FOUNDATIONS FOR TEMPORAL REASONING USING LOWER PREVISIONS WITHOUT A POSSIBILITY SPACE

MATTHIAS C. M. TROFFAES AND MICHAEL GOLDSTEIN

ABSTRACT. We introduce a new formal mathematical framework for probability theory, taking random quantities to be the fundamental objects of interest, without reference to a possibility space, in spirit of de Finetti's treatment of probability, Goldstein's Bayes linear analysis, and Williams's treatment of lower and upper previsions. The aim of our framework is to formalize temporal reasoning, where we treat future beliefs as random quantities themselves. We do this by taking random quantities to form a linear space of expressions, which we endow with structure through a linear projection operator.

We then use a version of the temporal sure preference principle as a basis for inference over time. We formulate the principle in terms of desirability, and explore its implications for lower previsions. We derive an explicit expression for the natural extension of a lower prevision under the temporal sure preference principle. We establish consistency of the temporal sure preference principle with any given collection of assessments. We also derive various bounds on the natural extension. Finally, we show how we can recover standard Bayes linear calculus from our framework.

1. INTRODUCTION

As argued in [12], probabilistic inference has two components, one static and one dynamic. The static component is a description of probabilistic judgements now, where we are free to make any allocations of uncertainty that we consider to be appropriate, expressed, for example, through buying and selling prices on appropriate gambles, subject only to the constraints imposed by coherence over the collection of uncertainty judgements, precise or imprecise, that we choose now to make. The dynamic component describes how these uncertainty statements may change over time, as we receive further information, reflect further on the information that is currently available to us, and so forth.

Aspects of the dynamic component are expressed within the static component, for example through conditioning statements, which express our current buying and selling prices given various called-off bets which describe conditions under which the bets will or will not take place. Such conditioning is informative for our future judgements, but does not determine them, partly as our future experiences will not be summarisable as the observation of membership of a partition that we could specify in advance of our inferences, partly because we are always free to reflect further on the information that we have already received and change our judgements to those that we feel are in closer accord with the prior evidence, and partly because, in any case, there is nothing in the usual probabilistic formalism that forces an equivalence between current views on certain called-off bets, and actual future uncertainty assessments about the relevant quantities. This should not be seen as a failure of conditional reasoning itself—indeed, conditional reasoning is still a perfectly valid and extremely useful formalism for embedding the dynamic features of

Key words and phrases. updating; inference; temporal coherence; desirability; lower prevision.

inference strictly within our current static judgements as to how such an inference might proceed.

At this point, perhaps we should note that one might indeed not care about modelling future beliefs, and take the stance that all future decisions are fully determined solely by current beliefs about those random variables that affect these decisions. For example, normal form decision making is precisely concerned with such scenario: if a subject makes all future decisions right now, only his current beliefs count, and his future beliefs are completely irrelevant. In practice however, beliefs are revised over time, and it is rarely the case that future beliefs, which will determine future decisions, are determined solely on the basis of called-off bets with respect to current beliefs, say through repeated application of Bayes theorem. Analyzing our current beliefs about our future beliefs, as in this paper, is thus important if we now wish to know how we will act in the future based on the actual, but now still uncertain, beliefs that we will hold in the future.

One particular difficulty in trying to model future beliefs is that the possibility space is dynamic in itself. In [12], a possibility space for future beliefs was non-constructively assumed. In this paper, we provide a fully constructive and operational theory, in the sense that all random quantities belong to an explicitly constructed linear space. We replace the possibility space with a much simpler structure, namely, a bounding operator. This is similar to how Williams [14, 15] originally developed the theory of lower previsions. In addition, we also directly build in our knowledge about which random quantities will be known at time t. This was left implicit in [12]. We shall see that making these random quantities explicit in the theory impacts inferences, and therefore it is imperative that they are appropriately modelled.

Temporal coherence [3, 4, 5] is concerned with the careful description of the relationships between the static and dynamic features of probabilistic reasoning. We do not know what our future uncertainty judgements will be, but we may now express views about them. These views are, themselves, probabilistic. The temporal sure preference principle imposes reasonable constraints on our current judgements about our future judgements. In imprecise probability theory, preferences come about as a very natural way of modelling beliefs. The concept of desirability, that is, which gambles we (possibly marginally) prefer to the zero gamble, forms a natural foundation for imprecise probability [14, 15, 13]. As shown in [12], the temporal sure preference principle can be expressed directly in terms of desirability. Consequently, through natural extension, we can directly exploit these constraints when we perform inference.

In contrast, the traditional way of looking at updating in the subjective approach to imprecise probability goes by means of conditioning, that is, looking at called-off gambles. For instance, Zaffalon and Miranda [16] provided a justification for conditioning and conglomerability, through temporal reasoning, in a setting where future beliefs are assumed to be fixed now. However, in practice, future subjective beliefs rarely reflect past called-off gambles, and in fact there is no compelling reason for this to be so, simply because there is no compelling reason for them to be fixed now. Indeed, it seems far more natural to start out from the premise that future beliefs are inherently random, which leads to a more general theory, but of course we also risk it to be far less tractable—interestingly, in the precise case, the generality gained leads to updating rules which are far more efficient than computing with called-off gambles, particularly for large scale problems (for instance, see [1]). Having preference, in the form of desirability, at its foundations, imprecise probability is a natural candidate for temporal coherence.

The aims of this paper are: (i) to develop a theory of lower previsions that abandons the assumption of a possibility space, (ii) to directly model future beliefs as random quantities, and (iii) to explicitly account for temporal reasoning, not through conditioning, but by expressing that we will know certain random quantities at a future time, and by making an explicit assumption about temporal coherence through the temporal sure preference principle. Such framework can provide a basis not only for Bayes linear analysis, but also for imprecise generalisations thereof.

This paper is organised as follows. Section 2 develops a theory of lower previsions without possibility spaces, but instead starting from a bounding operator. In this section, we also recover all results for lower previsions that we will need later. Section 3 explores how we can define, constructively, a set of gambles that includes future beliefs as random quantities, and we derive a convenient representation theorem for this set. Section 4 discusses the temporal sure preference principle. We derive an explicit constructive expression for the natural extension under the temporal sure preference principle, and explore its implications. Section 5 looks at the theory when we additionally assume that future beliefs are precise. There, we allow multiplication of gambles to a limited degree, to allow us to talk about variance as well. We recover the main results from standard Bayes linear calculus. Section 6 concludes the paper. Proofs that are omitted from the main text can be found in Appendix A. Some proofs have been left in the main text, especially where the proofs may have some interest in themselves.

2. Lower Previsions and Desirability

One of the premises of temporal reasoning is that we cannot specify in advance what the possibility space ought to be. Therefore, in this paper, following [2, 14, 15], we consider random quantities as fundamental objects, without referring to a possibility space.

A *gamble* is simply a bounded random quantity. We will denote gambles by capital letters X, Y, ... By $\mathscr{L}$, we denote the linear span of an arbitrary collection of gambles of interest. Following [14, 15], we assume that there is a least upper bound operator on $\mathscr{L}$, denoted by sup. This operator induces a greatest lower bound operator inf on $\mathscr{L}$, defined as $\inf(X) := -\sup(-X)$.

In our treatment, sup reflects our subjective judgements about the logical boundaries on all gambles in $\mathscr{L}$. For example, if X_1, ..., X_n is a basis of $\mathscr{L}$, and we deem all X_i to be logically independent, then we might define

$$\sup\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \sup(\operatorname{sign}(a_i) X_i) \tag{1}$$

where $\operatorname{sign}(a_i)$ is 1, 0, or -1 depending on whether $a_i > 0$, $a_i = 0$, or $a_i < 0$ (so we only need to specify sup on each X_i and $-X_i$). As another example, if all gambles are defined as bounded functions on some common possibility space, then we can define sup as the supremum over functions.

We assume that sup satisfies the following conditions:

(S1) $\sup(X+Y) \leq \sup(X) + \sup(Y)$ for all $X, Y \in \mathscr{L}$.
(S2) $\sup(\lambda X) = \lambda \sup(X)$ for all $X \in \mathscr{L}$ and $\lambda \geq 0$.

From sup, we can define a relation $\succeq$ on $\mathscr{L}$, defined as $X \succeq Y$ whenever $\inf(X-Y) \geq 0$. It is easily seen that $\succeq$ is a preorder (i.e. it is reflexive and transitive). We write $X \simeq Y$ whenever $X \succeq Y$ and $Y \succeq X$, or in other words, whenever $\sup(X-Y) = \inf(X-Y) = 0$.

Clearly, $\simeq$ is an equivalence relation on $\mathscr{L}$. If $X \succeq Y$ then $\sup X \geq \sup Y$ and $\inf X \geq \inf Y$.[1] If $X \simeq Y$ then $\sup X = \sup Y$ and $\inf X = \inf Y$.

We also assume:

(S3) There is a gamble $\Omega \in \mathscr{L}$ such that $\sup(\Omega) = \inf(\Omega) = 1$.

One can easily show that, for all $a \in \mathbb{R}$, $\sup(a\Omega) = \inf(a\Omega) = a$. We will denote the gamble $a\Omega$ simply by a, if no confusion is possible.

By going to the quotient set of the relation $\simeq$ on $\mathscr{L}$, we can also assume, without loss of generality:

(S4) There is a unique gamble X such that $\sup(X) = \inf(X) = 0$.

