# Evidence and the Epistemic Betterness

Ilho Park (Jeonbuk National University)

A Penultimate Version, Forthcoming in *Synthese*



**Abstract**

It seems intuitive that our credal states are improved if we obtain evidence favoring truth over any falsehood. In this regard, Fallis and Lewis have recently provided and discussed some formal versions of such an intuition, which they name 'the Monotonicity Principle' and 'Elimination'. They argue, with those principles in hand, that the Brier rule, one of the most popular rules of accuracy, is not a good measure, and that accuracy-firsters cannot underwrite both probabilism and conditionalization. In this paper, I will argue that their conclusions are somewhat hasty. Specifically, I will demonstrate that there is another version of the Monotonicity Principle that can be satisfied by some additive rules of accuracy, such as the Brier rule. Moreover, it will also be argued that their version of the principle has some undesirable features regarding the epistemic betterness. Therefore, their criticisms can hardly jeopardize accuracy-firsters until any further justification of their versions of the Monotonicity Principle and Elimination is provided.

## 1 Introduction

It is often taken for granted that evidence, though not necessary, epistemically improves our credal state. Some epistemic norms like the reflection principle, which is suggested by van Fraassen (1984), depend on this kind of intuition about the relationship between evidence and epistemic betterness. The epistemic plausibility of the principle may well be bolstered by treating an agent's future self as an expert who is epistemically superior to her current self. Why do we take it that such epistemic superiority should hold? Some authors may say that this is because the future self undergoes more courses of experience than the current one—in other words, credal states have more pieces of evidence than their past states, thereby epistemically improving over time.[1]

Admittedly, there may be some misleading evidence that makes a credal state epistemically worse.[2] In particular, it is entirely possible that one obtains evidence favoring a false hypothesis over a true one. For example, you may get 8 heads in a row in tossing a fair coin 10 times. This evidence obviously favors the false hypothesis that the coin toss is biased, over the true hypothesis that the coin toss is fair—that is, such evidence raises your credence in the false hypothesis to a greater degree than it does your credence in the true one. In this case, the evidence may be said to make our credal state worse.

We don't have to say, though, that evidence has little to do with epistemic progress. By the same token as the above example, it is also intuitive that evidence makes your credal state better if the evidence raises your credences in a true hypothesis to a greater degree than it does your credence in any false one. Having the Jamesian commandment of epistemology "Believe the truth and avoid error!" in mind, we should take it as a constitutive principle of epistemology that your credal state is improved if you obtain evidence favoring some true hypotheses over any false hypothesis (Joyce, 2009).

Notably, Fallis and Lewis have recently provided and discussed some formal versions of such an intuition, which they name 'the Monotonicity Principle' and 'Elimination'.[3] Such a formal principle can be regarded as a bridge connecting two epistemic quantities. The first is a quantity representing the degree to which evidence has an impact on a credence. The more impact evidence has on a credence, the more the evidence raises the credence. Such a quantity, in this paper, will be called the 'evidential parameter'. The second is a quantity representing the *actual* epistemic betterness on the basis of which credal states are epistemically compared with each other at a particular world. Such a quantity may be dubbed the 'epistemic utility'.[4] Fallis and Lewis formulate the evidential parameter

[1]As is well known, this understanding of the reflection principle depends on the assumption that there is no cognitive malfunction like memory loss in between the current and the past credal states.

[2]In this paper, the evaluative concepts of betterness and improvement of a credal state are all epistemic, rather than practical in nature. I will occasionally omit the modifiers like 'epistemic' and 'epistemically' if there is no danger of confusion.

[3]This principle and its variants can be found in Fallis and Lewis (2016). Their other paper considers a weak version of the principle, which is called 'Elimination' (Lewis and Fallis, 2021).

[4]I would like to emphasize that the *actual* epistemic utility (or betterness) should be distinguished from

using a ratio between two credences, and measure the epistemic utility using the concept of accuracy.

With such principles in hand, they raise some interesting criticisms about the so-called 'accuracy-first epistemology'. Accuracy-firsters assume that *accuracy* is the most fundamental epistemic virtue, and other virtues such as verisimilitude are overwhelmed by accuracy.[5] Under this assumption, they identify the epistemic utility with accuracy, and formulate several rules measuring the accuracy of a credal state at a given world. Moreover, accuracy-firsters are to vindicate probabilism and conditionalization using such rules and some decision-theoretic maxims.

Fallis and Lewis argue in their 2016 paper that the Brier rule, which is a very popular rule of accuracy, is not a good measure. In a subsequent paper of 2021, they also argue that accuracy-firsters cannot underwrite both probabilism and conditionalization. At least at first blush, their criticisms appear to be very serious for accuracy-firsters. If their arguments are sound, then accuracy-firsters may lose one of the most popular measures, and may abandon at least one of their main two epistemological projects—that is, vindicating probabilism and conditionalization.

However, I argue in this paper that their criticisms of the accuracy-first epistemology are somewhat hasty. Specifically, I show that there is another version of the Monotonicity Principle that can be satisfied by some additive rules of accuracy, such as the Brier rule, and that Fallis and Lewis's version of the principle has some undesirable features regarding the epistemic betterness. As a result, I conclude that their criticisms can hardly jeopardize accuracy-firsters until any further justification of their versions of the Monotonicity Principle and Elimination is provided. For this purpose, this paper is structured as follows: Section 2 is devoted to formulating Fallis and Lewis's Monotonicity Principle and its weak variant. In particular, I will introduce and explain evidential parameters and

the *expected* epistemic utility (or betterness). The former is the epistemic utility of a credal state *at a particular world*, whereas the second is a kind of *average over those actual epistemic utilities* each of which may be different across possible worlds. It is noteworthy that some works employ expected epistemic utility to investigate the relationship between evidence and the epistemic betterness. For example, see Good (1967) and Myrvold (2012). Thanks to an anonymous reviewer for informing me of such references. My main concern in this paper, however, is the actual epistemic utility, not the expected epistemic utility.

[5] Many works address the accuracy-first approach including Greaves and Wallace (2006), Joyce (2009, 1998), Leitgeb and Pettigrew (2010*a*,*b*) and Pettigrew (2016).

accuracy measures, both of which are connected by the Monotonicity Principle. In Section 3, I will provide several mathematical results related to the Monotonicity Principle. Some have already been given by Fallis and Lewis, but some are new. Especially, it will be shown that there is another version of the Monotonicity Principle that can be satisfied by some additive rules of accuracy like the Brier rule. It will be argued in Section 4 that the law of likelihood, which Fallis and Lewis rely on to justify their version of the principle, yields a problem when we are to epistemically compare one credence function with another function. Furthermore, it will also be shown that Fallis and Lewis's arguments cannot help losing their way if they resort to a weak version of the law in order to circumvent the problem in question.

Before I proceed further, some preliminary remarks are in order. I will assume that our credal state could be identified by a function assigning a credence, which is represented by a real number in $[0, 1]$, to a proposition. Such a function will be called a 'credence function', and denoted by '**c**' , '**s**' , '**r**' , etc. While credence functions are defined over various sets of propositions, I will restrict the discussions to the functions over a *finite partition* whose members are mutually exclusive and collectively exhaustive. I will often use an $n$-tuple to represent a credence function over a partition consisting of $n$ members. Suppose, for instance, that a credence function **c** is defined over a partition $\mathbb{H} = \{H_1, \cdots, H_n\}$. This can be represented as $(c_1, \cdots, c_n)$ such that $\mathbf{c}(H_i) = c_i \in [0, 1]$ for any $H_i \in \mathbb{H}$. If there is no danger of confusion, I will assume, without any explicit explanation about the relevant partition, that $c_i$ is the credence of a hypothesis $H_i$ in $\mathbb{H}$. Similar stipulations apply to other credence functions like **s**, and **r**. Most credence functions that appear in this paper are probabilistically coherent. A credence function **c** over a finite partition is a probabilistically coherent if and only if $c_i \in [0, 1]$ for any $i$, and $\sum_i c_i = 1$.

## 2 Fallis and Lewis's Monotonicity Principle

As stated, Fallis and Lewis's Monotonicity Principle can be thought of as a bridge connecting the evidential parameter and the actual epistemic utility. In this section, I will explain some features of the parameters and measures, and formulate Fallis and Lewis's principle using them.

### 2.1 Evidential Parameters

It seems intuitive that some evidence, which has a stronger impact on the credences in truths than the credences in falsehoods, leads us to epistemic progress. Regarding this, we should pay attention to what factors determine our credences. Suppose that your old credence function is updated to new function after some evidence is obtained. It is well accepted that, in such a case, your new credences are determined by two factors—that is, your old credences and the evidence itself. Evidential parameters represent the impact of evidence itself on a new credence *with its old credence factored out*.[6]

Fallis and Lewis use the ratios of new to old credences to formulate such a parameter. Suppose that two credence functions **s** and **r** are all defined over a partition $\mathbb{H} = \{H_1, \cdots, H_n\}$. Then, the ratio parameter of $H_i (\in \mathbb{H})$ with respect to the credence updating from **s** to **r**, which will be denoted by $\pi_i^{\mathbf{s},\mathbf{r}}$, is defined as follows:

**Ratio Parameter**. $\pi_i^{\mathbf{s},\mathbf{r}} = r_i/s_i$, for any $i$ such that $H_i \in \mathbb{H}$.

Can we take $\pi_i^{\mathbf{s},\mathbf{r}}$ as an adequate parameter representing *only* the impact of the evidence itself? In particular, how can we ascertain that the ratio parameter has such a feature?

One quick way is to look into whether they satisfy what is called 'Commutativity'.[7]

**Commutativity**. Suppose that **s**, **p**, **q**, **r**, and $\mathbf{r}^*$ are all defined over a finite partition $\mathbb{H}$. Suppose also that $\alpha_i^{\mathbf{x},\mathbf{y}}$ is an adequate evidential parameter of $H_i \in \mathbb{H}$ with respect to the credence updating from a credence function **x** to another function **y**. Then, it holds that $\mathbf{r} = \mathbf{r}^*$ if $\alpha_i^{\mathbf{s},\mathbf{q}} = \alpha_i^{\mathbf{p},\mathbf{r}^*}$ and $\alpha_i^{\mathbf{q},\mathbf{r}} = \alpha_i^{\mathbf{s},\mathbf{p}}$ for any $i$.

The following diagram, which is borrowed from Wagner (2002, 2003), may be of help in understanding the plausibility of this adequacy condition.

[6] Some relevant discussions can be found in Field (1978), Jeffrey (2004), and Wagner (2002, 2003). In those works, evidential parameters are often called 'input parameters', 'probabilistic observational reports', 'indices of probability change', and so on.

[7] Commutativity has been mainly discussed in the context related to the credence updating by Jeffrey conditionalization. See Domotor (1980), Döring (1999), Lange (2000), and Wagner (2002, 2003).

$$
\begin{tikzcd}
\mathbf{s} \arrow[r, "\alpha^{\mathbf{s},\mathbf{q}}"] \arrow[dd, "\alpha^{\mathbf{s},\mathbf{p}}"'] & \mathbf{q} \arrow[d, "\alpha^{\mathbf{q},\mathbf{r}}"] \\
 & \mathbf{r} \\
\mathbf{p} \arrow[r, "\alpha^{\mathbf{p},\mathbf{r}^*}"] & \mathbf{r}^*
\end{tikzcd}
$$

This displays two kinds of successive credence updating whose initial credence function is $\mathbf{s}$. On the one hand, $\mathbf{s}$ is first updated to $\mathbf{q}$, and then updated to $\mathbf{r}$. The first updating is led by the evidence whose impact is represented by an evidential parameter $\alpha^{\mathbf{s},\mathbf{q}}$, and the second updating is led by another evidential impact $\alpha^{\mathbf{q},\mathbf{r}}$. On the other hand, the diagram displays another kind of successive belief updating in which $\mathbf{s}$ is first updated to $\mathbf{p}$ under the impact of $\alpha^{\mathbf{s},\mathbf{p}}$, and then updated to $\mathbf{r}^*$ under the impact of $\alpha^{\mathbf{p},\mathbf{r}^*}$.

