



A scoring rule and global inaccuracy measure for contingent varying importance

Pavel Janda¹

Accepted: 11 August 2023
© The Author(s) 2023

Abstract

Levinstein recently presented a challenge to accuracy-first epistemology. He claims that there is no strictly proper, truth-directed, additive, and differentiable scoring rule that recognises the contingency of varying importance, i.e., the fact that an agent might value the inaccuracy of her credences differently at different possible worlds. In my response, I will argue that accuracy-first epistemology can capture the contingency of varying importance while maintaining its commitment to additivity, propriety, truth-directedness, and differentiability. I will construct a scoring rule — a weighted scoring rule — and a global inaccuracy measure that has all four required properties and recognises the contingency of varying importance. I will show that Levinstein’s and my results coexist without contradicting each other

Keywords Contingent importance · Weighted scoring rules · Strict propriety · Accuracy

1 Introduction

In a recent paper “An objection of varying importance to epistemic utility theory” published in *Philosophical Studies*, Levinstein presented the following challenge to accuracy-first epistemology (see Theorem A.4 in Levinstein, 2019, p. 2929). He claims that no scoring rule with four properties (additivity, propriety, truth-directedness, and differentiability) usually assumed by accuracy-firsters (see Levinstein, 2019, p. 2928 and, for example, compare to Joyce, 2009 or Pettigrew, 2016) recognises the contingency of varying importance, i.e., the fact that an agent might value the inaccuracy of her credences differently at different possible worlds. In my response, I will argue that accuracy-first epistemology can accommodate the contingency of varying importance while maintaining its commitment to additivity,

✉ Pavel Janda
pavel.janda.early@gmail.com

¹ Department of Architecture and Urban Studies (DASu), Politecnico di Milano, via Bonardi 3, 20133 Milan, Italy

propriety, truth-directedness, and differentiability. I will construct an inaccuracy measure of individual credences, s_*^X , and an inaccuracy measure of entire credence functions, \mathcal{I}_{s_*} , that has all four required properties and recognises the contingency of varying importance. I will also show that Levinstein's and my results coexist without contradiction and explain why it is so.

One of the main goals of accuracy-first epistemology is to quantify epistemic value of one's credences (degrees of belief). The idea behind such quantification is alethic (Levinstein 2019, p. 2921), i.e., the higher one's credences in truths and the lower one's credences in falsehoods, the more epistemic value those credences have. In this context, one often speaks of the accuracy of one's credences and assumes (which I will follow in this paper) that accuracy is what is ultimately of epistemic value (e.g., see Levinstein, 2019, p. 2920 or Pettigrew's veritism in Pettigrew, 2016, p. 8). However, from the formal point of view, it is often more convenient to quantify epistemic disvalue (i.e., inaccuracy) rather than epistemic value (i.e., accuracy). In this paper, I will quantify inaccuracy and work with measures of inaccuracy such as scoring rules; this brings no significant change since the accuracy of a credence is the negative of its inaccuracy (see Pettigrew, 2019, p. 141) or is similarly related.¹ So, I will say that the lower the credences in truths and the higher the credences in falsehoods, the more *inaccurate* credences one has. I will assume that one aims to minimise (expected) inaccuracy of one's credences since an inaccuracy score is a form of penalty.

One immediate question is what properties the measures of inaccuracy should have, and a natural follow-up question is whether there exists a measure of inaccuracy that has all the desired properties. Levinstein's Theorem A.4 lists four properties (additivity, propriety, truth-directedness, and differentiability) that accuracy-firsters usually want measures of inaccuracy to have, but it also adds a fifth property to the list, recognising the contingency of varying importance or contingent importance, for short. This last property, roughly speaking, means that the degree to which the inaccuracy of one's credence in a proposition matters can differ from one proposition to another and from one possible world to another (see Levinstein, 2019, pp. 2925–2927). To me, this sounds like a desirable additional property. Theorem A.4 (Levinstein, 2019, p. 2929),² however, claims that no scoring rule with the four aforementioned properties recognises the contingency of varying importance.

¹ When Levinstein discusses accuracy-first epistemology, he often uses an epistemic utility measure U (see Levinstein, 2019, pp. 2922–2924). Since, by assumption, an agent wants to maximise (expected) utility, Levinstein's discussion is often presented in the context of maximising one's (expected) epistemic utility. Levinstein, however, expresses U in terms of a measure of inaccuracy \mathcal{I} , i.e., $U = 1 - \mathcal{I}$ (see Levinstein, 2019, p. 2924 or Levinstein, 2019, p. 2921, for a construction of a specific utility function, the Brier utility). Moreover, in the formulation and the proof of Theorem A.4, he uses \mathcal{I} instead of U . Since the central part of my paper works with Theorem A.4 and its proof, this provides additional motivation to frame my discussion in terms of minimising expected inaccuracy rather than maximising expected epistemic utility.

² Names in brackets next to results, e.g., theorems, indicate those results' original author/authors. Sometimes, I will make minor changes, e.g., notational, to such results.

Theorem A.4 (Levinstein) *There does not exist a scoring rule \mathcal{I} that is additive, proper, truth-directed, and differentiable, such that \mathcal{I} recognises contingent importance.*

Levinstein discusses possible ways out of the predicament established by Theorem A.4 (see Levinstein, 2019, p. 2928). He says that accuracy-firsters cannot give up truth-directedness and propriety, and one also does not want to deny that propositions vary in importance across worlds (see Levinstein, 2019, p. 2928). Levinstein also takes no issue with differentiability (Levinstein 2019, p. 2924), which is a technical property. But he considers denying additivity as a possible way out. On the one hand, such a solution sounds possible since non-additive measures of inaccuracy exist, as shown in Pettigrew (2022). On the other hand, Levinstein himself questions the legitimacy of such a move since, as he says, additivity is a nice property (see Levinstein, 2019, p. 2924). It captures — at least in some cases — a reasonable property that the epistemic value of any of one’s credences in a given world is separable from the other credences one has in that world (Levinstein 2019, p. 2928). Finally, Levinstein tentatively concludes that accuracy-first epistemology cannot capture all facts about epistemic value, including the contingency of varying importance, in an attractive way while maintaining its commitment to propriety and truth-directedness. Thus, according to Levinstein, accuracy-first epistemology fails to capture the entirety of the epistemic value of an agent’s credence function (Levinstein 2019, p. 2928).

I am not sure what constitutes an “attractive way” and whether, in general, accuracy-first epistemology can capture the entirety of the epistemic value of an agent’s credence function. But, in what follows, I will argue that accuracy-first epistemology can capture the contingency of varying importance while maintaining its commitment to additivity, propriety, truth-directedness, and differentiability. That is, I will prove that there exists a weighted scoring rule s_*^X (a weighted inaccuracy measure of individual credences) and a global inaccuracy measure \mathcal{I}_{s_*} (an inaccuracy measure of entire credence functions) constructed from s_*^X that possesses each of those properties listed in Theorem A.4. But since Theorem A.4 is correct (and there are other similar results, e.g., see Ranjan & Gneiting, 2011 or Douven, 2020), the question is how Levinstein’s and my results can coexist without contradiction. The central difference between Levinstein’s and my approach, which I will elaborate on later, in short, goes as follows. Levinstein took a strictly proper inaccuracy measure of individual credences, added non-constant weights to it (to recognise contingent importance), and showed in Theorem A.4 that this combination is not strictly proper. I will combine non-constant weights with an inaccuracy measure of individual credences, i.e., a scoring rule s^X , that is not strictly proper such that the whole combination – a weighted scoring rule s_*^X – is strictly proper (and has all the other properties from Theorem A.4). So, I give up the strict propriety of s^X to circumvent Theorem A.4 but, in the process, construct s_*^X and \mathcal{I}_{s_*} that accommodates all five properties listed in Theorem A.4.

The structure of this paper is as follows. In Sect. 1, I introduce terminology and notation.³ In Sect. 2, I reconstruct Levinstein's proof of Theorem A.4. In Sect. 3, I define s_*^X , construct \mathcal{I}_{s_*} , and show that it has all the properties listed in Theorem A.4. In Sect. 4, I discuss an application of s_*^X to a concrete example.

2 Terminology and notation

2.1 Technical preliminaries

In what follows, assume that Ω is always a finite set of possible worlds ω ; this assumption is consistent with Levinstein's account and accuracy-first epistemology (e.g., see Levinstein, 2019, p. 2921 or Pettigrew, 2016, p. 222). Let \mathcal{F} be either a σ -algebra defined on Ω (i.e., a set of subsets of Ω that is closed under countable unions, complementation in Ω , and contains Ω) or a set that is extendible to such a σ -algebra. So, (Ω, \mathcal{F}) is a measurable space and let \mathcal{P} be a convex set of probabilistic measures defined on that space. For a finite Ω , I will write $P(\omega)$ instead of $P(\{\omega\})$ for $\omega \in \Omega$ and $P \in \mathcal{P}$ (e.g., see 2.2 in Halpern, 2005 for details).

One can interpret elements of \mathcal{F} (i.e., subsets of Ω) as propositions; I will use capital letters, e.g., X , to denote propositions. A proposition $X \in \mathcal{F}$ is modelled as a set of possible worlds from Ω (e.g., see Halpern, 2005, p. 12). Let a function $c : \mathcal{F} \rightarrow [0, 1]$ be an agent's credence function; this follows Levinstein (see Levinstein, 2019, p. 2921). The value from $[0, 1]$ of one's credence $c(X)$ in $X \in \mathcal{F}$ is one's subjective degree of belief that X is true. For example, assume that X says that it is raining. Then, $c(X) = 0.8$ means that one believes to the degree of 0.8 that it is raining. Further, let $\neg X$ be a set of $\omega \in \Omega$ such that $\omega \notin X$, that is, a set of ω where X is false. So, X and $\neg X$ partition Ω , i.e., $X \cap \neg X = \emptyset$ and $X \cup \neg X = \Omega$. Sometimes, for the sake of simplicity, I will write only ω , X , $\neg X$, P , or c without specifying that it belongs to some Ω , \mathcal{F} , \mathcal{P} , or is defined on (Ω, \mathcal{F}) , respectively.

One's ideal credences (i.e., omniscient, perfectly accurate, or vindicated credences) at a given possible world ω assign the maximal degree of belief to all truths and the minimal degree of belief to all falsehoods at that ω . By convention (e.g., see Pettigrew, 2016, p. 2), I will assume that the maximal credence is 1 and the minimal credence is 0. That is, the ideal credence is 1 if X is true at ω and 0 if X is false at ω . I will use a characteristic function $v_\omega(X)$ to indicate an ideal credence in X at ω , where "v" stands for "vindicated" (e.g., see Pettigrew, 2013, p. 899 for similar terminology and notation). So, $v_\omega(X) = 1$ if $\omega \in X$ and $v_\omega(X) = 0$ if $\omega \notin X$.

³ I try to stay close to Levinstein's terminology and notation, but it will sometimes be convenient to change it to make later sections of the paper notationally manageable while maintaining clarity and to help me compare Levinstein's definitions and results to other definitions and results in the field. I will always point out major divergences from Levinstein's terminology and the notation that I use.

2.2 Inaccuracy measures

Assume, for now, that one wants to quantify the epistemic disvalue of an agent's credence in a single proposition X . Accuracy-firsters use functions called scoring rules to measure the inaccuracy of one's credence in a single proposition at a possible world ω . A scoring rule s takes the omniscient credence in X at ω together with one's credence $c(X)$ and returns an inaccuracy score $s(v_\omega(X), c(X))$ of having credence $c(X)$ in X at that ω , see (Pettigrew, 2016, p. 36).⁴ Formally, one can define a scoring rule as follows (see Definition I.B.3 in Pettigrew, 2016, p. 86).

Definition 1 (*Scoring Rule*) A function $s : \{0, 1\} \times [0, 1] \rightarrow [0, \infty]$ such that $s(0, 0) = s(1, 1) = 0$ is a scoring rule.

Note that a particular credence in a proposition gets a score at a world that depends only on the value of that credence in that proposition and the truth-value of that proposition (i.e., its vindicated credence) at that world. An example of a scoring rule is the squared Euclidean distance: $s(v_\omega(X), c(X)) = (v_\omega(X) - c(X))^2$, e.g., see Pettigrew (2016, pp. 4–5). Sometimes, the requirement that $s(0, 0) = s(1, 1) = 0$ is not a part of the definition of a scoring rule, but I will include it since accuracy-firsters require it anyway. If the context is clear, I will drop the argument of a scoring rule and write only s instead of $s(v_\omega(X), c(X))$.

In Footnote 4 (Levinstein 2019, p. 2924), Levinstein introduces a superscript X , which allows one to use different scoring rules to evaluate credences in different propositions. Following Levinstein, I will use the same superscript to indicate that a given scoring rule is used to score one's credences in a given proposition. For example, I will write $s^X(v_\omega(X), c(X))$ to indicate that s is used to score one's credences in X . But I will assume that one uses the same s to evaluate any of one's credences in a single proposition at any possible world. So, for one's credence $c(X)$ and any two possible worlds $\omega_1, \omega_2 \in \Omega$, $s^X(v_{\omega_1}(X), c(X))$ and $s^X(v_{\omega_2}(X), c(X))$ use the same s .

One usually has credences in more than a single proposition. If one has credences in multiple propositions, one can use scoring rules to measure the inaccuracy score of each of one's individual credences at ω and combine those scores into an overall (or global) score of one's credences at that ω . Accuracy-firsters (e.g., see Pettigrew, 2016, p. 36) call such measures of overall inaccuracy scores "(global) inaccuracy measures". I will use \mathcal{I} to denote such global measures. So, $\mathcal{I}(c, \omega)$ gives an overall inaccuracy score of a credence function c (representing all the credences one has) at a world ω . It will be helpful to use a subscript with \mathcal{I} to indicate what type of scoring rule one uses to construct that global inaccuracy measure. For example, I will write $\mathcal{I}_s(c, \omega)$ to indicate that one uses a scoring rule – possibly different scoring rules for different propositions – satisfying Definition 1 without further qualification.

