Probabilism, Entropies and Strictly Proper Scoring Rules

Jürgen Landes

Department of Philosophy
Rutherford Building
University of Kent
Canterbury
CT2 7NZ
United Kingdom

Abstract

Accuracy arguments are the en vogue route in epistemic justifications of probabilism and further norms governing rational belief. These arguments often depend on the fact that the employed inaccuracy measure is strictly proper. I argue controversially that it is ill-advised to assume that the employed inaccuracy measures are strictly proper and that strictly proper statistical scoring rules are a more natural class of measures of inaccuracy. Building on work in belief elicitation I show how strictly proper statistical scoring rules can be used to give an epistemic justification of probabilism.

An agent's evidence does not play any role in these justifications of probabilism. Principles demanding the maximisation of a generalised entropy depend on the agent's evidence. In the second part of the paper I show how to simultaneously justify probabilism and such a principle. I also investigate scoring rules which have traditionally been linked with entropies.

Keywords: Accuracy, scoring rule, probabilism, strict propriety, entropy, principle of indifference

---

Introduction and Notation

1. Introduction

All Bayesians agree on one basic norm governing strength of rational belief

> **Probabilism:** Any rational agent's subjective belief function ought to satisfy the axioms of probability and every probability function is, in principle, permissible. *Prob*

The question arises as to how to justify this norm. Traditionally, axiomatic justifications [6, 41], justifications on logical grounds [22] and Dutch Book Arguments [12, 50] were given to this end. Dutch Book Arguments have been widely regarded as the most persuasive justification, however, they have recently begun losing some of their once widespread appeal [21].[1]

Recent epistemic justifications of probabilism are accuracy-based arguments [24, 25, 30, 31, 49], which all build on [11]. The latter three arguments employ Inaccuracy Measures (IMs) which are assumed to be strictly proper. These IMs are closely related to the notion of a *Scoring Rule* (SR) which the statistical community has a long tradition of studying, see [10] in the Encyclopedia of Statistics.

In the first part of this paper, we argue that statistical SRs, properly understood, are better suited than IMs to justify *Prob*. The argument will be along the following lines: the most convincing justifications of *Prob* relying on IMs require these IMs to be strictly proper (Section 4.1). However, for the purposes of justifying *Prob*, assuming that an IM is strictly proper is ill-advised (Section 4.3). On the contrary, assuming that a SR is strictly proper is not only defensible but a desideratum (Section 3.2).

In Theorem 5.6 we show how strictly proper IMs give rise to strictly proper SRs in a canonical way. We demonstrate in Theorem 6.2 how the class of so-constructed SRs can be used to justify *Prob*.

[1] We are joining the debate concerning rational belief formation assuming that degrees of beliefs are best represented by real numbers in the unit interval $[0, 1] \subset \mathbb{R}$. Anyone who rejects this premise will have to carefully assess whether the here presented account has implications on her line of thinking. Some of our results also hold true for degrees of belief represented by arbitrary positive real numbers.

The justifications in the first part of this paper do not take the agent's evidence into account. In all realistic cases rational agents do possess some evidence and this evidence ought to influence their degrees of belief, in some way. Maximum (generalised) entropy principles require an agent to adopt the probability function which maximises (a generalised) entropy among those probability functions which satisfy constraints imposed by her evidence.

In the second part of this paper we show how to simultaneously justify *Prob* and a such principle (Theorem 7.1 and Theorem 7.2). The usual argument here consists of a two-stage justification – first one justifies *Prob* and then one justifies the entropy principle – and a story explaining why and how the justification of *Prob* trumps that of the entropy principle. The advantage of the simultaneous justification given here is that no such story needs to be told.

Taken together, *Prob* and such a principle entail the Principle of Indifference (PoI) in a large number of cases (Theorem 7.5, Corollary 7.6).

The logarithmic SR is well-known to be the only local SR which is strictly proper when applied to belief functions which are probability functions. Furthermore, this SR is at the heart of the maximum entropy principle. Since we here do not presuppose *Prob*, we investigate notions of locality applied to SRs for general belief functions (Section 8 and Section 9). We prove a non-existence result for such SRs in Theorem 8.4. Furthermore, we investigate how to weaken our assumptions to obtain strictly proper statistical SRs which are local in some sense, see Proposition 9.1 and Proposition 9.2.