Then, what happens if the order in which the evidential impacts are incorporated into the credal state is reversed? In particular, could the final credence functions $\mathbf{r}$ and $\mathbf{r}^*$ be different from each other even if $\alpha_i^{\mathbf{s},\mathbf{q}} = \alpha_i^{\mathbf{p},\mathbf{r}^*}$ and $\alpha_i^{\mathbf{q},\mathbf{r}} = \alpha_i^{\mathbf{s},\mathbf{p}}$ for any $i$? According to Commutativity, this cannot be the case if $\alpha^{\mathbf{x},\mathbf{y}}$ is an adequate evidential parameter factoring out the impact of the old credences. As stated, Bayesians accept that a new credence is determined by the evidence itself and its old credence. So, it can be said that the final credence function $\mathbf{r}$ in the above diagram is determined by the evidential parameters $\alpha^{\mathbf{s},\mathbf{q}}$ and $\alpha^{\mathbf{q},\mathbf{r}}$, and the old credence function $\mathbf{s}$. Similarly, $\mathbf{r}^*$ in the diagram can be said to be determined by $\alpha^{\mathbf{s},\mathbf{p}}$, $\alpha^{\mathbf{p},\mathbf{r}^*}$, and $\mathbf{s}$. Thus, it is natural, as Commutativity says, that $\mathbf{r} = \mathbf{r}^*$ when $\alpha_i^{\mathbf{s},\mathbf{q}} = \alpha_i^{\mathbf{p},\mathbf{r}^*}$ and $\alpha_i^{\mathbf{q},\mathbf{r}} = \alpha_i^{\mathbf{s},\mathbf{p}}$ for any $i$.

The ratio parameter $\pi$ satisfies Commutativity.[8] Hence, we cannot conclude, appealing to Commutativity, that $\pi$ is not an adequate evidential parameter. Admittedly, it is not the case that nothing but the ratio parameter satisfies Commutativity. In what follows, I will consider another parameter that also satisfies it.

[8]Here is the proof. For any two credence functions $\mathbf{x}$ and $\mathbf{y}$ defined over a finite partition $\mathbb{H}$, it holds that $y_i = x_i \cdot \pi_i^{\mathbf{x},\mathbf{y}}$. Then, we have that $r_i = q_i \cdot \pi_i^{\mathbf{q},\mathbf{r}} = s_i \cdot \pi_i^{\mathbf{s},\mathbf{q}} \cdot \pi_i^{\mathbf{q},\mathbf{r}}$ and $r_i^* = p_i \cdot \pi_i^{\mathbf{p},\mathbf{r}^*} = s_i \cdot \pi_i^{\mathbf{s},\mathbf{p}} \cdot \pi_i^{\mathbf{p},\mathbf{r}^*}$. Suppose now that $\pi_i^{\mathbf{s},\mathbf{q}} = \pi_i^{\mathbf{p},\mathbf{r}^*}$ and $\pi_i^{\mathbf{q},\mathbf{r}} = \pi_i^{\mathbf{s},\mathbf{p}}$. Then it follows from the above equations that $r_i = r_i^*$. This holds for any member in $\mathbb{H}$. Hence, we can conclude that the ratio parameter satisfies Commutativity. $\square$

## 2.2 Accuracy Measures

Now, let us turn our attention to the rules measuring the accuracy. Following accuracy-firsters, Fallis and Lewis assume that the epistemic utility of a credence function is identified with the accuracy of the function, and consider some rules of measuring such accuracy. In what follows, I will use '$\mathfrak{A}$' , '$\mathfrak{B}^S$' ,'$\mathfrak{S}^A$' etc. to refer to such rules—namely, the accuracy measures.

Let $\mathfrak{A}$ be an accuracy measure. Then, '$\mathfrak{A}_w(\mathbf{c})$' refers to the accuracy of a credence function **c** at a world $w$ relative to $\mathfrak{A}$. (I will often omit clauses like 'relative to $\mathfrak{A}$' in what follows if there is no danger of confusion.) Accuracy is often thought of as something related to a distance between the credence function **c** and the truth function $\mathbf{v}_w$. (Here, '$\mathbf{v}_w$' refers to the truth-function at $w$ such that $\mathbf{v}_w(H_i) = 1$ when $H_i$ is true at $w$, and $\mathbf{v}_w(H_i) = 0$ otherwise.) As noted, I consider in this paper only credence functions defined over a finite partition. So, the accuracy of a credence function depends on what member of such a partition is true. I will use '$\mathfrak{A}_i(\mathbf{c})$' to denote the accuracy of **c** at worlds where $H_i$ is true.

Accuracy-firsters suggest several constraints that should be met by any legitimate accuracy measure, and narrow the class of the measures down using such constraints. For the discussion that follows, we need to take a look at the following constraint.[9]

**Strict Propriety.** Suppose that **r** is a coherent credence function defined over a partition $\mathbb{H}$, and that $\mathfrak{A}$ is a legitimate accuracy measure. Then, for any coherent credence function **c** over $\mathbb{H}$,

$$\sum_i r_i \mathfrak{A}_i(\mathbf{r}) > \sum_i r_i \mathfrak{A}_i(\mathbf{c}).$$

Most accuracy-firsters agree that Strict Propriety is a fundamental constraint on accuracy measures. The accuracy measures that are not strictly proper could make an agent evaluate her own credence function as being worse than the other functions in that the former has a lesser expected epistemic utility than the latter. In such a case, the rational agent may be required to discard her own function and adopt the new one. However, this kind of cre-

[9]The following formulation of Strict Propriety is restricted to coherent credence functions. However, many accuracy-firsters do not impose such a restriction on Strict Propriety. I will revisit this issue in the next section.

dence updating is epistemically implausible since it occurs in the absence of any relevant new evidence. Strict Propriety helps us to avoid such ill-motivated credence updating.[10]

### 2.2.1 Simple Accuracy Measures

Here are some rules of accuracy that satisfy Strict Propriety.

**Simple Accuracy Measures.**

- Simple Brier rule: $\mathfrak{B}_k^S(\mathbf{c}) = 2c_k - \sum_i c_i^2$.
- Simple Spherical rule: $\mathfrak{S}_k^S(\mathbf{c}) = c_k / \sum_i c_i^2$.
- Simple Logarithmic rule: $\mathfrak{L}_k^S(\mathbf{c}) = \ln(c_k)$.

The modifier 'simple' and the superscript '$S$' are added in order to distinguish these rules from another kind of rules that will be introduced below.

We should note that these rules are of little help in vindicating probabilism. This is because the simple accuracy measures are available only to the credence functions over a *partition*. They cannot say anything about the accuracy of the function defined over a set that is not a partition. However, accuracy-firster's vindication of probabilism hinges on the rules that can measure the accuracy of the credence functions defined over various sets—in particular, the so-called $\sigma$-algebra, which is a set that is closed under a (countable) truth-functional combination.

This problem is more serious for the simple logarithmic rule than the others. The above version of Strict Propriety is restricted to coherent credence functions. However, accuracy-firster's vindication requires any legitimate accuracy measure to satisfy a general version that does not have such a restriction. In particular, accuracy-firsters, who are to vindicate probabilism, should make use of the measures satisfying the constraint that is reformulated as follows: When $\mathbf{r}$ is a coherent credence function over a partition $\mathbb{H}$, $\sum_i r_i \mathfrak{U}_i(\mathbf{r}) > \sum_i r_i \mathfrak{U}_i(\mathbf{c})$ for any *coherent* or *incoherent* credence function $\mathbf{c}$ over $\mathbb{H}$. However, if Strict Propriety is reformulated in this way, then the simple logarithmic rule must

[10]Indeed, Weak Propriety, according to which $\sum_i r_i \mathfrak{A}_i(\mathbf{r}) \geq \sum_i r_i \mathfrak{A}_i(\mathbf{c})$ for any coherent credence function $\mathbf{c}$, is sufficient to avoid such ill-motivated updating. However, Weak Propriety entails Strict Propriety under some plausible assumptions. See Campbell-Moore and Levinstein (2021).

violate the constraint, and so be of no use to vindicate probabilism.[11]

### 2.2.2 Additive Accuracy Measures

Then, are there any other accuracy measures that are free from the aforementioned problems? Of course, yes. In vindicating probabilism and conditionalization, several accuracy-firsters like Joyce (2009), Leitgeb and Pettigrew (2010*a*), and Pettigrew (2016) classify the accuracy measures into two types, and provide a way of connecting them. The first type is what may be called a 'local accuracy measure'. This type measures only one particular credence at a given world. Let '$\mathfrak{a}$' be such a measure. Then, '$\mathfrak{a}_w(c_i)$' denotes the accuracy of a credence of $c_i$ at a world $w$.

The second type, which is dubbed a 'global accuracy measure', measures the overall accuracy of credence functions. Some global accuracy measures are generated from their local counterparts.[12] Here is a way of generating a global rule $\mathfrak{A}$ from a local rule $\mathfrak{a}$.

**Additivity.** Suppose that **c** is defined over a set $\mathcal{F}$. Then, a global accuracy of **c** at a world $w$, i.e., $\mathfrak{A}_w(\mathbf{c})$, is generated from a local rule of accuracy $\mathfrak{a}$, as follows:

$$\mathfrak{A}_w(\mathbf{c}) = \sum_{X \in \mathcal{F}} \mathfrak{a}_w(\mathbf{c}(X)).$$

In words, this says that the global accuracy of a credence function is a simple sum of the local accuracies of all credences.[13] This kind of global rule is often called an 'additive (global) accuracy measure'.

What kinds of additive global rule are there? To formulate such global rules, the corresponding local rules should be provided in advance. Here are such local measures:

[11]This point and some relevant discussions can be found in Pettigrew (2022). Consider two credence functions defined over a partition $\{H_1, H_2, H_3\}$: $\mathbf{r} = (1/3, 1/3, 1/3)$ and $\mathbf{c} = (1, 1, 1)$. Note that **r** is coherent but **c** is not. Then, $\sum_i r_i \mathfrak{L}_i^S(\mathbf{r}) = -\ln 3 < 0 = \sum_i r_i \mathfrak{L}_i^S(\mathbf{c})$, which conflicts with the reformulated version of Strict Propriety.

[12]Not all the global measures can be generated from a local measure. The simple spherical and logarithmic rules can be taken as global, but they have no local counterpart from which they are generated.

[13]There is another way of generating the global measure from the local one, which can be called 'Averaging'. This way can be formulated as follows: $\mathfrak{A}_k^*(\mathbf{c}) = (1/N) \sum_i \mathfrak{u}_k(c_i)$. Here, $N$ refers to the size of $\mathbb{H}$. For some relevant discussions, see Joyce (2009) and Carr (2015). The discussion that follows remains the same even if we adopt Averaging, instead of Additivity, as a way of generating the global measure from the local one. This is because, in this context, the rankings based on Additivity are ordinally equivalent to the rankings based on Averaging.

**Local Accuracy Measures.**

- Local Brier Rule: $\mathfrak{b}_w(\mathbf{c}(X)) = -(\mathbf{v}_w(X) - \mathbf{c}(X))^2$.
- Local Spherical Rule: $\mathfrak{s}_w(\mathbf{c}(X)) = \frac{|\mathbf{v}_w(X)-(1-\mathbf{c}(X))|}{\sqrt{\mathbf{c}(X)^2+(1-\mathbf{c}(X))^2}}$.
- Local Logarithmic Rule: $\mathfrak{l}_w(\mathbf{c}(X)) = \ln|\mathbf{v}_w(X) - (1 - \mathbf{c}(X))|$.

Here, $\mathbf{v}_w$ is a truth-function at the world $w$ such that $\mathbf{v}_w(X) = 1$ when $X$ is true at $w$, and $\mathbf{v}_w(X) = 0$ otherwise.

Now we can provide the additive version of each simple measure using Additivity.