⁴ In his paper, Levinstein uses function $w(X)$, e.g., see Levinstein (2019, p. 2921), instead of $v_\omega(X)$, but I prefer the other notational standard. Levinstein also switches the order of the arguments of s and writes $s^X(c(X), w)$, where w stands for a possible world, i.e., the ideal credence in X at a possible world w (e.g., see Levinstein, 2019, p. 2924). I will follow Pettigrew's notation (e.g., see Pettigrew, 2016, p. 86) and use $s(v_\omega(X), c(X))$ instead. I will soon discuss the superscripted X .

I will sometimes omit the predicate “global” when talking about global inaccuracy measures, and, if the context is clear, I will drop the argument and write only \mathcal{I}_s instead of $\mathcal{I}_s(c, \omega)$.

2.3 Four properties

I will use \mathcal{I}_s and s^X for now, but the following definitions also hold for inaccuracy measures and scoring rules introduced later. One method of constructing a global inaccuracy measure is by adding the inaccuracy scores of one’s individual credences at ω , which gives an **additive** inaccuracy measure (compare to Levinstein, 2019, p. 2924).⁵

Definition 2 (*Additivity*) Let (Ω, \mathcal{F}) , a credence function c , a world $\omega \in \Omega$, and Definition 1 be given. A global measure of inaccuracy $\mathcal{I}_s(c, \omega)$ is additive just in case $\mathcal{I}_s(c, \omega) = \sum_{X \in \mathcal{F}} s^X(v_\omega(X), c(X))$.

Truth-directedness captures the alethic idea of accuracy-first epistemology that as one’s credence $c(X)$ changes and gets closer to the ideal credence in X at ω , the inaccuracy score of $c(X)$ will get better, i.e. smaller (e.g., see Levinstein, 2017, p. 617). A measure of inaccuracy is truth-directed if it always assigns lower inaccuracy to one credence function than to another when each credence assigned by the first is closer to the ideal credence than the corresponding credence assigned by the second (see Pettigrew, 2016, p. 62); compare Definition 3 to Levinstein (2019, p. 2922), Joyce (2009, p. 269), or Pettigrew (2016, p. 40).

Definition 3 (*Truth-Directedness*) Let (Ω, \mathcal{F}) , credence functions c and \hat{c} , a world $\omega \in \Omega$, and a global measure of inaccuracy \mathcal{I}_s be given. If

- (i) for all $X \in \mathcal{F}$, $|v_\omega(X) - c(X)| \leq |v_\omega(X) - \hat{c}(X)|$, and
- (ii) for some $X \in \mathcal{F}$, $|v_\omega(X) - c(X)| < |v_\omega(X) - \hat{c}(X)|$,

then $\mathcal{I}_s(c, \omega) < \mathcal{I}_s(\hat{c}, \omega)$.

In Definition 3, condition (i) says that, at ω , c always assigns credences at least as close to the ideal credence $v_\omega(X)$ as \hat{c} . That is, $\hat{c}(X) \leq c(X) \leq 1$ for all X true at ω and $0 \leq c(X) \leq \hat{c}(X)$ for all X false at ω . Condition (ii) says that, at ω , c assigns a strictly higher credence than \hat{c} to at least one truth or a strictly lower credence than \hat{c} to at least one falsehood. That is, $\hat{c}(X) < c(X) \leq 1$ for some X true at ω or $0 \leq c(X) < \hat{c}(X)$ for some X false at ω . If conditions (i) and (ii) hold, then c gets a strictly lower inaccuracy score than \hat{c} at that ω according to any legitimate measure of inaccuracy \mathcal{I}_s .

⁵ In his paper, Levinstein sometimes calls $\mathcal{I}_s(c, \omega)$ a scoring rule (e.g., see Definition A.1 Levinstein, 2019, p. 2929) and talks about the additivity of scoring rules (e.g., see Theorem A.4 or Definition A.3 in Levinstein, 2019, p. 2929). In what follows, I will keep calling $\mathcal{I}_s(c, \omega)$ a global measure of inaccuracy and talk about the additivity of such measures. To predicate that scoring rules – as defined in Definition 1 – are additive is redundant since it only says that one can add the individual scores to give the overall score, which Definition 2 already expresses.

Differentiability is a technical requirement to which Levinstein pays little attention (see Levinstein, 2019, p. 2924). Differentiability means that if one differentiates $s^X(v_\omega(X), c(X))$ or $\mathcal{I}_s(c, \omega)$ with respect to $c(X)$ – where $X \in \mathcal{F}$ is in the domain of c – that derivative will exist; notice that X in $c(X)$ is fixed while differentiating. Differentiability implies **continuity** (e.g., see Pug, 2015, p. 149), so if $s^X(v_\omega(X), c(X))$ and $\mathcal{I}_s(c, \omega)$ are differentiable, then they are continuous functions of $c(X)$ and c , respectively, for all worlds ω . Roughly speaking, continuity means that there are no “jumps” in inaccuracy as credences change, so small changes in one’s credence will give rise to small changes in inaccuracy (see Pettigrew, 2016, p. 52). I mention this relation because accuracy-firsters often require continuity of scoring rules/inaccuracy measures (see Pettigrew, 2016, pp. 51–57 for a more detailed discussion).

Finally, *strict propriety* says that every probability function expects itself to be least inaccurate. For a scoring rule, it is defined as follows (see Pettigrew, 2016, p.66).

Definition 4 (*Strictly Proper Scoring Rule*) Let (Ω, \mathcal{F}) , $X \in \mathcal{F}$, a credence function c , and Definition 1 be given. A scoring rule s^X is strictly proper only if, for all $0 \leq p \leq 1$, it holds that:

$$p s^X(1, c(X)) + (1 - p) s^X(0, c(X)) \tag{1}$$

is uniquely minimised as a function of $c(X)$ at $c(X) = p$.

Strict propriety can be generalised for global inaccuracy measures if one has a notion of expected inaccuracy for a whole credence function c , which amounts to weighting the inaccuracy of c at each world ω by probabilistic weights (possibly interpreted as one’s credences that a given ω is the actual world). Let $E_P \mathcal{I}_s(c)$ stand for the expected inaccuracy of a credence function c with respect to an inaccuracy measure \mathcal{I}_s and weights given by a probability function P . Strict propriety of global inaccuracy measures means that every probability function P assigns itself the lowest expected inaccuracy. Definition 5 corresponds to Levinstein’s definition of propriety in Levinstein (2019, p. 2923); Levinstein considers $c \in \mathcal{P}$ because a stronger dominance condition holds for $c \notin \mathcal{P}$ (see Theorem 1 in Predd, 2009, p. 4788 or Pettigrew, 2016, p. 65), but the inequality in Definition 5 holds for any credence function $c \neq P$.

Definition 5 (*Strictly Proper Global Measure of Inaccuracy*) Let (Ω, \mathcal{F}) , \mathcal{P} , and an inaccuracy measure \mathcal{I}_s be given. \mathcal{I}_s is strictly proper just in case, for any distinct credence functions c and $P \in \mathcal{P}$, it holds that $E_P \mathcal{I}_s(P) < E_P \mathcal{I}_s(c)$.

Notice that, in Definition 5, the inequality $E_P \mathcal{I}_s(P) < E_P \mathcal{I}_s(c)$ compares only the overall expected inaccuracy scores of functions P and c . Beside propriety (what I call strict propriety), Levinstein defines **strong propriety** (see Definition A.1 in Levinstein, 2019, p. 2929). Strong propriety makes the overall comparison expressed by the inequality $E_P \mathcal{I}_s(P) < E_P \mathcal{I}_s(c)$ but, in addition, looks at the expected inaccuracy of single credences. An inaccuracy measure is strongly proper if it expects each

of the credences (not only the whole credence function) to be least inaccurate (Levinstein 2019, p. 2929). That is, if one's optimal credence in X is the probability of X , i.e., $P(X)$. To notationally differentiate the global (concerning the whole credence function) perspective from the local (concerning credence in a single proposition) perspective, let $E_P s^X(c(X)) = \sum_{\omega \in \Omega} P(\omega) s^X(v_\omega(X), c(X))$ be the expected inaccuracy of one's credence $c(X)$ in a single proposition X with respect to a probability function P and a scoring rule s^X .⁶

Definition 6 (Strong Propriety) Let (Ω, \mathcal{F}) , \mathcal{P} , a credence function c , a world $\omega \in \Omega$, and Definition 1 be given. An additive global inaccuracy measure $\mathcal{I}_s(c, \omega) = \sum_{X \in \mathcal{F}} s^X(v_\omega(X), c(X))$ is strongly proper just in case $E_P \mathcal{I}_s(c)$ is uniquely minimised at $P(X) = c(X)$ for all $X \in \mathcal{F}$ and all probability functions $P \in \mathcal{P}$. So, given s^X , it holds that $E_P s^X(P(X)) < E_P s^X(c(X))$ for all $X \in \mathcal{F}$, all probability functions $P \in \mathcal{P}$, and all credence functions c such that $P(X) \neq c(X)$.

2.4 The contingency of varying importance

The contingency of varying importance is based on two claims. First, varying importance says that the degree to which the inaccuracy of one's credence in a proposition matters can differ from one proposition to another. In other words, one can value the inaccuracy of one's credence in an important proposition more than in an unimportant one. For example, having low inaccuracy in propositions concerning fundamental laws of nature is better than having low inaccuracy in the claim that one wore wool socks on January 8th, 2004 (Levinstein (2019, p. 2925). Secondly, varying importance is contingent, according to Levinstein. That is, the degree to which the inaccuracy of one's credence in a proposition matters in one world might differ from the degree to which it matters in another world (Levinstein 2019, p. 2926).

Levinstein differentiates two ways in which varying importance is contingent. First, the level of importance differs at worlds where a given proposition is true from worlds where it is false (Levinstein 2019, p. 2926). Levinstein believes that accuracy-firsters can accommodate at least some of these cases; he refers to an approach from Merkle and Steyvers (2013) as a possible but complicated solution (see Levinstein, 2019, p. 2926). In this paper, I do not discuss these first-type cases since they are not the focus of Theorem A.4.

I will exclusively focus on the following second way of recognising the contingency of varying importance: the levels of importance differ at worlds where a given proposition has the same truth-value. Levinstein interprets these second-type cases as situations when the importance of one proposition depends on the truth-value of another proposition (Levinstein 2019, p. 2926). Consider one of his examples: "Bill is obsessed with popular music, but he's also terribly elitist. He wants to know everything about the lives of the singers who are actually the most talented musicians. As it turns out, in one world, Beyoncé meets the cut, but in a very distant one,

⁶ $E_P s^X(c(X))$ corresponds to Levinstein's expected inaccuracy of one's credence x in X with respect to a probability function, e.g., see $f_{P_r(x)}$ in Levinstein (2019, p. 2929). Levinstein drops the superscript X in calculating the expected inaccuracy of one's credence in a single proposition.

she doesn't. In only some worlds, Bill places high importance on knowing where Beyoncé was born, the name of her high school, and the sales figures of her first album" (Levinstein 2019, p. 2927). For example, the inaccuracy of Bill's credence about where Beyoncé was born is highly important in the worlds where Beyoncé makes the cut and has low importance in the worlds where she does not make it. So, how much the inaccuracy of that credence matters changes from one possible world to another depending on the truth-value of another proposition, i.e., whether Beyoncé makes the cut. To formally capture these levels of importance, let me introduce a weight function $\lambda(X, \omega)$ that expresses how epistemically valuable it is not to be inaccurate (i.e., to be accurate) about X at ω and restrict its range to $\mathbb{R}_{>0}$.⁷

Definition 7 (*Weight Function for Varying Importance*) Given a space (Ω, \mathcal{F}) , let $\lambda(X, \omega) := \mathcal{F} \times \Omega \rightarrow (0, \infty)$ be a weight function for some $X \in \mathcal{F}$ and $\omega \in \Omega$.

The value of $\lambda(X, \omega)$ from Definition 7 will be high in worlds where it is important to be accurate (i.e., not to be inaccurate) about X and low in worlds where it is not that important. The range of a weight function may differ. For example, in Holzmann & Klar (2017, pp. 2409–2410), weights are restricted to $[0, 1]$. But one can be less restrictive (e.g., see Theorem 1 in Ranjan & Gneiting, 2011, p. 413). I will assume that $\lambda(X, \omega)$ takes values strictly between 0 and ∞ . So, one always cares about the inaccuracy of one's credence in X at any $\omega \in \Omega$ at least a little bit, and the inaccuracy of one's credence in no proposition is infinitely important at any $\omega \in \Omega$. One can define a weighted scoring rule by joining a weight function with a scoring rule from Definition 1, e.g., compare to Ranjan & Gneiting (2011, p. 413) or Pelenis (2014, p. 9) (but I assume a strictly proper scoring rule s^X).

Definition 8 (*Weighted Scoring Rule*) Let (Ω, \mathcal{F}) , $X \in \mathcal{F}$, and s^X satisfying Definition 1 be given. A weighted scoring rule is a function $s_\lambda^X : \{0, 1\} \times [0, 1] \times (0, \infty) \rightarrow [0, \infty]$ such that $s_\lambda^X(\cdot, \cdot, \lambda)$ is a scoring rule for each weight $\lambda \in (0, \infty)$.