2. The Formal Framework

Throughout, we work with a fixed, non-empty and finite set Ω, which is interpreted as the set possible worlds or elementary events. The power set of Ω, $\mathcal{P}\Omega$, is the set of events or the set of propositions. We shall assume throughout that $|\Omega| \geq 2$ and for $X \subseteq \Omega$ let $\bar{X} := \Omega \setminus X$.

The set of probability functions $\mathbb{P}$ is the set of functions $P : \mathcal{P}\Omega \to [0,1]$ such that $\sum_{\omega\in\Omega} P(\{\omega\}) = 1$ and whenever $X \subseteq \Omega$ is such that $X = Y \cup Z$ with $Y \cap Z = \emptyset$, then $P(X) = P(Y) + P(Z)$. We shall use $P(\omega)$ as shorthand for $P(\{\omega\})$.

Note that for all probability functions $P \in \mathbb{P}$ we have that $P(X)+P(\bar{X}) = 1$ and hence $2\sum_{X\subseteq\Omega} P(X) = \sum_{X\subseteq\Omega} P(X) + P(\bar{X}) = |\mathcal{P}\Omega|$.

The set of belief functions is the set of functions $Bel : \mathcal{P}\Omega \to [0,1]$ and shall be denoted by $\mathbb{B}$. Throughout, we assume that all belief and probability functions are *total*, i.e. defined on every $X \subseteq \Omega$. Trivially, since $|\Omega| \geq 2$ we have $\mathbb{P} \subset \mathbb{B}$, where $\subset$ denotes strict inclusion. Of particular interest are the functions $v_\omega \in \mathbb{P}$ for $\omega \in \Omega$. A v_ω is the *at a world* $\omega \in \Omega$ *vindicated credence function*. A v_ω can also be thought of as the *indicator function of the elementary event* $\omega \in \Omega$. The v_ω are defined as follows:

$$v_\omega(X) := \begin{cases} 0 & \text{if } X \text{ is false at } \omega \\ 1 & \text{if } X \text{ is true at } \omega \end{cases} .$$

By "X is true at ω" we mean that $\omega \in X$; on the contrary, "X is false at ω", if and only if $\omega \notin X$.

In this paper we will stay within the classical framework of decision making developed in [53]. So, we assume act-state independence[2], we also only consider propositions which do not refer to themselves nor to their chances. Such propositions are well-known to cause problems for the classical decision making framework. Unsurprisingly, accuracy arguments based on the classical decision making framework are also troubled by such propositions, see [5, 18]. Decision making frameworks for accuracy arguments which can deal with such propositions are explored in [27].

Part 1

3. The Statistical Approach

3.1. Scoring Rules, Applications and Interpretations

Central to SRs and IMs is a measure function measuring the goodness or badness, in some sense, of a belief function Bel. In the statistical community this function is interpreted pragmatically as a loss incurred in a betting scenario, whereas the epistemic tradition interprets the goodness measure as a measure of (in)accuracy.

SRs have mainly been used to *elicit beliefs* or to *assess forecasts*. For belief elicitation it is widely assumed that the agent's belief function Bel^*

[2] In our context this means that neither the truth value nor the objective probability of a proposition $X \subseteq \Omega$ depends on the agent's belief function Bel.

is a probability function, i.e., $Bel^* \in \mathbb{P}$. Similarly, forecasted events are normally assumed to be ruled by an objective probability function P^*, often taken to be the distribution of one (or several) random variable(s). In both applications, there exists a canonical probability function $P \in \mathbb{P}$ (either Bel^* or P^*) which can be used to aggregate losses incurred in different elementary events.

Formally, L is a loss function $L : \Omega \times \mathbb{P} \to [0, +\infty]$ and is referred to as a SR. For a guide to the voluminous literature to SRs refer to [17]. Expected loss is computed in the usual way

$$S_L : \mathbb{P} \times \mathbb{P} \to [0, +\infty], \quad S_L(P, Bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, Bel) \ . \tag{1}$$

Statisticians consider degrees of belief which satisfy *Prob*. Their notion of loss is thus only defined for probabilistic belief functions. For $Bel \in \mathbb{P}$ we have that Bel is completely determined by $\{Bel(\omega) \,|\, \omega \in \Omega\}$. In this case we can regard $L(\omega, Bel)$ as only depending on the first argument, ω, and $\{Bel(\omega) \,|\, \omega \in \Omega\}$.