The conditions stated above are equivalent to the conditions given in [14, 15].

We also introduce the function $\|X\| := \max\{\sup(X), \sup(-X)\}$. This turns $\mathscr{L}$ into a normed vector space.

Proposition 1. *$\|\cdot\|$ is a norm on $\mathscr{L}$.*

Although we will not need to do so in this paper, using the standard completion through equivalence classes of Cauchy sequences, we could in principle use this norm to turn $\mathscr{L}$ into a Banach space.

As mentioned in the introduction, we will take desirability to be the basic concept, and will use it for studying the implications of temporal coherence on lower previsions. To keep the treatment as simple as possible, however, we will restrict ourselves to sets of almost desirable gambles induced by lower previsions.

The following serves to introduce the basic ideas behind desirability and lower previsions, and to fix the notation and conventions used in the paper. We refer to [13, 11] for much more information on the topic. In particular, throughout the paper, we will use the properties of coherent lower previsions extensively [13, Sec. 2.6.1] [11, Sec. 4.3]. Note that most of the literature, including [13, 11], assume that $\mathscr{L}$ are bounded functions on some possibility space, whereas we make no such assumption. However, many proofs in the literature only rely on the properties of sup listed above. Where that is the case, we will use those results but explicitly point to the relevant proof, allowing the reader to verify carefully that we only rely on the properties of sup.

A lower prevision $\underline{P}$ is a function mapping from any subset of $\mathscr{L}$ to the real numbers. For every gamble X in the domain of $\underline{P}$, the number $\underline{P}(X)$ is interpreted as a subjective assessment of a subject's supremum buying price. That is, for all $\varepsilon > 0$, the subject would be willing to pay $\underline{P}(X) - \varepsilon$ in exchange for X, or in other words, the subject would be willing to accept $X - \underline{P}(X) + \varepsilon$. Any gamble Y such that $Y + \varepsilon$ is acceptable for all $\varepsilon > 0$ is called *almost desirable*. Concluding, a subject's specification of a lower prevision $\underline{P}$ means that the subject declares $X - \underline{P}(X)$ to be almost desirable for all X in the domain of $\underline{P}$.

In de Finetti's approach [2], a subject specifies a prevision P, and where this is understood to mean that both $X - P(X)$ and $P(X) - X$ are almost desirable to the subject. De Finetti's previsions are thus a special case of lower previsions. With lower previsions, if both X and $-X$ are in the domain of $\underline{P}$, then $X - \underline{P}(X)$ is almost desirable and so is $-X - \underline{P}(-X)$. However it is not assumed that $-\underline{P}(-X) = \underline{P}(X)$, and generally $-\underline{P}(-X) \geq \underline{P}(X)$. We can interpret $-\underline{P}(-X)$ as the infimum buying price for X. For convenience of notation, we define

$$\overline{P}(X) := -\underline{P}(-X). \tag{2}$$

[1] To prove these implications, use $\sup X \geq \inf(X - Y) + \sup Y$ and $\inf X \geq \inf(X - Y) + \inf Y$.

A lower prevision is said to avoid sure loss if and only if [11, p. 43, Def. 4.6(E)]

$$\sup\left(\sum_{i=1}^{n}\lambda_i(X_i-\underline{P}(X_i))\right)\geq 0 \tag{3}$$

for all $n\in\mathbb{N}$, all $\lambda_1\geq 0$, ..., $\lambda_n\geq 0$, and all X_1, ..., X_n in the domain of $\underline{P}$. One can show that this condition implies that there is no combination of acceptable gambles that leads to a strictly positive loss; see for instance [11, p. 44–45, Sec. 4.2.2] and note that only the properties of sup are required to complete the argument.

The *natural extension* $\underline{E}$ of a lower prevision $\underline{P}$ is defined, for every gamble $X\in\mathscr{L}$, by [13, p. 122, Sec. 3.1.1] [11, p. 47, Def. 4.8]:

$$\begin{aligned}\underline{E}(X) := \sup\Big\{\alpha\in\mathbb{R}\colon X-\alpha\succeq\sum_{i=1}^{n}\lambda_i(X_i-\underline{P}(X_i)),\\ n\in\mathbb{N},\,\lambda_1\geq 0,\dots,\lambda_n\geq 0,\,X_1,\dots,X_n\in\operatorname{dom}\underline{P}\Big\}.\end{aligned} \tag{4}$$

The value $\underline{E}(X)$ represents the supremum buying price that can be inferred from the subject's specifications $\underline{P}$. If $\underline{P}$ avoids sure loss, then $\underline{E}$ is a lower prevision on $\mathscr{L}$ satisfying:

C1 $\underline{E}(X)\geq\inf(X)$
C2 $\underline{E}(X+Y)\geq\underline{E}(X)+\underline{E}(Y)$
C3 $\underline{E}(\lambda X)=\lambda\underline{E}(X)$

for all X, $Y\in\mathscr{L}$ and all $\lambda\geq 0$. Any lower prevision on $\mathscr{L}$ satisfying these conditions is called *coherent*.

Proposition 2. *If $\underline{P}$ avoids sure loss, then $\underline{E}$ is a coherent lower prevision on $\mathscr{L}$.*

This is a well known result, however for the sake of completeness, we give a proof in the appendix to ensure that only the properties of sup are used.

By $\underline{\mathbb{E}}(\mathscr{L})$ we denote the set of all coherent lower previsions on $\mathscr{L}$. The upper prevision $\overline{E}$ corresponding to $\underline{E}$ is defined as:

$$\overline{E}(X) := -\underline{E}(-X). \tag{5}$$

Coherence implies that $\underline{E}(X)\leq\sup(X)$ and $\underline{E}(X+a)=\underline{E}(X)+a$ for all $a\in\mathbb{R}$; we will use these properties further.

We can easily construct a set of almost desirable gambles for $\underline{P}$ through $\underline{E}$:

$$\mathscr{D} := \{X\in\mathscr{L}\colon \underline{E}(X)\geq 0\}. \tag{6}$$

We can show that $\mathscr{D}$ satisfies the following conditions [13, p. 152, Sec. 3.7.3]:[2]

D1 if $X\succeq 0$ then $X\in\mathscr{D}$,
D2 if $\sup X<0$ then $X\notin\mathscr{D}$,
D3 if $X\in\mathscr{D}$ and $Y\in\mathscr{D}$ then $X+Y\in\mathscr{D}$,
D4 if $\lambda\geq 0$ and $X\in\mathscr{D}$ then $\lambda X\in\mathscr{D}$, and
D5 if $X+\varepsilon\in\mathscr{D}$ for all $\varepsilon>0$, then $X\in\mathscr{D}$.

Any subset of $\mathscr{L}$ that satisfies these conditions is said to be a *coherent* set of almost desirable gambles.

Proposition 3. *$\mathscr{D}$ is a coherent set of almost desirable gambles.*

Note that we can recover $\underline{E}$ from $\mathscr{D}$:

[2]Our axiom $\inf X\geq 0\implies X\in\mathscr{D}$ follows from Walley's (D1) and (D4): if $X\succeq 0$, or equivalently, $\inf X\geq 0$, then $X+\varepsilon\in\mathscr{D}$ for all $\varepsilon>0$ by Walley's (D1), and consequently $X\in\mathscr{D}$ by Walley's (D4).

Proposition 4. *For every coherent lower prevision $\underline{E} \in \underline{\mathbb{E}}(\mathscr{L})$, and every $X \in \mathscr{L}$, we have that*

$$\underline{E}(X) = \max\{a \in \mathbb{R} \colon X - a \in \mathscr{D}\} \tag{7}$$

So, in the following, we can use $\underline{E}$ and $\mathscr{D}$ interchangeably.

A lower prevision is called a prevision when it is self-conjugate, that is, when $\underline{E} = \overline{E}$, in which case we simply denote it by E. By $\mathbb{E}(\mathscr{L})$ we denote the set of all coherent previsions on $\mathscr{L}$. It is well known that coherent lower previsions correspond to lower envelopes of sets of coherent previsions. This still holds in our framework. Specifically, let

$$\mathscr{M}(\underline{E}) \coloneqq \{E \in \mathbb{E}(\mathscr{L}) \colon E \geq \underline{E}\}. \tag{8}$$

Then we have the following.

Proposition 5. *For every coherent lower prevision $\underline{E} \in \underline{\mathbb{E}}(\mathscr{L})$, and every $X \in \mathscr{L}$, we have that*

$$\underline{E}(X) = \min_{P \in \mathscr{M}(\underline{E})} P(X). \tag{9}$$

3. Future Beliefs as Gambles

We will consider lower previsions at different points in time—and in this paper, at just two points in time, 0 and $t > 0$. As mentioned before, by $\mathscr{L}$ we denote the (linear span of the) collection of gambles that we consider now, at time 0: it represents our subjective judgement, now, about what gambles are relevant to the statistical problem at hand.