Definition 8 is a general definition, but, for the purpose of this paper, I will restrict Definition 8 to the following interpretation. Let $\{0, 1\}$ be ideal credences, the values from $[0, 1]$ one's credences, and weights λ are values of function $\lambda(X, \omega)$ from Definition 7. Also, I will assume that if $\lambda(X, \omega) > \lambda'(X, \omega)$ and one's credence $c(X)$ is in $(0, 1)$, then $s_\lambda^X(v_\omega(X), c(X), \lambda(X, \omega)) > s_{\lambda'}^X(v_\omega(X), c(X), \lambda'(X, \omega))$. In this paper, weights will scale the score by multiplying it, i.e., $s_\lambda^X(v_\omega(X), c(X), \lambda(X, \omega)) = \lambda(X, \omega)s^X(v_\omega(X), c(X))$. One can then define what it means to recognise the contingency of varying importance (compare Definition 9 to Definition A.3 in Levinstein, 2019, p. 2929). I will write $\mathcal{I}_{s_\lambda^X}$ to indicate that one uses weighted scoring rules s_λ^X in a global inaccuracy measure.

⁷ Levinstein did not formally define $\lambda(X, \omega)$ or restrict its range. But, in Definition 7, I try to follow Levinstein's idea (see Levinstein, 2019, p. 2927) and his use of $\lambda(X, \omega)$. For example, he seems to assume that constant weights take positive real values (see Levinstein, 2019, p. 2925), so he considers the possibility of restricting the range to $\mathbb{R}_{>0}$.

Definition 9 (*Recognising Contingent Importance*) Assume that (Ω, \mathcal{F}) , a credence function c , and a weight function $\lambda(X, \omega)$ from Definition 7 is given. Let then $\mathcal{I}_{s_\lambda}(c, \omega) = \sum_{X \in \mathcal{F}} \lambda(X, \omega) s^X(v_\omega(X), c(X))$ be an additive inaccuracy measure. $\mathcal{I}_{s_\lambda}(c, \omega)$ recognises contingent importance just in case there exist worlds ω_1 and ω_2 from Ω and a proposition X from \mathcal{F} such that $v_{\omega_1}(X) = v_{\omega_2}(X)$ but $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$.

For example, let X stand for the proposition that Beyoncé was born in Houston, Texas. Let $c(X)$ be Bill's credence in X and $\omega_1, \omega_2 \in \Omega$ be two possible worlds. If X is true at both ω_1 and ω_2 , i.e., $v_{\omega_1}(X) = v_{\omega_2}(X)$, then $s^X(v_{\omega_1}(X), c(X)) = s^X(v_{\omega_2}(X), c(X))$. But suppose that Beyoncé makes the cut only at ω_1 and not at ω_2 . It means that, to Bill, the inaccuracy of his credence in X matters more at ω_1 than at ω_2 , that is, $\lambda(X, \omega_1) > \lambda(X, \omega_2)$. Consequently, $\lambda(X, \omega_1) s^X(v_{\omega_1}(X), c(X)) > \lambda(X, \omega_2) s^X(v_{\omega_2}(X), c(X))$.

3 The proof and idea behind Theorem A.4

It will be useful to briefly discuss Levinstein's proof of Theorem A.4 to understand the idea behind it. Following Levinstein (see Levinstein, 2019, p. 2929), let me formulate a *reductio* assumption and then use it to prove that \mathcal{I}_{s_λ} is not a strictly proper inaccuracy measure.⁸

Assumption 1 (*Reductio Assumption for \mathcal{I}_{s_λ}*) Suppose \mathcal{I}_{s_λ} is additive, strictly proper, truth-directed, differentiable, and recognises contingent importance.

First let me note that Levinstein makes two assumptions about s^X in s_λ^X that are important for understanding his proof of Theorem A.4. First, s^X is differentiable, see Levinstein, 2019, p. 2930). Secondly, s^X is strictly proper (for example, he considers the strictly proper Brier score, see Levinstein, 2019, p. 2927).⁹ So, let me make the same assumptions about s^X in s_λ^X .

Assumption 2 (*Assumptions about s^X in s_λ^X*) Given Definition 8, a scoring rule $s^X(v_\omega(X), c(X))$ in a weighted scoring rule $s_\lambda^X(v_\omega(X), c(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(v_\omega(X), c(X))$ is strictly proper and differentiable with respect to its second argument, $c(X)$.

⁸ Levinstein formulates his *reductio* assumption in terms of a scoring rule \mathcal{I} and shows that \mathcal{I} is not proper (see Levinstein, 2019, pp. 2929–2930). Following my notational standards and terminology in Assumption 1 does not affect the final result and its message.

⁹ If one wants further confirmation of this second assumption, see Subsection 3.1 in Levinstein (2019, p. 2925), where Levinstein claims that accuracy-first epistemology can accommodate a situation when the weights of importance for each proposition X are constant, i.e., $\lambda(X, \omega)$ is a constant function with respect to X at any ω (see also the first paragraph of Levinstein, 2019, p. 2926). This means that if X is important in one world just as it is important in every other world, there is no clash between what I call strict propriety and the contingency of varying importance. One can check that there is no such clash if and only if scoring rules for individual credences are strictly proper.

By Lemma A.1 in Schervish et al. (1989, p. 1874) (here stated as Lemma 1), strictly proper scoring rules are truth-directed. By Assumption 2, s^X is truth-directed, i.e., $s^X(1, c(X))$ is strictly decreasing in $c(X)$ and $s^X(0, c(X))$ is strictly increasing in $c(X)$.

Lemma 1 (Schervish) *Let (g_0, g_1) be a (strictly) proper scoring rule, possibly attaining infinite values on the closed interval $[0, 1]$. Then, $g_1(x)$ is (strictly) decreasing in x and $g_0(x)$ is (strictly) increasing in x .*

Let me now discuss Levinstein’s proof of Theorem A.4. Following Levinstein (see Levinstein, 2019, p. 2930 for the same step), let me fix a probability function Pr defined on some (Ω, \mathcal{F}) such that $0 < Pr(X) < 1$ for $X \in \mathcal{F}$ and use Pr to define another probability function, Pr' .

Definition 10 Given (Ω, \mathcal{F}) , $\omega_1, \omega_2 \in \Omega$, and $Pr \in \mathcal{P}$, define Pr' as follows:

$$Pr'(\omega) = \begin{cases} Pr(\omega_1) + \epsilon & \omega = \omega_1 \\ Pr(\omega_2) - \epsilon & \omega = \omega_2 \\ Pr(\omega) & \text{otherwise,} \end{cases}$$

where $\epsilon < \min\{Pr(\omega_1), Pr(\omega_2), 1 - Pr(\omega_1), 1 - Pr(\omega_2)\}$ to guarantee that $Pr' \in \mathcal{P}$.

Notice that if X is true (or false) at both ω_1 and ω_2 , then $Pr(X) = Pr'(X)$ (and $Pr'(\neg X) = Pr(\neg X)$) since for a finite number n of $\omega \in \Omega$, one has that:

$$\begin{aligned} Pr'(X) &= \sum_{\omega \in X} Pr'(\omega) = Pr'(\omega_1) + Pr'(\omega_2) + \dots + Pr'(\omega_n) \\ &= [Pr(\omega_1) + \epsilon] + [Pr(\omega_2) - \epsilon] + \dots + Pr(\omega_n) = \sum_{\omega \in X} Pr(\omega) = Pr(X). \end{aligned}$$

If, following Definition 9, weighting strictly proper s^X by non-constant weights $\lambda(X, \omega)$ preserves strict propriety, then the expected inaccuracy of $c(X)$ with respect to s^X_λ and Pr' , i.e., $E_{Pr', s^X_\lambda}(c(X))$, is minimised at $c(X) = Pr'(X)$. Similarly, $E_{Pr, s^X_\lambda}(c(X))$ is then minimised at $c(X) = Pr(X)$. So, both expectations are minimised at the same point since $Pr(X) = Pr'(X)$. But Levinstein’s Theorem A.4 (see Theorem 1 below) shows that $E_{Pr', s^X_\lambda}(c(X))$ and $E_{Pr, s^X_\lambda}(c(X))$ are not minimised at the same point if one assumes the contingency of varying importance. Thus, $\mathcal{I}_{s^X_\lambda}$ is not strictly proper, which contradicts Assumption 1.

Theorem 1 (Levinstein) *Let $\mathcal{I}_{s^X_\lambda}$ from Assumption 1 and Assumption 2 be given. If the contingency of varying importance holds, then $E_{Pr', s^X_\lambda}(c(X))$ and $E_{Pr, s^X_\lambda}(c(X))$ are not minimised at the same point. Thus, s^X_λ and $\mathcal{I}_{s^X_\lambda}$ is not strictly proper.*

4 Constructing s_*^X and \mathcal{I}_{s_*}

4.1 Minimisers for weighted scoring rules

By Theorem 1, $E_{p_r} s_\lambda^X(c(X))$ and $E_{p_{r'}} s_\lambda^X(c(X))$ have different minimisers if the contingency of varying importance holds. It will be useful for the construction of s_*^X to know what those minimisers are. But, instead of finding minimisers for $E_{p_r} s_\lambda^X(c(X))$ and $E_{p_{r'}} s_\lambda^X(c(X))$ separately, let me find one for a general expectation $E_p s_\lambda^X(c(X))$, so I can overlook the use of any specific weight function (Definition 7 still holds) or a probability function. Theorem 2 follows from restricting Gneiting and Ranjan's Theorem 1 in Ranjan & Gneiting (2011, pp. 413–414) to finite cases. The proof strategy is the same for both the continuous and discrete cases. Take weighted scores and find the normalising constant (I call it γ). Since γ is a constant, it can be taken out or placed inside the sums and integrals with no issue and used to find the result.

Theorem 2 (Gneiting and Ranjan) *Let (Ω, \mathcal{F}) , $X \in \mathcal{F}$, a credence function c , $P \in \mathcal{P}$, and Assumption 2 be given. $E_p s_\lambda^X(c(X))$ is uniquely minimised at*

$$c^*(X) = \frac{\sum_{\omega \in X} P(\omega) \lambda(X, \omega)}{\sum_{\omega \in \Omega} P(\omega) \lambda(X, \omega)}.$$

The strict propriety of s^X in s_λ^X (and of s^{-X} in s_λ^{-X}) means that $c^*(X)$ (and $c^*(\neg X)$) are unique minimisers, where, by Theorem 2, $c^*(\neg X) = \frac{\sum_{\omega \in \neg X} P(\omega) \lambda(\neg X, \omega)}{\sum_{\omega \in \Omega} P(\omega) \lambda(\neg X, \omega)}$. In what follows, the existence and the uniqueness of minimisers $c^*(X)$ and $c^*(\neg X)$ from Theorem 2 is an important corollary of Assumption 2 and Definition 7. But, in general, one must be careful because uniqueness might not hold, or a minimiser might not exist, for example, when $\lambda(X, \omega)$ is always 0; see (Brehmer & Gneiting, 2020) for further discussion about the uniqueness and existence of minimisers. Since $c^*(X)$ and $c^*(\neg X)$ represent one's optimal credences, one might want to know under what conditions they are probabilistic. The numeric bound and the probability of the entire space and the empty set follow directly.

Lemma 2 *It holds that $0 \leq c^*(X) \leq 1$ (and $0 \leq c^*(\neg X) \leq 1$) and if $X = \Omega$, then $c^*(X) = 1$ (and $c^*(\neg X) = 0$).*

Additivity of $c^*(X)$ and $c^*(\neg X)$ does not come so easily and requires, for example, the following additional assumption.

Assumption 3 Given (Ω, \mathcal{F}) , $\lambda(X, \omega_i) = \lambda(\neg X, \omega_i)$ for any $X \in \mathcal{F}$ and $\omega_i \in \Omega$.

Assumption 3 says that, at any $\omega \in \Omega$, an agent values the inaccuracy of her credence in any $X \in \mathcal{F}$ to the same degree as the inaccuracy of her credence in $\neg X$. For example, assume that X says that it is raining. According to Assumption 3, if it is raining at ω , one values the closeness of her credence in X to the ideal credence of 1 to the same degree as the closeness of her credence in $\neg X$ to the ideal credence of 0. Notice that Assumption 3 considers weights of importance for two propositions (X and $\neg X$) and requires that those weights are equal at the same ω , which I indicated

by the subscript i . It does not require the weights of importance for X and $\neg X$ to be the same across different possible worlds. So, Assumption 3 does not conflict with the contingency of varying importance from Definition 9, which considers weights for a single proposition and requires that those weights are different at different possible worlds.

Lemma 3 *Given Assumption 3, then $c^*(X) + c^*(\neg X) = 1$.*

4.2 Scoring rule s_*^X

I will now use the minimiser $c^*(X)$ from Theorem 2 to define a function s_*^X and prove that s_*^X is a strictly proper, truth-directed, and differentiable weighted scoring rule.¹⁰

Definition 11 Let (Ω, \mathcal{F}) , $X \in \mathcal{F}$, $P \in \mathcal{P}$, s^X satisfying Definition 1, $\lambda(X, \omega)$ satisfying Definition 7, and a credence function c be given. Let $c^*(X)$ be a unique minimiser with respect to X, P , and a weighted scoring rule s_λ^X satisfying Definition 8. A function s_*^X for one's credence $c(X)$ is then defined as follows:

$$s_*^X(0, c(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(0, c^*(X) + \epsilon k(\epsilon))$$

and

$$s_*^X(1, c(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(1, c^*(X) + \epsilon k(\epsilon)),$$

where $\epsilon = c(X) - P(X)$ and

$$k(\epsilon) = \begin{cases} \frac{1-c^*(X)}{1-P(X)} & \text{if } \epsilon > 0, \\ \frac{c^*(X)}{P(X)} & \text{if } \epsilon < 0, \\ 1 & \text{if } \epsilon = 0. \end{cases}$$

By Definition 11, the score assigned by s_*^X to $c(X)$ is the score that s^X assigns to a value of $c^*(X) + \epsilon k(\epsilon)$ that is then weighted by $\lambda(X, \omega)$. Notice that one can rewrite $c^*(X) + \epsilon k(\epsilon)$ as $k(\epsilon)c(X) + [c^*(X) - P(X)k(\epsilon)]$, which is a linear function of $c(X)$. So, formally, $(0, c^*(X) + \epsilon k(\epsilon))$ and $(1, c^*(X) + \epsilon k(\epsilon))$ are results of an affine transformation of $(v_\omega(X), c(X))$, see Proposition 1 for details. The product $\epsilon k(\epsilon)$ determines the degree of punishment one receives for not aligning one's credence with a probability function P . That is, the more $c(X)$ diverges from $P(X)$, the bigger inaccuracy score s_*^X assigns to $c(X)$. So, s_*^X is always defined with respect to some $P \in \mathcal{P}$.