**Additive Accuracy Measures.**

- Additive Brier Rule: $\mathfrak{B}^A_w(\mathbf{c}) = -\sum_{X\in\mathcal{F}}(\mathbf{v}_w(X) - \mathbf{c}(X))^2$.
- Additive Spherical Rule: $\mathfrak{S}^A_w(\mathbf{c}) = \sum_{X\in\mathcal{F}} \frac{|\mathbf{v}_w(X)-(1-\mathbf{c}(X))|}{\sqrt{\mathbf{c}(X)^2+(1-\mathbf{c}(X))^2}}$.
- Additive Logarithmic Rule : $\mathfrak{L}^A_w(\mathbf{c}) = \sum_{X\in\mathcal{F}} \ln|\mathbf{v}_w(X) - (1 - \mathbf{c}(X))|$.

Here $\mathcal{F}$ is an arbitrary set of propositions. Similar to the simple rules, the superscript '$A$' is intended to express that each rule is additive.[14]

The additive rules, unlike the simple ones, can measure the accuracy of any credence function whether or not the function is defined over a partition. Thus, such a rule can be said to be free from the problem besetting the simple rule, and thus useful to vindicate probabilism. It is also noteworthy that all of those additive rules satisfy Strict Propriety. In particular, contrary to the simple logarithmic rule, its additive counterpart $\mathfrak{L}^A$ satisfies Strict Propriety even if the constraint is extended to incoherent credence functions.[15]

## 2.3 Fallis and Lewis's Monotonicity Principle

Let me now provide Fallis and Lewis's Monotonicity Principle and its consequences. Here is the principle suggested in their 2016 paper.

[14] In regard to the credence functions over a partition, the additive Brier rule $\mathfrak{B}^A$ is a positive linear transformation of the simple Brier rule $\mathfrak{B}^S$. It is not hard to see this. Note that $\mathfrak{B}^A_k(\mathbf{c}) = -(1-c_k)^2 + (c_k)^2 - \sum_i (c_i)^2 = 1 + \mathfrak{B}^S_k(\mathbf{c})$. However, such a relationship between an additive rule and its simple counterpart dose not hold for the spherical and logarithmic rules.

[15] Revisit the example in the previous section: $\mathbf{r} = (1/3, 1/3, 1/3)$ and $\mathbf{c} = (1, 1, 1)$. Note that $\sum_i r_i \mathfrak{L}^A_i(\mathbf{r})$ is negatively finite while $\sum_i r_i \mathfrak{L}^A_i(\mathbf{c})$ is negatively infinite, and so that $\sum_i r_i \mathfrak{L}^A_i(\mathbf{r}) > \sum_i r_i \mathfrak{L}^A_i(\mathbf{c})$. (Recall that $\sum_i r_i \mathfrak{L}^S_i(\mathbf{r}) = -\ln 3 < 0 = \sum_i r_i \mathfrak{L}^S_i(\mathbf{c})$.)

**Monotonicity$^\pi$.** Suppose that **s** and **r** are credence functions over a finite partition $\mathbb{H}$.

Then, for any $k$, it holds that: $\mathfrak{A}_k(\mathbf{s}) \le \mathfrak{A}_k(\mathbf{r})$ if $\pi_i^{\mathbf{s},\mathbf{r}} \le \pi_k^{\mathbf{s},\mathbf{r}}$ for any $i$.

The superscript '$\pi$' is added to express that this principle is formulated with the ratio parameter $\pi$.

Suppose that the credence function is updated from **s** to **r** after some evidence is obtained. Suppose also that the ratio parameter $\pi_i^{\mathbf{s},\mathbf{r}}$ adequately represents the impact of the evidence itself. Consider a world where $H_k$ is true. The clause "$\pi_i^{\mathbf{s},\mathbf{r}} \le \pi_k^{\mathbf{s},\mathbf{r}}$ for any $i$" means that the evidence has an impact on the credence in truth at least as strongly as it does the credence in any other falsehoods. In this case, it is natural that the relevant credal state does not get epistemically worse at that world, as the principle says.

In addition, Lewis and Fallis (2021) consider a weak version of the principle.[16]

**Elimination$^\pi$**. Suppose that **s** and **r** are credence functions over a finite partition $\mathbb{H}$. Then,

it holds that: $\mathfrak{A}_k(\mathbf{s}) < \mathfrak{A}_k(\mathbf{r})$ for any $k\ (\neq n)$ if $s_n > 0 = r_n$ and $s_i/s_j = r_i/r_j$ for any $i, j\ (\neq n)$ .

Suppose that **s** assigns a positive credence to a false hypothesis $H_n$ whereas **r** assigns a zero credence to that hypothesis.[17] Then, it seems natural that **r** should not be evaluated as being epistemically worse than **s**, *other things being the same as each other*. In this regard, Fallis and Lewis take it that the proviso 'other things being the same as each other' is well met by the condition that the credences in the remaining propositions stay in *the same ratio*—that is, $s_i/s_j = r_i/r_j$ for any $i, j\ (\neq n)$, and so $\pi_i^{\mathbf{s},\mathbf{r}} = \pi_j^{\mathbf{s},\mathbf{r}}$ for any $i, j\ (\neq n)$. (I will revisit this issue in the following section.)

As said, the Monotonicity Principle may be regarded as a bridge principle connecting the evidential parameters and the actual epistemic utilities. Similarly, Elimination$^\pi$ may be thought of as a bridge principle connecting conditionalization and the actual epistemic

[16] Fallis and Lewis (2016, 582-583) say, without any explicit proof, that Monotonicity$^\pi$ entails Elimination$^\pi$. M3 and M4 appearing in that paper correspond to Monotonicity$^\pi$ and Elimination$^\pi$, respectively. However, for the proof at issue, we need an assumption entailing that, for any $k (\neq n)$, it holds that $\mathfrak{A}_k(\mathbf{s}) \neq \mathfrak{A}_k(\mathbf{r})$ if $s_n > r_n$ and $\pi_i^{\mathbf{s},\mathbf{r}} = \pi_k^{\mathbf{s},\mathbf{r}}$ for any $i (\neq n)$. Without such an assumption, we can prove just a weak version of Elimination$^\pi$—that is, $\mathfrak{A}_k(\mathbf{s}) \le \mathfrak{A}_k(\mathbf{r})$ for any $k\ (\neq n)$ if $s_n > 0 = r_n$ and $s_i/s_j = r_i/r_j$ for any $i, j\ (\neq n)$ . I think the assumption in question is very plausible although Fallis and Lewis do not explicitly mention it.

[17] In other words, suppose that evidence *eliminates* the epistemic possibility that $H_n$ is true, and so leads to assign a zero credence to that hypothesis. This is why Fallis and Lewis call this version of the monotoncity principle 'Elimination'.

utilities.[18] Conditionalization is typically defined as follows:

**Conditionalization.** A credence function $\mathbf{r}$ is updated from another credence function $\mathbf{s}$ in accordance with conditionalization on $E$ if and only if $\mathbf{r}(\cdot) = \mathbf{s}(\cdot|E)$ when $\mathbf{s}(\cdot|E)$ is well defined.

Suppose that $\mathbf{s}$ and $\mathbf{r}$ are credence functions over a finite partition $\mathbb{H}$. Then, it holds that $s_n > 0 = r_n$ and $s_i/s_j = r_i/r_j$ for any $i, j$ $(\neq n)$ if $\mathbf{r}$ is updated from $\mathbf{s}$ in accordance with conditionalization on the evidence that $H_n$ is false. Thus, Elimination$^\pi$ can be thought of as a claim that our credal states are epistemically improved if the states are updated in accordance with conditionalization on evidence. So, if Elimination$^\pi$ is a plausible constraint on accuracy measures and conditionalization is a rational credence updating rule, then the accuracy measures should guarantee that conditionalization epistemically improves our actual credal sates. For this reason, Fallis and Lewis require that accuracy-firsters who want to ground conditionalization should use the accuracy measures satisfying Elimination$^\pi$.

# 3 Accuracy-firsters in Crisis and Evidential Parameters

Fallis and Lewis argue in their 2016 and 2021 papers that nothing but $\mathfrak{S}^S$ and $\mathfrak{L}^S$ of the aforementioned rules satisfies Monotonicity$^\pi$ (and Elimination$^\pi$)—in particular, none of the additive measures under consideration satisfies it. On the basis of these results, they conclude that the Brier rule is not a good measure of accuracy, and that the additive accuracy measures, which are needed to vindicate probabilism, are not adequate to vindicate conditionalization, and so accuracy-firster cannot underwrite both probabilism and conditionalization. These conclusions seem to be very serious for accuracy-firsters. If these conclusions are correct, accuracy-firsters may lose one of the most popular accuracy measures, and may abandon at least one of their main two epistemological projects—that is, vindicating both probabilism and conditionalization.

[18]Note that Elimination$^\pi$ can be thought of as connecting conditionalization with the *actual* epistemic utilities, not the expected epistemic utilities. See footnote 4. Indeed, there are many attempts to elucidate the relationship betwen conditionalization and the expected epistemic utilities. One of the most relevant works is given by Greaves and Wallace (2006).

Then, are there any ways of rescuing accuracy-firsters from Fallis and Lewis's criticisms? In this regard, I would like to emphasize that the epistemic plausibility of Monotonicity$^\pi$ (and Elimination$^\pi$) depends on how to formulate the degree to which evidence raises credences—in other words, how to represent the impact of evidence itself with old credences factored out. Fallis and Lewis use the ratio parameter to represent such an impact.

Interestingly, they also explore the possibility that the difference between two credences may play a role as an evidential parameter (Fallis and Lewis, 2016, 585). The difference parameter of $H_i (\in \mathbb{H} = \{H_1, \cdots, H_n\})$ with respect to the credence updating from **s** to **r**, which will be denoted by $\delta_i^{\mathbf{s},\mathbf{r}}$, is defined as follows:

**Difference Parameter**. $\delta_i^{\mathbf{s},\mathbf{r}} = r_i - s_i$, for any $i$ such that $H_i \in \mathbb{H}$.

By replacing the ratio parameter in Monotonicity$^\pi$ and Elimination$^\pi$ with the difference parameter, we obtain alternative formulations of the Monotonicity Principle and Elimination.

**Monotonicity$^\delta$**. Suppose that **s** and **r** are credence functions over a finite partition $\mathbb{H}$. Then, it holds that $\mathfrak{A}_k(\mathbf{s}) \leq \mathfrak{A}_k(\mathbf{r})$ if $\delta_i^{\mathbf{s},\mathbf{r}} \leq \delta_k^{\mathbf{s},\mathbf{r}}$ for any $i$.

**Elimination$^\delta$**. Suppose that **s** and **r** are credence functions over a finite partition $\mathbb{H}$. Then, it holds that $\mathfrak{A}_k(\mathbf{s}) \leq \mathfrak{A}_k(\mathbf{r})$ for any $k$ $(\neq n)$ if $s_n > 0 = r_n$ and $s_i - s_j = r_i - r_j$ for any $i, j$ $(\neq n)$.

Here, the superscript '$\delta$' is added to express that this principle is formulated with the difference parameter $\delta$.[19]

Which rules of accuracy satisfy this version of the Monotonicity Priniciple? If no additive rule of accuracy satisfies it, then taking another evidential parameter into account may be of little help in rescuing accuracy-firsters from Fallis and Lewis's criticisms. Unfortunately, Fallis and Lewis seem to make a mistake regarding this point. In particular,

[19] Similar to the relationship between Monotonicity$^\pi$ and Elimination$^\pi$, Monotonicity$^\delta$ entails Elimination$^\delta$ under the assumption that, for any $k (\neq n)$, it holds that $\mathfrak{A}_k(\mathbf{s}) \neq \mathfrak{A}_k(\mathbf{r})$ if $s_n > r_n$ and $\delta_i^{\mathbf{s},\mathbf{r}} = \delta_k^{\mathbf{s},\mathbf{r}}$ for any $i (\neq n)$. See footnote 16.

their example, which is intended to show that the simple (and additive) Brier rule violates Monotonicty$^{\delta}$, misses the target. They write,

> "Unfortunately though, these other [evidential parameters] do not vindicate all of the verdicts of the Brier rule. For instance, the Brier rule says that $\mathbf{r} = (3/7, 4/7, 0)$ is epistemically better than $\mathbf{s} = (3/10, 4/10, 3/10)$ if $H_1$ is true even though $r_2 - s_2 > r_1 - s_1$." (Fallis and Lewis (2016, 585). My brackets.)

