On coherent immediate prediction: Connecting two theories of imprecise probability

Gert de Cooman and **Filip Hermans**
SYSTeMS Research Group
Ghent University
Technologiepark – Zwijnaarde 914, 9052 Zwijnaarde, Belgium
{gert.decooman,filip.hermans}@UGent.be

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 immediate prediction in the framework of Walley's theory, and we prove an interesting and quite general version of the weak law of large numbers.

Keywords. Game-theoretic probability, imprecise probabilities, coherence, conglomerability, event tree, lower prevision, immediate prediction, Prequential Principle, law of large numbers, Hoeffding's inequality.

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 [18], which is introduced in Section 2, and Peter Walley's behavioural theory [20], 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 [10], Bruno de Finetti [4] and Peter Williams [24] in trying to establish a rational model for a subject's beliefs in terms of her behaviour. Shafer and Vovk follow an approach that is strongly coloured by ideas about gambling systems and martingales. They use Cournot's Principle to interpret lower and upper probabilities (see [17]; and [18, 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, and in particular in Sections 4 and 5, is to show that in many practical situations, the two approaches are strongly connected.[1] 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* [1, 2].

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

In their game-theoretic approach to probability [18], Shafer and Vovk consider a game with two players, World and Skeptic, who play according to a certain *protocol*. They obtain the most interesting results for what they call *coherent probability protocols*. This section is devoted to explaining what this means.

G1. The first player, World, can make 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 Skeptic.

This means that we can represent his game-play by an event tree (see also [14, 16] for more information about event trees). We restrict ourselves here to the discussion of *bounded protocols*, where World makes only a finite and bounded number of moves from the beginning to the end of the game, whatever happens. But we do not exclude the possibility that at some point in the tree, World has the choice between an infinite number of next moves.

[1] Our line of reasoning here should be compared to the one in [17], where Shafer *et al.* use the game-theoretic framework developed in [18] 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.

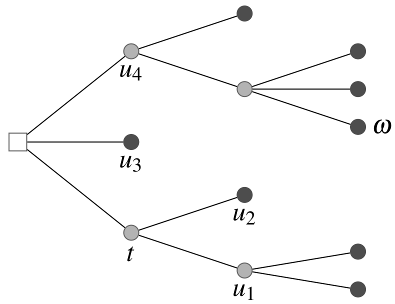

Figure 1: A simple event tree for World, displaying the initial situation $\square$, other non-terminal situations (such as t) as grey circles, and paths, or terminal situations, (such as ω) as black circles. Also depicted is a cut of $\square$, consisting of the situations u_1, u_2, u_3 and u_4.

Let us establish some terminology related to World's event tree. A *path* in the tree represents a possible sequence of moves for World 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 World 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 $\square$, which represents the empty initial segment. See Figure 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 $s \sqsubseteq t$. 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 $s \neq t$. Denote by $\uparrow t := \{\omega \in \Omega \colon t \sqsubseteq \omega\}$ the set of all paths that go through t. If we call any subset of Ω an *event*, then $\uparrow t$ is the event that corresponds to World getting to a situation t. It is clear that not all events will be of the type $\uparrow t$.[2]

Any (partial) function on $\Omega^\diamond$ is called a *process*, and any process whose domain includes all situations that follow a situation t is called a *t-process*. A special 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. In the bounded protocols we are considering here, there is a natural number D such that $d(t,s) \leq D$ for all $s \sqsupseteq t$.

Similarly, any (partial) function on 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 $\mathscr{F}$ to the set $\uparrow t$ of all terminal situations that follow t, we obtain a t-variable, which we denote by $\mathscr{F}_\Omega$.

Call a *cut* U of a situation t any set of situations that (i) follow t, and (ii) such that for all paths ω through t [$t \sqsubseteq \omega$], there is a unique $u \in U$ that ω goes through [$u \sqsubseteq \omega$]; see also Figure 1. A set U of situations is a cut of t if and only if the corresponding set $\{\uparrow u \colon u \in U\}$ is a partition of $\uparrow t$. 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 the *terminal cut*. $\uparrow t$ is the corresponding terminal cut of a situation t.