¹⁰ Function s_*^X in Definition 11 is not a random choice. It is a modification of a scoring rule defined by Gneiting and Brehmer in their Theorem 1 (see Brehmer & Gneiting, 2020, p. 660). Their general approach tells us how to properise scoring rules (including the weighted scoring rules), so truth-telling becomes an optimal strategy (Brehmer and Gneiting (2020, p. 660). One needs to modify Gneiting and Brehmer's properisation strategy to be applicable to accuracy-first epistemology, which is my goal with Definition 11. My modification is not a unique option, but it works.

Note that function $k(\epsilon)$ in Definition 11 (and thus s_*^X) is not well-defined in the following two cases: (1). $k(\epsilon) = \frac{c^*(X)}{P(X)}$ and $P(X) = 0$ or (2). $k(\epsilon) = \frac{1-c^*(X)}{1-P(X)}$ and $P(X) = 1$. Lemma 4 shows that these two cases cannot happen, so s_*^X is always well-defined on its respective domain.

Lemma 4 *If $P(X) = 0$, then $k(\epsilon) = 1 - c^*(X)$ or 1 and if $P(X) = 1$, then $k(\epsilon) = c^*(X)$ or 1.*

By Definition 11, s^X in s_*^X satisfies Definition 1, i.e., $s^X = \{0, 1\} \times [0, 1] \rightarrow [0, \infty]$. Clearly, the affine transformation maps the vindicated credence 0 to 0 and 1 to 1, so the set $\{0, 1\}$ is preserved. In other words, Definition 11 changes nothing about the assumption that 0 and 1 are the vindicated credences in false and true propositions, respectively. But I need to check that the values of $c^*(X) + \epsilon k(\epsilon)$ always come from the unit interval.

Lemma 5 *Any value of $c^*(X) + \epsilon k(\epsilon)$ comes from $[0, 1]$.*

If one pairs $s^X(v_\omega(X), c^*(X) + \epsilon k(\epsilon))$ with a weight function $\lambda(X, \omega)$ from Definition 7, one forms a weighted scoring rule s_*^X that is strictly proper, truth-directed, and differentiable with respect to $c(X)$ such that the fully accurate credences get the zero inaccuracy score.

Proposition 1 *s_*^X is a strictly proper, truth-directed, continuous, and differentiable (with respect to $c(X)$) weighted scoring rule such that $s_*^X(0, 0) = s_*^X(1, 1) = 0$.*

There is an important difference between s_*^X and s_λ^X from Theorem A.4 (i.e., s_λ^X paired with Assumption 2). By Assumption 2, s^X in s_λ^X from Theorem A.4 is strictly proper and differentiable, but, by Theorem 1, s_λ^X is not strictly proper. In contrast, by Proposition 1, s_*^X is strictly proper but, by the following Lemma 6, s^X in s_*^X is not strictly proper. In other words, from Assumption 2, I only keep the technical requirement that s^X is differentiable.

Lemma 6 *s^X in s_*^X is not strictly proper.*

In Lemma 5 and Proposition 1, I showed that the values of $c^*(X) + \epsilon k(\epsilon)$ always come from the unit interval such that if $c(X) = 1$, then $c^*(X) + \epsilon k(\epsilon) = 1$ and if $c(X) = 0$, then $c^*(X) + \epsilon k(\epsilon) = 0$. Given these results, one might wonder about additivity. Assuming that $c(\neg X) - P(\neg X) = \epsilon'$, a sufficient and necessary condition for the additivity of $c^*(X) + \epsilon k(\epsilon)$ and $c^*(\neg X) + \epsilon' k(\epsilon')$ is that one's credence function is additive, i.e., $c(X) + c(\neg X) = 1$; see point 4 of Lemma 7.

Lemma 7 *Given Assumption 3, $P \in \mathcal{P}$, and $c(\neg X) - P(\neg X) = \epsilon'$, then*

1. $\epsilon = -\epsilon'$ iff $c(X) + c(\neg X) = 1$,
2. if $c^*(X) = 0$, then $P(X) = 0$ and if $c^*(\neg X) = 0$, then $P(\neg X) = 0$,

3. if $c^*(X) = 1$, then $P(X) = 1$ and if $c^*(\neg X) = 1$, then $P(\neg X) = 1$, and
4. $[c^*(X) + \epsilon k(\epsilon)] + [c^*(\neg X) + \epsilon' k(\epsilon')] = 1$ iff $c(X) + c(\neg X) = 1$.

Finally, assume that a weighted scoring rule s_λ^X (e.g., the one from Theorem 1) is strictly proper, i.e., let the value of $\lambda(X, \omega)$ be constant for the given X at any ω . Then, s_λ^X assigns the same inaccuracy score to one's credence $c(X)$ as that strictly proper s_λ^X (see Brehmer & Gneiting, 2020, p. 660 for a more general discussion).

Lemma 8 (Gneiting and Brehmer) *If s_λ^X is strictly proper, then s_*^X and s_λ^X assign the same score to one's credence $c(X)$.*

4.3 Global inaccuracy measure \mathcal{I}_{s_*}

I will now use s_*^X from Definition 11 to construct a global inaccuracy measure \mathcal{I}_{s_*} and show that \mathcal{I}_{s_*} has all the properties listed in Theorem A.4, without reaching a contradiction.

Proposition 2 \mathcal{I}_{s_*} is additive, proper (strictly and strongly), truth-directed, differentiable, and recognises the contingency of varying importance.

Let me start with **additivity** which follows directly from the assumption that one can add weighted scores (e.g., see Levinstein, 2019, p. 2929) and the fact that s_*^X is a weighted scoring rule. **Differentiability** also follows easily. By Proposition 1, s_*^X is differentiable with respect to $c(X)$ and summing differentiable functions preserves differentiability, so an additive \mathcal{I}_{s_*} is differentiable. Since differentiability implies continuity, $\mathcal{I}_{s_*}(c, \omega)$ is a continuous function of c for all worlds ω . Let me now move to **strict/strong propriety** and **truth-directedness**.

Lemma 9 (Strict and Strong Propriety) \mathcal{I}_{s_*} is strongly and strictly proper, thus truth-directed.

Finally, consider the **contingency of varying importance**. I will show that, $E_{P_r} s_*^X(c(X))$ and $E_{P_r'} s_*^X(c(X))$ have a common minimiser while recognising the contingency of varying importance from Definition 9.

Lemma 10 (Contingency of Varying Importance) *Given \mathcal{I}_{s_*} , $E_{P_r} s_*^X(c(X))$ and $E_{P_r'} s_*^X(c(X))$ are uniquely minimised at the same point.*

So, the contingency of varying importance does not clash with strict propriety (or any other property listed in Proposition 2). Therefore, Proposition 2 holds without contradiction.

5 Example and discussion

Let me show how some of the abstract results work on a concrete example. I will consider the expected total inaccuracy discussed by Kierland and Monton in Kierland and Monton (2005). Kierland and Monton introduced expected total inaccuracy as a possible expected-inaccuracy-minimising approach to solving the Sleeping Beauty problem. The Sleeping Beauty problem has many variations, but following the major part of Kierland and Monton (2005), I will consider its basic version that goes as follows¹¹: “On Sunday Sleeping Beauty is put to sleep, and she knows that on Monday researchers will wake her up, and then put her to sleep with a memory-erasing drug that causes her to forget that waking-up. She also knows that the researchers will then flip a fair coin; if the result is Heads, they will allow her to continue to sleep, and if the result is Tails, they will wake her up again on Tuesday. Thus, when she is awakened, she will not know whether it is Monday or Tuesday. On Sunday, she assigns probability $\frac{1}{2}$ to the proposition H that the coin lands Heads. What probability should she assign to H on Monday, when she wakes up?” (Kierland & Monton (2005, p. 389). Note that there are three moments at which Beauty can be awake: 1.) Monday and Heads, 2.) Monday and Tails, and 3.) Tuesday and Tails. In other words, there is one awakening when Heads (i.e., H is true) and two awakenings when Tails (i.e., H is false).

To find the expected total inaccuracy $S_{ET}(H)$ of Beauty’s credence $c(H)$ in H , first, find the inaccuracy score of $c(H)$ for each awakening. Kierland and Monton use the Brier score (see Kierland & Monton, 2005, p. 385) to find the inaccuracy score of $c(H)$ for each awakening, and I will do the same. Then, weight those scores by the probability of reaching the given awakening and sum those weighted scores, which gives the following formula to minimise (see Kierland & Monton, 2005, p. 389):

$$\begin{aligned} S_{ET}(H) &= \frac{1}{2}(1 - c(H))^2 + \frac{1}{2}[(0 - c(H))^2 + (0 - c(H))^2] \\ &= \frac{1}{2}(1 - c(H))^2 + 2\frac{1}{2}(0 - c(H))^2, \end{aligned} \quad (2)$$

where $\frac{1}{2}$ is the probability of Heads/Tails since the coin is fair by assumption. Note that the number of awakenings where H is true/false serves as a weight. Since there is one awakening where H is true, $(1 - c(H))^2$ is weighted by 1, and $(0 - c(H))^2$ is weighted by 2 since H is false for two awakenings. Since the Brier score is strictly proper and $S_{ET}(H)$ uses unequal positive weights, by Levinstein’s Theorem A.4, the weighted Brier score used in $S_{ET}(H)$ is not strictly proper. One can confirm this by showing that $S_{ET}(H)$ is minimised at $c(H) = \frac{1}{3}$ instead of $c(H) = \frac{1}{2}$ (see Kierland & Monton, 2005, p. 389). But if the weights were equal/constant, then $S_{ET}(H)$ would be minimised at $c(H) = \frac{1}{2}$ (see footnote 9 or Ranjan & Gneiting, 2011, p. 413 for details).

¹¹ So, for example, I assume that only one round is played (i.e., a fair coin is flipped only once), there are no Beauty’s perfect duplicates, etc. See Elga (2000) for the original formulation of the problem.

I will now consider $S_{ET}(H)$ but use it with s^X from Definition 11. Let X be H and $S_{ET}^*(H)$ be $S_{ET}(H)$ using s^H . I assume that s^H in s^H works as the Brier score. So, for example, $s^H(1, c^*(H) + \epsilon k(\epsilon)) = (1 - [c^*(H) + \epsilon k(\epsilon)])^2$. If ω_H is a world where H is true and $\omega_{\neg H}$ is a world where H is false, then weights are $\lambda(H, \omega_H) = 1$ and $\lambda(H, \omega_{\neg H}) = 2$ (i.e., one counts centres within the given uncentred world). We know that $S_{ET}(H)$ is minimised at $c(H) = \frac{1}{3}$, so $c^*(H) = \frac{1}{3}$. For the sake of notational simplicity, let Beauty's credence in Heads $c(H) = x$, so $\epsilon = x - P(H)$. By Definition 11, if $P(H) = \frac{1}{2}$ (i.e., the coin is fair), then the values of $k(\epsilon)$ in our example are as follows:

$$k(\epsilon) = \begin{cases} \frac{1-c^*(X)}{1-P(X)} = \frac{1-\frac{1}{3}}{1-\frac{1}{2}} = \frac{4}{3} & \text{if } \epsilon > 0, \\ \frac{c^*(X)}{P(X)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3} & \text{if } \epsilon < 0, \\ 1 & \text{if } \epsilon = 0. \end{cases}$$

If s^H is strictly proper, then $S_{ET}^*(H)$ will be minimised at $c(H) = \frac{1}{2}$ for a fair coin. So, my goal is to show that for each of the three values of $k(\epsilon)$, $S_{ET}^*(H)$ is minimised at $c(H) = \frac{1}{2}$. I will go case by case, but to avoid repetition, let me make a general observation for $x = c(H)$ and $\epsilon = x - P(H) = x - \frac{1}{2}$:

$$\begin{aligned} S_{ET}^*(H) &= P(H)s^H(1, x, \lambda(H, \omega_H)) + (1 - P(H))s^H(0, x, \lambda(H, \omega_{\neg H})) \\ &= \frac{1}{2}\lambda(H, \omega_H)s^H(1, c^*(H) + \epsilon k(\epsilon)) + \frac{1}{2}\lambda(H, \omega_{\neg H})s^H(0, c^*(H) + \epsilon k(\epsilon)) \\ &= \frac{1}{2}s^H\left(1, \frac{1}{3} + \left(x - \frac{1}{2}\right)k(\epsilon)\right) + s^H\left(0, \frac{1}{3} + \left(x - \frac{1}{2}\right)k(\epsilon)\right). \end{aligned}$$