The calculations in their example are all correct. Nevertheless, the two credence functions **r** and **s** in the example cannot be taken as functions showing that the Brier rule violates Monotonicity$^{\delta}$. Rather, they are just an example showing that the rule violates the *converse* Monotonicity$^{\delta}$—$\mathfrak{A}_1(\mathbf{s}) \leq \mathfrak{A}_1(\mathbf{r})$ only if $\delta_i^{\mathbf{s},\mathbf{r}} \leq \delta_1^{\mathbf{s},\mathbf{r}}$ for any $i$.[20]

In this regard, someone may argue that the converse Monotonicity Principle is another compelling constraint governing the relationship between the evidential parameter and the actual accuracy, and thus that the Brier rule still cannot be a good measure. However, this response is just pointless. It is entirely unclear whether our credal state cannot be improved if evidence raises the credence in even just one falsehood to a greater degree than the credence in a truth. To see this, suppose that $H_1$ is true, and that $\mathbf{s} = (3/12, 3/12, 3/12, 3/12)$ is updated to $\mathbf{r} = (4/12, 2/12, 5/12, 0)$ after some evidence is obtained. While the evidence raises the credence in the false hypothesis $H_3$ to only a greater degree than the credence in the true hypothesis $H_1$, the other credences are all farther from falsehoods than before. In this case, can we say that **r** is not epistemically better than **s**? It is not clear.

Moreover, we can prove, contrary to what they attempted to argue, that several additive rules including the Brier rule satisfy Monotonicity$^{\delta}$. Table 1 summarizes the relevant results.[21] With these results in hand, we can arrive at the different conclusions from Fallis

[20]The additive Brier rule is a positive linear transformation of the simple Brier rule. Note that, in order for **s** and **r** in the above quotation to be taken as an example to show that the simple (and additive) Brier rule violates Monotonicity$^{\delta}$, it should be shown that $\mathfrak{B}_2^S(\mathbf{s})$ is greater than $\mathfrak{B}_2^S(\mathbf{r})$, but it is not the case since $\mathfrak{B}_2^S(\mathbf{s}) = 0.460 < 0.632 = \mathfrak{B}_2^S(\mathbf{r})$.

[21]Table 1 includes the results given in Fallis and Lewis (2016) and Lewis and Fallis (2021). 'OK' means that the relevant pair of an accuracy measure and an evidential parameter satisfies the Monotonicity Principle, and 'NG' means that such a pair violates the principle. The relevant proofs and comments are given in Appendix.

| | Simple Rules | | | Additive Rules | | |
|---|---|---|---|---|---|---|
| | $\mathfrak{B}^S$ | $\mathfrak{S}^S$ | $\mathfrak{L}^S$ | $\mathfrak{B}^A$ | $\mathfrak{S}^A$ | $\mathfrak{L}^A$ |
| $\delta$ | OK | NG | OK | OK | NG | OK |
| $\pi$ | NG | OK | OK | NG | NG | NG |

Table 1: Monotonicity and Accuracy Measures

and Lewis's, assuming that *the difference parameter, rather than the ratio parameter, is an adequate evidential parameter*. In particular, we can conclude, under the assumption, that only the simple and additive *spherical rules* of the aforementioned rules violate the Monotonicity Principle, and so such versions of spherical rule are not a good accuracy measure. On the other hand, the condition '$s_n > 0 = r_n$ and $s_i - s_j = r_i - r_j$ for any $i$, $j$ $(\neq n)$' in Elimination$^\delta$ is not equivalent to **s** being updated to **r** in accordance with conditionalization on the evidence that $H_n$ is false. So, under the assumption at issue, we don't have to require accuracy measures to guarantee that conditionalization epistemically improves our credal states. Furthermore, the assumption at issue prevents us from arriving at the conclusion that accuracy-firsters cannot underwrite both the Monotonicity Principle and probabilism.[22] This is because, as shown in Table 1, some additive measures like $\mathfrak{B}^A$ and $\mathfrak{L}^A$ satisfy Monotonicity$^\delta$.

As a result, Fallis and Lewis's criticisms may lose their cogency unless it is justified that the ratio parameter, rather than the difference parameter, is an adequate evidential parameter. In this regard, we first need to note that, like the ratio parameter, the difference parameter also satisfies Commutativity, which any adequate evidential parameter should satisfy in order to represent only the impact of evidence itself with old credences factored out.[23] Thus, the difference parameter can be regarded as being at least as adequate as the ratio parameter.[24]

[22] Note that Fallis and Lewis's claim can be thought of as the claim that accuracy-firsters cannot underwrite both the Monotonicity Principle and probabilism.

[23] Here is the proof. For any two credence functions **x** and **y** defined over a finite partition $\mathbb{H}$, it holds that $y_i = x_i + \delta_i^{\mathbf{x},\mathbf{y}}$. Then, we have that $r_i = q_i + \delta_i^{\mathbf{q},\mathbf{r}} = s_i + \delta_i^{\mathbf{s},\mathbf{q}} + \delta_i^{\mathbf{q},\mathbf{r}}$ and $r_i^* = p_i + \delta_i^{\mathbf{p},\mathbf{r}^*} = s_i + \delta_i^{\mathbf{s},\mathbf{p}} + \pi_i^{\mathbf{p},\mathbf{r}^*}$. Suppose now that $\delta_i^{\mathbf{s},\mathbf{q}} = \delta_i^{\mathbf{p},\mathbf{r}^*}$ and $\delta_i^{\mathbf{q},\mathbf{r}} = \delta_i^{\mathbf{s},\mathbf{p}}$. Thus, it follows from the above equations that $r_i = r_i^*$. This holds for any member in $\mathbb{H}$. Therefore, the difference parameter satisfies Commutativity. □

[24] A caveat needs to be stated. I just say here that Commutativity is a necessary, but not sufficient, condition for an evidential parameter to represent only the impact of evidence itself. Other conditions have also been suggested and utilized to epistemically compare various evidence parameters. See Jeffrey (2004)

Admittedly, some readers may argue that there being an alternative formulation of the Monotonicity Principle, which is satisfied by some additive rules, is not sufficient to undermine Fallis and Lewis's criticisms. In particular, such readers might say that any plausible response to them should disclose some drawbacks of their version of the Monotonicity Principle.[25] I agree. And I will show in the next section that their Monotonicity$^\pi$ has some undesirable features regarding the epistemic betterness.

For this purpose, I will pay attention to the so-called *law of likelihood*. By appealing to the law, Fallis and Lewis (2016) argue that conditionalization leads us to the ratio parameter and Monotonicity$^\pi$. If this argument is sound, then it can be said that accuracy-firsters, who aim to vindicate conditionalization, should advocate the ratio parameter instead of the difference parameter, and thus Fallis and Lewis's criticisms about accuracy-first epistemology remain intact, regardless of the results in Table 1. However, I will argue in the next section that the law of likelihood renders their Monotonicity$^\pi$ counterintuitive regarding the epistemic betterness. Moreover, it will also be argued that if we accept a weak version of the law to circumvent this difficulty, then conditionalization cannot be used to vindicate the ratio parameter and Fallis and Lewis's criticisms do not go through.

# 4 The Law of Likelihood and Accuracy

Let me start with formulating a version of the law of likelihood:[26]

**Law of Likelihood**. Evidence $E$ supports a hypothesis $H_i$ at least as much as it does another hypothesis $H_j$ relative to a credence function $\mathbf{s}$ if and only if $\mathbf{s}(E|H_i) \geq \mathbf{s}(E|H_j)$.

Note that $\mathbf{s}(E|H_i) \geq \mathbf{s}(E|H_j)$ if and only if $\mathbf{s}(H_i|E)/\mathbf{s}(H_i) \geq \mathbf{s}(H_j|E)/\mathbf{s}(H_j)$. And, suppose

and Wagner (2002, 2003), for example. Using such conditions, those authors advocate what is known as the Bayes factor. The Bayes factor of a proposition $H_i$ against another proposition $H_j$ with respect to the credence updating from $\mathbf{s}$ to $\mathbf{r}$, which may be denoted by $\beta^{\mathbf{s},\mathbf{r}}_{i,j}$, is defined as follows: $\beta^{\mathbf{s},\mathbf{r}}_{i,j} = (r_i/r_j)/(s_i/s_j)$. (Here, $r_i$ and $r_j$ are the new credence in $H_i$ and in $H_j$, respectively. These two propositions $H_i$ and $H_j$ are members of a partition $\mathbb{H}$. Similarly, $s_i$ and $s_j$ are the old credence in $H_i$ and in $H_j$, respectively.) In words, the Bayes factor of $H_i$ against $H_j$ is the ratio of new-to-old odds, not probabilities. Thanks to an anonymous reviewer for helping me clarify this point.

[25] I am grateful to an anonymous reviewer for helping me clarify this point.

[26] Some formulations of the law and the relevant discussions can be found in Hacking (2016), Royall (1997), and Sober (2008), for example.

that **s** is updated to **r** in accordance with conditionalization when evidence $E$ is obtained. Then, we can say, appealing to the law of likelihood, that $E$ raises the credence in $H_i$ at least as much as it does the credence in $H_j$ if and only if $\mathbf{r}(H_i)/\mathbf{s}(H_i) \geq \mathbf{r}(H_j)/\mathbf{s}(H_j)$. In this manner, the law and conditionalization jointly provide a rationale to adopt the ratio between two credences as an adequate evidential parameter, and so espouse Monotonicity$^{\pi}$.

However, the law of likelihood is not beyond all doubt. Especially, what is called 'the problem of irrelevant conjunction'—also known as 'the tacking problem'—has been regarded as one of the most serious problems of the law. Suppose that a hypothesis $H$ entails evidence $E$, and so that the likelihood of $H$ on $E$ should be 1. Then, the law of likelihood says that, for any proposition $X$, $E$ supports the conjunction $H\&X$ to the same degree to which it does $H$. This is because $H\&X$ also entails $E$, and thus the likelihood of $H\&X$ on $E$ is the same as the likelihood of $H$ on $E$. However, this is somewhat counterintuitive. The more information a proposition contains, the harder the proposition is confirmed. Note that $H\&X$ may contain much more information than $H$ (unless $H$ entails $X$). Thus, the degree to which $E$ supports $H\&X$ may be less than the degree to which it supports $H$.[27]

This undesirable feature of the law of likelihood seems to jeopardize Monotonicity$^{\pi}$ as well. To illustrate, consider a well-shuffled deck of cards consisting of 52 regular cards. Someone has drawn a card from the deck at random, and you know this. Let $H$ be the proposition that the card is a heart, $D$ be the proposition that the card is a diamond, and $B$ be the proposition that the card is black. In addition, let $X$ be an arbitrary contingent proposition that is not entailed by $H$. The evidence that the drawn card is not black is entirely uninformative between a heart and a diamond drawn—namely, between $H$ and $D$. Thus, it can be said that the evidence that $B$ is false supports $H$ as much as it supports $D$. On the other hand, it can also be said that the evidence in question supports $H\&X$ to a lesser degree than it supports $H$. This is because $H\&X$ is more informative than $H$. As a

[27]The problem of irrelevant conjunction is related to how to formulate the degree to which evidence incrementally supports a hypothesis—i.e., the degree of confirmation. In particular, the problem besets the attempt to formulate the degree of confirmation by means of $P(H|E)/P(H)$. (Here $P$ is a probability function.) For the problem of irrelevant conjunction and the law of likelihood, see Earman (1992), Fitelson (1999, 2007), Steel (2007), and Rosenkrantz (1994), for example.

result, we can say that the evidence that $B$ is false raises the credence in $H\&X$ to a lesser degree than it raises the credence in $D$.[28]

To make our discussions concrete, let me consider a case in which you have a lottery ticket. Let $X$ be the proposition that your ticket is a winning ticket. The lottery in question will be canceled when a heart card is not drawn, and you know this. So, you currently have credences in only the members of the partition $\{H\&X, H\&\neg X, D, B\}$. In other words, your current credence function **s** is defined over that partition. After a while, you obtain the evidence that $B$ is false and nothing else, and update your credence function from **s** to **r** using conditionalization on $\neg B$. Note that $H\&X$ is much more informative than $D$. In light of the above consideration, thus, we can say that the evidence raises the credence in $H\&X$ to a lesser degree than it raises the credence in $D$.