However, we might also consider gambles about future beliefs about gambles in $\mathscr{L}$. Specifically, $\underline{P}_t$ denotes our (currently unknown) assessments at time t, in the form of supremum buying prices for gambles in $\operatorname{dom}\underline{P}_t$, at time t. If we assume that $\underline{P}_t$ avoids sure loss (which seems reasonable), then we can consider its natural extension $\underline{E}_t$, which is a coherent lower prevision. For any $X \in \mathscr{L}$, the value $\underline{E}_t(X)$ represents the supremum buying price for X, based on the assessments $\underline{P}_t$ at time t. This value $\underline{E}_t(X)$ is only known to us at time t and beyond.

By $\underline{E}_0(X)$, we denote our current supremum buying price for X. It embodies our current assessments concerning the problem domain at hand, based on some lower prevision $\underline{P}_0$ that we specify now. However, $\underline{E}_t(X)$ is a gamble by itself, whose value is only realised at time t.

For some gambles, we may know that we will learn their values at time t. We assume that $\mathscr{L}$ is a direct sum of a set $\mathscr{L}_1$ of gambles whose values are not known at time t, and a set $\mathscr{L}_2$ of gambles whose values are known at time t. Such separation is natural in a statistical setting: $\mathscr{L}_1$ will typically contain quantities that we wish to learn about, and $\mathscr{L}_2$ will contain quantities that represent the data that we know at time t. For example, when planning the construction of a new wind farm, we may wish to learn about how the wind will behave at potential locations. In this case, $\mathscr{L}_2$ may contain past wind measurements near locations of interest until time t, whilst $\mathscr{L}_1$ may contain future wind observations at each location when the farm is producing energy at times beyond t. Those future wind observation are obviously unknown when the planning decision is made at time t.

For now, we only consider sums of quantities in $\mathscr{L}_1$ and $\mathscr{L}_2$, and obviously that is a restriction. In Section 5 we relax the domain restriction in exchange for an additional restriction on future beliefs.

The corresponding projection operators are denoted by π_1 and π_2, so for any gamble X we have that $X = \pi_1 X + \pi_2 X$ with $\pi_1 X \in \mathscr{L}_1$ and $\pi_2 X \in \mathscr{L}_2$. Although constants are obviously known at time t, for convenience, *we assume that the constants are in* $\mathscr{L}_1$ *rather than in* $\mathscr{L}_2$, as we will consider coherent lower previsions on just $\mathscr{L}_1$ later on.

We denote the linear space of all gambles, including gambles that represent beliefs about other gambles, by $\mathscr{L}^*$. We define $\mathscr{L}^*$ to be the set of expressions generated through the following axioms:

(L1) If $X \in \mathscr{L}$ then $X \in \mathscr{L}^*$.
(L2) If U and $V \in \mathscr{L}^*$ then $U+V \in \mathscr{L}^*$.
(L3) If $a \in \mathbb{R}$ and $U \in \mathscr{L}^*$ then $aU \in \mathscr{L}^*$.
(L4) If $U \in \mathscr{L}^*$ then $\underline{E}_t(U) \in \mathscr{L}^*$.

So, $\mathscr{L}^*$ is the set of all expressions containing gambles in $\mathscr{L}$, linear combinations, and applications of $\underline{E}_t$. Note that $\mathscr{L}^*$ is much larger than $\mathscr{L}$. We will usually denote gambles in $\mathscr{L}$ by X, Y, and so on, and gambles in $\mathscr{L}^*$ by U, V, and so on.

For any coherent lower prevision $\underline{Q}$ on $\mathscr{L}_1$ and any $U \in \mathscr{L}^*$, we can project U onto a gamble $\pi_{\underline{Q}}U \in \mathscr{L}$, simply by by replacing $\underline{E}_t$ by $\underline{Q}$ in the expression of U, where it is understood that, in the remaining expression, we can treat any $Z \in \mathscr{L}_2$ inside of $\underline{Q}$ as a constant. Because these terms always occur additively, they can always be pulled out of $\underline{Q}$, and therefore the resulting expression can indeed be expressed as an element of $\mathscr{L}$. For example, if $U := Y + 2Z + \underline{E}_t(2Y - 3\underline{E}_t(Y+Z))$, then

$$\pi_{\underline{Q}}U = Y + 2Z + \underline{Q}(2Y - 3\underline{Q}(Y+Z)) = Y - Z - \underline{Q}(Y) \tag{10}$$

provided that $Y \in \mathscr{L}_1$ and $Z \in \mathscr{L}_2$. One can think of this as the the gamble U after $\underline{E}_t = \underline{Q}$ has realised.

With our projection operator from $\mathscr{L}^*$ to $\mathscr{L}$, we can extend the sup operator from $\mathscr{L}$ to $\mathscr{L}^*$ as follows.

Definition 6. *For every $U \in \mathscr{L}^*$, let*

$$\sup(U) := \sup_{\underline{Q} \in \mathbb{E}(\mathscr{L}_1)} \sup(\pi_{\underline{Q}}U). \tag{11}$$

Proposition 7. *The operator* sup *on $\mathscr{L}^*$ satisfies (S1)–(S3).*

In the same way as before, this sup operator induces an equivalence relation $\simeq$ on $\mathscr{L}^*$. We study some of its properties in the next few theorems.

Proposition 8. *For every $a \in \mathbb{R}$ and $Z \in \mathscr{L}_2$, we have that $a + Z \simeq \underline{E}_t(a+Z)$.*

More generally, under the equivalence relation $\simeq$, $\mathscr{L}^*$ has the following structure:

Proposition 9. *For every $U \in \mathscr{L}^*$, there are $n \in \mathbb{N}$, a_1, ..., $a_n \in \mathbb{R}$, Y_0, Y_1, ..., $Y_n \in \mathscr{L}_1$, and $Z \in \mathscr{L}_2$ such that*

$$U \simeq Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z. \tag{12}$$

Here, Z is unique, and Y_0 is unique up to an additive constant. Moreover, we have that

$$\underline{E}_t(U) \simeq \underline{E}_t(Y_0) + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z \tag{13}$$

To see that Y_0 in the representation is only unique up to a constant, note that for instance $Y_0 + \underline{E}_t(Y_1) \simeq Y_0 + c + \underline{E}_t(Y_1 - c)$: we can arbitrarily shift constants between Y_0 and any of the Y_1, ..., Y_n. We could force a unique representation by explicitly projecting out the constants in all Y_i. For simplicity, we have chosen not to do so in the current paper.

The benefit of this representation is that to apply $\pi_{\underline{Q}}$, we can simply replace $\underline{E}_t$ by $\underline{Q}$, and no other operations are required. Another consequence is:

Proposition 10. *If $U \simeq V$ then $\underline{E}_t(U) \simeq \underline{E}_t(V)$.*

We write $\mathscr{D}_t$ for the set of almost desirable gambles corresponding to $\underline{E}_t$.

Definition 11.

$$\mathscr{D}_t := \{U \in \mathscr{L}^* : \underline{E}_t(U) \succeq 0\} \tag{14}$$

The set $\mathscr{D}_t$ naturally contains all non-negative gambles in $\mathscr{L}$. It also contains gambles such as $U - \underline{E}_t(U)$, for any $U \in \mathscr{L}^*$. Up to equivalence, we can write $\mathscr{D}_t$ as follows:

Proposition 12. *We have that*

$$\mathscr{D}_t \simeq \left\{ Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z \colon \inf_{\underline{Q} \in \mathbb{E}(\mathscr{L}_1)} \left(\underline{Q}(Y_0) + \sum_{i=1}^{n} a_i \underline{Q}(Y_i) \right) + \inf Z \geq 0, \right.$$
$$\left. n \in \mathbb{N}, a_1, \dots, a_n \in \mathbb{R}, Y_0, \dots, Y_n \in \mathscr{L}_1, Z \in \mathscr{L}_2 \right\} \tag{15}$$

For example, to see that $U - \underline{E}_t(U) \in \mathscr{D}_t$ for $U \simeq Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z$, note that

$$U - \underline{E}_t(U) \simeq Y_0 - \underline{E}_t(Y_0) \tag{16}$$

and obviously $\underline{Q}(Y_0) - \underline{Q}(Y_0) \geq 0$ for every $\underline{Q}$.

4. Temporal Sure Preference

4.1. **The Temporal Sure Preference Principle.** In order to establish relationships between current and future beliefs, we must impose conditions that go beyond coherence at a single time point. These conditions should be sufficiently weak and compelling to be widely applicable, while leading to a meaningful account of inference.

Any principle which asserts that beliefs now are compelling for beliefs in the future is, by its nature, unconvincing, as we cannot know what future information we may receive or what the outcome of our future reflections may be. The converse, however, is that we may often view our future beliefs as compelling for our current beliefs, as all such future reflections and information will be taken into account in such future judgements. In order for future judgements to influence our current judgements, we must know what such future judgements are. We therefore introduce the notion of a sure prefererence, at a future time, as one which we are now sure that we will hold at that time.

It may seem unreasonable, now, to think that we hold any such sure preferences. However, it so happens that we do indeed hold many such, and recognising them explicitly, and formalising their implications for our current judgements, provides a natural account of temporal reasoning. For this reason, Goldstein introduced the temporal sure preference principle (see [3], [4], [5, Sec. 3.5]), which we discuss next.