I will now crunch the numbers for the three cases (i.e., $\epsilon > 0$, $\epsilon < 0$, and $\epsilon = 0$). For $\epsilon > 0$, one has:

$$\begin{aligned} S_{ET}^*(H) &= \frac{1}{2}s^H\left(1, \frac{1}{3} + \left(x - \frac{1}{2}\right)\frac{4}{3}\right) + s^H\left(0, \frac{1}{3} + \left(x - \frac{1}{2}\right)\frac{4}{3}\right) \\ &= \frac{1}{2}\left(1 - \left(\frac{4}{3}x - \frac{1}{3}\right)\right)^2 + \left(0 - \left(\frac{4}{3}x - \frac{1}{3}\right)\right)^2 = \frac{8x^2 - 8x + 3}{3}. \end{aligned}$$

One can now easily verify that $S_{ET}^*(H)$ is minimised at $c(H) = x = \frac{1}{2}$. For $\epsilon < 0$, one has:

$$\begin{aligned} S_{ET}^*(H) &= \frac{1}{2}s^H\left(1, \frac{1}{3} + \left(x - \frac{1}{2}\right)\frac{2}{3}\right) + s^H\left(0, \frac{1}{3} + \left(x - \frac{1}{2}\right)\frac{2}{3}\right) \\ &= \frac{1}{2}\left(1 - \frac{2}{3}x\right)^2 + \left(0 - \frac{2}{3}x\right)^2 = \frac{4x^2 - 4x + 3}{6}. \end{aligned}$$

One can easily verify that $S_{ET}^*(H)$ is again minimised at $c(H) = x = \frac{1}{2}$. Finally, for $\epsilon = 0$, one has:

$$S_{ET}^*(H) = \frac{1}{2}s^H\left(1, \frac{1}{3}\right) + s^H\left(0, \frac{1}{3}\right) = \frac{1}{2}\left(1 - \frac{1}{3}\right)^2 + \left(0 - \frac{1}{3}\right)^2.$$

For a variable $z \in \mathbb{R}$, we know that $\frac{1}{2}(1 - z)^2 + (0 - z)^2$ is minimised at $z = \frac{1}{3}$ (one can also check Eq. (2)), so $\frac{1}{2}(1 - \frac{1}{3})^2 + (0 - \frac{1}{3})^2$ gives the minimum. This is achieved for $\epsilon = 0$, i.e., $\frac{1}{2} = P(H) = c(H) = x$, as required. I have verified that $S_{ET}^*(H)$ is always minimised at $c(H) = \frac{1}{2}$, as it should be if one uses a strictly proper scoring rule and a fair coin.

Note that I used Levinstein’s Theorem A.4 to show that the weighted Brier score in $S_{ET}(H)$ is not strictly proper and Definition 11 to show that s_*^H is strictly proper without reaching any contradiction along the way. $S_{ET}(H)$ uses the Brier score, a strictly proper scoring rule, and unequal positive weights. In this case, Levinstein’s Theorem A.4 applies and says that the weighted Brier score is not strictly proper. My results about Definition 11 say nothing about a situation in which one weights a strictly proper scoring rule (well, Theorem 2 agrees with Theorem A.4). My result says that one can use unequal positive weights with a scoring rule such that the combination is strictly proper. But it is not done by weighting a strictly proper scoring rule. It is done by weighting a scoring rule that is not strictly proper. Specifically, note that strictly proper s_*^H is a combination of weights and a scoring rule s^H . But s^H is not strictly proper. To see it, consider a situation with $\lambda(X, \omega_H) = \lambda(H, \omega_{-H}) > 0$, so weights are positive and equal:

$$\begin{aligned}
 S_{ET}^*(H) &= \frac{1}{2} s_*^H(1, x, \lambda(H, \omega_H)) + \frac{1}{2} s_*^H(0, x, \lambda(H, \omega_{-H})) \\
 &= \frac{1}{2} \lambda(H, \omega_H) \left[\left(1 - \left[\frac{1}{3} + \left(x - \frac{1}{2} \right) k(\epsilon) \right] \right)^2 + \left(\frac{1}{3} + \left(x - \frac{1}{2} \right) k(\epsilon) \right)^2 \right].
 \end{aligned}
 \tag{3}$$

If s^H is strictly proper, then equal positive weights will not interfere with its strict propriety (see footnote 9 or Ranjan & Gneiting, 2011, p. 413 for details). Differentiating Eq. (3) with respect to x and setting it equal to 0 gives:

$$\frac{\partial S_{ET}^*(H)}{\partial x} = \frac{1}{2} \lambda(H, \omega_H) \left(4k(\epsilon)^2 x - 2k(\epsilon)^2 - \frac{2k(\epsilon)}{3} \right) = 0.$$

Separating x gives (note that $k(\epsilon) \neq 0$, so the following equation is well-defined):

$$c(H) = x = \frac{3k(\epsilon) + 1}{6k(\epsilon)} = \frac{\frac{1}{2} - \frac{1}{3}}{k(\epsilon)} + \frac{1}{2} = \frac{P(H) - c^*(H)}{k(\epsilon)} + P(H),
 \tag{4}$$

which corresponds to the formula $c(X) = \frac{p - c^*(X)}{k(\epsilon)} + p$ from Lemma 6, where X is H and $p = P(H)$. By plugging values $k(\epsilon) = \frac{4}{3}$, $k(\epsilon) = \frac{2}{3}$, and $k(\epsilon) = 1$ into Eq. (4), one gets that $c(H) = \frac{5}{8}$, $c(H) = \frac{3}{4}$, and $c(H) = \frac{2}{3}$, respectively. But we know, by assumption, that $P(H) = \frac{1}{2}$. So, there is no value of $k(\epsilon)$ for which $S_{ET}^*(H)$ with equal weights is minimised at $c(H) = P(H)$, which would be the case if s_{ET}^* was strictly proper. Hopefully, this clarifies the difference between Levinstein’s result and my approach and why they coexist without contradicting each other.

Lastly, one of the reviewers expressed a worry about a case when P is not the agent’s credence function but is still an argument of s_*^X . This is potentially worrying, the reviewer says, because in many applications of scoring rules, we take the distribution P with respect to which expected inaccuracy is calculated to be the “true” distribution. One reason to require strict propriety is so that the expected value of a scoring rule measuring the inaccuracy of an agent’s credence function is minimised when that credence function matches the true data-generating distribution. In such a context, when we define a scoring rule such that the distribution with respect to which its expected value is calculated is also an argument of that scoring rule, we effectively assume that we have access to the true data-generating distribution when evaluating accuracy. But this is not always the case, such that our ability to actually estimate the expected value of such a scoring rule (i.e., s_*^X) may be significantly limited.

Given the discussed example, my understanding is that the reviewer is worried about what happens if one does not know the coin’s bias (i.e., $P(H)$, which one can see as a true data-generating distribution) since it is needed for constructing s_*^H . First, let me say that, in the context of accuracy-first epistemology, P is often interpreted as one’s probabilistic credence function (e.g., see Pettigrew, 2016, p. 24 or Pettigrew, 2016, pp.189–190). But more is needed to answer the point fully. I would argue that what limits our ability to estimate the expected value is that one formulates expectation with respect to an unknown probability function P rather than P being an argument of s_*^X . For example, assume that, in our example, $P(H)$ is the chance of the coin landing Heads, and one does not know the value of $P(H)$. Then, $S_{ET}(H) = P(H)(1 - c(H))^2 + 2(1 - P(H))(0 - c(H))^2$ is minimised at $c(H) = \frac{P(H)}{2-P(H)}$.¹² So, one is limited in estimating the expected value independently of s_*^H (since $S_{ET}(H)$ does not use it). But note that, for equal weights, $S_{ET}(H)$ is minimised at $c^*(H) = P(H)$ whether or not one knows the value of $P(H)$. Similarly, $S_{ET}^*(H)$ is minimised at $c(H) = P(H)$ whether or not one knows the value of $P(H)$. For $c^*(H) = \frac{P(H)}{2-P(H)}$ and unknown $P(H)$, one can still find the values of $k(\epsilon)$:

$$k(\epsilon) = \begin{cases} \frac{1 - \frac{P(H)}{2-P(H)}}{1 - P(H)} = \frac{2}{2-P(H)} & \text{if } \epsilon > 0, \\ \frac{\frac{P(H)}{2-P(H)}}{P(H)} = \frac{1}{2-P(H)} & \text{if } \epsilon < 0, \\ 1 & \text{if } \epsilon = 0. \end{cases}$$

One can now plug these values into $S_{ET}^*(H)$, but for the sake of brevity, let me write the formula only for $k(\epsilon) = \frac{2}{2-P(H)}$ (and leave the rest to the reader):

$$S_{ET}^*(H) = P(H) \left(1 - \frac{2x - P(H)}{2 - P(H)} \right)^2 + 2(1 - P(H)) \left(\frac{2x - P(H)}{2 - P(H)} \right)^2.$$

¹² Note that $c(H) = \frac{P(H)}{2-P(H)}$, where $2 - P(H) = P(H) + 2(1 - P(H))$ is the normalising constant. It is a concrete instance of $c^*(X) = \frac{\sum_{\omega \in X} P(\omega)\lambda(X,\omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(X,\omega)}$ from Theorem 2.

Taking the derivative with respect to x gives $\frac{8(x-P(H))}{2-P(H)}$, which is minimised at $x = c(H) = P(H)$. For $k(\epsilon) = \frac{1}{2-P(H)}$, the derivative w.r.t. x gives $\frac{2(x-P(H))}{2-P(H)}$ minimised at $c(H) = P(H)$.

Note that Gneiting and Brehmer in Brehmer and Gneiting (2020) do not use P as an argument in their properisation approach. But they use P to find a minimiser that I have called $c^*(X)$, which is an argument in their properisation approach (see Brehmer & Gneiting, 2020, p. 660). So, knowledge of P plays a crucial role also in their approach and work outside accuracy-first epistemology. That said, I admit that requiring $P(X)$ and $c^*(X)$ to construct and evaluate s_*^X and its expected value is a demanding assumption. There may be a way around it, but as it stands now, I do not know how to do it.

Let me briefly comment on the existence and uniqueness assumption of minimiser $c^*(X)$. In our example, $c^*(H)$ exists and is unique. But, in general, it is not guaranteed that $c^*(X)$ exists or is unique (see Brehmer & Gneiting, 2020, especially Sect. 3). There is a need for an argument justifying this assumption in the context of accuracy-first epistemology. Accuracy-first epistemology already uses assumptions that make such an argument possible. For example, scoring rules are generated by strictly convex functions (e.g., see Pettigrew, 2016, pp. 84–85) and are bounded from below (i.e., ideal credences get the zero inaccuracy score, e.g., see Definition 1 or (Schervish et al., 2009, p. 206) for details). But to formulate this argument properly, one needs a formal set-up that is too complicated to start here and now, so I will leave this question open in this paper.

6 Conclusion

Levinstein argued that there does not exist a scoring rule that is additive, proper, truth-directed, and differentiable such that it recognises the contingency of varying importance (see Theorem A.4 in Levinstein, 2019, p. 2929). He concluded that accuracy-first epistemology could not capture the contingency of varying importance while maintaining its commitment to propriety and truth-directedness. In this paper, I argue that accuracy-first epistemology can capture the contingency of varying importance while maintaining its commitment to additivity, propriety, truth-directedness, and differentiability. I argue that there exists a strictly proper, truth-directed, and differentiable weighted scoring rule s_*^X (an inaccuracy measure of individual credences) that recognises the contingency of varying importance and a global inaccuracy measure \mathcal{I}_{s_*} (an inaccuracy measure of entire credence functions) that also has all the required properties. That is, \mathcal{I}_{s_*} is truth-directed, differentiable, proper (strictly and strongly), additive (which, avoiding redundancy, I predicate only about global inaccuracy measures), and it recognises the contingency of varying importance. I also discuss how Levinstein's and my results coexist without contradicting each other and why it is so.

A proofs for Sect. 2 (the proof and idea behind Theorem A.4)

Theorem 1 (Levinstein) *Let \mathcal{I}_{s_λ} from Assumption 1 and Assumption 2 be given. If the contingency of varying importance holds, then $E_{Pr} s_\lambda^X(c(X))$ and $E_{Pr'} s_\lambda^X(c(X))$ are not minimised at the same point. Thus, s_λ^X and \mathcal{I}_{s_λ} is not strictly proper.*

Proof Following the notational convention from Subsection 1.3, one has:

$$E_{Pr} s_\lambda^X(c(X)) = \sum_{\omega \in \Omega} Pr(\omega) \lambda(X, \omega) s^X(v_\omega(X), c(X)), \text{ and}$$

$$E_{Pr'} s_\lambda^X(c(X)) = \sum_{\omega \in \Omega} Pr'(\omega) \lambda(X, \omega) s^X(v_\omega(X), c(X)).$$

By Assumption 2, $E_{Pr} s_\lambda^X(c(X))$ and $E_{Pr'} s_\lambda^X(c(X))$ are differentiable with respect to $c(X)$, so use the first derivative test to find the optima (let me drop the superscript X on the right-hand side of the following equations):

$$\frac{\partial E_{Pr} s_\lambda^X(c(X))}{\partial c(X)} = s'(1, c(X)) \sum_{\omega \in X} Pr(\omega) \lambda(X, \omega) + s'(0, c(X)) \sum_{\omega \notin X} Pr(\omega) \lambda(X, \omega) = 0,$$

$$\frac{\partial E_{Pr'} s_\lambda^X(c(X))}{\partial c(X)} = s'(1, c(X)) \sum_{\omega \in X} Pr'(\omega) \lambda(X, \omega) + s'(0, c(X)) \sum_{\omega \notin X} Pr'(\omega) \lambda(X, \omega) = 0,$$

which are equations that any minimiser of $E_{Pr} s_\lambda^X(c(X))$ and $E_{Pr'} s_\lambda^X(c(X))$, respectively, must satisfy. Using basic arithmetic operations, one gets that:

$$\frac{s'(1, c(X))}{s'(0, c(X))} = - \frac{\sum_{\omega \notin X} \lambda(X, \omega) Pr(\omega)}{\sum_{\omega \in X} \lambda(X, \omega) Pr(\omega)} \text{ and } \frac{s'(1, c(X))}{s'(0, c(X))} = - \frac{\sum_{\omega \notin X} \lambda(X, \omega) Pr'(\omega)}{\sum_{\omega \in X} \lambda(X, \omega) Pr'(\omega)}, \tag{5}$$

where, by assumption, $0 < Pr(X), Pr'(X) < 1$ and $0 < \lambda(X, \omega) < \infty$, so the fractions are well-defined. Assuming that $c^*(X)$ is the common minimiser of $E_{Pr} s_\lambda^X(c(X))$ and $E_{Pr'} s_\lambda^X(c(X))$, then $c^*(X)$ must satisfy both equalities in 5:

$$\frac{s'(1, c^*(X))}{s'(0, c^*(X))} = - \frac{\sum_{\omega \notin X} \lambda(X, \omega) Pr(\omega)}{\sum_{\omega \in X} \lambda(X, \omega) Pr(\omega)} = - \frac{\sum_{\omega \notin X} \lambda(X, \omega) Pr'(\omega)}{\sum_{\omega \in X} \lambda(X, \omega) Pr'(\omega)}. \tag{6}$$

It is, however, impossible for Eq. (6) to hold by the construction of Pr' and Definition 9; compare to Levinstein (2019, p. 2930). By Definition 9, $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$ and either $\omega_1, \omega_2 \in X$ or $\omega_1, \omega_2 \notin X$, thus consider two cases.