Consider now a world where $H\&X$ is true. Can we say in the above case that **r** is at least as good as **s** at that world? According to our intuition motivating the Monotonicity Principle, if the credence in a truth increases less than the credence in *some* falsehoods, then we cannot say *without any qualification* that the new credence function is at least as good as its old function. To make our intuition clear, let me assume that the chance of your ticket being the winning ticket is extremely low (and you know this). Then, it can be said that, after the evidence is obtained, the credence in the true proposition increases less than the credence in *every* false proposition that is a member of the partition and is compatible with the evidence.[29] In this case, can we say conclusively that **r** is at least as good as **s** at the world where $H\&X$ is true? It seems not.

As seen above, the verdict about the epistemic betterness between **s** and **r** may vary depending on how informative the proposition $X$ is. If $X$ is highly informative and so your credence in the true proposition $H\&X$ increases much less than the credence in the false propositions, then **r** may be worse than **s** at the world where $H\&X$ is true. On the other hand, if $X$ is only slightly more informative than a tautology, then **r** could be better than

[28] Note that, contrary to this consideration, the law of likelihood says that $\neg B$ raises the credence in $H\&X$ as much as it raises the credence in $D$. This is because each of the two propositions entails $\neg B$.

[29] Suppose that the chance of $X$ is extremely low and so the chance of $\neg X$ is extremely high. Then, while $H\&X$ is much more informative than $H$, $H\&\neg X$ is only slightly more informative than $H$. So, it can be said that the evidence $\neg B$ raises the credence in $H\&X$ to a lesser degree than it raises the credence in $H\&\neg X$ as well as $D$. Note that $H\&\neg X$ and $D$ are false at the world where $H\&X$ is true.

**s** at the world where $H\&X$ is true. Any rule of accuracy cannot be said to be legitimate if the epistemic comparison using the rule is *insensitive to the informativeness at issue*.

Are the rules of accuracy, which satisfy Monotonicity$^{\pi}$, sensitive to the informativeness of $X$? Unfortunately, they cannot be sensitive when **s** is updated to **r** in accordance with conditionalization. Such rules say, regardless of the informativeness of $X$, that **r** is at least as good as **s** at the world where $H\&X$ is true.[30]

What about the Brier rule that satisfies Monotonicity$^{\delta}$? Interestingly, we can find that the rule is sensitive to the informativeness in question. Suppose that **s** is updated to **r** in accordance with conditionalization on $\neg B$. Suppose also that $\mathbf{s}(H\&X) = x/4$, $\mathbf{s}(H\&\neg X) = (1-x)/4$, $\mathbf{s}(D) = 1/4$, and $\mathbf{s}(B) = 1/2$—that is, $\mathbf{s} = (x/4, (1-x)/4, 1/4, 1/2)$.[31] The value of $x$ may vary depending on the informativeness of the proposition $X$. In particular, it may be said that the more informative $X$ is, the less the value of $x$ is.

Now, we can find that the additive Brier rule $\mathfrak{B}^A$ produces different verdicts about the epistemic betterness between **s** and **r**, depending on the value of $x$. Consider a case where the chance of your ticket being a winning ticket is significantly low (say 1/100), and thus $x = 1/100$. In this case, the Brier rule says that $\mathbf{s} = (1/400, 99/400, 1/4, 1/2)$ is better than $\mathbf{r} = (1/200, 99/200, 1/2, 0)$ at the world $w$—that is, $\mathfrak{B}^A_w(\mathbf{s}) = -1.369 > -1.485 = \mathfrak{B}^A_w(\mathbf{r})$. (Here and below, '$w$' refers to the world where $H\&X$ is true.) Consider another case where the chance at issue is not significantly low (say 1/2), and thus $x = 1/2$—that is, $\mathbf{s} = (1/8, 1/8, 1/4, 1/2)$ and $\mathbf{r} = (1/4, 1/4, 1/2, 0)$. In this case, the rule says, contrary to the above case, that **r** is better than **s** at the world $w$—that is, $\mathfrak{B}^A_w(\mathbf{r}) = -0.875 > -1.094 = \mathfrak{B}^A_w(\mathbf{s})$. Moreover, with the help of a bit of mathematics, we have that $\mathfrak{B}^A_w(\mathbf{s}) > \mathfrak{B}^A_w(\mathbf{r})$ when $0 \leq x < 0.152$. The Brier rule says that your credence function may be epistemically worse off than before if the true propositions are too informative.

[30] Suppose that **s** is updated to **r** in accordance with conditionalization on $\neg B$. Then, for any contingent proposition $X$, $\mathbf{r}(H\&X)/\mathbf{s}(H\&X) \geq \mathbf{r}(Y)/\mathbf{s}(Y)$ for any proposition $Y$ in the partition at issue. Thus, if an accuracy measure satisfies Monotonicity$^{\pi}$, then the measure says, regardless of the informativeness at issue, that **r** is at least as good as **s** at the world where $H\&X$ is true.

[31] Here, I assume that $X$ is probabilistically independent of $H$. Someone might think that the credence function **s** is defined over the partition $\{H\&X, H\&\neg X, D, B\}$, and therefore we cannot use **s** to formulate the probabilisitic independence between $H$ and $X$. However, there is no need to worry about this. We can formulate the independence in question using some probabilistically coherent extensions of **s**, which are defined over a $\sigma$-algebra that includes $H$ and $X$ as its members.

This consideration seems to cast doubt on the plausibility of Monotonicity$^{\pi}$, and provide a reason to espouse Monotonicity$^{\delta}$ rather than Monotonicity$^{\pi}$. How can the proponents of Monotonicity$^{\pi}$ like Fallis and Lewis cope with this problem? They may appeal to a weak version of the law of likelihood, which can be formulated as follows.

**Weak Law of Likelihood**. Evidence $E$ supports a hypothesis $H_i$ at least as much as it does another hypothesis $H_j$ relative to a credence function $\mathbf{s}$ if $\mathbf{s}(E|H_i) \geq \mathbf{s}(E|H_j)$ and $\mathbf{s}(\neg E|\neg H_i) \geq \mathbf{s}(\neg E|\neg H_j)$.

Note that the weak law of likelihood is free from the problem of irrelevant conjunction, and so many Bayesians prefer this weak version to the original law.[32]

Can Fallis and Lewis's arguments against accuracy-first epistemology still go through when the Monotonicity Principle is reformulated by appealing to this weak law of likelihood? Unfortunately, not. To see this, let me first assume that $\mathbf{s}$ is updated to $\mathbf{r}$ in accordance with conditionalization when evidence $E$ is obtained. Note also that $\mathbf{s}(\neg E|\neg H_i) \geq \mathbf{s}(\neg E|\neg H_j)$ if and only if $\mathbf{s}(\neg H_i|E)/\mathbf{s}(\neg H_i) \leq \mathbf{s}(\neg H_j|E)/\mathbf{s}(\neg H_j)$. So, we can say, with the help of the weak law of likelihood, that $E$ raises the credence in $H_i$ at least as much as it does the credence in $H_j$ if $\mathbf{r}(H_i)/\mathbf{s}(H_i) \geq \mathbf{r}(H_j)/\mathbf{s}(H_j)$ and $\mathbf{r}(\neg H_i)/\mathbf{s}(\neg H_i) \leq \mathbf{r}(\neg H_j)/\mathbf{s}(\neg H_j)$. As mentioned, the law of likelihood and conditionalization jointly provide a rationale to espouse Monotonicity$^{\pi}$. In a similar vein, it can be said that the weak law of likelihood and conditionalization jointly undergird the following weak version of Monotonicity$^{\pi}$.

**Weak Monotonicity.** Suppose that $\mathbf{s}$ and $\mathbf{r}$ are credence functions over a finite partition $\mathbb{H}$. Then, it holds that $\mathfrak{A}_k(\mathbf{s}) \leq \mathfrak{A}_k(\mathbf{r})$ if $\pi_i^{\mathbf{s},\mathbf{r}} \leq \pi_k^{\mathbf{s},\mathbf{r}}$ and $\bar{\pi}_k^{\mathbf{s},\mathbf{r}} \leq \bar{\pi}_i^{\mathbf{s},\mathbf{r}}$ for any $i$.

Here, '$\bar{\pi}_i^{\mathbf{s},\mathbf{r}}$' denotes $\mathbf{r}(\neg H_i)/\mathbf{s}(\neg H_i)$.

Interestingly, Monotonicity$^{\delta}$ also entails this weak version of Monotonicity$^{\pi}$, but the converse does not hold.[33] So, every accuracy rule that satisfies at least one of Monotonicity$^{\pi}$ and Monotonicity$^{\delta}$ also satisfies Weak Monotonicity. (This is a reason why I name this version 'Weak Monotonicity' without attaching any superscript $\pi$ and $\delta$.) Then, it can

[32]For example, see Fitelson (2007) and Joyce (2021). The name 'weak law of likelihood' is borrowed from Fitelson (2007). Joyce (2021) calls it the 'weak likelihood principle'.

[33]A proof is given in Appendix II.

be concluded, with the help of Table 1 in Section 3, that there are several additive rules, including the additive Brier rule, that satisfy Weak Monotonicity.[34]

Moreover, it can be proved that all of the strictly proper rules of accuracy satisfy the following weak version of Elimination, which corresponds to Weak Monotonicity.

**Weak Elimination.** Suppose that **s** and **r** are credence functions over a finite partition $\mathbb{H}$. Then, it holds that $\mathfrak{A}_k(\mathbf{s}) < \mathfrak{A}_k(\mathbf{r})$ for any $k$ $(\neq n)$ if $s_n > 0 = r_n$, and $s_i/s_j = r_i/r_j$ and $(1-s_i)/(1-s_j) = (1-r_i)/(1-r_j)$ for any $i, j$ $(\neq n)$.

Notably, the condition that $s_i/s_j = r_i/r_j$ and $(1-s_i)/(1-s_j) = (1-r_i)/(1-r_j)$ for any $i, j$ $(\neq n)$ entails that $s_i = s_j$ for any $i, j$ $(\neq n)$. That is, the condition in question cannot be met unless the credences in question are evenly distributed over all of the hypotheses except for $H_n$. And, it follows from this fact that all of the strictly proper rules satisfy Weak Elimination.[35]

To sum up, conditionalization does not lead us to the ratio parameter and Monotononicity$^\pi$ when the weak version of the law of likelihood is assumed. Instead, conditionalization and the weak law jointly lead us to espouse Weak Monotonicity and Weak Elimination, which undermine Fallis and Lewis's criticisms about accuracy-first epistemology. In other words, accuracy-firsters can rescue themselves from the criticisms by invoking the weak law of likelihood.

# 5 Conclusions

The above considerations obviously suggest some ways out for accuracy-first epistemology. We can address Fallis and Lewis's criticisms by reformulating the condition for when

[34]How about the additive spherical rule $\mathfrak{S}^A$? Example A.2 in Appendix, which shows $\mathfrak{S}^A$ violates Monotonicity$^\delta$, does not prove that the rule also violates Weak Monotonicity. This is because the credence functions **s** and **r** in that example violate the condition that $\bar{\pi}_k^{\mathbf{s},\mathbf{r}} \leq \bar{\pi}_i^{\mathbf{s},\mathbf{r}}$ for any $i$. Similarly, Fallis and Lewis's example, which is used to show that $\mathfrak{S}^A$ violates Monotonicity$^\pi$, cannot show that $\mathfrak{S}^A$ violates Weak Monotonicity. (In that example, $\mathbf{s} = (1/7, 3/7, 3/7)$ and $\mathbf{r} = (1/4, 3/4, 0)$.) Admittedly, if it can be proved in this paper whether $\mathfrak{S}^A$ satisfies Weak Monotonicity or not, my discussions will be more complete. However, I will not prove such a thing. This is because it is proved here that Weak Elimination is satisfied by all of the strictly proper accuracy measures, and the conclusions in this paper do not depend on whether $\mathfrak{S}^A$ satisfies Weak Monotonicity.