We shall here be interested in *justifying Prob*. We thus consider a more general loss function L that also depends on degrees of belief in all non-elementary events $X \subseteq \Omega$. We thus consider a loss function $L : \Omega \times \mathbb{B} \to [0, +\infty]$ and define expected loss by

$$S_L : \mathbb{P} \times \mathbb{B} \to [0, +\infty], \quad S_L(P, Bel) := \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, Bel) \ . \tag{2}$$

In general, such a loss function $L : \Omega \times \mathbb{B} \to [0, +\infty]$ is *not* determined by the first argument, ω, and $\{Bel(\omega) \,|\, \omega \in \Omega\}$. Rather, $L(\omega, Bel)$ depends on the elementary event ω and $\{Bel(X) \,|\, X \subseteq \Omega\}$. So, although (1) and (2) appear at first glance to be the same expressions, they do differ in important aspects.

We shall tacitly assume that $L(\omega, Bel)$ in (1) and (2) may also depend on Ω throughout. That is, L may explicitly refer to the elementary events $\nu \in \Omega \setminus \{\omega\}$ or the the events $X \subseteq \Omega$ which contain ω. An example of the former kind of dependence can be found in (3) and of the latter kind in (14).

For ease of reading, we shall use the term *statistical SR* to refer to $S_L(\cdot, \cdot)$ as in (2), rather than the long-winded "expectation of a SR $L : \Omega \times \mathbb{B} \to$

$[0, +\infty]$".

The most famous SR is the Brier Score [3]:

Definition 3.1. *The Brier Score S_{Brier} takes the following form:*[3]

$$S_{Brier}(P, Bel) := \sum_{\omega\in\Omega} P(\omega) \cdot \Big(\sum_{\mu\in\Omega}(v_\omega(\mu) - Bel(\mu))^2\Big) \tag{3}$$

$$= \sum_{\omega\in\Omega} P(\omega)\cdot\Big((1 - Bel(\omega))^2 + \sum_{\mu\in\Omega\setminus\{\omega\}} Bel(\mu)^2\Big) \tag{4}$$

$$= \sum_{\omega\in\Omega} P(\omega)\cdot\Big(1 - 2Bel(\omega) + \sum_{\mu\in\Omega} Bel(\mu)^2\Big) \tag{5}$$

$$= 1 + \sum_{\mu\in\Omega} Bel(\mu)^2 - \sum_{\omega\in\Omega} P(\omega) \cdot 2Bel(\omega). \tag{6}$$

See [57] for an axiomatic characterization of S_{Brier}.

3.2. Strict Propriety for statistical Scoring Rules

We now turn to the key property:

Definition 3.2 (Strict $\mathbb{X}$-propriety)**.** *For any set of belief functions $\mathbb{P} \subseteq \mathbb{X} \subseteq \mathbb{B}$, a statistical SR S_L is strictly $\mathbb{X}$-proper*[4]*, if and only if for all $P \in \mathbb{P}$*

$$\arg\inf_{Bel\in\mathbb{X}} S_L(P, Bel) = \{P\} \ . \tag{7}$$

In plain English, strictly $\mathbb{X}$-proper statistical SRs track probabilities, whatever these probabilities are.

[3]The original definition in [3] does not contain the formal expectation operator $\sum_{\omega\in\Omega} P(\omega)\cdot$. Rather, Brier envisioned a series of n forecasts which would all be scored by $\sum_{\omega\in\Omega}(Bel_i(\omega) - E_{i,\omega})^2$ where $Bel_i(\omega)$ notates the i-th forecast in ω and $E_{i,\omega}$ denotes indicator function for ω on the i-th occasion. The final score is then computed by dividing this sum by n. In essence, this amounts to taking expectations.

[4]Our notion of strict $\mathbb{X}$-propriety notably differs from Γ-strictness, see [20]. A SR is Γ-strict, if and only if for all $P \in \Gamma \subseteq \mathbb{P}$ it holds that $\arg\inf_{Bel\in\mathbb{P}} S_L(P, Bel) = \{P\}$; Γ-strictness is thus a weakening of strict $\mathbb{P}$-propriety. Strict $\mathbb{B}$-propriety is a strengthening of strict $\mathbb{P}$-propriety. $\Gamma \subseteq \mathbb{P}$ constraints the set of probability functions according to which expectations are computed, $\mathbb{X}$ is a set of belief functions containing $\mathbb{P}$.