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 ω that go through u. In that case we can also consider g as a variable on U, which we can denote as g_U.

If $\mathscr{F}$ is a t-process, then with any cut U of t we can associate a t-variable $\mathscr{F}_U$, which assumes the same value $\mathscr{F}_U(\omega) := \mathscr{F}(u)$ 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 $\mathscr{F}_\Omega$ introduced earlier. Similarly, we can associate with $\mathscr{F}$ a new, *U-stopped*, t-process $U(\mathscr{F})$, as follows:

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

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

We call a *move* $\mathbf{w}$ for World in a non-terminal situation t any arc that connects t to one of its children $s \in C(t)$, meaning that $s = t\mathbf{w}$ is the concatenation of the segment t and the arc $\mathbf{w}$. World's *move space* in t is the set $\mathbf{W}_t$ of those moves $\mathbf{w}$ that World can make in t: $\mathbf{W}_t = \{\mathbf{w} \colon t\mathbf{w} \in C(t)\}$. We have already mentioned that $\mathbf{W}_t$ may be infinite. But it should contain at least two elements (otherwise there is no choice for World to make).

We now turn to the other player, Skeptic. His possible moves may well depend on the previous moves that World 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 Skeptic's *move space* in t.

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

[2] Shafer [15] 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 only they may occur as part of World's game-play.

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 Skeptic's capital in situation t when he makes move $\mathbf{s}$ and World makes move $\mathbf{w}$.

Let us introduce some further notions and terminology related to Skeptic's game-play. A *strategy* $\mathscr{P}$ for Skeptic is a partial process defined on the set $\Omega^\Diamond \setminus \Omega$ of non-terminal situations, such that $\mathscr{P}(t) \in \mathbf{S}_t$ is the move that Skeptic will make in each non-terminal situation t. With each such strategy $\mathscr{P}$ there corresponds a *capital process* $\mathscr{K}^\mathscr{P}$, whose value in each situation t gives us Skeptic's capital accumulated so far, when he starts out with zero capital and plays according to the strategy $\mathscr{P}$. It is given by the recursion relation

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

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

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

Assumptions G1 and G2 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. Then the *upper price* $\overline{\mathbb{E}}_t(f)$ *for* f *in* t is defined as the infimum capital α that Skeptic has to start out with in t in order that there would be some t-strategy $\mathscr{P}$ such that his accumulated capital $\alpha + \mathscr{K}^\mathscr{P}$ allows him, at the end of the game, to hedge f, whatever moves World makes after t:

$$\overline{\mathbb{E}}_t(f) := \inf\left\{\alpha\colon\, \alpha + \mathscr{K}^\mathscr{P}_\Omega \geq f \text{ for some } t\text{-strategy } \mathscr{P}\right\}, \tag{1}$$

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

$$\underline{\mathbb{E}}_t(f) := \sup\left\{\alpha\colon\, \alpha - \mathscr{K}^\mathscr{P}_\Omega \leq f \text{ for some } t\text{-strategy } \mathscr{P}\right\}, \tag{2}$$

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

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, Skeptic'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, Skeptic'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*[3] 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 $\lambda_t(\mathbf{s},\mathbf{w}) \leq 0$.

It is clear what this last requirement means: in each non-terminal situation, World has a strategy for playing from t onwards such that Skeptic cannot (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 real 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, $f_1 \leq f_2$ is taken to mean that $f_1(\omega) \leq f_2(\omega)$ for all paths ω that go through t.

Proposition 1 (Properties of lower and upper prices in a coherent probability protocol [18])**.** *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. $\inf_{\omega\in\uparrow t} f(\omega) \leq \underline{\mathbb{E}}_t(f) \leq \overline{\mathbb{E}}_t(f) \leq \sup_{\omega\in\uparrow t} f(\omega)$ *[convexity];*
2. $\underline{\mathbb{E}}_t(f_1 + f_2) \geq \underline{\mathbb{E}}_t(f_1) + \underline{\mathbb{E}}_t(f_2)$ *[super-additivity];*
3. $\underline{\mathbb{E}}_t(\lambda f) = \lambda\underline{\mathbb{E}}_t(f)$ *for all real* $\lambda \geq 0$ *[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. $f_1 \leq f_2$ *implies that* $\underline{\mathbb{E}}_t(f_1) \leq \underline{\mathbb{E}}_t(f_2)$ *[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.

