1 + \delta\big(m_s - h_s(u)\big)\right] \geqslant \sum_{t \sqsubseteq s \sqsubset u} \delta\big(m_s - h_s(u)\big) - \sum_{t \sqsubseteq s \sqsubset u} \left[\delta\big(m_s - h_s(u)\big)\right]^2$$

$$\geqslant \delta \sum_{t \sqsubseteq s \sqsubset u} \left[m_s - h_s(u)\right] - \delta^2 n_U(u) B^2$$

$$= n_U(u)\delta\left[-G_U(u) - B^2\delta\right].$$

But for all $u \in \Delta_{t,\epsilon}^c$, $-G_U(u) > \epsilon$, so for all such u

$$\sum_{t \sqsubseteq s \sqsubset u} \ln\left[1 + \delta\big(m_s - h_s(u)\big)\right] > n_U(u)\delta(\epsilon - B^2\delta).$$

If we therefore choose α such that for all $u \in U$, $\ln\alpha + n_U(u)\delta(\epsilon - B^2\delta) \geqslant 0$, or equivalently $\alpha \geqslant \exp(-n_U(u)\delta(\epsilon - B^2\delta))$, then the above condition (A.8) will indeed be satisfied for all $u \in \Delta_{t,\epsilon}^c$, and then α is an upper bound for $\overline{P}(\Delta_{t,\epsilon}^c|t)$. The tightest (smallest) upper bound is always (for all $u \in U$) achieved for $\delta = \frac{\epsilon}{2B^2}$. Replacing n_U by its minimum N_U allows us to get rid of the u-dependence, so we see that $\overline{P}(\Delta_{t,\epsilon}^c|t) \leqslant \exp(-\frac{N_U\epsilon^2}{4B^2})$. We previously required that $\delta < \frac{1}{2B}$, so if we use this value for δ, we find that we have indeed proved this inequality for $\epsilon < B$. $\quad\square$

References

[1] G. Boole, The Laws of Thought, Dover Publications, New York, 1847, reprint 1961.

[2] M.A. Campos, G.P. Dimuro, A.C. da Rocha Costa, V. Kreinovich, Computing 2-step predictions for interval-valued finite stationary Markov chains, Technical Report UTEP-CS-03-20a, University of Texas at El Paso, 2003.

[3] F.G. Cozman, Credal networks, Artificial Intelligence 120 (2000) 199–233.

[4] F.G. Cozman, Graphical models for imprecise probabilities, International Journal of Approximate Reasoning 39 (2–3) (2005) 167–184.

[5] A.Ph. Dawid, Statistical theory: The prequential approach, Journal of the Royal Statistical Society, Series A 147 (1984) 278–292.

[6] A.Ph. Dawid, V.G. Vovk, Prequential probability: Principles and properties, Bernoulli 5 (1999) 125–162.

[7] G. de Cooman, F. Hermans, On coherent immediate prediction: Connecting two theories of imprecise probability, in: G. de Cooman, J. Vejnarova, M. Zaffalon (Eds.), ISIPTA '07—Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, 2007, pp. 107–116.

[8] G. de Cooman, E. Miranda, Symmetry of models versus models of symmetry, in: W.L. Harper, G.R. Wheeler (Eds.), Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr., King's College Publications, 2007, pp. 67–149.

[9] G. de Cooman, M. Zaffalon, Updating beliefs with incomplete observations, Artificial Intelligence 159 (1–2) (2004) 75–125.

[10] B. de Finetti, Teoria delle Probabilità, Einaudi, Turin, 1970.

[11] B. de Finetti, Theory of Probability: A Critical Introductory Treatment, John Wiley & Sons, Chichester, 1974–1975. English translation of [10], two volumes.

[12] P. Gärdenfors, N.-E. Sahlin, Decision, Probability, and Utility, Cambridge University Press, Cambridge, 1988.

[13] M. Goldstein, The prevision of a prevision, Journal of the American Statistical Society 87 (1983) 817–819.

[14] W. Hoeffding, Probability inequalities for sums of bounded random variables, Journal of the American Statistical Association 58 (1963) 13–30.

[15] Ch. Huygens, Van Rekeningh in Spelen van Geluck, 1656–1657. Reprinted in vol. XIV of [16].

[16] Ch. Huygens, Œuvres complètes de Christiaan Huygens, Martinus Nijhoff, Den Haag, 1888–1950. Twenty-two volumes. Available in digitised form from the Bibliothèque nationale de France, http://gallica.bnf.fr.

[17] I.O. Kozine, L.V. Utkin, Interval-valued finite Markov chains, Reliable Computing 8 (2) (2002) 97–113.

[18] H.E. Kyburg Jr., H.E. Smokler (Eds.), Studies in Subjective Probability, Wiley, New York, 1964. Second edition (with new material) 1980.

[19] C. Manski, Partial Identification of Probability Distributions, Springer-Verlag, New York, 2003.

[20] E. Miranda, G. de Cooman, Marginal extension in the theory of coherent lower previsions, International Journal of Approximate Reasoning 46 (1) (2007) 188–225.

[21] S.B. Needleman, C.D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology 48 (1970) 443–453.

[22] F.P. Ramsey, Truth and probability (1926), in: R.B. Braithwaite (Ed.), The Foundations of Mathematics and other Logical Essays, Kegan, Paul, Trench, Trubner & Co., London, 1931, pp. 156–198. Chapter VII, reprinted in [18] and [12].

[23] G. Shafer, Bayes's two arguments for the rule of conditioning, The Annals of Statistics 10 (1982) 1075–1089.

[24] G. Shafer, A subjective interpretation of conditional probability, Journal of Philosophical Logic 12 (1983) 453–466.

[25] G. Shafer, Conditional probability, International Statistical Review 53 (1985) 261–277.

[26] G. Shafer, The Art of Causal Conjecture, The MIT Press, Cambridge, MA, 1996.

[27] G. Shafer, P.R. Gillett, R. Scherl, The logic of events, Annals of Mathematics and Artificial Intelligence 28 (2000) 315–389.

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

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

[30] V. Vovk, A. Gammerman, G. Shafer, Algorithmic Learning in a Random World, Springer, New York, 2005.

[31] D. Škulj, Finite discrete time Markov chains with interval probabilities, in: J. Lawry, E. Miranda, A. Bugarin, S. Li, M.A. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.), Soft Methods for Integrated Uncertainty Modelling, Springer, Berlin, 2006, pp. 299–306.

[32] D. Škulj, Regular finite Markov chains with interval probabilities, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), ISIPTA '07—Proceedings of the Fifth International Symposium on Imprecise Probability: Theories and Applications, SIPTA, 2007, pp. 405–413.

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

[34] P. Walley, Measures of uncertainty in expert systems, Artificial Intelligence 83 (1) (1996) 1–58.

[35] P. Walley, Towards a unified theory of imprecise probability, International Journal of Approximate Reasoning 24 (2000) 125–148.

[36] P. Walley, R. Pelessoni, P. Vicig, Direct algorithms for checking consistency and making inferences from conditional probability assessments, Journal of Statistical Planning and Inference 126 (2004) 119–151.

[37] L. Wasserman, All of Statistics, Springer, New York, 2004.

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

[39] P.M. Williams, Notes on conditional previsions, International Journal of Approximate Reasoning 44 (2007) 366–383. Revised journal version of [38].