Recall from when we introduced statistical SRs that losses are usually interpreted pragmatically as losses in a betting scenario. For our purposes we will interpret the function S_L as a *measure of inaccuracy*. The intended interpretation is that $S_L(P, Bel)$ scores the inaccuracy of Bel with respect to the probability function P. By convention, score is an *in*accuracy measure, a *low* score thus means low inaccuracy.

Now consider a function $P \in \mathbb{P}$ and a statistical SR $S_L(P, Bel)$. If $S_L(P, Bel)$ is strictly $\mathbb{B}$-proper, then $Bel = P$ is the unique belief function for which $S_L(P, \cdot)$ is minimal. So, $Bel = P$ is the unique function which minimises inaccuracy. On the other hand, if $S_L(P, Bel)$ is not strictly $\mathbb{B}$-proper, then there exists a $P \in \mathbb{P}$ and a $Bel' \in \mathbb{B} \setminus \{P\}$ with $Bel' \in \arg\inf_{Bel \in \mathbb{B}} S_L(P, Bel)$. Arguably, then

> The class of strictly $\mathbb{B}$-proper statistical SRs is the class of inaccuracy measures in the class of statistical SRs.

Plausibly, one might want to demand further desiderata (such as continuity of L) an inaccuracy measure ought to satisfy. However, it is not clear which other desideratum stands out in the class of further desiderata. Moreover, our approach covers the entire class of strictly $\mathbb{B}$-proper statistical SRs. We will henceforth take it that the class of statistical SRs which measure inaccuracy is the class of strictly $\mathbb{B}$-proper statistical SRs.

While S_{Brier} is well-known to be strictly $\mathbb{P}$-proper it is not strictly $\mathbb{B}$-proper since it does not depend at all on beliefs in non-elementary events and general belief functions $Bel \in \mathbb{B}$ are not determined by their values on elementary events. Thus, S_{Brier} cannot be the SR of choice for rational belief formation approaches that do not presuppose *Prob*.

To the best of our knowledge, strictly $\mathbb{B}$-proper SRs have, surprisingly, not been studied in the literature. So far, only strictly $\mathbb{P}$-proper SRs and strictly proper IMs (see Definition 4.2) have been investigated. In [29], Landes & Williamson use "strictly $\mathbb{B}$-proper SR" to refer to a function which computes expected losses of *normalised* belief functions. Their notion and our notion are thus not the same.

4. The Epistemic Approach

4.1. Ingredients

To highlight that we are now working within the epistemic framework we refer to the $\omega \in \Omega$ as possible worlds, Ω is now called the set of possible worlds and the $X \subseteq \Omega$ are referred to as propositions. This change in terminology is, of course, purely cosmetic.

In recent epistemic approaches, the basic unit of inaccuracy is the inaccuracy of $Bel(X)$ at a world $\omega \in \Omega$, where proposition X is either true or false at ω. Formally, the inaccuracy is represented by an inaccuracy function $I(X, v_\omega(X), Bel(X))$. Since there may be reasons to treat different propositions $X \subseteq \Omega$ differently, the inaccuracy of $Bel(X)$ at world ω may depend on the proposition $X \subseteq \Omega$. For example, different (additive or multiplicative) weights may be attached to different propositions. The basic inaccuracy units, $I(X, v_\omega(X), Bel(X))$, are then aggregated to an overall IM IM_I which measures the inaccuracy of $Bel \in \mathbb{B}$ with respect to a world $\omega \in \Omega$.

Definition 4.1 (Inaccuracy Measure)**.** *Let I be a function $I : \mathcal{P}\Omega \times \{0,1\} \times [0,1] \to [0,\infty]$. An IM IM_I is a map $IM_I : \Omega \times \mathbb{B} \to [0,\infty]$ such that*

$$IM_I(\omega, Bel) := \sum_{X \subseteq \Omega} I(X, v_\omega(X), Bel(X)) \ . \tag{8}$$

So, for a given world ω and a given belief function Bel, IM_I sums the inaccuracies over all propositions $X \subseteq \Omega$ of all beliefs $Bel(X)$ with respect to ω (or, depending on one's point of view, with respect to the at ω vindicated credence function v_ω).

It is natural to think of I as some measure of distance between $v_\omega(X)$ and $Bel(X)$. For example, measuring inaccuracy in Euclidean terms one could consider

$$\begin{aligned} I(X, v_\omega(X), Bel(X)) =& (1 - Bel(X))^2, \text{ if } \omega \notin X \\ I(X, v_\omega(X), Bel(X)) =& Bel(X)^2, \text{ if } \omega \notin X \ . \end{aligned}$$

Such an IM will formally be introduced in Definition 4.4.