[3] For a discussion of the use of 'coherent' here, we refer to [17, Appendix C].

3 Walley's behavioural approach to probability

In his book on the behavioural theory of imprecise probabilities [20], 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 [21].

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*[4] on Ω is a real-valued map on Ω. 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 or negative. If a subject *accepts* a gamble f, this means that she is willing to engage in the transaction where, (i) first it is determined which ω obtains, and then (ii) she receives the reward $f(\omega)$. We can try and model the subject's beliefs about Ω by considering which gambles she accepts.

Suppose our subject specifies some set $\mathscr{R}$ 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 $f \notin \mathscr{R}$ [avoiding partial loss];

D2. if $f \geq 0$ then $f \in \mathscr{R}$ [accepting partial gain];

D3. if f_1 and f_2 belong to $\mathscr{R}$ then their (point-wise) sum $f_1 + f_2$ also belongs to $\mathscr{R}$ [combination];

D4. if f belongs to $\mathscr{R}$ then its (point-wise) scalar product λf also belongs to $\mathscr{R}$ for all non-negative real numbers λ [scaling].

Here '$f < 0$' means '$f \leq 0$ and not $f = 0$'. Walley has also argued that sets of really desirable gambles should satisfy an additional axiom, where I_B denotes the *indicator* of the event B [a gamble that assumes the value one on B and zero elsewhere]:

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

Full conglomerability is a very strong requirement, and it is not without controversy. If a model $\mathscr{R}$ is $\mathscr{B}$-conglomerable, this means that certain inconsistency problems when conditioning on elements B of $\mathscr{B}$ are avoided; see [20, Section 6.8] for more details and examples. Conglomerability of belief models was not required by forerunners of Walley, such as Williams [24],[5] or de Finetti [4]. 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.[6] This is the path we shall follow in Section 4.

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,

$$\overline{P}(f|B) := \inf\{\alpha \colon I_B(\alpha - f) \in \mathscr{R}\} \quad (3)$$
$$\underline{P}(f|B) := \sup\{\alpha \colon I_B(f - \alpha) \in \mathscr{R}\}, \quad (4)$$

so $\underline{P}(f|B) = -\overline{P}(-f|B)$, and $\underline{P}(f|B)$ 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. 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. (4)], $\underline{P}(f|B)$ is a conditional lower prevision on what Walley [20, 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 $I_B(f - \underline{P}(f|B) + \varepsilon)$, $\varepsilon > 0$, 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* [20, 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 that B* has occurred. For precise conditional previsions, this seems to be the interpretation considered in [6, 11, 12, 17]. It is

[4] Walley [20] assumes gambles to be bounded. We make no such assumption here. It seems the concept of a really desirable gamble (at least formally) allows for such a generalisation, because the coherence axioms for real desirability, as opposed to those for Walley's related notions of almost- and strict desirability, nowhere hinge on such a boundedness assumption, at least not from a technical mathematical point of view.

[5] Axioms (D1)–(D4), but not (D5), were actually suggested by Williams. 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.

[6] The view expressed here seems related to Shafer's, as sketched near the end of [13, Appendix 1].

far from obvious that there should be a relation between the first two and the third interpretations.[7] We shall briefly come back to this distinction in the following sections.

For a partition $\mathscr{B}$ of Ω, we let $\underline{P}(f|\mathscr{B}) := \sum_{B\in\mathscr{B}} I_B \underline{P}(f|B)$ 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 $\mathscr{B}$.

The following properties of conditional lower and upper previsions associated with a coherent set of really desirable gambles were (essentially) proven by Walley.