[35]Some proofs related to the discussion in this paragraph are given in Appendix II.

evidence favors one hypothesis over another. The first reformulation is to use the difference parameter. The second reformulation is to appeal to the weak law of likelihood. Both ways out lead us to the conclusion that some additive rules, including the Brier rule, satisfy the Monotonicity Principle. In addition, the second way out makes us conclude that all of the strictly proper rules satisfy Elimination. In light of these considerations, we can say that it is somewhat hasty to conclude that the Brier rule is not a good measure of epistemic utility (Fallis and Lewis, 2016), and that Accuracy-firsters cannot underwrite both probabilism and conditionalization (Lewis and Fallis, 2021). Fallis and Lewis's criticisms can hardly be taken as serious threats to accuracy-firsters until it is explained why Monotonicity$^\pi$, rather than Monotonicity$^\delta$ and/or Weak Monotonicity, should be expoused even if Monotonicity$^\pi$ suffers from a problem similar to what is known as the problem of irrelevant conjunction.

## Appendix I: Proofs of the Results in Table 1

'OK' in Table 1 means that the corresponding accuracy measure satisfies the relevant version of the Monotonicty Principle. On the other hand, 'NG' in the table means that the corresponding measure violates the principle in question. Some proofs have already been provided in Fallis and Lewis's papers of 2016 and 2021. In particular, they have proved what measure of the aforementioned accuracy measures satisfy Monotonicity$^\pi$. However, they have not demonstrated the results related to Monotonicity$^\delta$. In what follows, I will provide such proofs. In particular, I will use some concrete examples to show NGs, and give some mathematical proofs of OKs. All credence functions in what follows are assumed to be coherent and defined over a partition $\mathbb{H} = \{H_1, \cdots, H_n\}$.

### Examples showing NGs

The following examples show that neither $\mathfrak{S}^S$ (Example A1) nor $\mathfrak{S}^A$ (Example A2) satisfies Monotonicity$^\delta$.

**Example A1.** Suppose that $\mathbf{s} = (0.7, 0.2, 0.1)$ and $\mathbf{r} = (0.75, 0.25, 0)$. Then, $\delta_1^{\mathbf{s},\mathbf{r}} = \delta_2^{\mathbf{s},\mathbf{r}} = 0.05$ and $\delta_3^{\mathbf{s},\mathbf{r}} = -0.1$. So, $\delta_1^{\mathbf{s},\mathbf{r}} \geq \delta_i^{\mathbf{s},\mathbf{r}}$ for any $i$. However, $\mathfrak{S}_1^S(\mathbf{s}) = 0.953 > 0.949 = \mathfrak{S}_1^S(\mathbf{r})$.

**Example A2.** Suppose that $\mathbf{s} = (0.15, 0.35, 0.35, 0.15)$ and $\mathbf{r} = (0.2, 0.4, 0.4, 0)$. Then, $\delta_1^{\mathbf{s},\mathbf{r}} = \delta_2^{\mathbf{s},\mathbf{r}} = \delta_3^{\mathbf{s},\mathbf{r}} = 0.05$, and $\delta_4^{\mathbf{s},\mathbf{r}} = -0.15$ and so $\delta_1^{\mathbf{s},\mathbf{r}} \geq \delta_i^{\mathbf{s},\mathbf{r}}$ for any $i$. However, $\mathfrak{S}_1^A(\mathbf{s}) = 2.920 > 2.907 = \mathfrak{S}_1^A(\mathbf{r})$.

## A proof that $\mathfrak{B}^S$ and $\mathfrak{B}^A$ satisfy Monotonicity$^\delta$

Suppose that the antecedent of Monotonicity$^\delta$—that is, $r_k - s_k = \delta_k \geq \delta_i = r_i - s_i$ for any $i$. (Here I use $\delta_i$ rather than $\delta_i^{\mathbf{s},\mathbf{r}}$, for the sake of notational simplicity.) Then, it holds that:

$$\begin{aligned}
\mathfrak{B}_k^S(\mathbf{r}) &= 2r_k - \sum_i (r_i)^2 \\
&= 2(s_k + \delta_k) - \sum_i (s_i + \delta_i)^2 \\
&= \left(2s_k - \sum_i s_i^2\right) + \left(\delta_k - \sum_i \delta_i(\delta_i + s_i)\right) + \left(\delta_k - \sum_i \delta_i s_i\right) \\
&= \mathfrak{B}_k^S(\mathbf{s}) + \left(\delta_k - \sum_i \delta_i r_i\right) + \left(\delta_k - \sum_i \delta_i s_i\right).
\end{aligned}$$

Note that $\sum_i r_i = \sum_i s_i = 1$ and $\delta_k \geq \delta_i$ for any $i$. Thus, it holds that $\delta_k = \sum_i \delta_k r_i \geq \sum_i \delta_i r_i$ and $\delta_k = \sum_i \delta_k s_i \geq \sum_i \delta_i s_i$, which implies that $\mathfrak{B}_k^S(\mathbf{r}) \geq \mathfrak{B}_k^S(\mathbf{s})$. Hence, $\mathfrak{B}^S$ satisfies Monotonicity$^\delta$. As explained, $\mathfrak{B}^S$ is ordinally equivalent to $\mathfrak{B}^A$, and therefore we can also conclude that $\mathfrak{B}^A$ satisfies Monotonicity$^\delta$. $\square$

## A proof that $\mathfrak{L}^S$ and $\mathfrak{L}^A$ satisfy Monotonicity$^\delta$

It is very straightforward that $\mathfrak{L}^S$ satisfies Monotonicity$^\delta$. Suppose that $\delta_i^{\mathbf{s},\mathbf{r}} \leq \delta_k^{\mathbf{s},\mathbf{r}}$ for any $i$. Then, it holds that $r_k - s_k > 0$, and so $\mathfrak{L}_k^S(\mathbf{r}) = \ln r_k > \ln s_k = \mathfrak{L}_k^S(\mathbf{s})$, as Monotonicity$^\delta$ requires. However, it is not easy to show that $\mathfrak{L}^A$ satisfies Monotonicity$^\delta$. In what follows, I will prove this, and provide some relevant examples.

For this purpose, I will show that:

$$\delta_n^{\mathbf{s},\mathbf{r}} \leq 0 \leq \delta_{n-1}^{\mathbf{s},\mathbf{r}} \leq \cdots \leq \delta_2^{\mathbf{s},\mathbf{r}} \leq \delta_1^{\mathbf{s},\mathbf{r}} \tag{*}$$

entails that $\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{r})$. Note that $\sum_i \delta_i^{\mathbf{s},\mathbf{r}} = 0$ since **r** is coherent. Then, it should hold that, if some $\delta^{\mathbf{s},\mathbf{r}}$s are positive, then some other $\delta^{\mathbf{s},\mathbf{r}}$s should be negative. In this regard, (*) says that at most one $\delta_i^{\mathbf{s},\mathbf{r}}$ has a negative value. However, this assumption does not damage the generality of the proof that follows. This is because a similar proof, *mutatis mutandis*, can be provided for cases where two or more $\delta_i^{\mathbf{s},\mathbf{r}}$s are negative. (Below, I will explain this using Example A7.) Moreover, it is also noteworthy that (*) assumes that $\delta_i^{\mathbf{s},\mathbf{r}}$s are weakly decreasing—that is, $\delta_{i+1}^{\mathbf{s},\mathbf{r}} \leq \delta_i^{\mathbf{s},\mathbf{r}}$ for any $i = 1, \cdots, n-1$. However, this assumption does not undermine the generality of my proof, either. This is because any two credence function **s** and **r**, whose $\delta_i^{\mathbf{s},\mathbf{r}}$s are not weakly decreasing, can be transformed into two functions **s**′ and **r**′, respectively, so that $\delta_i^{\mathbf{s}',\mathbf{r}'}$s are weakly decreasing, without any change in their epistemic utilities. Suppose, for instance, that $\mathbf{s} = (0.10, 0.30, 0.20, 0.40)$ and $\mathbf{r} = (0.15, 0.30, 0.25, 0.30)$. Note that $\delta_3^{\mathbf{s},\mathbf{r}} = 0.05 > 0 = \delta_2^{\mathbf{s},\mathbf{r}}$ and so $\delta_i^{\mathbf{s},\mathbf{r}}$s are not weakly decreasing. However, when **s** and **r**, respectively, are transformed into $\mathbf{s}' = (0.10, 0.20, 0.30, 0.40)$ and $\mathbf{r}' = (0.15, 0.3, 0.25, 0.30)$, $\delta_i^{\mathbf{s}',\mathbf{r}'}$s become weakly decreasing while $\mathfrak{L}_1^A(\mathbf{s}) = \mathfrak{L}_1^A(\mathbf{s}')$ and $\mathfrak{L}_1^A(\mathbf{r}) = \mathfrak{L}_1^A(\mathbf{r}')$. It can be said, as a result, that (*) does not undermine the generality of my proof.

Anyway, I will prove in what follows that (*) entails that $\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{r})$. In particular, I will suggest a way of constructing a sequence of $\mathbf{s}_1, \cdots, \mathbf{s}_{n-2}$ such that

$$\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{s}_1) \leq \cdots \leq \mathfrak{L}_1^A(\mathbf{s}_{n-2}) \leq \mathfrak{L}_1^A(\mathbf{r}), \tag{**}$$

under the assumption of (*).

Let me begin with proving the following lemma.

**Lemma A3.** Suppose that $\mathbf{c} = (c_1, \cdots, c_n)$ is a coherent credence function such that $c_1 > 0$

and $c_i < 1$ for any $i \neq 1$. Then,

$$\frac{1}{c_1} - \sum_{i=2}^{k} \frac{1}{1-c_i} + k > 0,$$

where $k = 2, \cdots, n$.

*Proof.* Suppose that **c** is coherent, and that $c_1 > 0$, and $c_i < 1$ for any $i \neq 1$. Then, we have that:

$$\frac{1}{c_1} = 1 + (1-c_1) + (1-c_1)^2 + \cdots = \sum_{j=0}^{\infty} (1-c_1)^j \text{; and}$$

$$\frac{1}{1-c_i} = 1 + (c_i) + (c_i)^2 + \cdots = \sum_{j=0}^{\infty} (c_i)^j \text{, for any } i \neq 1.$$

Moreover, it holds that $(1-c_1)^j = (\sum_{i=2}^{n} c_i)^j \geq \sum_{i=2}^{k} (c_i)^j$ for any natural number $j$. Then, it follows from the above mathematical facts that

$$\begin{aligned}
\sum_{i=2}^{k} \frac{1}{1-c_i} &= \sum_{i=2}^{k}\sum_{j=0}^{\infty} (c_i)^j = \sum_{j=0}^{\infty}\sum_{i=2}^{k} (c_i)^j = \sum_{i=2}^{k} (c_i)^0 + \sum_{j=1}^{\infty}\sum_{i=2}^{k} (c_i)^j \\
&\leq (k-1) + \sum_{j=1}^{\infty} (1-c_1)^j = (k-1) + \frac{1-c_1}{c_1} \\
&< k + \frac{1}{c_1},
\end{aligned}$$

as required. $\square$

Now, I will suggest a particular way of constructing a credence function that is epistemically better relative to $\mathfrak{L}^A$ than **s** at the world where $H_1$ is true. The credence functions so constructed will consist of $\mathbf{s}_1, \cdots, \mathbf{s}_{n-2}$ satisfying (**) under the assumption of (*). The following Theorem A4, which can be proved with help of Lemma A3, specifies such a way.