The terminology in the literature has not yet converged. The function I has been called an (local) "inaccuracy measure" in [30, 43], whereas Predd et al. call I a SR and refer to IM_I as a "penalty function", while Joyce calls it a "component function" in [25]. Groves (private communications) refers to I as "proposition-specific inaccuracy measure" which is more to the point but quite a mouthful.

In principle, it would be desirable to measure inaccuracy by some function

$f : \Omega \times \mathbb{B} \to [0, +\infty]$ (possibly satisfying further conditions) without assuming that f can be written as a sum over the $X \subseteq \Omega$. For further discussion on this point see [30, Section 5.2.1]. For the purposes of this paper we shall be interested in the set-up of Definition 4.1.

Conceptually, statistical SRs and IMs formalise notions of *inaccuracy*. While they share a common idea they measure inaccuracy differently. Statistical SRs measure inaccuracy between a belief function Bel and a probability function $P \in \mathbb{P}$, strictly $\mathbb{B}$-proper statistical SRs track probabilities. Whereas IMs measure inaccuracy between a belief function Bel and a possible world $\omega \in \Omega$, strictly proper IMs track the actual world, as we will see shortly. For some further discussion see Section 6.1.

One final difference of note is that $S_L(P, Bel)$ is a single real number, whereas $IM_I(\omega, Bel)$ is a tuple of real numbers, one real number for each $\omega \in \Omega$.

Definition 4.2 (Strict Propriety)**.** *An IM IM_I is called* strictly proper*, if and only if the following two conditions are satisfied*

- *for all $p \in [0, 1]$ and all $\emptyset \subset X \subset \Omega$ it holds that $pI(X, 1, x) + (1 - p)I(X, 0, x)$ is uniquely minimized by $x = p$*
- *$I(\Omega, 1, x) + I(\emptyset, 0, y)$ is uniquely minimised by $x = 1$ and $y = 0$.*

Intuitively, strict propriety ensures that setting degrees of belief in X equal to the probability of X is the only way to minimise expected inaccuracy, see further Section 4.3.

In general, the second condition above is required because $P(\emptyset) = 0$ and $P(\Omega) = 1$ for all $P \in \mathbb{P}$ and later on we want p to equal the probability of X.

Some authors do not allow I to depend on X, see for instance [44]. For such a loss function the requirement that $I(1, x) + I(0, y)$ is uniquely minimised by $x = 1$ and $y = 0$ is simply an instance of the first condition. For such an I, the second condition follows from the first.

If IM_I is strictly proper, then for all $\omega \in \Omega$ and all $X \subseteq \Omega$ such that $\omega \in X$ it holds that $I(X, 1, Bel(X)) + I(\bar{X}, 0, Bel(\bar{X}))$ is minimised, if and only if $Bel(X) = 1$ and $Bel(\bar{X}) = 0$. That is, Bel and v_ω agree on X and $\bar{X}$. Hence, $IM_I(\omega, Bel)$ is uniquely minimized by $Bel = v_\omega$. So, if $\omega^* \in \Omega$ is the actual world, then the strictly least inaccurate belief function is $Bel = v_{\omega^*}$. In this sense, strictly proper IMs track the actual world.

Strict propriety as a desideratum for IMs has been argued for in various contexts in which *Prob* is pre-supposed, see [14, 16, 19, 38]. We shall not advance arguments for strict propriety here; in Section 4.3 we shall argue *against* the use of strictly proper IMs in the current context.

The following condition strikes us as a sensible property an IM should satisfy:

Definition 4.3. *An IM IM_I is called* continuous*, if and only if I is continuous in $Bel(X)$.*

Continuity is here taken in the usual sense: For all $X \subseteq \Omega$, for all $i \in \{0,1\}$ and for all sequences $(Bel_n(X))_{n\in\mathbb{N}}$ converging to $Bel(X) \in [0,1]$ it holds that $\lim_{n\to\infty} I(X,i,Bel_n(X)) = I(X,i,Bel(X))$, where both sides of this equation may be equal to $+\infty$.