Proposition 2 (Properties of conditional lower and upper previsions [20])**.** *Consider a coherent set of really desirable gambles $\mathscr{R}$, let B be any non-empty subset of Ω, and let f, f_1 and f_2 be gambles on Ω. Then*[8]

1. $\inf_{\omega\in B} f(\omega) \leq \underline{P}(f|B) \leq \overline{P}(f|B) \leq \sup_{\omega\in B} f(\omega)$ *[convexity];*
2. $\underline{P}(f_1+f_2|B) \geq \underline{P}(f_1|B)+\underline{P}(f_2|B)$ *[super-additivity];*
3. $\underline{P}(\lambda f|B) = \lambda\underline{P}(f|B)$ *for all real* $\lambda \geq 0$ *[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. $f_1 \leq f_2$ *implies that* $\underline{P}(f_1|B) \leq \underline{P}(f_2|B)$ *[monotonicity];*
7. *if* $\mathscr{B}$ *is a partition of* Ω *that refines the partition* $\{B, B^c\}$ *and* $\mathscr{R}$ *is* $\mathscr{B}$*-conglomerable, then* $\underline{P}(f|B) \geq \underline{P}(\underline{P}(f|\mathscr{B})|B)$ *[conglomerative property].*

The analogy between Propositions 1 and 2 is striking, even if there is an equality in Proposition 1.7 and only an inequality in Proposition 2.7.[9] We now set out to identify the exact correspondence between the two models.[10]

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 Subject, who, *in situation* $\square$, has certain *piece-wise* beliefs about what moves World will make.

More specifically, for each non-terminal situation $t \in \Omega^\lozenge \setminus \Omega$, she has beliefs (in situation $\square$) about which move $\mathbf{w}$ World will choose from the set $\mathbf{W}_t$ of moves available to him in t. We suppose she represents those beliefs in the form of a *coherent*[11] set $\mathscr{R}_t$ of really desirable gambles on $\mathbf{W}_t$. These beliefs are conditional on the updating interpretation, in the sense that they represent Subject's beliefs in situation $\square$ about what World will do *immediately after he gets to situation* t. We call any specification of such coherent $\mathscr{R}_t$, $t \in \Omega^\lozenge \setminus \Omega$, an *immediate prediction model* for Subject. It should be stressed here that $\mathscr{R}_t$ should *not* be interpreted dynamically, i.e., as a set of gambles on $\mathbf{W}_t$ that Subject accepts in situation t.

We can now ask ourselves what the behavioural implications of these conditional assessments $\mathscr{R}_t$ in the immediate prediction model are. For instance, what do they tell us about whether or not Subject should accept certain gambles[12] on Ω, the set of possible paths for World? In other words, how can these beliefs (in $\square$) about which next move World will make in each non-terminal situation t be combined coherently into beliefs (in $\square$) about World'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 shall construct, using the local pieces of information $\mathscr{R}_t$, a set of really desirable gambles on Ω for Subject 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.

First, we collect the pieces. Consider any non-terminal situation $t \in \Omega^\lozenge \setminus \Omega$ and any gamble h_t in $\mathscr{R}_t$. Then with h_t we can associate a t-gamble,[13] also denoted by h_t, and defined by

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

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 write $h_t(s) := h_t(\omega(t))$, for any $t \sqsubset s$, where ω is any terminal situation that follows s.

$I_{\uparrow t} h_t$ represents the gamble on Ω that is called off unless World ends up in situation t, and which, when it is not called off, depends only on World's move immediately after t, and gives the same value $h_t(\mathbf{w})$ to all paths ω that go through

[7] We may be wrong, but it seems to us that in [17], the authors 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".

[8] 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 $\underline{P}(f|\mathscr{B})$ is real-valued, and nowhere infinite. Shafer and Vovk do not seem to mention the need for this.

[9] Concatenation inequalities for lower prices do appear in the more general context described in [17].

[10] We shall find a specific situation where applying Walley's theory leads to equalities rather than the more general inequalities of Proposition 2.7. 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 [20, Theorem 6.7.2] and [9].

[11] Since we do not 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.