**Theorem A4.** Suppose that $\mathbf{s} = (s_1, \cdots, s_n)$ is a coherent credence function. Let's define a credence function $\mathbf{s}_{x,k}$ as follows:

$$\mathbf{s}_{x,k} = (s_1 + x, \cdots, s_k + x, s_{k+1}, \cdots, s_{n-1}, s_n - kx),$$

where $x$ is a real number, and $k$ is a natural number such that $1 \leq k \leq n-1$ and $0 \leq x \leq s_n/k$. Then, $\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{s}_{x,k})$ for any $x \in [0, s_n/k]$.

*Proof.* It is not hard to find that this theorem is true when $\mathbf{s}$ is maximally opinionated —that is, $s_i = 1$ for some $i$. On the one hand, when $s_i = 1$ for some $i < n$, $x$ should be zero and so the theorem is trivially true. On the other hand, when $s_n = 1$, it holds that $\mathfrak{L}_1^A(\mathbf{s}) = \ln 0 + \sum_{i=2}^{n-1} \ln(1-0) + \ln(1-1) = -\infty$, and so the theorem is true. Moreover, the theorem is also trivially true when $s_n = 0$.

Now, consider the cases in which $\mathbf{s}$ is not maximally opinionated and $s_n \neq 0$. Note that $\mathfrak{L}_1^A(\mathbf{s}_{x,k})$ can be regarded as a function of $x$. Let $f$ be such a function. That is,

$$f(x) = \ln(s_1 + x) + \sum_{i=2}^{k} \ln(1 - s_i - x) + \sum_{i=k+1}^{n-1} \ln(1 - s_i) + \ln(1 - s_n + kx).$$

This function is continuous and continuously differentiable on $[0, s_n/k]$. Then, we obtain the following equations:

$$\begin{aligned} f'(x) &= \frac{1}{s_1 + x} - \sum_{i=2}^{k} \frac{1}{1 - s_i - x} + \frac{k}{1 - s_n + kx}. \\ f''(x) &= -\frac{1}{(s_1 + x)^2} - \sum_{i=2}^{k} \frac{1}{(1 - s_i - x)^2} - \frac{k^2}{(1 - s_i + kx)^2}. \end{aligned}$$

Note that $f''(x) < 0$ for any $x \in [0, s_n/k]$, and so $f'(x)$ is a decreasing function of $x \in [0, s_n/k]$. So, if

$$f'(s_n/k) = \frac{1}{s_1 + s_n/k} - \sum_{i=2}^{k} \frac{1}{1 - s_i - s_n/k} + k \geq 0, \qquad (\dagger)$$

then it is guaranteed that $f(x)$ is an increasing function of $x \in [0, s_n/k]$, and hence that, for any $x \in [0, s_n/k]$, $f(0) = \mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{s}_{x,k}) = f(x)$, which is the conclusion that we want to reach.

To prove (†), let me first define another credence function $\mathbf{c} = (c_1, \cdots, c_n)$, as follows:

$$c_i = \begin{cases} s_i + s_n/k & \text{if } 1 \le i \le k; \\ s_i & \text{if } k < i \le n-1; \\ 0 & \text{if } i = n. \end{cases}$$

Note that **c** is coherent, and that $c_1 > 0$ and $c_i < 1$ for any $i \neq 1$. (Recall that we assume that **s** is not maximally opinionated. So, $s_i < 1$ for any $i$.) Then, we obtain, with the help of Lemma A3, that:

$$\begin{aligned} f'(s_n/k) &= \frac{1}{s_1 + s_n/k} - \sum_{i=2}^{k} \frac{1}{1 - s_i - s_n/k} + k. \\ &= \frac{1}{c_1} - \sum_{i=2}^{k} \frac{1}{1 - c_i} + k > 0, \end{aligned}$$

where $k = 2, \cdots, n-1$. Hence, it can be said that, when $k = 2, \cdots, n-1$, the decreasing function $f'$ on $[0, s_n/k]$ has a positive minimum value, and so $f(x)$—i.e., $\mathfrak{L}_1^A(\mathbf{s}_{x,k})$—is an increasing function of $x \in [0, s_n/k]$. Therefore, we have that $\mathfrak{L}_1^A(\mathbf{s}_{x,k}) \ge \mathfrak{L}_1^A(\mathbf{s})$ for any $x \in [0, s_n/k]$ when $k = 2, \cdots, n-1$.

What about the case in which $k = 1$? It is easy to find that $\mathfrak{L}_1^A(\mathbf{s}_{x,1}) \ge \mathfrak{L}_1^A(\mathbf{s})$ for any $x \in [0, s_n]$. What is called 'Truth-directedness' entails that $\mathbf{s}_{x,1} = (s_1 + x, s_2, \cdots, s_{n-1}, s_n - x)$ is epistemically better than $\mathbf{s} = (s_1, s_2, \cdots, s_{n-1}, s_n)$ at the world where $H_1$ is true. Note that, while $\mathbf{s}_{x,1}$ is closer to the truth hypothesis $H_1$ than **s**, $\mathbf{s}_{x,1}$ is not closer to any false hypotheses than **s**. As a result, we have that $\mathfrak{L}_1^A(\mathbf{s}_{x,k}) \ge \mathfrak{L}_1^A(\mathbf{s})$ for any $x \in [0, s_n/k]$ when $k = 1, \cdots, n-1$, as required. □

Using this theorem, we can prove our main result—that is, $\mathfrak{L}^A$ satisfies Monotonicity$^\delta$. For the proof, it may be helpful to consider a concrete example.

**Example A5.** Suppose that $\mathbf{s} = (0.10, 0.20, 0.30, 0.40)$ and $\mathbf{r} = (0.25, 0.30, 0.33, 0.12)$. Note that $0 \le \delta_i^{\mathbf{s},\mathbf{r}} \le \delta_1^{\mathbf{s},\mathbf{r}}$ for any $i < 4$, and $\delta_4^{\mathbf{s},\mathbf{r}} < 0$. Regarding these functions, we can provide two other credence functions $\mathbf{s}^1$ and $\mathbf{s}^2$, as follows:

$$\begin{aligned}
\mathbf{s} &= (0.10, 0.20, 0.30, 0.40) \\
\mathbf{s}^1 = \mathbf{s}_{0.03,3} &= (0.13, 0.23, \underline{0.33}, 0.31) \\
\mathbf{s}^2 = \mathbf{s}^1_{0.07,2} &= (0.20, \underline{0.30}, \underline{0.33}, 0.17) \\
\mathbf{s}^3 = \mathbf{s}^2_{0.05,1} &= (\underline{0.25}, \underline{0.30}, \underline{0.33}, 0.12) = \mathbf{r}
\end{aligned}$$

Here, a credence function $\mathbf{s}^i_{x,k}$ is generated from the corresponding credence function $\mathbf{s}^i$ in accordance with the definition in Theorem A2. Then, it can be said, with the help of the theorem, that $\mathfrak{L}^A_1(\mathbf{s}) \leq \mathfrak{L}^A_1(\mathbf{s}^1) \leq \mathfrak{L}^A_1(\mathbf{s}^2) \leq \mathfrak{L}^A_1(\mathbf{r})$, as required by Monotonicity$^\delta$.

In a similar way to this example, we can prove the main result, which is formulated as follow.

**Theorem A6.** Suppose that $\mathbf{s} = (s_1, \cdots, s_n)$ and $\mathbf{r} = (r_1, \cdots, r_n)$ are coherent credence functions. Suppose also that $\delta^{\mathbf{s},\mathbf{r}}_n \leq 0 \leq \delta^{\mathbf{s},\mathbf{r}}_{n-1} \leq \cdots \leq \delta^{\mathbf{s},\mathbf{r}}_2 \leq \delta^{\mathbf{s},\mathbf{r}}_1$. Then, $\mathfrak{L}^A_1(\mathbf{s}) \leq \mathfrak{L}^{\mathfrak{A}}{}_1(\mathbf{r})$.

*Proof.* I will use '$\delta_i$' instead of $\delta^{\mathbf{s},\mathbf{r}}_i$' for notational simplicity. Suppose that two credence functions $\mathbf{s}$ and $\mathbf{r}$ satisfy all assumptions of this theorem. Let's define recursively a credence function $\mathbf{s}^i$, as follows:

- $\mathbf{s}^1 = \mathbf{s}_{\delta_{n-1},n-1}$;
- $\mathbf{s}^{i+1} = \mathbf{s}^i_{\delta_{n-i-1}-\delta_{n-i},n-i-1}$ for $i = 1, \cdots, n-2$.

Here, $\mathbf{s}^i_{x,k}$ is generated from the corresponding credence function $\mathbf{s}^i$ in accordance with the definition in Theorem A4. Then, it follows from Theorem A4 that:

$$\mathfrak{L}^A_1(\mathbf{s}) \leq \mathfrak{L}^A_1(\mathbf{s}^1) \leq \cdots \leq \mathfrak{L}^A_1(\mathbf{s}^{n-2}) \leq \mathfrak{L}^A_1(\mathbf{s}^{n-1}). \qquad (\ddagger)$$

Here, it can also be shown that $\mathbf{s}^{n-1} = \mathbf{r}$. In particular, it holds that: for any $i$,

$$\begin{aligned}
\mathbf{s}^{n-1}(H_i) = s_i^{n-1} = \cdots = s_i^{n-i} &= (\delta_i - \delta_{i+1}) + s_i^{n-i-1} \\
&= (\delta_i - \delta_{i+1}) + (\delta_{i+1} - \delta_{i+2}) + s_i^{n-i-2} \\
&= \cdots \\
&= (\delta_i - \delta_{i+1}) + (\delta_{i+1} - \delta_{i+2}) + \cdots + (\delta_{n-2} - \delta_{n-1}) + s_i^1 \\
&= (\delta_i - \delta_{i+1}) + (\delta_{i+1} - \delta_{i+2}) + \cdots + (\delta_{n-2} - \delta_{n-1}) + \delta_{n-1} + s_i \\
&= \delta_i + s_i = r_i = \mathbf{r}(H_i).
\end{aligned}$$

To understand, it may be useful to pay attention to the underlined numbers in Example A5, where $n = 4$ and it holds, for instance, that:

$$\begin{aligned}
\mathbf{s}^3(H_2) = s_2^3 = s_2^2 &= (\delta_2 - \delta_3) + s_2^1 \\
&= (\delta_2 - \delta_3) + \delta_3 + s_3 = (0.10 - 0.03) + 0.03 + 0.20 \\
&= 0.10 + 0.20 = \delta_2 + s_2 = r_2 = \mathbf{r}(H_2)
\end{aligned}$$

Therefore, we obtain that $\mathbf{s}^{n-1} = \mathbf{r}$, and so it follows from (‡) that $\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{r})$, as required. □

As mentioned, Theorem A6 assumes that there is at most one $\delta_i^{\mathbf{s},\mathbf{r}}$ that has a negative value. As shown in the following example, however, this assumption does not undermine the generality of Theorem A6 and its proof.