The most popular IM is an epistemic version of the Brier Score S_{Brier}:

Definition 4.4 (Brier IM)**.** *The Brier IM is defined as*

$$IM_{Brier}(\omega, Bel) := \sum_{X\subseteq\Omega} (v_\omega(X) - Bel(X))^2 \ . \tag{9}$$

In other words: $IM_{Brier}(\omega, Bel)$ is the square of the Euclidean distance in $\mathbb{R}^{|\mathcal{P}\Omega|}$ between v_ω and Bel. It is well-known that IM_{Brier} is strictly proper and continuous. Recently, quadratic IMs, such as IM_{Brier}, have been advocated in [30, 31] on the grounds that they are the only class of measures which keep an agent out of certain epistemic dilemmas.

Compare this measure IM_{Brier} to S_{Brier} (Definition 3.1) and observe that $IM_{Brier}(\omega, Bel)$ depends on the entire belief function while $S_{Brier}(P, Bel)$ only depends on beliefs in elementary events. In Definition 5.2, we will see how to associate IM_{Brier} and a statistical SR. For now, we simply observe the following structural similarity

$$\begin{aligned} S_{Brier}(v_\omega, Bel) &= \sum_{\mu\in\Omega} (v_\omega(\mu) - Bel(\mu))^2 \\ IM_{Brier}(\omega, Bel) &= \sum_{X\subseteq\Omega} (v_\omega(X) - Bel(X))^2 \ . \end{aligned}$$

4.2. Justifications of Probabilism

In justifications of norms of rational belief formation employing IMs it is normally assumed that the agent has no information as to which world is the actual one. How is one then to aggregate inaccuracies $IM_I(\omega, Bel)$ in different worlds? Surely, one could simply add the inaccuracies up, $\sum_{\omega\in\Omega} IM_I(\omega, Bel)$. But why should one not multiply the inaccuracies, $\prod_{\omega\in\Omega} IM_I(\omega, Bel)$, or consider the sum of the logarithms of the inaccuracies, $\sum_{\omega\in\Omega} \log(IM_I(\omega, Bel))$? Apparently, there is no canonical way to aggregate the inaccuracies $IM_I(\omega, Bel)$ for the possible worlds $\omega \in \Omega$.

The Decision Theoretic Norm (DTN) which is widely applied in such a situation is dominance. Historically, the first justification of *Prob* applying dominance was:

Theorem 4.5 (De Finetti [11])**.**

- *For all* $Bel \in \mathbb{B} \setminus \mathbb{P}$ *there exists some* $P \in \mathbb{P}$ *such that for all* $\omega \in \Omega$ $IM_{Brier}(\omega, Bel) > IM_{Brier}(\omega, P)$.
- *For all* $Bel \in \mathbb{P}$ *and all* $Bel' \in \mathbb{B} \setminus \{Bel\}$ *there exists an* $\omega \in \Omega$ *such that* $IM_{Brier}(\omega, Bel') > IM_{Brier}(\omega, Bel)$.

De Finetti's result relies on IM_{Brier} to measure inaccuracy. Plausibly, there are other IMs which measure inaccuracy. Recently, the following generalisation has been proved in the context of belief *elicitation*:

Theorem 4.6 (Predd et al. [49])**.** *If* IM_I *is a continuous and strictly proper IM, then:*

- *For all* $Bel \in \mathbb{B} \setminus \mathbb{P}$ *there exists some* $P \in \mathbb{P}$ *such that for all* $\omega \in \Omega$ $IM_I(\omega, Bel) > IM_I(\omega, P)$.
- *For all* $Bel \in \mathbb{P}$ *and all* $Bel' \in \mathbb{B} \setminus \{Bel\}$ *there exists an* $\omega \in \Omega$ *such that* $IM_I(\omega, Bel') > IM_I(\omega, Bel)$.

Predd et al. credit Lindley (see [34]) for a precursor of their result.

The first parts of these theorems say that every non-probabilistic belief function $Bel \in \mathbb{B} \setminus \mathbb{P}$ is strongly accuracy dominated by some probability function and thus impermissible. The second parts mean that every probabilistic belief function $Bel \in \mathbb{P}$ is permissible, because no $Bel \in \mathbb{P}$ is weakly accuracy dominated.

The two other main justifications of *Prob* along similar lines are due to Joyce, see [24] and [25]. Both justifications apply dominance as DTN in the same way as de Finetti and Predd et al.

The former justification in [24], does not require that a measure of inaccuracy $f(\omega, Bel)$ can be written as a sum over the propositions $X \subseteq \Omega$. In order to prove the theorem Joyce has to assume a number of properties f has to satisfy. The assumed symmetry property has been objected to in [16, 35], Maher also objected to the convexity property. In his 2009 paper, Joyce concedes that the objections raised have merit and that it would be best to do without these properties [25, p. 285].