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

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

$t\mathbf{w}$. The fact that Subject, in situation $\square$, accepts h_t on $\mathbf{W}_t$ conditional on World's getting to t, translates immediately to the fact that Subject accepts the contingent gamble $I_{\uparrow t}h_t$ on Ω, by Walley's Updating Principle. We thus end up with a set of gambles on Ω

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

that Subject accepts in situation $\square$. The only thing left to do now, is to find the smallest coherent set $\mathscr{E}_\mathscr{R}$ of really desirable gambles that includes $\mathscr{R}$ (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.

These partitions are what we call *cut partitions*. Consider any cut U of the initial situation $\square$. Then the set of events $\mathscr{B}_U := \{\uparrow u \colon u \in U\}$ 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 $\mathscr{B}_U$.[14]

Why do we only require conglomerability for cut partitions? Simply because we are interested in *predictive inference*: we eventually will want to find out about the gambles on Ω that Subject accepts in situation $\square$, conditional (contingent) on World getting to a situation t. This is related to finding lower previsions for Subject conditional on the corresponding events $\uparrow t$. A collection $\{\uparrow t \colon t \in T\}$ 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 $\mathscr{E}_\mathscr{R}$ 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 $h_u \in \mathscr{R}_u$, $u \in U$. This is because $I_{\uparrow u}g = I_{\uparrow u}h_u \in \mathscr{R}$ for all $u \in U$. Because moreover $\mathscr{E}_\mathscr{R}$ 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 $\mathscr{E}_\mathscr{R}$. But, since in the case of bounded protocols we are discussing here, World 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 $\mathscr{E}_\mathscr{R}$ for all choices $h_u \in \mathscr{R}_u$, $u \in \Omega^\Diamond \setminus \Omega$.

Call therefore, for any non-terminal situation t, a *t-selection* any partial process $\mathscr{S}$ defined on the non-terminal situations $s \sqsupseteq t$ such that $\mathscr{S}(s) \in \mathscr{R}_s$. With such a t-selection, we can associate a t-process, called a *gamble process* $\mathscr{G}^\mathscr{S}$, with value

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

in all situations s that follow t, where it should be recalled that $\mathscr{S}(u)(s) = \mathscr{S}(u)(\omega(u))$ for all $\omega \sqsupseteq s$ (see above). Alternatively, $\mathscr{G}^\mathscr{S}$ is given by the recursion relation $\mathscr{G}^\mathscr{S}(s\mathbf{w}) = \mathscr{G}^\mathscr{S}(s) + \mathscr{S}(s)(\mathbf{w})$ for all non-terminal $s \sqsupseteq t$ and all $\mathbf{w} \in \mathbf{W}_s$, with initial value $\mathscr{G}^\mathscr{S}(t) = 0$. In particular, this leads to the t-gamble $\mathscr{G}^\mathscr{S}_\Omega$ defined on all terminal situations ω that follow t, by letting

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

We have just argued that the gambles $\mathscr{G}^\mathscr{S}_\Omega$ should belong to $\mathscr{E}_\mathscr{R}$ for all non-terminal situations t and all t-selections $\mathscr{S}$. As before for strategy and capital processes, we call a $\square$-selection $\mathscr{S}$ simply a *selection*, and a $\square$-gamble process simply a *gamble process*. 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,[15] of the Marginal Extension Theorem first proven by Walley [20, Theorem 6.7.2] and subsequently extended by De Cooman and Miranda [9].

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

$$\mathscr{E}_\mathscr{R} := \left\{g \colon g \geq \mathscr{G}^\mathscr{S}_\Omega \text{ for some selection } \mathscr{S}\right\}.$$

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

We now use the coherent set of really desirable gambles $\mathscr{E}_\mathscr{R}$ to define special lower (and upper) previsions $\underline{P}(\cdot|t) := \underline{P}(\cdot|\uparrow t)$ for Subject in situation $\square$, conditional on an event $\uparrow t$, i.e., on World getting to situation t, indicated in Section 3.[16] We shall call such conditional lower previsions *predictive* lower previsions. We then get, using Theorem 3, that for any non-terminal situation t,

$$\underline{P}(f|t) := \sup\left\{\alpha \colon I_{\uparrow t}(f-\alpha) \in \mathscr{E}_\mathscr{R}\right\} \tag{5}$$

$$= \sup\left\{\alpha \colon f-\alpha \geq \mathscr{G}^\mathscr{S}_\Omega \text{ for some } t\text{-selection } \mathscr{S}\right\}. \tag{6}$$

Eq. (5) is also valid in terminal situations t, whereas Eq. (6) 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.