Mathematically, we would like to extend $\underline{E}_0$, defined on $\mathscr{L}$, to a coherent lower prevision on the much larger set of gambles $\mathscr{L}^*$. We will do so through desirability. Remember, $\mathscr{D}_0 \subseteq \mathscr{L}$ denotes the set of gambles that are almost desirable now according to $\underline{E}_0$ (and thereby, according to our initial assessments $\underline{P}_0$). By $\mathscr{D}_0^* \subseteq \mathscr{L}^*$ we denote the set of gambles in $\mathscr{L}^*$ that we deem almost desirable now, according to whichever principles that we choose to adopt.

Clearly, we must have that $\mathscr{D}_0 \subseteq \mathscr{D}_0^*$. But we can add many more gambles to $\mathscr{D}_0^*$, by adopting the following variant of the temporal sure preference principle:

Principle 1 (The Temporal Sure Preference Principle)**.** *For any gamble $U \in \mathscr{L}^*$, if U is certain to be almost desirable for us at future time t, i.e. if $\underline{E}_t(U) \succeq 0$, then U should be almost desirable for us now:*

$$\mathscr{D}_t \subseteq \mathscr{D}_0^* \tag{17}$$

This principle is proposed for lower previsions in [12, Principle III]. In that work, we compared this principle with other possible variants, and we proved that it is equivalent with the standard temporal sure preference for previsions introduced in [3], [4], and [5, Sec. 3.5]. A key difference between the above principle and the principles proposed in [12] is that the above principle is slightly simpler in its formulation, because we do not need to refer to the possibility space.

We note that there can be situations where the principle might not hold. For example, we might consider that, at the future time, we could undergo personality changes which render our future judgements suspect to us now (the Doctor Jekyll and Mister Hyde scenario). More prosaically, we might just recognise situations where our future judgements are likely to be less reliable than our current judgements (for example, the problem of forgetting). Therefore, the intention of the temporal sure preference principle is that it should be viewed as a very weak, and widely applicable principle, whose relevance we should consider for the problem at hand. If we consider the temporal sure preference principle applicable in our problem, then we may draw on the strong implications of the principle to provide an account of temporal coherence for this situation.

4.2. **Temporal Natural Extension.** We can now constructively define $\mathscr{D}_0^*$ through natural extension, namely, we let $\mathscr{D}_0^*$ be the smallest closed convex cone containing both $\mathscr{D}_0$ and $\mathscr{D}_t$. Or, equivalently, in terms of lower previsions:

Definition 13. *For any gamble $U \in \mathscr{L}^*$, we define*

$$\underline{E}_0^*(U) := \sup_{\substack{\alpha\in\mathbb{R} \\ X\in\mathscr{L}\colon \underline{E}_0(X)\geq 0 \\ V\in\mathscr{L}^*\colon \underline{E}_t(V)\succeq 0}} \{\alpha \colon U-\alpha \succeq X+V\} \tag{18}$$

In this way, $\underline{E}_0^*$ captures all inferences both from our initial assessments $\underline{P}_0$ (remember that $\underline{E}_0$ is the natural extension of $\underline{P}_0$), as well as all inferences that we can make from the temporal sure preference principle. Note that this form of natural extension is fully constructive, in contrast to the natural extension proposed in [12].

The next theorem provides a more convenient formula:

Proposition 14. *For any gamble $U \in \mathscr{L}^*$,*

$$\underline{E}_0^*(U) = \sup_{X\in\mathscr{L}, Y\in\mathscr{L}_1} \inf[U - X + \underline{E}_0(X) - Y + \underline{E}_t(Y)] \tag{19}$$

$$= \sup_{X\in\mathscr{L}, Y\in\mathscr{L}_1} \inf_{\underline{Q}\in\mathbb{E}(\mathscr{L}_1)} \inf\left[\pi_{\underline{Q}}U - X + \underline{E}_0(X) - Y + \underline{Q}(Y)\right] \tag{20}$$

Proof. The equality of the two expressions follows immediately from the definition of sup (and inf) on $\mathscr{L}^*$.

To see that $\underline{E}_0^*$ is at least as large as the given expression, note that for any $X \in \mathscr{L}$ and $Y \in \mathscr{L}_1$, we have that $\underline{E}_0(X - \underline{E}_0(X)) \geq 0$, $\underline{E}_t(Y - \underline{E}_t(Y)) \succeq 0$, and

$$U - \alpha \succeq X - \underline{E}_0(X) + Y - \underline{E}_t(Y) \tag{21}$$

is satisfied for $\alpha = \inf[U - X + \underline{E}_0(X) - Y + \underline{E}_t(Y)]$.

Conversely, consider any $\alpha \in \mathbb{R}$, $X \in \mathscr{L}$, $V \in \mathscr{L}^*$, such that $\underline{E}_0(X) \geq 0$, $\underline{E}_t(V) \succeq 0$, and $U - \alpha \succeq X + V$. Then it must be that

$$\inf(U - \alpha - X - V) \geq 0 \tag{22}$$

and consequently,

$$\alpha \leq \inf[U - X - V] \leq \inf[U - X + \underline{E}_0(X) - V + \underline{E}_t(V)] \tag{23}$$

Now note that

$$V - \underline{E}_t(V) \simeq Y - \underline{E}_t(Y) \tag{24}$$

for some $Y \in \mathscr{L}_1$, by Theorem 9. Consequently $\underline{E}_0^*(U)$ is less or equal to the given expression. □ □

Before we proceed investigating actual inferences from the above expression for natural extension, we need to address a few concerns. First, there is no guarantee that Principle 1 is consistent with our initial assessments $\underline{E}_0$. Eq. (18) provides us with a means to verify this: we merely have to check that $\underline{E}_0^*(0) < +\infty$ [13, p. 123, ll. 4–7]. Secondly, there is no guarantee that $\underline{E}_0^*$ coincides with $\underline{E}_0$ on $\mathscr{L}$. The next theorem answers these concerns.

Proposition 15. *For all $X \in \mathscr{L}$, we have that*

$$\underline{E}_0^*(X) = \underline{E}_0(X). \tag{25}$$

Proof. Consider any $X \in \mathscr{L}$. Clearly,

$$\underline{E}_0(X) = \sup_{X' \in \mathscr{L}} \inf\left[X - X' + \underline{E}_0(X')\right] \tag{26}$$

$$\leq \sup_{X' \in \mathscr{L}, Y \in \mathscr{L}_1} \inf\left[X - X' + \underline{E}_0(X') - Y + \underline{E}_t(Y)\right] = \underline{E}_0^*(X) \tag{27}$$

Conversely,

$$\underline{E}_0^*(X) = \sup_{X' \in \mathscr{L}, Y \in \mathscr{L}_1} \inf_{\underline{Q} \in \underline{\mathbb{E}}(\mathscr{L}_1)} \inf\left[X - X' + \underline{E}_0(X') - Y + \underline{Q}(Y)\right] \tag{28}$$

and because every coherent lower prevision $\underline{Q}$ on $\mathscr{L}_1$ can be extended to a coherent lower prevision $\underline{Q}$ on $\mathscr{L}$ (e.g. through natural extension),

$$\leq \sup_{X' \in \mathscr{L}, Y \in \mathscr{L}_1} \inf_{\underline{Q} \in \underline{\mathbb{E}}(\mathscr{L})} \underline{Q}\left[X - X' + \underline{E}_0(X') - Y + \underline{Q}(Y)\right] \tag{29}$$

and now because $\underline{E}(X_1 - X_2) \leq \underline{E}(X_1) - \underline{E}(X_2)$ by coherence,

$$\leq \sup_{X' \in \mathscr{L}, Y \in \mathscr{L}_1} \inf_{\underline{Q} \in \underline{\mathbb{E}}(\mathscr{L})} \left\{\underline{Q}(X - X' + \underline{E}_0(X')) - \underbrace{\underline{Q}(Y - \underline{Q}(Y))}_{=0}\right\} \tag{30}$$

$$= \sup_{X' \in \mathscr{L}} \inf\left[X - X' + \underline{E}_0(X')\right] = \underline{E}_0(X) \tag{31}$$

□ □

Note that this proof is *much* simpler and arguably far more elegant than the proof given earlier in [12].

We also immediately have the following important result [12, Proposition 12], which effectively reformulates Principle 1 in terms of lower previsions:

Proposition 16. *For every gamble $U \in \mathscr{L}^*$,*

$$\underline{E}_0^*(U) \geq \inf \underline{E}_t(U) \tag{32}$$

Proof. By Theorem 14,

$$\underline{E}_0^*(U) = \sup_{X\in\mathscr{L},Y\in\mathscr{L}_1} \inf[U - X + \underline{E}_0(X) - Y + \underline{E}_t(Y)] \tag{33}$$

and because $U - \underline{E}_t(U) \simeq Y - \underline{E}_t(Y)$ for some $Y \in \mathscr{L}_1$ (see Theorem 9),

$$\geq \inf[U - 0 + \underline{E}_0(0) - U + \underline{E}_t(U)] = \inf \underline{E}_t(U) \tag{34}$$

□ □

4.3. **Further Implications.** First, we derive the following version of conglomerability [12, Corollary 13]:

Proposition 17. *For every $U \in \mathscr{L}^*$,*

$$\underline{E}_0^*(U - \underline{E}_t(U)) \geq 0. \tag{35}$$

Proof. By Eq. (32),

$$\underline{E}_0^*(U - \underline{E}_t(U)) \geq \inf \underline{E}_t(U - \underline{E}_t(U)). \tag{36}$$

Now note that $\underline{E}_t(U - \underline{E}_t(U)) \simeq 0$. □ □

Clearly, if we were to impose a conditioning interpretation, Eq. (35) corresponds to one of Walley's conditions for coherence [13, p. 303, (C11)].