The latter justification ([25, Theorem 2]) also does not require that the measure of inaccuracy $f(\omega, Bel)$ can be written as a sum over the propositions $X \subseteq \Omega$. It is only assumed that the measure of inaccuracy f satisfies a number of conditions one of which is that f has to be finitely-valued.

We feel that the main draw-back with [25, Theorem 2] is that it only applies for every partition of propositions and not to all propositions $X \subseteq \Omega$. For further discussions see [61, Section 1].

4.3. Strict Propriety for Justifications of Probabilism

We now argue that Theorem 4.6 does not provide a satisfactory justification of *Prob* for belief formation. The problem lies with the requirement that IM_I be strictly proper.

We fully agree with Joyce

> [..] we cannot hope to justify probabilism by assuming that rational agents should maximize the expected accuracy of their opinions because the concept of an expectation really only makes sense for agents whose partial beliefs already obey the laws of probability. [24, p. 590]

Proponents of strictly proper IMs may object that strict propriety guarantees that it is permissible to hold degrees of belief that agree with known probabilities.

This objection misses the mark in at least two decisive ways.

Firstly, a function f ought to be considered as a measure of inaccuracy in virtue of f measuring inaccuracy and emphatically not solely on the virtue of the belief functions it renders permissible given a certain DTN. This objection does not make clear why every appropriate measure of inaccuracy IM_I has to be strictly proper. Intuitively plausible properties such as $I(X, 1, x)$ has a unique minimum on $[0, 1]$ for $x = 1$ or that $I(X, 1, x)$ is a (strictly) decreasing function in $x \in [0, 1]$ do not feature in this objection.

Secondly, as Joyce already pointed out, why would an agent with a non-probabilistic belief function $Bel^* \in \mathbb{B} \setminus \mathbb{P}$ care for the following expectation $Bel^*(X)I(X, 1, Bel^*(X)) + (1 - Bel^*(X))I(X, 0, Bel^*(X))$? It seems that such an agent rather cares for the "expectation" $Bel^*(X)I(X, 1, Bel^*(X)) + Bel^*(\bar{X})I(X, 0, Bel^*(X))$. Since we are in the business of *justifying Prob*, an agent with degrees of belief $Bel^*(X) = 0$ for all $X \subseteq \Omega$ would not be threatened in her beliefs by strict propriety.

We conclude that assuming strict propriety for our purposes is ill-advised. So, Theorem 4.6 does *not* yield a satisfactory justification of *Prob* for belief *formation*.

4.4. Strict Propriety for Belief Elicitation

In the belief *elicitation* framework of Predd et al. it is assumed that the agent's belief function Bel^* is a probability function. Predd et al. [49, p. 4786] motivate strict propriety by "Our scoring rule thus encourages sincerity since your interest lies in announcing probabilities that conform to your beliefs." That is, a subjective Bayesian agent avoiding inaccurate beliefs has a clear impetus to minimise the expectation $Bel^*(X)I(X, 1, Bel'(X)) + Bel^*(\bar{X})I(X, 0, Bel'(X))$ by announcing $Bel'(X) = Bel^*(X)$. I hence find no fault with the requirement of "strict propriety" for eliciting beliefs from subjective Bayesian agents, although I do object to it for the purposes belief formation.

Belief elicitation is at heart an empirical problem, which is often tackled by employing questionnaires, by conducting interviews and/or by observational studies (of subjects playing [incentive compatible] games). SRs have made their way into the applied sciences [39, 65]. See [16, Section 3] for a recent philosophical treatment of belief elicitation.

5. Associating Inaccuracy Measures with Scoring Rules

5.1. Extended Scoring Rules

In this section we shall introduce a class of statistical SRs which allow us to connect IMs to the here introduced class of statistical SRs. We follow [29] and define:

Definition 5.1 (Extended Scoring Rule)**.** *A statistical SR* $S_L : \mathbb{P} \times \mathbb{B} \to [0, \infty]$ *is called* extended*, if and only if it can be written as*

$$
\begin{aligned}
S_L^{ext}(P, Bel) &= \sum_{\omega \in \Omega} P(\omega) \cdot L(\omega, Bel) &(10)\\
&= \sum_{X \subseteq \Omega} P(X) \cdot L'(X, Bel) &(11)\\
&= \sum_{\omega \in \Omega} P(\omega) \cdot \sum_{\substack{X \subseteq \Omega \\ \omega \in X}} L'(X, Bel) \ , &(12)
\end{aligned}
$$

for some function $L' : \mathcal{P}\Omega \times \mathbb{B} \to [0, \infty]$.