[14] When all of World's move spaces $\mathbf{W}_t$ are finite, cut conglomerability (D5') is a consequence of D3, and therefore needs no extra attention.

[15] 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.

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

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. *$\underline{P}(f|t) = \underline{P}(fI_{\uparrow t}|t)$ and $\overline{P}(f|t) = \overline{P}(fI_{\uparrow t}|t)$;*
3. *$f_1 \leq f_2$ (on $\uparrow t$) implies that $\underline{P}(f_1|t) \leq \underline{P}(f_2|t)$ [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 $\omega \sqsupseteq t$, where u is the unique element of U that ω goes through. 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 $\underline{P}(gf|U) = g_U\underline{P}(f|U)$.*

There appears to be a close correspondence between the expressions [such as (2)] for lower prices $\underline{\mathbb{E}}_t(f)$ associated with coherent probability protocols and those [such as (6)] for the predictive lower previsions $\underline{P}(f|t)$ based on an 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* $t \in \Omega^\lozenge \setminus \Omega$.

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.*

It is interesting to indicate here how matching is actually achieved. If we have a coherent probability protocol with move spaces $\mathbf{S}_t$ and gain functions λ_t for Skeptic, define the immediate prediction model for Subject to be (essentially) $\mathscr{R}_t := \{-\lambda(\mathbf{s},\cdot)\colon \mathbf{s} \in \mathbf{S}_t\}$. If, conversely, we have an immediate prediction model for Subject consisting of the sets $\mathscr{R}_t$, define the move spaces for Skeptic by $\mathbf{S}_t := \mathscr{R}_t$, and his gain functions by $\lambda_t(h,\cdot) := -h$ for all h in $\mathscr{R}_t$.

Theorem 7 (Concatenation Formula)**.** *Consider any two cuts U and V of a situation t such that U precedes V. Then for all t-gambles f on Ω,*[17]

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

This theorem, in combination with the following two propositions (8 and 9), 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 $\mathscr{R}_t$. To see this, observe in addition that in the above theorem, the t-gamble $\underline{P}(f|V)$ is V-measurable, and therefore actually a gamble on V.

To make clear what the following Proposition 8 implies, consider any t-selection $\mathscr{S}$, and define the *U-called off t-selection* $\mathscr{S}^U$ as the selection that mimics $\mathscr{S}$ until we get to U, where we begin to select the zero gambles: for any non-terminal situation $s \sqsupseteq t$, let $\mathscr{S}^U(s) := \mathscr{S}(s)$ if s strictly precedes (some element of) U, and let $\mathscr{S}^U(s) := 0 \in \mathscr{R}_s$ otherwise. Then

$$U(\mathscr{G}^{\mathscr{S}}) = \mathscr{G}^{\mathscr{S}^U} \quad \text{and therefore} \quad \mathscr{G}_U^{\mathscr{S}} = \mathscr{G}_\Omega^{\mathscr{S}^U}, \qquad (7)$$

so we see that stopped gamble processes are gamble processes themselves, that correspond to selections being 'called-off' after a cut. This also means that we can actually restrict ourselves to selections $\mathscr{S}$ 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, $I_{\uparrow t}f \in \mathscr{E}_{\mathscr{R}}$ if and only is there is some t-selection $\mathscr{S}$ such that $I_{\uparrow t}f \geq \mathscr{G}_\Omega^{\mathscr{S}^U}$, or equivalently, $f_U \geq \mathscr{G}_U^{\mathscr{S}}$. Consequently,*

$$\begin{aligned}\underline{P}(f|t) &= \sup\left\{\alpha\colon f-\alpha \geq \mathscr{G}_\Omega^{\mathscr{S}^U} \text{ for some } t\text{-selection } \mathscr{S}\right\}\\ &= \sup\left\{\alpha\colon f_U-\alpha \geq \mathscr{G}_U^{\mathscr{S}} \text{ for some } t\text{-selection } \mathscr{S}\right\}.\end{aligned}$$

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 modal $\mathscr{R}_t$.