1. Assume that $\omega_1, \omega_2 \in X$ and $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$. By Definition 10, $Pr(\omega) = Pr'(\omega)$ for every ω other than ω_1 and ω_2 , so $\sum_{\omega \notin X} \lambda(X, \omega) Pr'(\omega) = \sum_{\omega \notin X} \lambda(X, \omega) Pr(\omega)$. The denominator with Pr' contains $\lambda(X, \omega_1)[Pr(\omega_1) + \epsilon] + \lambda(X, \omega_2)[Pr(\omega_2) - \epsilon]$, but the denominator with Pr con-

tains only a term $\lambda(X, \omega_1)Pr(\omega_1) + \lambda(X, \omega_2)Pr(\omega_2)$. Since $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$, the denominators do not equal. Thus, Eq. (6) cannot hold.

- Assume that $\omega_1, \omega_2 \notin X$ and $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$. So, $\sum_{\omega \in X} \lambda(X, \omega)Pr'(\omega) = \sum_{\omega \in X} \lambda(X, \omega)Pr(\omega)$. The nominator with Pr' contains $\lambda(X, \omega_1)[Pr(\omega_1) + \epsilon] + \lambda(X, \omega_2)[Pr(\omega_2) - \epsilon]$, but the nominator with Pr contains only the term $\lambda(X, \omega_1)Pr(\omega_1) + \lambda(X, \omega_2)Pr(\omega_2)$. Since $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$, the nominators do not equal. Thus, Eq. (6) cannot hold.

So, there exists no $c^*(X)$ that satisfies all the equalities in Eq. (6), i.e., there is no common minimiser for $E_{Pr} s_\lambda^X(c(X))$ and $E_{Pr'} s_\lambda^X(c(X))$. But if s_λ^X is strictly proper, then $E_{Pr'} s_\lambda^X(c(X))$ and $E_{Pr} s_\lambda^X(c(X))$ are minimised at the same point. So, s_λ^X is not strictly proper, which implies that \mathcal{I}_{s_λ} is not strictly proper. For *reductio*, assume that s_λ^X is not strictly proper, but \mathcal{I}_{s_λ} is strictly proper. So, by Definition 5, $E_P \mathcal{I}_{s_\lambda}(P) < E_P \mathcal{I}_{s_\lambda}(c)$ for any distinct functions c and $P \in \mathcal{P}$. But since s_λ^X is not strictly proper, then, by Definition 4, there is a credence $\hat{c}(X) \neq P(X)$ such that $E_P s_\lambda^X(\hat{c}(X)) \leq E_P s_\lambda^X(P(X))$. Let then one's credence function c be such that $c(X) = \hat{c}(X)$ and, for any other proposition Y to which P assigns a value, let $c(Y) = P(Y)$. Since $c(X) \neq P(X)$, P and c are different functions. For an additive \mathcal{I}_{s_λ} , it holds that $E_P \mathcal{I}_{s_\lambda}(P) \geq E_P \mathcal{I}_{s_\lambda}(c)$, which violates Definition 5. Therefore, \mathcal{I}_{s_λ} is not strictly proper. \square

B proofs for Sect. 3 (Constructing s_*^X and \mathcal{I}_{s_*})

Theorem 2 (Gneiting and Ranjan) *Let $(\Omega, \mathcal{F}), X \in \mathcal{F}$, a credence function $c, P \in \mathcal{P}$, and Assumption 2 be given. $E_P s_\lambda^X(c(X))$ is uniquely minimised at $c^*(X) = \frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)}$.*

Proof For the sake of notational simplicity, let me call the following sum γ :

$$\gamma = \sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega) = \sum_{\omega \in X} P(\omega)\lambda(X, \omega) + \sum_{\omega \notin X} P(\omega)\lambda(X, \omega). \tag{7}$$

Note that γ is the normalising constant. It will be useful noting that if $0 < \gamma < \infty$, one can multiply both sides of Eq. (7) by $\frac{1}{\gamma}$ to get:

$$1 = \frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma} + \frac{\sum_{\omega \notin X} P(\omega)\lambda(X, \omega)}{\gamma}.$$

Assume that s_λ^X satisfies Definition 8, so, by Definition 7, $0 < \gamma < \infty$. I can then write that:

$$\begin{aligned} E_P s_\lambda^X(c(X)) &= \sum_{\omega \in X} P(\omega)\lambda(X, \omega) s^X(v_\omega(X), c(X)) + \sum_{\omega \notin X} P(\omega)\lambda(X, \omega) s^X(v_\omega(X), c(X)) \\ &= \gamma \left[\frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma} s^X(1, c(X)) + \frac{\sum_{\omega \notin X} P(\omega)\lambda(X, \omega)}{\gamma} s^X(0, c(X)) \right]. \end{aligned}$$

Note that $c^*(X) = \frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)} = \frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma}$. Assume that $\hat{c}(X)$ is a credence in X such that $\hat{c}(X) \neq c^*(X)$, so:

$$\begin{aligned} E_{p s_\lambda^X}(c^*(X)) &= \gamma \left[\frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma} s^X(1, c^*(X)) + \frac{\sum_{\omega \notin X} P(\omega)\lambda(X, \omega)}{\gamma} s^X(0, c^*(X)) \right] \\ &< \gamma \left[\frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma} s^X(1, \hat{c}(X)) + \frac{\sum_{\omega \notin X} P(\omega)\lambda(X, \omega)}{\gamma} s^X(0, \hat{c}(X)) \right] \\ &= E_{p s_\lambda^X}(\hat{c}(X)), \end{aligned}$$

since, by Assumption 2, s^X in s_λ^X is strictly proper. That is, by Definition 4, the unique minimiser of $E_{p s_\lambda^X}(c(X))$ is at $c(X) = p$, where p is a weight. By Definition 4, this works only if the weights sum up to 1. But we have already established that $\frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma} + \frac{\sum_{\omega \notin X} P(\omega)\lambda(X, \omega)}{\gamma} = 1$. So, $E_{p s_\lambda^X}(c(X))$ is uniquely minimised at:

$$p = \frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\gamma} = \frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)} = c^*(X).$$

□

Lemma 2 *It holds that $0 \leq c^*(X) \leq 1$ (and $0 \leq c^*(\neg X) \leq 1$) and if $X = \Omega$, then $c^*(X) = 1$ (and $c^*(\neg X) = 0$).*

Proof By Definition 7, $0 < \lambda(X, \omega) < \infty$. By Theorem 2, $0 \leq \sum_{\omega \in X} P(\omega)\lambda(X, \omega) \leq \sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)$ and so $0 \leq c^*(X) \leq 1$. One can use similar reasoning to show that $0 \leq c^*(\neg X) \leq 1$. Now, assume that $X = \Omega$, so $\sum_{\omega \in X} P(\omega)\lambda(X, \omega) = \sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)$. Thus, by Theorem 2, $c^*(X) = 1$. Moreover, the set of $\omega \notin X$, i.e., $\neg X$, is empty. Probability P of the empty set is 0, so $P(\omega) = 0$ for any $\omega \in \neg X$. Thus, $\sum_{\omega \in \neg X} P(\omega)\lambda(\neg X, \omega) = 0$ and, by Theorem 2, $c^*(\neg X) = 0$. □

Lemma 3 *Given Assumption 3, then $c^*(X) + c^*(\neg X) = 1$.*

Proof By Assumption 3, $\lambda(X, \omega) = \lambda(\neg X, \omega)$ for any $X \in \mathcal{F}$ and $\omega \in \Omega$. Since the set of $\omega \in \neg X$ equals the set of $\omega \notin X$, Theorem 2 gives that:

$$c^*(\neg X) = \frac{\sum_{\omega \in \neg X} P(\omega)\lambda(\neg X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(\neg X, \omega)} = \frac{\sum_{\omega \notin X} P(\omega)\lambda(X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)}.$$

By Theorem 2, it then follows that $c^*(X) + c^*(\neg X) = 1$. □

Lemma 4 *If $P(X) = 0$, then $k(\epsilon) = 1 - c^*(X)$ or 1 and if $P(X) = 1$, then $k(\epsilon) = c^*(X)$ or 1.*

Proof Assume that $P(X) = 0$. By Definition 11, $c(X) = \epsilon$, so $0 \leq \epsilon \leq 1$ since $c : \mathcal{F} \rightarrow [0, 1]$. By Definition 11, if $\epsilon = 0$, then $k(\epsilon) = 1$ and if $0 < \epsilon \leq 1$, then $k(\epsilon) = 1 - c^*(X)$. Next, assume that $P(X) = 1$. By Definition 11,

$c(X) - P(X) = \epsilon \leq 0$. By Definition 11, if $\epsilon = 0$, then $k(\epsilon) = 1$ and if $\epsilon < 0$, then $k(\epsilon) = c^*(X)$. □

Lemma 5 Any value of $c^*(X) + \epsilon k(\epsilon)$ comes from $[0, 1]$.

Proof For $c^*(X) + \epsilon k(\epsilon)$ to be from $[0, 1]$, one needs to check three possible cases: $\epsilon > 0$, $\epsilon < 0$, and $\epsilon = 0$; remember that $c(X) - P(X) = \epsilon$.

1. Let $\epsilon > 0$, so assume that $0 \leq c^*(X) + \frac{1-c^*(X)}{1-P(X)}\epsilon \leq 1$, which is well-defined by Lemma 4. Multiply the inequality by $1 - P(X)$ and subtract ϵ to get $-\epsilon \leq c^*(X)[1 - P(X) - \epsilon] \leq 1 - P(X) - \epsilon$. If $1 - P(X) - \epsilon = 0$ (i.e., $c(X) = P(X) + \epsilon = 1$) or $c^*(X) = 0$, then the inequality becomes $-\epsilon \leq 0$ or $-\epsilon \leq 0 \leq 1 - c(X)$. These hold for $\epsilon > 0$ and thus $0 < c(X) \leq 1$. If $1 - P(X) - \epsilon \neq 0$ (i.e., $0 \leq c(X) < 1$) and $c^*(X) \neq 0$, then $\frac{-\epsilon}{1-c(X)} \leq c^*(X) \leq 1$. Both sides of this inequality hold. First, by Lemma 2, $0 \leq c^*(X) \leq 1$. Secondly, $\frac{-\epsilon}{1-c(X)} < 0$ for $\epsilon > 0$ and $0 \leq c(X) < 1$.
2. Let $\epsilon < 0$, so assume that $0 \leq c^*(X) + \frac{c^*(X)}{P(X)}\epsilon \leq 1$, which is well-defined by Lemma 4. Multiply by $P(X)$ to get $0 \leq c^*(X)[\epsilon + P(X)] \leq P(X)$. If $P(X) + \epsilon = 0$ (i.e., $c(X) = 0$) or $c^*(X) = 0$, the inequality reduces to $0 \leq P(X)$, which holds for a probabilistic P . If $P(X) + \epsilon \neq 0$ (i.e., $0 < c(X) \leq 1$) and $c^*(X) \neq 0$, then one has $0 \leq c^*(X) \leq \frac{P(X)}{\epsilon + P(X)}$. Both sides of this inequality hold. First, by Lemma 2, $0 \leq c^*(X) \leq 1$. Secondly, $1 < \frac{P(X)}{\epsilon + P(X)}$ for $\epsilon < 0$ and $0 < \epsilon + P(X) = c(X) \leq 1$.
3. Assume that $\epsilon = 0$. So, $\epsilon k(\epsilon) = 0$ and $c^*(X) + \epsilon k(\epsilon) = c^*(X)$. By Lemma 2, $0 \leq c^*(X) \leq 1$. □

Proposition 1 s_*^X is a strictly proper, truth-directed, continuous, and differentiable (with respect to $c(X)$) weighted scoring rule such that $s_*^X(0, 0) = s_*^X(1, 1) = 0$.