Theorem 17 immediately implies a number of interesting inequalities:

Proposition 18. *For every $U \in \mathscr{L}^*$,*

$$\underline{E}_0^*(\underline{E}_t(U)) \leq \underline{E}_0^*(U) \leq \underline{E}_0^*(\overline{E}_t(U)), \tag{37}$$

$$\overline{E}_0^*(\underline{E}_t(U)) \leq \overline{E}_0^*(U) \leq \overline{E}_0^*(\overline{E}_t(U)). \tag{38}$$

and consequently, for all $X \in \mathscr{L}$,

$$\underline{E}_0^*(\underline{E}_t(X)) \leq \underline{E}_0(X) \leq \underline{E}_0^*(\overline{E}_t(X)), \tag{39}$$

$$\overline{E}_0^*(\underline{E}_t(X)) \leq \overline{E}_0(X) \leq \overline{E}_0^*(\overline{E}_t(X)). \tag{40}$$

Proof. See [12, Corollary 14]. □ □

Unfortunately, we cannot derive a non-trivial lower bound on $\underline{E}_0^*(\underline{E}_t(U))$, and similarly, we cannot derive a non-trivial upper bound on $\overline{E}_0^*(\overline{E}_t(U))$, due to the possibility of dilation [9]. Some further inequalities for imprecise variance can be found in [12, Proposition 16]. This requires $\mathscr{L}^*$ to include squares of gambles, which complicates the analysis considerably. Additionally, only the lower variance can be effectively bounded, and therefore these inequalities are not particularly interesting. So we will not discuss them here.

5. Temporal Coherence for Previsions

5.1. **Basic Structure.** In this section, we revisit the existing theory of temporal coherence for previsions, from the point of view of the theory that we have so far developed. Stronger properties can be derived if $\underline{E}_t$ is a coherent prevision. To study these properties, we need to be able to square gambles. For this reason, we assume that $\mathscr{L}$ is the set of all constant, linear, and quadratic polynomials in the variables $\mathscr{B}_1 \cup \mathscr{B}_2$, where the gambles in $\mathscr{B}_1$ are assumed to be unknown at time t, and the gambles in $\mathscr{B}_2$ are assumed to be known at time t. We assume that sup is defined on $\mathscr{L}$ and satisfies (S1)–(S4) as well as:

(S5) $\sup(X^2) \geq 0$ for all $X \in \mathscr{L}$ such that $X^2 \in \mathscr{L}$.

We then construct $\mathscr{L}^*$ as the set of expressions generated through the following axioms, where we simultaneously also define the degree function $\deg^*\colon \mathscr{L}^* \to \mathbb{N}$:

(L'1) If $a \in \mathbb{R}$ then $a \in \mathscr{L}^*$ and $\deg^*(a) := 0$.
(L'2) If $X \in \mathscr{B}_1 \cup \mathscr{B}_2$ then $X \in \mathscr{L}^*$ and $\deg^*(X) := 1$.
(L'3) If U and $V \in \mathscr{L}^*$ then $U + V \in \mathscr{L}^*$ and $\deg^*(U+V) := \max\{\deg^*(U), \deg^*(V)\}$.
(L'4) If U and $V \in \mathscr{L}^*$ then $UV \in \mathscr{L}^*$ provided that $\deg^*(UV) := \deg^*(U) + \deg^*(V) \leq 2$.
(L'5) If $U \in \mathscr{L}^*$ then $E_t(U) \in \mathscr{L}^*$ and $\deg^*(E_t(U)) := \deg^*(U)$.

The smallest set $\mathscr{L}^*$ satisfying (L'1)–(L'5) is the set of all expressions containing gambles in $\mathscr{B}_1$ and $\mathscr{B}_2$, linear combinations, products, and applications of E_t, whose degree is less or equal than 2. We could obviously relax the degree requirement, and let $\mathscr{L}^*$ be a full polynomial ring, however doing so would require us to assume that we can specify a future coherent prevision E_t on the full polynomial ring in the variables $\mathscr{B}_1$. But, the distribution of a variable is often (although not always) uniquely determined by its moments [10, Section 11]. So if we go for a full polynomial ring, we might as well assume that we can specify a possibility space, which is precisely what we set out to avoid.

We emphasize that $\deg^*(U)$ is the degree of U *as an expression*, which may be larger than the polynomial degree. For example, let $X \in \mathscr{B}_1$, and consider $U := X^2 - X^2$. As an expression, $\deg^*(X^2 - X^2) = 2$. As a polynomial, obviously the degree of $X^2 - X^2$ is 0.

In order to extend the sup operator from $\mathscr{L}$ to a sup operator on $\mathscr{L}^*$, we use a similar construction as before. Let $\mathscr{L}_1$ denote the set of all constant, linear, and quadratic polynomials in the variables $\mathscr{B}_1$, and let $\mathscr{L}_2$ denote the set of all constant, linear, and quadratic polynomials in the variables $\mathscr{B}_2$. For any element X of $\mathscr{L}_1$ or $\mathscr{L}_2$, we denote the polynomial degree of X by $\deg(X)$.

For any coherent prevision Q on $\mathscr{L}_1$ and any $U \in \mathscr{L}^*$, we can project U onto a gamble $\pi_Q U \in \mathscr{L}$ by replacing E_t by Q in the expression of U, where it is understood that we treat all appearances of $Z \in \mathscr{L}_2$ in the remaining expression as constants. We then define:

$$\sup(U) := \sup_{Q \in \mathbb{E}(\mathscr{L}_1)} \sup(\pi_Q U). \tag{41}$$

where $\mathbb{E}(\mathscr{L}_1)$ denotes the set of all coherent previsions on $\mathscr{L}_1$.

The following property is of particular interest:

Proposition 19. *For all $U \in \mathscr{L}^*$ with $\deg^*(U) \leq 1$, we have that $\sup(U^2) \geq 0$.*

Proof. By definition of π_Q, we have that $\pi_Q(U^2) = (\pi_Q U)^2$, and $\sup\big((\pi_Q U)^2\big) \geq 0$ by assumption. □

Because of the linearity of Q, we have the following result:

Proposition 20. *For every $U \in \mathscr{L}^*$, there are Y_i and $Y'_{ij} \in \mathscr{L}_1$, $Z_i \in \mathscr{L}_2$, and $k_j \in \mathbb{N}$, such that*

$$U \simeq \sum_{i=1}^{n} Z_i Y_i \prod_{j=1}^{k_i} E_t(Y'_{ij}) \tag{42}$$

where $\deg(Z_i) + \deg(Y_i) + \sum_{j=1}^{k_i} \deg(Y'_{ij}) \leq 2$, for all $i \in \{1, \dots, n\}$. In that case,

$$E_t(U) \simeq \sum_{i=1}^{n} Z_i E_t(Y_i) \prod_{j=1}^{k_i} E_t(Y'_{ij}) \tag{43}$$

This theorem says that $\mathscr{L}^*$ is, up to equivalence, equal to the set of all polynomials generated by $\mathscr{B}_1 \cup \mathscr{B}_2 \cup \{E_t(Y) \colon Y \in \mathscr{L}_1\}$ subject to a degree constraint. It also tells us that, when written in the above form, we can apply E_t simply by replacing the Y_i terms by $E_t(Y_i)$.

Note that this does not work for lower previsions, because we cannot treat terms of the form $\underline{E}_t(ZY)$: for any $z \in \mathbb{R}$, $\underline{Q}(zY)$ is equal to $z\underline{Q}(Y)$ when $z \geq 0$ and equal to $z\overline{Q}(X)$ when $z < 0$: we cannot simply pull constants out unless we know their sign. The linearity of E_t is really key to make Theorem 20 work.

5.2. **Temporal Natural Extension.** We can define again a natural extension, as before. Note that this natural extension is inherently imprecise, even if our current beliefs $\underline{E}_0$ are precise (i.e. when $\underline{E}_0$ is a prevision).

$$\underline{E}_0^*(U) \coloneqq \sup_{\substack{\alpha\in\mathbb{R} \\ X\in\mathscr{L}\colon \underline{E}_0(X)\geq 0 \\ V\in\mathscr{L}^*\colon E_t(V)\succeq 0}} \{\alpha \colon U - \alpha \succeq X + V\} \tag{44}$$

$$= \sup_{X\in\mathscr{L}, V\in\mathscr{L}^*} \inf\,[U - X + \underline{E}_0(X) - V + E_t(V)] \tag{45}$$

The proof of the second expression for the natural extension follows in essence the proof of Theorem 14. We can also show that, for all $X \in \mathscr{L}$,

$$\underline{E}_0^*(X) = \underline{E}_0(X) \tag{46}$$

(see the proof of Theorem 15), and that, for all $U \in \mathscr{L}^*$,

$$\underline{E}_0^*(U) \geq \inf E_t(U) \tag{47}$$

(see the proof of Theorem 16).

The next proposition, which is in essence due to Goldstein [4, Theorem 1], forms the basis for linking future beliefs about expectation and variance to current beliefs about expectation and variance.