[17] 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.

Proposition 9. *Let t be a non-terminal situation, and consider a $C(t)$-measurable gamble h. Then*

$$\underline{P}(h|t) = \underline{P}_t(h) := \sup\{\alpha\colon h-\alpha \in \mathscr{R}_t\}.$$

5 Interpretation

The Matching Theorem has a very interesting interpretation. In Shafer and Vovk's approach, World is sometimes decomposed into two players, Reality and Forecaster. It is Reality whose moves are characterised by the above-mentioned event tree, and Forecaster who determines what Skeptic's move space $\mathbf{S}_t$ and gain function λ_t are, in each non-terminal situation t. We now make Shafer and Vovk's model a bit more involved, by adding something to it.

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 $\mathscr{R}_t$ of really desirable gambles on $\mathbf{W}_t$. In other words, *we identify Forecaster with Subject*.[18]

When Forecaster specifies such a set, she is making certain behavioural commitments. In fact, she is committing herself to accepting, in situation $\square$, any gamble in $\mathscr{R}_t$, contingent on World 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 Forecaster can be made to buy the t-gamble f for, conditional on World's getting to t, and on the basis of the commitments she has made in the initial situation $\square$.

What Skeptic can now do, is take Forecaster up on her commitments. This means that in situation $\square$, he can use a selection $\mathscr{S}$, which for each non-terminal situation t, selects a gamble (or equivalently, any non-negative linear combination of gambles) $\mathscr{S}(t) = h_t$ in $\mathscr{R}_t$ and offer the corresponding gamble $\mathscr{G}^{\mathscr{S}}_{\Omega}$ 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 Skeptic's capital by (the positive or negative amount) $-h_t(\mathbf{w})$. In other words, his move space $\mathbf{s}_t$ can then be identified with the convex set of gambles $\mathscr{R}_t$ and his gain function λ_t is then given by $\lambda_t(h_t,\cdot) = -h_t$. But then the *selection* $\mathscr{S}$ can be identified with a *strategy* $\mathscr{P}$ for Skeptic, and $\mathscr{K}^{\mathscr{P}}_{\Omega} = -\mathscr{G}^{\mathscr{S}}_{\Omega}$ (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 Skeptic coincide with Forecaster's predictive lower previsions $\underline{P}(\cdot|t)$.

[18] The germ for this idea, in the case that Forecaster's beliefs can be expressed using precise probability models on the $\mathscr{L}(\mathbf{W}_t)$, is already present in Shafer's work, see for instance [18, Chapter 8] and [13, Appendix 1]. We extend this idea here to Walley's imprecise probability models.

In a very nice paper [17], 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 he 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 also on our approach, the game-theoretic framework is useful.

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 (or any Subject'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 $n_U(u) > 0$, 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 $h_s - m_s \in \mathscr{R}_s$, 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 $t \sqsubseteq s \sqsubset U$. Then it follows from the coherence of $\mathscr{R}_s$ [D1] that $m_s \leq \sup h_s$. To make things interesting, we shall also assume that $\inf h_s \leq m_s$, because otherwise $h_s - m_s \geq 0$ and accepting this gamble represents no real commitment on Forecaster's part. As a result, we see that $|h_s - m_s| \leq B$.

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 price 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

$$\{G_U \geq -\varepsilon\} := \{\omega \in \uparrow t \colon G_U(\omega) \geq -\varepsilon\},$$

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

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

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

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 will not be negative*. Of course, this is a very general version of the weak law of large numbers. It significantly extends the result mentioned in Section 5. Perhaps surprisingly, it can be seen as generalisation of Hoeffding's inequality for martingale differences [7] (see also [22, Chapter 4] and [19, Appendix A.7]) to coherent lower previsions on event trees.