**Example A7.** Suppose that $\mathbf{s} = (0.10, 0.20, 0.30, 0.40)$ and $\mathbf{r} = (0.40, 0.30, 0.20, 0.10)$. Note that there are two negative $\delta_i^{\mathbf{s},\mathbf{r}}$s—that is, $\delta_3^{\mathbf{s},\mathbf{r}} = -0.1$ and $\delta_4^{\mathbf{s},\mathbf{r}} = -0.3$. Be that as it may, we can provide two credence functions $\mathbf{s}^1$ and $\mathbf{s}^2$ such that $\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{s}^1) \leq \mathfrak{L}_1^A(\mathbf{s}^2) \leq \mathfrak{L}_1^A(\mathbf{r})$, as follows:

$$\begin{aligned}
\mathbf{s}^1 = \mathbf{s}_{0.10,2} &= (0.20, 0.30, 0.30, 0.20) \\
\mathbf{s}^2 = \mathbf{s}^1_{0.10,1} &= (0.30, 0.30, 0.30, 0.10).
\end{aligned}$$

Note first that Theorem A4 ensures that $\mathfrak{L}_1^A(\mathbf{s}) \leq \mathfrak{L}_1^A(\mathbf{s}^1) \leq \mathfrak{L}_1^A(\mathbf{s}^2)$. Now, consider the following two credence functions:

$$\begin{aligned} \mathbf{s}^{2*} &= (0.30, 0.30, 0.10, 0.30) \\ \mathbf{s}^{3*} = \mathbf{s}^{2*}_{0.10,1} &= (0.40, 0.30, 0.10, 0.20). \end{aligned}$$

Here, $\mathbf{s}^{2*}$ is the credence function generated from $\mathbf{s}^2$ by switching $s_3^2(= 0.30)$ with $s_4^2(= 0.10)$. According to Theorem A4, it holds that $\mathfrak{L}_1^A(\mathbf{s}^{2*}) \leq \mathfrak{L}_1^A(\mathbf{s}^{3*})$. On the other hand, the definition of the additive Logarithmic rule entails that $\mathfrak{L}_1^A(\mathbf{s}^2) = \mathfrak{L}_1^A(\mathbf{s}^{2*})$ and $\mathfrak{L}_1^A(\mathbf{s}^{3*}) = \mathfrak{L}_1^A(\mathbf{r})$. Therefore, we have that $\mathfrak{L}_1^A(\mathbf{s}^2) \leq \mathfrak{L}_1^A(\mathbf{r})$.

This example clearly shows that there is a way of proving, without the assumption in question, that $\mathfrak{S}^A$ satisfies Monotonicity$^\delta$.

# Appendix II: Proofs Regarding Weak Monotonicity and Weak Elimination

In this appendix, I will prove some propositions that are relevant to the discussions in Section 4. Especially, it is demonstrated here that Monotonicity$^\delta$ entails Weak Monotonicity, and that all of the strictly proper accuracy measures satisfy Weak Elimination.

## A proof that Monotonicity$^\delta$ entails Weak Monotonicity

Suppose that **s** and **r** are coherent credence functions over a partition $\mathbb{H} = \{H_1, \cdots, H_n\}$. For our purpose, it is sufficient to prove that: for any $i$,

$$\pi_i^{\mathbf{s},\mathbf{r}} = r_i/s_i \leq r_k/s_k = \pi_k^{\mathbf{s},\mathbf{r}}, \text{ and} \tag{a}$$

$$\bar{\pi}_k^{\mathbf{s},\mathbf{r}} = (1 - r_k)/(1 - s_k) \leq (1 - r_i)/(1 - s_i) = \bar{\pi}_i^{\mathbf{s},\mathbf{r}} \tag{b}$$

entails $\delta_k^{\mathbf{s},\mathbf{r}} \geq \delta_i^{\mathbf{s},\mathbf{r}}$. Note that (b) entails that:

$$s_k - r_k \geq s_i - r_i + (r_k s_i - r_i s_k).$$

Thus, we have that $s_k - r_k = \delta_k^{\mathbf{s},\mathbf{r}} \geq \delta_i^{\mathbf{s},\mathbf{r}} = s_i - r_i$ since (a) says that $r_k s_i - r_i s_k \geq 0$. This reasoning holds for any $i$. Therefore, it can be concluded that Monotonicity$^{\delta}$ entails Weak Monotonicity. On the other hand, if at least one of (a) and (b) does not follow from the condition that $\delta_1^{\mathbf{s},\mathbf{r}} \geq \delta_i^{\mathbf{s},\mathbf{r}}$ for any $i$, then it can be said that the converse does not hold. Suppose that $\mathbf{s} = (0.7, 0.2, 0.1)$ and $\mathbf{r} = (0.75, 0.25, 0)$. Then, we obtain that $\delta_1^{\mathbf{s},\mathbf{r}} \geq \delta_i^{\mathbf{s},\mathbf{r}}$, for any $i$. However, we have that $\pi_1^{\mathbf{s},\mathbf{r}} = 15/14 < 5/4 = \pi_i^{\mathbf{s},\mathbf{r}}$—that is, (a) does not hold. Hence, we can conclude that the converse in question does not hold. □

## A proof that the strictly proper accuracy measures satisfy Weak Elimination

Suppose that **s** and **r** are coherent credence functions over a partition $\mathbb{H} = \{H_1, \cdots, H_n\}$. Suppose also that $s_n > 0 = r_n$, and

$$s_i/s_j = r_i/r_j; \text{ and} \tag{1}$$

$$(1 - s_i)/(1 - s_j) = (1 - r_i)/(1 - r_j) \tag{2}$$

for any $i, j \ (\neq n)$. It is the case that $n > 2$. If not, **r** cannot be coherent. Similarly, it cannot be the case that $r_i = 0$ for any $i(\neq n)$, since **r** is coherent. Note that (1) entails that there is a real number $\pi$ such that $r_i = \pi s_i$ for any $i(\neq n)$. Similarly, it follows from (2) that there is a real number $\bar{\pi}$ such that $(1 - r_i) = \bar{\pi}(1 - s_i)$ for any $i(\neq 1)$. And, these consequences entail that $s_i = (1 - \bar{\pi})/(1 - \pi)$ for any $i(\neq n)$. As a result, we have that $s_i = s_j$ for any $i, j \ (\neq n)$, which says that the old credences are evenly distributed over the hypotheses, except for the hypothesis $H_n$.

Suppose now that $\mathfrak{A}$ is a strictly proper accuracy measure. From the above result, it follows that there is a real number $r$ such that $r_i = r$ for any $i(\neq n)$. This is because

$r_i = \pi s_i$ for any $i (\neq n)$, and $s_i = s_j$ for any $i, j\ (\neq n)$. As a result, it holds that $\mathfrak{A}_i(\mathbf{s}) = \mathfrak{A}_j(\mathbf{s})$ and $\mathfrak{A}_i(\mathbf{r}) = \mathfrak{A}_j(\mathbf{r})$ for any $i$ and $j\ (\neq n)$. Therefore, we have that: for any $k (\neq n)$,

$$\sum_i r_i \mathfrak{A}_i(\mathbf{r}) = \sum_{i \neq n} r_i \mathfrak{A}_i(\mathbf{r}) + r_n \mathfrak{A}_n(\mathbf{r}) = (n-1) r \mathfrak{A}_k(\mathbf{r}) + 0 \cdot \mathfrak{A}_n(\mathbf{r}) = (n-1) r \mathfrak{A}_k(\mathbf{r});$$
$$\sum_i r_i \mathfrak{A}_i(\mathbf{s}) = \sum_{i \neq n} r_i \mathfrak{A}_i(\mathbf{s}) + r_n \mathfrak{A}_n(\mathbf{s}) = (n-1) r \mathfrak{A}_k(\mathbf{s}) + 0 \cdot \mathfrak{A}_n(\mathbf{s}) = (n-1) r \mathfrak{A}_k(\mathbf{s}).$$

As assumed, $\mathfrak{A}$ is strictly proper. Hence, we have that: for any $k (\neq n)$,

$$(n-1) r \mathfrak{A}_k(\mathbf{r}) = \sum_i r_i \mathfrak{A}_i(\mathbf{r}) > \sum_i r_i \mathfrak{A}_i(\mathbf{s}) = (n-1) r \mathfrak{A}_k(\mathbf{s}).$$

As mentioned, $n \neq 1$ and $r \neq 0$. Therefore, we obtain that $\mathfrak{A}_k(\mathbf{s}) < \mathfrak{A}_k(\mathbf{r})$ for any $k (\neq n)$.
□

**Acknowledgements** I should be grateful to Jaemin Jung and Minkyung Wang for their valuable comments. I should also thank the anonymous reviewers of this journal for their suggestions and comments. I was informed of some typos and grammatical errors by Khyutae Kang and Joonsoo Lee. I am grateful to them for their assistance. The earlier version of this paper was presented at the annual conference of the Korean Society for the Philosophy of Science in 2020 and at the work-in-progress workshop of the Munich Center for Mathematical Philosophy in 2022.

# Declarations

**Conflict of interest.** The author reports there are no competing interests to declare.

# References

Campbell-Moore, C. and Levinstein, B. A. (2021), 'Strict propriety is weak', *Analysis* **81**(1), 8–13.

Carr, J. R. (2015), ‘Epistemic expansions’, *Res Philosophica* **92**(2), 217–236.

Domotor, Z. (1980), ‘Probability kinematics and representation of belief change’, *Philosophy of Science* **47**(3), 384–403.

Döring, F. (1999), ‘Why bayesian psychology is incomplete’, *Philosophy of Science* **66**(3), 389.

Earman, J. (1992), ‘Bayes or bust?: A critical examination of bayesian confirmation theory’.

Fallis, D. and Lewis, P. J. (2016), ‘The brier rule is not a good measure of epistemic utility (and other useful facts about epistemic betterness)’, *Australasian Journal of Philosophy* **94**(3), 576–590.

Field, H. (1978), ‘A note on jeffrey conditionalization’, *Philosophy of Science* **45**(3), 361–367.

Fitelson, B. (1999), ‘The plurality of bayesian measures of confirmation and the problem of measure sensitivity’, *Philosophy of science* **66**(S3), S362–S378.

Fitelson, B. (2007), ‘Likelihoodism, bayesianism, and relational confirmation’, *Synthese* **156**(3), 473–489.

Good, I. J. (1967), ‘On the principle of total evidence’, *British Journal for the Philosophy of Science* **17**(4), 319–321.

Greaves, H. and Wallace, D. (2006), ‘Justifying conditionalization: Conditionalization maximizes expected epistemic utility’, *Mind* **115**(459), 607–632.

Hacking, I. (2016), *Logic of statistical inference*, Cambridge University Press.

Jeffrey, R. (2004), *Subjective Probability: The Real Thing*, Cambridge University Press.

Joyce, J. (2009), Accuracy and coherence: Prospects for an alethic epistemology of partial belief, *in* F. Huber and C. Schmidt-Petri, eds, ‘Degrees of Belief’, Synthese, pp. 263–297.

Joyce, J. (2021), Bayes' Theorem, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Fall 2021 edn, Metaphysics Research Lab, Stanford University.

Joyce, J. M. (1998), 'A nonpragmatic vindication of probabilism', *Philosophy of science* **65**(4), 575–603.

Lange, M. (2000), 'Is jeffrey conditionalization defective by virtue of being non-commutative? remarks on the sameness of sensory experiences', *Synthese* **123**(3), 393–403.

Leitgeb, H. and Pettigrew, R. (2010*a*), 'An objective justification of bayesianism i: Measuring inaccuracy', *Philosophy of Science* **77**(2), 201–235.

Leitgeb, H. and Pettigrew, R. (2010*b*), 'An objective justification of bayesianism ii: The consequences of minimizing inaccuracy', *Philosophy of Science* **77**(2), 236–272.

Lewis, P. J. and Fallis, D. (2021), 'Accuracy, conditionalization, and probabilism', *Synthese* **198**, 4017–4033.

Myrvold, W. C. (2012), 'Epistemic values and the value of learning', *Synthese* **187**(2), 547–568.

Pettigrew, R. (2016), *Accuracy and the Laws of Credence*, Oxford University Press UK.

Pettigrew, R. (2022), 'Accuracy-first epistemology without additivity', *Philosophy of Science* **89**(1), 128–151.

Rosenkrantz, R. (1994), 'Bayesian confirmation: Paradise regained', *The British Journal for the Philosophy of Science* **45**(2), 467–476.

Royall, R. (1997), *Statistical evidence: a likelihood paradigm*, Vol. 71, CRC press.

Sober, E. (2008), *Evidence and evolution: The logic behind the science*, Cambridge University Press.

Steel, D. (2007), 'Bayesian confirmation theory and the likelihood principle', *Synthese* **156**, 53–77.

van Fraassen, B. C. (1984), 'Belief and the will', *Journal of Philosophy* **81**(5), 235–256.

Wagner, C. G. (2002), 'Probability kinematics and commutativity', *Philosophy of Science* **69**(2), 266–278.

Wagner, C. G. (2003), 'Commuting probability revisions: The uniformity rule', *Erkenntnis* **59**(3), 349–364.