Proposition 21. *For every $U \in \mathscr{L}^*$ and $V \in \mathscr{L}^*$ with $\deg^*(U) \leq 1$ and $\deg^*(V) \leq 1$, and such that $E_t(V) \simeq V$, we have that*

$$\underline{E}_0^*((U-V)^2 - (U - E_t(U))^2)) \geq 0 \tag{48}$$

Proof. First note that, by Theorem 20, if $E_t(V) \simeq V$, then we must have the following representation for V:

$$V \simeq \sum_{i=1}^{n} Z_i \prod_{j=1}^{k_i} E_t(Y'_{ij}) \tag{49}$$

with $Y'_{ij} \in \mathscr{L}_1$ and $Z_i \in \mathscr{L}_2$. Consequently, we must also have a similar representation for V^2 (with different gambles of course). It follows, again by Theorem 20, that,

$$E_t(-2VU + V^2) \simeq -2V E_t(U) + V^2. \tag{50}$$

Note that, by Eq. (47),

$$\underline{E}_0^*((U-V)^2 - (U - E_t(U))^2)) \tag{51}$$

$$\geq \inf E_t((U-V)^2 - (U - E_t(U))^2) \tag{52}$$

$$= \inf E_t(-2VU + V^2 + 2U E_t(U) - E_t(U)^2) \tag{53}$$

and now by Eq. (50) and Theorem 20,

$$= \inf\left(-2VE_t(U)+V^2+2E_t(U)^2-E_t(U)^2\right) \tag{54}$$

$$= \inf\left((E_t(U)-V)^2\right) \geq 0 \tag{55}$$

where we used Theorem 19. □ □

The following is a well known result:

Lemma 22. *For every $U \in \mathscr{L}^*$ with $\deg^*(U) \leq 1$ and every coherent prevision E on $\mathscr{L}^*$, we have that*

$$\arg\min_{a\in\mathbb{R}} E((U-a)^2) = E(U) \tag{56}$$

Proof. Simply note that

$$E((U-a)^2) = E((U-E(U))^2) + (E(U)-a)^2. \tag{57}$$

□ □

If we combine this with Theorem 21, we get:

Proposition 23. *For every $U \in \mathscr{L}^*$ with $\deg(U) \leq 1$ we have that*

$$\underline{E}_0^*(U-E_t(U)) = \overline{E}_0^*(U-E_t(U)) = 0. \tag{58}$$

Proof. Fix any $E \in \mathscr{M}(\underline{E}_0^*)$.

We now follow the argument presented in [4, Corollary 1]. By Theorem 21, with $V := E_t(U)+a$,

$$E((U-E_t(U)-a)^2) \geq E((U-E_t(U))^2)) \tag{59}$$

for all $a \in \mathbb{R}$. Consequently, it follows that

$$\arg\min_{a\in\mathbb{R}} E((U-E_t(U)-a)^2) = 0. \tag{60}$$

By Eq. (56), this means that $E(U-E_t(U)) = 0$.

Since this holds for all $E \in \mathscr{M}(\underline{E}_0^*)$, by Theorem 5, it must be that $\underline{E}_0^*(U-E_t(U)) = \overline{E}_0^*(U-E_t(U)) = 0$. □ □

Interestingly, we can drop the degree condition on U, if we use a slightly different method of proof:

Proposition 24. *For every $U \in \mathscr{L}^*$ we have that*

$$\underline{E}_0^*(U-E_t(U)) = \overline{E}_0^*(U-E_t(U)) = 0. \tag{61}$$

Proof. By Eq. (47),

$$\underline{E}_0^*(U-E_t(U)) \geq \inf E_t(U-E_t(U)) = 0 \tag{62}$$

and similarly,

$$\overline{E}_0^*(U-E_t(U)) = -\underline{E}_0^*(-U-E_t(-U)) \tag{63}$$

$$\leq -\inf E_t(-U-E_t(-U)) = 0 \tag{64}$$

Putting everything together, and using $\underline{E}_0^* \leq \overline{E}_0^*$,

$$0 \leq \underline{E}_0^*(U-E_t(U)) \leq \overline{E}_0^*(U-E_t(U)) \leq 0 \tag{65}$$

establishing the desired equality. □ □

If we apply Theorem 24 any $X \in \mathscr{L}$ with $\underline{E}_0(X) = \overline{E}_0(X) = E_0(X)$, then we recover [4, Corollary 1]: we find that $\underline{E}_0^*(E_t(X)) = \overline{E}_0^*(E_t(X)) = E_0(X)$ (we also relied on Eq. (46)). This is very similar to the usual definition of conglomerability as in for instance [13, p. 305, (C15)]. It is worth emphasizing that this is *not* your usual conglomerability, because $E_t(X)$ is not necessarily obtained through conditioning. We also emphasize that Theorem 24 applies also to situations where $\underline{E}_0(X) \neq \overline{E}_0(X)$, so it is more general than [4, Corollary 1]. Because our framework uses a much weaker structure, we also side-step the non-conglomerability issues one has when relying on finitely additive probability measures [7, 8].

The next lemma is also well known; see [4] or [5, p. 55–57]:

Lemma 25. *For every coherent prevision E on $\mathscr{L}^*$, and every U and $V \in \mathscr{L}^*$ with $\deg^*(U) \leq 1$ and $\deg^*(V) \leq 1$, we have that*

$$\operatorname{var}_V(U) := \min_{a,b\in\mathbb{R}} E((U-a-bV)^2) = \operatorname{var}(U) - \frac{\operatorname{cov}(U,V)^2}{\operatorname{var}(V)} \tag{66}$$

where the minimum is achieved for $a + bV$ equal to

$$E_V(U) := E(U) + \frac{\operatorname{cov}(U,V)}{\operatorname{var}(V)}(V - E(V)) \tag{67}$$

Using this, we can also say something about the expected future variance, that is, $\operatorname{var}_t(U)$.

Proposition 26. *For every $U \in \mathscr{L}^*$ and $V \in \mathscr{L}^*$ with $\deg^*(U) \leq 1$ and $\deg^*(V) \leq 1$, and such that $E_t(V) \simeq V$, and every $E \in \mathscr{M}(\underline{E}_0^*)$, we have that*

$$E(\operatorname{var}_t(U)) \leq \operatorname{var}_V(U), \tag{68}$$

Proof. By Theorem 24, we have that $E(U') = E(E_t(U'))$, for $U' := (U - E_t(U))^2$. So,

$$E(\operatorname{var}_t(U)) = E(E_t((U - E_t(U))^2)) = E((U - E_t(U))^2) \tag{69}$$

and now, by Theorem 21,

$$\leq E((U - E_V(U))^2) = \operatorname{var}_V(U) \tag{70}$$

$\square$ $\square$

Note that we could not have used Theorem 23 in the proof, because we would have that $\deg^*(U') = 2$ when $\deg^*(U) = 1$.

In particular, if $U \in \mathscr{L}_1$ and $V \in \mathscr{L}_2$, and our assessments at time 0 constitute a prevision E_0, then $\operatorname{var}_V(U)$ is precisely known, and we have an upper bound on the expected variance of our future beliefs.

All these results demonstrate that our framework fully encapsulates standard Bayes linear calculus.

6. Conclusion

We have discussed lower previsions without possibility spaces, and used this framework to set up a theory of lower previsions that fully treats future beliefs as random quantities. First, we revisited the theory of lower previsions but starting from a bounding operator instead of a possibility space. We proved that many results carry over immediately to this more general framework. Next, we constructed a set of gambles as a set of expressions that involved also future beliefs as random quantities. We used a linear projection operator to induce a logical bounding operator on these expressions. This enabled us to characterize coherent lower previsions on such set.

For simplicity, in the current paper, for the case where we work with lower previsions, we limited ourselves to sums of variables that are known at time t and variables that remain unknown. This separation is quite natural: one class corresponds to data and one to quantities for which we will use the data to make inferences. Linearity allows us to explore how the linear operations of Bayes linear theory extend to lower previsions. Nevertheless, it would be interesting to see if this limitation could be lifted. To allow products, as in the full Bayes linear case, a minimal requirement would be to be able to consider the positive and negative parts of any gamble Z whose value is known at time t, to allow for instance:

$$\underline{E}_t(YZ) \simeq (Z \vee 0)\underline{E}_t(Y) + (Z \wedge 0)\overline{E}_t(Y) \tag{71}$$

We can know that $Z \vee 0$ and $Z \wedge 0$ belong to $\mathscr{L}_2$ if $\mathscr{L}_2$ forms a Riesz space (and assuming that $\mathscr{L}_2$ also contains all constants). This is a much larger structure than just a linear space, and requires a lot more detail to be specified.

Following [12], we used the temporal sure preference principle in the context of desirability and lower previsions, however, our framework did so in a more constructive way, and in a way that explicitly models random quantities that are known at a future time t.

We identified an expression for natural extension under the temporal sure preference principle, and established consistency of the temporal sure preference principle with prior specifications, which also guarantees that those prior specifications are not modified by adopting the temporal sure preference principle, so we can still use the usual (non-temporal) form of natural extension for gambles as far as our current beliefs are concerned. We have also derived a host of bounds on lower and upper expectations of future lower and upper expecations.

We then extended the framework to sets of gambles that form constant, linear, and quadratic polynomials. We noted that a reduced representation is possible when future beliefs are precise. From this representation, we recovered the basic properties of standard Bayes linear calculus.