Proof Let $T = A\mathbf{x} + \mathbf{b}$ be a map such that $A = \begin{bmatrix} 1 & 0 \\ 0 & k(\epsilon) \end{bmatrix}$, $\mathbf{b} = (0, c^*(X) - P(X)k(\epsilon))$, and \mathbf{x} is a transpose of a vector $(v_\omega(X), c(X))$. Then, T is an affine transformation of a vector $(v_\omega(X), c(X))$ to $(v_\omega(X), c^*(X) + \epsilon k(\epsilon))$. Notice that $c^*(X)$, $P(X)$, and $k(\epsilon)$ are constant for any given $c(X)$. One can then use s^X from Definition 1 to score $(v_\omega(X), c^*(X) + \epsilon k(\epsilon))$ and weight that score by $\lambda(X, \omega)$ from Definition 7. Remember that, by Lemma 5, the value of $c^*(X) + \epsilon k(\epsilon)$ always comes from $[0, 1]$, so s^X in s_*^X satisfies Definition 1, i.e., $s^X = \{0, 1\} \times [0, 1] \rightarrow [0, \infty]$, and, by Definition 7, $\lambda(X, \omega) := \mathcal{F} \times \Omega \rightarrow (0, \infty)$. So, s_*^X gives values from $[0, \infty]$. Thus, one has $s_*^X : \{0, 1\} \times [0, 1] \times (0, \infty) \rightarrow [0, \infty]$ such that $s_*^X(v_\omega(X), c(X), \lambda(X, \omega))$, so s_*^X satisfies Definition 8 and is a weighted scoring rule.

Now, pick any X from \mathcal{F} and P from \mathcal{P} . If s_*^X is strictly proper, then, by Definition 4, $E_P s_*^X(P(X)) < E_P s_*^X(\hat{c}(X))$ holds for any $\hat{c}(X) \neq P(X)$. Assume that one's credence in X is $P(X)$, i.e., $c(X) = P(X)$. By Definition 11, $\epsilon = 0$ and so $s_*^X(v_\omega(X), P(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(v_\omega(X), c^*(X))$.

If one's credence in X is $\hat{c}(X) \neq P(X)$, then $\hat{c}(X) - P(X) = \hat{\epsilon} \neq 0$ and so $s_*^X(v_\omega(X), \hat{c}(X), \lambda(X, \omega)) = \lambda(X, \omega)s_*^X(v_\omega(X), c^*(X) + \hat{\epsilon}k(\hat{\epsilon}))$, where $\hat{\epsilon}k(\hat{\epsilon}) \neq 0$. By Theorem 2 and Definition 11, the inequality $E_{P_*} s_*^X(P(X)) < E_{P_*} s_*^X(\hat{c}(X))$ holds, as required:

$$\begin{aligned} E_{P_*} s_*^X(P(X)) &= \sum_{\omega \in X} P(\omega) s_*^X(1, P(X), \lambda(X, \omega)) + \sum_{\omega \notin X} P(\omega) s_*^X(0, P(X), \lambda(X, \omega)) \\ &= \sum_{\omega \in X} P(\omega) \lambda(X, \omega) s_*^X(1, c^*(X)) + \sum_{\omega \notin X} P(\omega) \lambda(X, \omega) s_*^X(0, c^*(X)) \\ &= E_{P_*} s_*^X(c^*(X)) \\ &< E_{P_*} s_*^X(c^*(X) + \hat{\epsilon}k(\hat{\epsilon})) \\ &= \sum_{\omega \in X} P(\omega) \lambda(X, \omega) s_*^X(1, c^*(X) + \hat{\epsilon}k(\hat{\epsilon})) + \sum_{\omega \notin X} P(\omega) \lambda(X, \omega) s_*^X(0, c^*(X) + \hat{\epsilon}k(\hat{\epsilon})) \\ &= \sum_{\omega \in X} P(\omega) s_*^X(1, \hat{c}(X), \lambda(X, \omega)) + \sum_{\omega \notin X} P(\omega) s_*^X(0, \hat{c}(X), \lambda(X, \omega)) \\ &= E_{P_*} s_*^X(\hat{c}(X)). \end{aligned}$$

Truth-directedness follows directly from Lemma 1. Now, consider differentiability. One can assume that s^X satisfying Definition 1 is differentiable w.r.t. $c(X)$, e.g., as Levinstein does (see Assumption 2). Let a function g take $c(X)$ to $c^*(X) + \epsilon k(\epsilon)$, which is a linear function of the form $ax + b$, where x is given by $c(X)$, $a = k(\epsilon)$, and $b = c^*(X) - P(X)k(\epsilon)$. So, g is differentiable with respect to $c(X)$. Since s^X and g are differentiable, then $s^X(v_\omega(X), g(c(X)))$ is differentiable because a composite function of differentiable functions is differentiable. Moreover, multiplying a differentiable function by a scalar, i.e., $\lambda(X, \omega)$, preserves differentiability, so s_*^X is differentiable with respect to $c(X)$. It follows that s_*^X is a continuous function of $c(X)$ for all worlds ω .

Finally, there are two cases when an agent has a fully accurate credence $c(X)$ at ω . First, when $c(X) = 1$ and X is true at that ω , see 1. Secondly, when $c(X) = 0$ and X is false at that ω , see 2.

1. Assume that X is true at ω and $c(X) = 1$. If $c^*(X) + \epsilon k(\epsilon) = 1$, then, by Definition 1, $s^X(1, c^*(X) + \epsilon k(\epsilon)) = s^X(1, 1) = 0$. Thus, by Definition 11, $s_*^X(1, 1) = 0$ independently of the value of a weight function. Let me show that if $c(X) = 1$, then $c^*(X) + \epsilon k(\epsilon) = 1$.
 - (a) Assume that $P(X) = c(X) = 1$. By Definition 11, $\epsilon = 0$ and so $c^*(X) + \epsilon k(\epsilon) = c^*(X)$. Since $P(X) = 1$, $P(\omega) = 0$ for any $\omega \notin X$, so $\sum_{\omega \notin X} \lambda(X, \omega) P(\omega) = 0$ and $\sum_{\omega \in X} \lambda(X, \omega) P(\omega) = \sum_{\omega \in \Omega} \lambda(X, \omega) P(\omega)$. Thus, by Theorem 2, $c^*(X) = 1$.
 - (b) Assume that $P(X) < c(X) = 1$. By Definition 11, $\epsilon = 1 - P(X) > 0$, so $c^*(X) + \epsilon k(\epsilon) = c^*(X) + \epsilon \frac{1 - c^*(X)}{1 - P(X)} = 1$.
2. Assume that X is false at ω and $c(X) = 0$. If $c^*(X) + \epsilon k(\epsilon) = 0$, then, by Definition 1, $s^X(0, c^*(X) + \epsilon k(\epsilon)) = s^X(0, 0) = 0$. Thus, by Definition 11, $s_*^X(0, 0) = 0$

independently of the value of a weight function. Let me show that if $c(X) = 0$, then $c^*(X) + \epsilon k(\epsilon) = 0$.

- (a) Assume that $P(X) = c(X) = 0$. By Definition 11, $\epsilon = 0$ and so $c^*(X) + \epsilon k(\epsilon) = c^*(X)$. Since $P(X) = 0$, $P(\omega) = 0$ for any $\omega \in X$, i.e., $\sum_{\omega \in X} \lambda(X, \omega)P(\omega) = 0$. By Theorem 2, $c^*(X) = 0$.
- (b) Assume that $P(X) > c(X) = 0$. By Definition 11, $\epsilon = -P(X)$, so $c^*(X) + \epsilon k(\epsilon) = c^*(X) + \frac{c^*(X)}{P(X)}\epsilon = 0$.

□

Lemma 6 s^X in s_*^X is not strictly proper.

Proof Given Definition 4, take the following expectation $ps^X(1, c^*(X) + \epsilon k(\epsilon)) + (1 - p)s^X(0, c^*(X) + \epsilon k(\epsilon))$ and proceed with the first derivative test by differentiating with respect to $c^*(X)$. Cancelling $k(\epsilon)$ gives $p(s^X)'(1, c^*(X) + \epsilon k(\epsilon)) + (1 - p)(s^X)'(0, c^*(X) + \epsilon k(\epsilon)) = 0$, which is minimised at $p = c^*(X) + \epsilon k(\epsilon)$ if s^X is strictly proper. By Definition 4, for a scoring rule to be strictly proper it must be minimised at $c(X) = p$, but $p = c^*(X) + \epsilon k(\epsilon)$ holds for $c(X) = \frac{p - c^*(X)}{k(\epsilon)} + p$, where $p = P(X)$.

Lemma 7 Given Assumption 3, $P \in \mathcal{P}$, and $c(\neg X) - P(\neg X) = \epsilon'$, then

1. $\epsilon = -\epsilon'$ iff $c(X) + c(\neg X) = 1$,
2. if $c^*(X) = 0$, then $P(X) = 0$ and if $c^*(\neg X) = 0$, then $P(\neg X) = 0$,
3. if $c^*(X) = 1$, then $P(X) = 1$ and if $c^*(\neg X) = 1$, then $P(\neg X) = 1$, and
4. $[c^*(X) + \epsilon k(\epsilon)] + [c^*(\neg X) + \epsilon' k(\epsilon')] = 1$ iff $c(X) + c(\neg X) = 1$.

Proof

1. (\Rightarrow) First, assume that $\epsilon = -\epsilon'$, which gives $c(X) - P(X) = -[c(\neg X) - P(\neg X)] = c(X) - P(X) = -[c(\neg X) - P(\neg X)] = -[c(\neg X) - (1 - P(X))]$. It follows that $c(X) + c(\neg X) = 1$. (\Leftarrow) Assuming that $c(X) + c(\neg X) = 1$ gives $\epsilon + P(X) + \epsilon' + P(\neg X) = \epsilon + P(X) + \epsilon' + 1 - P(X) = 1$. It follows that $\epsilon = -\epsilon'$.
2. Assuming that $c^*(X) = 0$, by Theorem 2, gives $\sum_{\omega \in X} P(\omega)\lambda(X, \omega) = 0$. By Definition 7, $\lambda(X, \omega) > 0$ for any X and ω . So, $P(\omega)$ must be 0 for any $\omega \in X$, i.e., $P(X) = 0$. Similarly, assume that $c^*(\neg X) = 0$. By Theorem 2, $\sum_{\omega \in \neg X} P(\omega)\lambda(\neg X, \omega) = 0$. Since $\lambda(\neg X, \omega)$ is always positive, $P(\omega)$ must be 0 for any $\omega \in \neg X$, i.e., $P(\neg X) = 0$.
3. Assuming that $c^*(X) = 1$, by Theorem 2, gives $\frac{\sum_{\omega \in X} P(\omega)\lambda(X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)} = 1$. Thus, $\sum_{\omega \in X} P(\omega)\lambda(X, \omega) = \sum_{\omega \in \Omega} P(\omega)\lambda(X, \omega)$ and $\sum_{\omega \notin X} P(\omega)\lambda(X, \omega) = 0$. So, by Definition 7, $P(\omega) = 0$ for any $\omega \notin X$ (i.e., any ω from $\neg X$) and thus $P(\neg X) = 0$ and $P(X) = 1$ since P is a probability function. Similarly, assuming that $c^*(\neg X) = 1$ gives $\frac{\sum_{\omega \in \neg X} P(\omega)\lambda(\neg X, \omega)}{\sum_{\omega \in \Omega} P(\omega)\lambda(\neg X, \omega)} = 1$. So, $\sum_{\omega \in \neg X} P(\omega)\lambda(\neg X, \omega) = \sum_{\omega \in \Omega} P(\omega)\lambda(\neg X, \omega)$ and $\sum_{\omega \notin \neg X} P(\omega)\lambda(\neg X, \omega) = 0$. So, by Definition 7, $P(\omega) = 0$ for any $\omega \notin \neg X$ (i.e., any ω from X). Thus, $P(X) = 0$ and $P(\neg X) = 1$ since P is a probability function.

4. (\Rightarrow) Assume that $[c^*(X) + \epsilon k(\epsilon)] + [c^*(\neg X) + \epsilon' k(\epsilon')] = 1$, which reduces to $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ if, by Lemma 3, one assumes that $c^*(X) + c^*(\neg X) = 1$. To prove that $c(X) + c(\neg X) = 1$, I will show that either if $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ holds, then $\epsilon = -\epsilon'$ (cases i - ii) or $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ does not hold (cases iii - iv). Since ϵ and ϵ' is positive, negative, or 0, there are nine options, but I will cluster together similar cases.

i Assume that $\epsilon = 0$ and $\epsilon' = 0$. So, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $0 = 0$, and $\epsilon = -\epsilon'$ holds.

ii Assume that $\epsilon > 0$ and $\epsilon' < 0$. So, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $\epsilon \frac{1-c^*(X)}{1-P(X)} = -\epsilon' \frac{c^*(\neg X)}{P(\neg X)}$. Since $P(\neg X) = 1 - P(X)$ and, by Lemma 3, $c^*(\neg X) = 1 - c^*(X)$, one has that $\epsilon = -\epsilon'$. Similarly, if $\epsilon < 0$ and $\epsilon' > 0$, then $\epsilon \frac{c^*(X)}{P(X)} = -\epsilon' \frac{1-c^*(\neg X)}{1-P(\neg X)}$ and $\epsilon = -\epsilon'$ follows.

iii Assume that $\epsilon < 0$ and $\epsilon' < 0$. So, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $\epsilon \frac{c^*(X)}{P(X)} = -\epsilon' \frac{c^*(\neg X)}{P(\neg X)}$. Now, $\frac{c^*(X)}{P(X)}$ and $\frac{c^*(\neg X)}{P(\neg X)}$ is either 0 or positive. Clearly, the equality does not hold if one is 0 and the other positive. Now, if $\frac{c^*(X)}{P(X)} > 0$ and $\frac{c^*(\neg X)}{P(\neg X)} > 0$, then $\epsilon k(\epsilon) < 0$ but $-\epsilon' k(\epsilon') > 0$ and $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ cannot hold. So, try $\epsilon \frac{c^*(X)}{P(X)} = -\epsilon' \frac{c^*(\neg X)}{P(\neg X)} = 0$. Then, both $c^*(X)$ and $c^*(\neg X)$ must be 0. If $c^*(X) = 0$, then $P(X) = 0$ and so, by Definition 11, $\epsilon \geq 0$, which contradicts that $\epsilon < 0$. If $c^*(\neg X) = 0$, then $P(\neg X) = 0$. So, $\epsilon' \geq 0$, which contradicts that $\epsilon' < 0$.