7 Scoring a predictive model

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: $G_U(u_o) < 0$.

Then we see that $\uparrow u_o \subseteq \{G_U < -\varepsilon\}$ for all $0 < \varepsilon < -G_U(u_o)$, and therefore all these events $\{G_U < -\varepsilon\}$ have actually occurred (because $\uparrow u_o$ has). On the other hand, Forecaster's upper probability (in $\square$) for their occurrence satisfies $\overline{P}(\{G_U < -\varepsilon\}) \leq \exp(-\frac{N_U \varepsilon^2}{4B^2})$, 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

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

By assumption, $\gamma_U(u_o)$ is a number in $[-1, 0)$. Coherence requires that Forecaster, because of her local predictive commitments, can be forced (by Skeptic, 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 Figure 2.

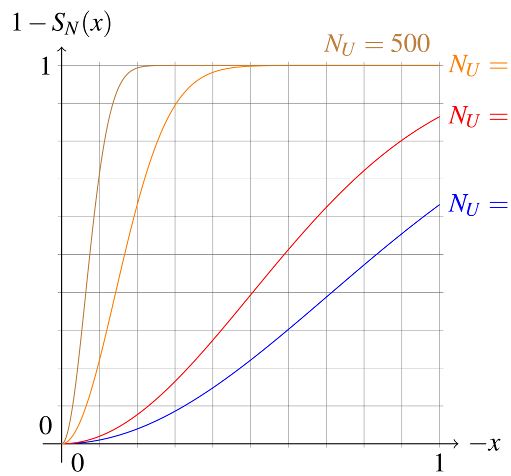

Figure 2: What Forecaster can be made to pay, $1 - S_N(x)$, as a function of $x = \gamma_U(u)$, for different values of $N = N_U$.

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. (6), this value would generally depend on Forecaster's predictive assessments $\mathscr{R}_s$ for situations s that do not 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.

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 is not 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 $\mathscr{R}'_s$ coincide with the $\mathscr{R}_s$ for all s that precede u_o. For other s, the $\mathscr{R}'_s$ 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. Then 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

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

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 $\mathscr{R}_s$ 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* [1, 2].

8 Additional Remarks

We have proven 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.

Acknowledgements

We would like to thank Enrique Miranda, Marco Zaffalon, Glenn Shafer, Vladimir Vovk and Didier Dubois for discussing and questioning the views expressed here, even though some of these discussions took place more than a few years ago. Sébastien Destercke and Erik Quaeghebeur have read and commented on earlier drafts of this paper.

References

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

[2] A. Ph. Dawid and V. G. Vovk. Prequential probability: principles and properties. *Bernoulli*, 5:125–162, 1999.

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

[4] B. de Finetti. *Theory of Probability*. John Wiley & Sons, Chichester, 1974–1975. English translation of [3], two volumes.

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

[6] M. Goldstein. The prevision of a prevision. *Journal of the American Statistical Society*, 87:817–819, 1983.

[7] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the Americal Statistical Association*, 58:13–30, 1963.

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

[9] E. Miranda and G. de Cooman. Marginal extension in the theory of coherent lower previsions. *International Journal of Approximate Reasoning*, 2006. In press.

[10] F. P. Ramsey. Truth and probability (1926). In R. B. Braithwaite, editor, *The Foundations of Mathematics and other Logical Essays*, chapter VII, pages 156–198. Kegan, Paul, Trench, Trubner & Co., London, 1931. Reprinted in [8] and [5].

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

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

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

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

[15] G. Shafer. The significance of Jacob Bernoulli's *Ars Conjectandi* for the philosophy of probability today. *Journal of Econometrics*, 75:15–32, 1996.

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

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

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

[19] V. Vovk, A. Gammerman, and G. Shafer. *Algorithmic learning in a Random World*. Springer, New York, 2005.

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

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

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

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

[24] P. M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44:366–383, 2007. Revised journal version of [23].