APPENDIX A. PROOFS

Proof of Theorem 1. Remember that we defined $\|X\| := \max\{\sup(X), \sup(-X)\}$.

First, for all $a \in \mathbb{R}$,

$$\begin{aligned} \|aX\| = \max\{\sup(|a|X), \sup(-|a|X)\} &= \max\{|a|\sup(X), |a|\sup(-X)\} \\ &= |a|\max\{\sup(X), \sup(-X)\} = |a|\|X\|. \end{aligned} \tag{72}$$

Next,

$$\|X+Y\| = \max\{\sup(X+Y), \sup(-X-Y)\} \tag{73}$$

$$\leq \max\{\sup(X)+\sup(Y), \sup(-X)+\sup(-Y)\} \tag{74}$$

$$\leq \max\{\sup(X), \sup(-X)\} + \max\{\sup(Y), \sup(-Y)\} = \|X\| + \|Y\|. \tag{75}$$

And, because $\inf(X) \leq \sup(X)$,

$$\sup(X) \leq 0 \implies \inf(X) = -\sup(-X) \leq 0 \implies \sup(-X) \geq 0 \tag{76}$$

and therefore $\|X\| \geq 0$. Finally, we also have that

$$\|X\| = 0 \implies \sup(X) = -\sup(-X) = 0 \implies X = 0 \tag{77}$$

because of the uniqueness of the zero gamble. So, $\|\cdot\|$ is a norm. □ □

Proof of Theorem 2. C1. Take all $\lambda_i = 0$ in Eq. (4).

C2. Consider any n and $m \in \mathbb{N}$ with $n < m$, any $\lambda_1 \geq 0, \dots, \lambda_m \geq 0$, and any $X_1, \dots, X_m \in \operatorname{dom}\underline{P}$. Assume that

$$X - \alpha \succeq \sum_{i=1}^{n} \lambda_i (X_i - \underline{P}(X_i)) \tag{78}$$

$$Y - \beta \succeq \sum_{i=n+1}^{m} \lambda_i (X_i - \underline{P}(X_i)) \tag{79}$$

By definition of $\succeq$, this means that

$$\inf\left(X - \sum_{i=1}^{n} \lambda_i (X_i - \underline{P}(X_i))\right) \geq \alpha \tag{80}$$

$$\inf\left(Y - \sum_{i=n+1}^{m} \lambda_i (X_i - \underline{P}(X_i))\right) \geq \beta \tag{81}$$

Because $\inf(V + W) \geq \inf(V) + \inf(W)$, it follows that

$$\inf\left(X + Y - \sum_{i=1}^{m} \lambda_i (X_i - \underline{P}(X_i))\right) \geq \alpha + \beta \tag{82}$$

or, in other words,

$$X + Y - \alpha - \beta \succeq \sum_{i=1}^{m} \lambda_i (X_i - \underline{P}(X_i)) \tag{83}$$

Consequently, for every α and β such that Eqs. (78) and (79) are satisfied, we can find a $\gamma \geq \alpha + \beta$ (actually, we can even take it to be equal) such that

$$X + Y - \gamma \succeq \sum_{i=1}^{m} \lambda_i (X_i - \underline{P}(X_i)) \tag{84}$$

By the definition of $\underline{E}$ (Eq. (4)), it thus follows that $\underline{E}(X+Y) \geq \underline{E}(X) + \underline{E}(Y)$.

C3. For $\lambda = 0$, this follows from the avoiding sure loss condition. For $\lambda > 0$, note that $X \succeq Y$ if and only if $\lambda X \succeq \lambda Y$, because $\inf(\lambda X - \lambda Y) = \lambda \inf(X - Y)$. By the definition of $\underline{E}$ (Eq. (4)), it now easily follows that $\underline{E}(\lambda X) = \lambda \underline{E}(X)$. □ □

Proof of Theorem 3. We establish each of the properties.

- D1 If $X \succeq 0$ then $\underline{E}(X) \geq \inf(X) \geq 0$ so $X \in \mathscr{D}$.
- D2 If $\sup X < 0$ then $\underline{E}(X) \leq \sup(X) < 0$ so $X \notin \mathscr{D}$.
- D3 If $X \in \mathscr{D}$ and $Y \in \mathscr{D}$ then $\underline{E}(X+Y) \geq \underline{E}(X) + \underline{E}(Y) \geq 0$ so $X + Y \in \mathscr{D}$.
- D4 If $\lambda \geq 0$ and $X \in \mathscr{D}$ then $\underline{E}(\lambda X) = \lambda \underline{E}(X) \geq 0$ so $\lambda X \in \mathscr{D}$.
- D5 If $X + \varepsilon \in \mathscr{D}$ for all $\varepsilon > 0$, then $\underline{E}(X + \varepsilon) = \underline{E}(X) + \varepsilon \geq 0$ for all $\varepsilon > 0$, so it must be that $\underline{E}(X) \geq 0$ as well, and thus $X \in \mathscr{D}$.

□ □

Proof of Theorem 4. Because $\underline{E}(X - a) = \underline{E}(X) - a$ for all $a \in \mathbb{R}$, it follows that

$$\max\{a \in \mathbb{R} \colon X - a \in \mathscr{D}\} = \max\{a \in \mathbb{R} \colon \underline{E}(X - a) \geq 0\} \tag{85}$$

$$= \max\{a \in \mathbb{R} \colon a \leq \underline{E}(X)\} = \underline{E}(X) \tag{86}$$

□ □

Proof of Theorem 5. This proof is a simplified version of the proof given in [11, p. 71–72, proof of Theorem 4.38(iii)].

Clearly $\underline{E}(X) \geq \inf_{P\in\mathscr{M}(\underline{E})} P(X)$ since $P(X) \geq \underline{E}(X)$ for all $P \in \mathscr{M}(\underline{E})$. We are left to show that $\underline{E}(X) = P(X)$ for some $P \in \mathscr{M}(\underline{E})$.

Consider the linear functional Λ defined on the linear space $\{aX : a \in \mathbb{R}\}$ by

$$\Lambda(aX) := a\underline{E}(X) \tag{87}$$

Note that

$$\Lambda(aX) = a\underline{E}(X) = \begin{cases} \underline{E}(aX) \leq \overline{E}(aX) & \text{if } a \geq 0 \\ \overline{E}(aX) & \text{if } a < 0 \end{cases} \tag{88}$$

so $\overline{E}$ dominates Λ on $\operatorname{dom}\Lambda$. Consequently, since $\overline{E}$ is sublinear, by the Hahn-Banach theorem [6, Section 12.31, (HB3)], Λ can be extended to a linear functional P on $\mathscr{L}$ such that $P(Y) \leq \overline{E}(Y)$ for all $Y \in \mathscr{L}$. In particular, for all Y and $Z \in \mathscr{L}$,

(i) $P(Y) \leq \overline{E}(Y) \leq \sup(Y)$,
(ii) $P(\lambda Y) = \lambda P(Y)$ for all $\lambda \geq 0$, and
(iii) $P(Y+Z) = P(Y) + P(Z)$.

Also $P(Y) = -P(-Y)$. Consequently, $P \in \mathscr{M}(\underline{E})$, and $P(X) = \underline{E}(X)$ by construction, establishing the desired equality. □ □

Proof of Theorem 7. For (S1), note that, by (P2):

$$\sup(U+V) = \sup_{\underline{Q}\in\mathbb{E}(\mathscr{L}_1)} \sup(\pi_{\underline{Q}}U + \pi_{\underline{Q}}V) \tag{89}$$

$$\leq \sup_{\underline{Q}\in\mathbb{E}(\mathscr{L}_1)} \left(\sup \pi_{\underline{Q}}U + \sup \pi_{\underline{Q}}V\right) \tag{90}$$

$$\leq \left(\sup_{\underline{Q}\in\mathbb{E}(\mathscr{L}_1)} \sup \pi_{\underline{Q}}U\right) + \left(\sup_{\underline{Q}\in\mathbb{E}(\mathscr{L}_1)} \sup \pi_{\underline{Q}}V\right) = \sup U + \sup V \tag{91}$$

(S2) follows similarly by (P3). (S3) follows because sup already satisfies (S3) on $\mathscr{L}$ and by (P1). □ □

Proof of Theorem 8. Note that, by definition,

$$\pi_{\underline{Q}}(\underline{E}_t(a+Z)) = \underline{Q}(a+Z) = a+Z = \pi_{\underline{Q}}(a+Z). \tag{92}$$

So $\inf((a+Z) - \underline{E}_t(a+Z)) = \sup((a+Z) - \underline{E}_t(a+Z)) = 0$, establishing the desired indifference. □ □

Proof of Theorem 9. We first show the last equivalence. Consider any $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{R}$, $Y_0, Y_1, \dots, Y_n \in \mathscr{L}_1$, and $Z \in \mathscr{L}_2$, such that $U \simeq Y_0 + \sum_{i=1}^n a_i\underline{E}_t(Y_i) + Z$. Then, for any $\underline{Q} \in \mathbb{E}(\mathscr{L}_1)$,

$$\pi_{\underline{Q}}(U) = \pi_{\underline{Q}}\left(Y_0 + \sum_{i=1}^n a_i\underline{E}_t(Y_i) + Z\right) \tag{93}$$

$$= Y_0 + \sum_{i=1}^n a_i\underline{Q}(X_i) + Z \tag{94}$$

Then, by Eq. (94),

$$\pi_{\underline{Q}}(\underline{E}_t U) = \underline{Q}\left(Y_0 + \sum_{i=1}^{n} a_i \underline{Q}(Y_i) + Z\right) \tag{95}$$

$$= \underline{Q}(Y_0) + \sum_{i=1}^{n} a_i \underline{Q}(Y_i) + Z \tag{96}$$

$$= \pi_{\underline{Q}}\left(\underline{E}_t(Y_0) + \sum_{i=1}^{n} a_i \underline{E}_t(X_i) + Z\right) \tag{97}$$

Consequently, indeed, it must be that $\underline{E}_t(U) \simeq \underline{E}_t(Y_0) + \sum_{i=1}^{n} a_i \underline{E}_t(X_i) + Z$.