Now, assume that $\epsilon > 0$ and $\epsilon' > 0$, so one has $\epsilon \frac{1-c^*(X)}{1-P(X)} = -\epsilon' \frac{1-c^*(\neg X)}{1-P(\neg X)}$. Again, $\frac{1-c^*(X)}{1-P(X)}$ and $\frac{1-c^*(\neg X)}{1-P(\neg X)}$ is either 0 or positive. The equality $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ does not hold for one 0 and the other positive or if $\frac{1-c^*(X)}{1-P(X)} > 0$ and $\frac{1-c^*(\neg X)}{1-P(\neg X)} > 0$ since then $\epsilon k(\epsilon) > 0$ but $-\epsilon' k(\epsilon') < 0$. Considering $\epsilon \frac{1-c^*(X)}{1-P(X)} = -\epsilon' \frac{1-c^*(\neg X)}{1-P(\neg X)} = 0$ means that $1 - c^*(X) = 1 - c^*(\neg X) = 0$. If $1 - c^*(X) = 0$, then $c^*(X) = 1$ and $P(X) = 1$. So, $\epsilon \leq 0$, which contradicts that $\epsilon > 0$. If $1 - c^*(\neg X) = 0$, then $c^*(\neg X) = 1$ and $P(\neg X) = 1$. So, $\epsilon' \leq 0$, which contradicts that $\epsilon' > 0$.

iv Assume that $\epsilon = 0$ and $\epsilon' > 0$. Then, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $0 = -\epsilon' \frac{1-c^*(\neg X)}{1-P(\neg X)}$ and so $1 - c^*(\neg X)$ must be 0, i.e., $c^*(\neg X) = 1$. Thus, $P(\neg X) = 1$ and so $\epsilon' \leq 0$, which contradicts that $\epsilon' > 0$. Similarly, assuming that $\epsilon = 0$ and $\epsilon' < 0$, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $0 = -\epsilon' \frac{c^*(\neg X)}{P(\neg X)}$ and $c^*(\neg X)$ must be 0. Thus, $P(\neg X) = 0$ and so $\epsilon' \geq 0$, which contradicts that $\epsilon' < 0$.

Likewise, assume that $\epsilon > 0$ and $\epsilon' = 0$. So, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $\epsilon \frac{1-c^*(X)}{1-P(X)} = 0$. It must be that $1 - c^*(X) = 0$, i.e., $c^*(X) = 1$. Thus, $P(X) = 1$ and so $\epsilon \leq 0$, which contradicts that $\epsilon > 0$. Now, assume that $\epsilon < 0$ and $\epsilon' = 0$. So, $\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$ becomes $\epsilon \frac{c^*(X)}{P(X)} = 0$, so it must be that $c^*(X) = 0$. Thus, $P(X) = 0$ and so $\epsilon \geq 0$, which contradicts that $\epsilon < 0$.

(\Leftarrow) Assume that $c(X) + c(\neg X) = 1$, so $\epsilon = -\epsilon'$. Let also Assumption 3 hold, so I can apply Lemma 3, which says that $c^*(X) + c^*(\neg X) = 1$. So, $[c^*(X) + \epsilon k(\epsilon)] + [c^*(\neg X) + \epsilon' k(\epsilon')] = 1 + \epsilon k(\epsilon) + \epsilon' k(\epsilon')$, which is 1 only if

$\epsilon k(\epsilon) = -\epsilon' k(\epsilon')$. If $\epsilon = \epsilon' = 0$, it holds. If $\epsilon > 0$, then $\epsilon' < 0$ and $e^{\frac{1-c^*(X)}{1-P(X)}} = -\epsilon' \frac{c^*(\neg X)}{P(\neg X)}$, that is, $\epsilon \frac{c^*(\neg X)}{P(\neg X)} = e^{\frac{c^*(\neg X)}{P(\neg X)}}$. If $\epsilon < 0$, then $\epsilon' > 0$ and $e^{\frac{c^*(X)}{P(X)}} = -\epsilon' \frac{1-c^*(\neg X)}{1-P(\neg X)}$, that is, $\epsilon \frac{c^*(X)}{P(X)} = e^{\frac{c^*(X)}{P(X)}}$. \square

Lemma 8 (Gneiting and Brehmer) *If s_λ^X is strictly proper, then s_*^X and s_λ^X assign the same score to one's credence $c(X)$.*

Proof Given Definition 8, let $s_\lambda^X(v_\omega(X), c(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(v_\omega(X), c(X))$ be a strictly proper scoring rule, i.e., let weights $\lambda(X, \omega)$ be constant for the given X at any ω . By Definition 4 and Theorem 2, if s_λ^X is strictly proper, then the unique minimiser $c^*(X) = P(X)$. By Definition 11, if $c^*(X) = P(X)$, then $k(\epsilon) = 1$ and so $c^*(X) + \epsilon k(\epsilon) = P(X) + \epsilon = c(X)$. So, by Definition 11, one has that $s_*^X(0, c(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(0, c(X)) = s_\lambda^X(0, c(X), \lambda(X, \omega))$ and $s_*^X(1, c(X), \lambda(X, \omega)) = \lambda(X, \omega) s^X(1, c(X)) = s_\lambda^X(1, c(X), \lambda(X, \omega))$. Thus, if s_λ^X is strictly proper, then s_λ^X and s_*^X assign the same score to one's credence $c(X)$. \square

Lemma 9 (Strict and Strong Propriety) *\mathcal{I}_{s_*} is strongly and strictly proper, thus truth-directed.*

Proof By Definition 6, $\mathcal{I}_{s_*}(c, \omega) = \sum_{X \in \mathcal{F}} s_*^X(v_\omega(X), c(X), \lambda(X, \omega))$ is strongly proper just in case $E_P \mathcal{I}_{s_*}(c)$ is uniquely minimised at $P(X) = c(X)$ for all $X \in \mathcal{F}$ and all probability functions $P \in \mathcal{P}$. Pick any $X \in \mathcal{F}$ and differentiate $E_P \mathcal{I}_{s_*}(c)$ with respect to $c(X)$ as follows; I omit the superscript X in the middle part of the following equation:

$$\frac{\partial E_P \mathcal{I}_{s_*}(c)}{\partial c(X)} = \sum_{\omega \in X} P(\omega) s_*'(1, c(X), \lambda(X, \omega)) + \sum_{\omega \notin X} P(\omega) s_*'(0, c(X), \lambda(X, \omega)) = \frac{d E_P s_*^X(c(X))}{d c(X)}.$$

By Proposition 1, s_*^X is strictly proper, so $\frac{\partial E_P \mathcal{I}_{s_*}(c)}{\partial c(X)} = 0 = \frac{d E_P s_*^X(c(X))}{d c(X)}$ holds at $P(X) = c(X)$. That is, $E_P \mathcal{I}_{s_*}(c)$ is uniquely minimised at $P(X) = c(X)$ for all $X \in \mathcal{F}$ and all probability functions $P \in \mathcal{P}$, as required. In other words, for any $X \in \mathcal{F}$, any probability function $P \in \mathcal{P}$, and any credence function c such that $P(X) \neq c(X)$, it holds that $E_P s_*^X(P(X)) < E_P s_*^X(c(X))$. Strict propriety of \mathcal{I}_{s_*} follows. By Definition 5, \mathcal{I}_{s_*} is strictly proper if for any distinct credence functions $P \in \mathcal{P}$ and c it holds that (I use the fact that \mathcal{I}_{s_*} is additive):

$$E_P \mathcal{I}_{s_*}(P) < E_P \mathcal{I}_{s_*}(c) \tag{8}$$

$$\sum_{X \in \mathcal{F}} E_P s_*^X(P(X)) < \sum_{X \in \mathcal{F}} E_P s_*^X(c(X)).$$

By the strict propriety of s_*^X , the inequality $E_P s_*^X(P(X)) < E_P s_*^X(c(X))$ holds for any $c(X) \neq P(X)$. So, for any distinct P and c , there is $X \in \mathcal{F}$ such that $c(X) \neq P(X)$ and,

by additivity, $E_P \mathcal{I}_{s_*}(P) < E_P \mathcal{I}_{s_*}(c)$. Truth-directedness of \mathcal{I}_{s_*} follows from the truth-directedness of s_*^X and additivity of \mathcal{I}_{s_*} . \square

Lemma 10 (*Contingency of Varying Importance*) Given \mathcal{I}_{s_*} , $E_{Pr} s_*^X(c(X))$ and $E_{Pr'} s_*^X(c(X))$ are uniquely minimised at the same point.

Proof By Proposition 2, assume that \mathcal{I}_{s_*} recognises the contingency of varying importance. That is, by Definition 9, assume that there are worlds $\omega_1, \omega_2 \in \Omega$ and a proposition $X \in \mathcal{F}$ such that $v_{\omega_1}(X) = v_{\omega_2}(X)$ but $\lambda(X, \omega_1) \neq \lambda(X, \omega_2)$. Assume there are probability functions Pr and Pr' as defined in Definition 10. Focusing on X , $E_{Pr} s_*^X(c(X))$ is a part of $E_{Pr} \mathcal{I}_{s_*}(c)$ and $E_{Pr'} s_*^X(c(X))$ is a part of $E_{Pr'} \mathcal{I}_{s_*}(c)$. By Proposition 1, s_*^X is strictly proper (or one can use strong propriety of \mathcal{I}_{s_*}), so $E_{Pr} s_*^X(c(X))$ is uniquely minimised at $c(X) = Pr(X)$ and $E_{Pr'} s_*^X(c(X))$ is uniquely minimised at $c(X) = Pr'(X)$. By construction, if $v_{\omega_1}(X) = v_{\omega_2}(X)$, then $Pr(X) = Pr'(X)$. Therefore, $E_{Pr} s_*^X(c(X))$ is minimised at the same point as $E_{Pr'} s_*^X(c(X))$ while \mathcal{I}_{s_*} recognises the contingency of varying importance. \square

Acknowledgements I am grateful to my colleagues from the Czech Academy of Sciences and the University of Bristol for their comments on various stages of the paper. Especially, I want to thank to Richard Pettigrew whose help with the paper was enormous. Thanks to anonymous referees for their very useful comments, to the editors, and to Jonáš Gray for linguistic advice.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement. I confirm that the work on this paper was supported by the Formal Epistemology – the Future Synthesis grant, in the framework of the Praemium Academicum programme of the Czech Academy of Sciences. These funding sources have no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. The work on this paper was supported by the Formal Epistemology – the Future Synthesis grant, in the framework of the Praemium Academicum programme of the Czech Academy of Sciences.

Data availability Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The author declares that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Brehmer, J. R., & Gneiting, T. (2020). Properization: Constructing proper scoring rules via Bayes acts. *Annals of the Institute of Statistical Mathematics*, 72(3), 659–673.
- Douven, I. (2020). Scoring in context. *Synthese*, 197(4), 1565–1580.
- Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis*, 60(2), 143–147.
- Halpern, J. Y. (2005). *Reasoning about uncertainty* (1st ed.). MIT Press.
- Holzmann, H., & Klar, B. (2017). Focusing on regions of interest in forecast evaluation. *The Annals of Applied Statistics*, 11(4), 2404–2431.
- Joyce, J. M. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In F. Huber & C. Schmidt-Petri (Eds.), *Degrees of Belief* (pp. 263–297). Springer.
- Kierland, B., & Monton, B. (2005). Minimizing inaccuracy for self-locating beliefs. *Philosophy and Phenomenological Research*, 70(2), 384–395.
- Levinstein, B. A. (2017). A pragmatist's guide to epistemic utility. *Philosophy of Science*, 84(4), 613–638.
- Levinstein, B. A. (2019). An objection of varying importance to epistemic utility theory. *Philosophical Studies*, 176(11), 2919–2931.
- Merkle, E. C., & Steyvers, M. (2013). Choosing a strictly proper scoring rule. *Decision Analysis*, 10(4), 292–304.
- Pelenis, J. (2014). Weighted scoring rules for comparison of density forecasts on subsets of interest. <https://api.semanticscholar.org/CorpusID:160001729>, Accessed 26-August-2022.
- Pettigrew, R. (2013). Epistemic utility and norms for credences. *Philosophy Compass*, 8(10), 897–908.
- Pettigrew, R. (2016). *Accuracy and the laws of credence* (1st ed.). OUP.
- Pettigrew, R. (2019). On the Accuracy of Group Credences. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology* (Vol. 6, pp. 137–160). OUP: Oxford.
- Pettigrew, R. (2022). Accuracy-first epistemology without additivity. *Philosophy of Science*, 89(1), 128–151.
- Predd, J. B., Seiringer, R., Lieb, E. H., Osherson, D. N., Poor, H. V., & Kulkarni, S. R. (2009). Probabilistic coherence and proper scoring rules. *IEEE Transactions on Information Theory*, 55(10), 4786–4792.
- Pug, C. C. (2015). *Real mathematical analysis* (2nd ed.). Springer.
- Ranjan, R., & Gneiting, T. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3), 411–422.
- Schervish, M. J. (1989). A general method for comparing probability assessors. *The Annals of Statistics*, 17(4), 1856–1879.
- Schervish, M. J., Seidenfeld, T., & Kadane, J. B. (2009). Proper scoring rules, dominated forecasts, and coherence. *Decision Analysis*, 6(4), 202–221.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.