For the first equivalence, we proceed by structural induction. Clearly, for any $X \in \mathscr{L}$, $X = \pi_1 X + \pi_2 X$ so the the statement holds for $n = 0$, $Y_0 = \pi_1 X$, and $Z = \pi_2 X$. If the statement holds for U and V, then obviously it will also hold for $U + V$ and for aU for all $a \in \mathbb{R}$. We are left to show that if the statement holds for U, then it also holds for $\underline{E}_t(U)$. This follows immediately from the first part of the proof.

To show the uniqueness of Z, simply note that we can obtain the Z component of U by applying the π_2 operator to $\pi_{\underline{Q}} U$.

To show the uniqueness of Y_0 up to a positive constant, assume that

$$Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z \simeq Y_0' + \sum_{j=1}^{m} b_i \underline{E}_t(Y_i') + Z' \tag{98}$$

We already know that $Z \simeq Z'$, so we can drop both terms from the above equivalence. We are left with

$$Y_0 - Y_0' \simeq \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) - \sum_{j=1}^{m} b_i \underline{E}_t(Y_i') \tag{99}$$

Equivalently, for every $\underline{Q} \in \underline{\mathbb{E}}(\mathscr{L}_1)$,

$$Y_0 - Y_0' = \sum_{i=1}^{n} a_i \underline{Q}(Y_i) - \sum_{j=1}^{m} b_i \underline{Q}(Y_i') \tag{100}$$

The left hand side does not depend on $\underline{Q}$, so it can only be that the right hand side is a constant. □ □

Proof of Theorem 10. Assume $U \simeq V$. It suffices to show that $\pi_{\underline{Q}}(\underline{E}_t U) = \pi_{\underline{Q}}(\underline{E}_t V)$ for all $\underline{Q} \in \underline{\mathbb{E}}(\mathscr{L}_1)$. Indeed, first note that, by Theorem 9, there are n, $m \in \mathbb{N}$, $a_1, \dots, a_n$, $b_1, \dots, b_m$, $c \in \mathbb{R}$, $Y_0, Y_1, \dots, Y_n, Y_1', \dots, Y_m' \in \mathscr{L}_1$, and $Z \in \mathscr{L}_2$ such that

$$U \simeq Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z \tag{101}$$

$$V \simeq Y_0 + c + \sum_{j=1}^{m} b_i \underline{E}_t(Y_i') + Z \tag{102}$$

Because $U \simeq V$,

$$Y_0 + \sum_{i=1}^{n} a_i \underline{Q}(Y_i) + Z = \pi_{\underline{Q}}\left(Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z\right) \tag{103}$$

$$= \pi_{\underline{Q}}\left(Y_0 + c + \sum_{j=1}^{m} b_i \underline{E}_t(Y_i') + Z\right) \tag{104}$$

$$= Y_0 + c + \sum_{j=1}^{m} b_i \underline{Q}(Y_i') + Z \tag{105}$$

So, again by Theorem 9, and by the above equality,

$$\pi_{\underline{Q}}(\underline{E}_t U) = \underline{Q}(Y_0) + \sum_{i=1}^{n} a_i \underline{Q}(Y_i) + Z \tag{106}$$

$$= \underline{Q}(Y_0) + c + \sum_{j=1}^{m} b_i \underline{Q}(Y_i') + Z \tag{107}$$

$$= \underline{Q}(Y_0 + c) + \sum_{j=1}^{m} b_i \underline{Q}(Y_i') + Z \tag{108}$$

$$= \pi_{\underline{Q}}(\underline{E}_t V) \tag{109}$$

$\square$ $\square$

Proof of Theorem 12. Let $U \simeq Y_0 + \sum_{i=1}^{n} a_i \underline{E}_t(Y_i) + Z$. Note that $\underline{E}_t(U) \succeq 0$ if and only if

$$0 \leq \inf(\underline{E}_t U) \tag{110}$$

$$= \inf_{\underline{Q} \in \mathbb{E}(\mathscr{L}_1)} \inf(\pi_{\underline{Q}}(\underline{E}_t U)) \tag{111}$$

$$= \inf_{\underline{Q} \in \mathbb{E}(\mathscr{L}_1)} \inf\left(\underline{Q}(Y_0) + \sum_{i=1}^{n} a_i \underline{Q}(Y_i) + Z\right) \tag{112}$$

$$= \inf_{\underline{Q} \in \mathbb{E}(\mathscr{L}_1)} \left(\underline{Q}(Y_0) + \sum_{i=1}^{n} a_i \underline{Q}(Y_i)\right) + \inf Z \tag{113}$$

$\square$ $\square$

Proof of Theorem 20. Let $U \in \mathscr{L}^*$. It suffices to show that there are Y_i and $Y_{ij}' \in \mathscr{L}_1$, $Z_i \in \mathscr{L}_2$, and $k_j \in \mathbb{N}$, such that for all $Q \in \mathbb{E}(\mathscr{L}_1)$,

$$\pi_Q U = \sum_{i=1}^{n} Z_i Y_i \prod_{j=1}^{k_i} Q(Y_{ij}') \tag{114}$$

We prove this by structural induction.

Clearly, the statement is satisfied for $X \in \{1\} \cup \mathscr{B}_1 \cup \mathscr{B}_2$ because then X easily can be written in the form $\sum_{i=1}^{n} Z_i Y_i$ by definition. Also, if the statement is satisfied for U and V, then it is satisfied for $U + V$. Since a product of sums can be rearranged into a sum of products, it will also be satisfied for UV, and $\deg^*(U) + \deg^*(V) \leq 2$ ensures that the polynomial degree condition is satisfied. Is it satisfied for $E_t(U)$? Indeed,

$$\pi_Q(E_t U) = Q\left(\sum_{i=1}^{n} Z_i Y_i \prod_{j=1}^{k_i} Q(Y_{ij}')\right) \tag{115}$$

and now, using the linearity of Q (remember that the terms Z_i are treated as constants),

$$= \sum_{i=1}^{n} Z_i Q(Y_i) \prod_{j=1}^{k_i} Q(Y'_{ij}) \tag{116}$$

This last equality also proves the equivalence expression for $E_t(U)$. □ □

REFERENCES

[1] Peter S. Craig, Michael Goldstein, Jonathan C. Rougier, and Allan H. Seheult. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454):717–729, June 2001.

[2] Bruno de Finetti. *Theory of Probability: A Critical Introductory Treatment*. Wiley, New York, 1974-5. Two volumes.

[3] Michael Goldstein. Temporal coherence. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, *Bayesian Statistics 2*. 1985.

[4] Michael Goldstein. Prior inferences for posterior judgements. In M. C. D. Chiara et al., editors, *Structures and Norms in Science*, pages 55–71. Kluwer, 1997.

[5] Michael Goldstein and David A. Wooff. *Bayes Linear Statistics: Theory and Methods*. Wiley, Chichester, 2007.

[6] Eric Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, 1997.

[7] M. J. Schervish, T. Seidenfeld, and J. B. Kadane. The extent of non-conglomerability of finitely additive probabilities. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, (66):205–226, 1984.

[8] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Non-conglomerability for finite-valued, finitely additive probability. *Sankhya: The Indian Journal of Statistics, Series A*, 60(3):476–491, 1998.

[9] Teddy Seidenfeld and Larry Wasserman. Dilation for sets of probabilities. *The Annals of Statistics*, 21(3):1139–1154, 1993.

[10] Jordan M. Stoyanov. *Counterexamples in Probability*. Dover, 3rd edition, 2013.

[11] Matthias C. M. Troffaes and Gert de Cooman. *Lower Previsions*. Wiley Series in Probability and Statistics. Wiley, 2014.

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

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

[14] Peter M. Williams. Notes on conditional previsions. Technical report, School of Math. and Phys. Sci., Univ. of Sussex, 1975.

[15] Peter M. Williams. Notes on conditional previsions. *International Journal of Approximate Reasoning*, 44(3):366–383, 2007.

[16] Marco Zaffalon and Enrique Miranda. Probability and time. *Artificial Intelligence*, 198:1–51, May 2013.

DURHAM UNIVERSITY, DEPARTMENT OF MATHEMATICAL SCIENCES, UK
Email address: `matthias.troffaes@durham.ac.uk`

DURHAM UNIVERSITY, DEPARTMENT OF MATHEMATICAL SCIENCES, UK
Email address: `michael.goldstein@durham.ac.uk`