The name *extended* is somewhat unfortunate. Originally, it was intended to capture the fact that the domain of the SR has been *extended* from $\mathbb{P} \times \mathbb{P}$ to $\mathbb{P} \times \mathbb{B}$ and that the sum in (10) is over all events $X \subseteq \Omega$ and not merely over the elementary events $\omega \in \Omega$ as in (1).

For our running example, Brier Scores, we give the following extended SR:

Definition 5.2 (Extended Brier Score)**.**

$$
\begin{aligned}
S_{Brier}^{ext}(P, Bel) :&= \sum_{X \subseteq \Omega} P(X) \cdot \Big((1 - Bel(X))^2 + Bel(\bar{X})^2\Big) &(13)\\
&= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subseteq \Omega \\ \omega \in X}} (1 - Bel(X))^2 + \sum_{\substack{Y \subseteq \Omega \\ \omega \notin Y}} Bel(Y)^2\Big) &(14)\\
&= \sum_{\omega \in \Omega} P(\omega) \cdot IM_{Brier}(\omega, Bel) \ . &(15)
\end{aligned}
$$

Proposition 5.3. *S_{Brier}^{ext} is strictly $\mathbb{B}$-proper.*

Proof. The idea is to decompose $S_{Brier}^{ext}(P, Bel)$ into pairs of summands, where each pair is of the form $P(X) \cdot ((1 - Bel(X))^2 + Bel(\bar{X})^2) + P(\bar{X}) \cdot ((1 - Bel(\bar{X}))^2 + Bel(X)^2)$. We then show that each such pair is uniquely minimised by $Bel(X) = P(X)$ and $Bel(\bar{X}) = 1 - P(X)$.

Consider the following minimization problem for fixed $P \in \mathbb{P}$, fixed $X \subseteq \Omega$ and $x := Bel(X)$, $y := Bel(\bar{X})$

$$
\text{minimize} \qquad P(X) \cdot ((1 - x)^2 + y^2) + (1 - P(X)) \cdot ((1 - y)^2 + x^2)
$$

subject to $\quad x, y \in [0,1]$.

Note that the objective function of this minimisation problem is equal to $x^2 - 2xP(X) + P(X) + y^2 - 2y(1-P(X)) + (1-P(X))$. The unique minimum obtains for $x = P(X)$ and $y = 1 - P(X)$.

Hence, $Bel = P$ uniquely minimizes $S^{ext}_{Brier}(P, \cdot)$. □

A version of de Finetti's Theorem (Theorem 4.5) for S^{ext}_{Brier} follows as a simple Corollary:

Corollary 5.4.

- *For all $Bel \in \mathbb{B} \setminus \mathbb{P}$ there exists some $P \in \mathbb{P}$ such that for all $Q \in \mathbb{P}$ $S^{ext}_{Brier}(Q, Bel) > S^{ext}_{Brier}(Q, P)$.*
- *For all $Bel \in \mathbb{P}$ and all $Bel' \in \mathbb{B} \setminus \{Bel\}$ there exists a $P \in \mathbb{P}$ such that $S^{ext}_{Brier}(P, Bel') > S^{ext}_{Brier}(P, Bel)$.*

Proof. 1) Let $Bel \in \mathbb{B} \setminus \mathbb{P}$. By Theorem 4.5 there exists a $P_{Bel} \in \mathbb{P}$ such that for all $\omega \in \Omega$ it holds that $IM_{Brier}(\omega, Bel) > IM_{Brier}(\omega, P_{Bel})$. Using (15), the fact that Ω is finite and that for all $Q \in \mathbb{P}$ there exists an $\omega \in \Omega$ with $Q(\omega) > 0$ we find that $S^{ext}_{Brier}(Q, Bel) > S^{ext}_{Brier}(Q, P_{Bel})$.

2) We saw in Proposition 5.3 that S^{ext}_{Brier} is strictly $\mathbb{B}$-proper. Hence, $S^{ext}_{Brier}(Bel, \cdot)$ is uniquely minimised by $Bel = Bel$. □

Note that de Finetti's Theorem applies dominance with respect to the possible worlds $\omega \in \Omega$ while the above corollary applies dominance with respect to the probability functions $Q \in \mathbb{P}$.

5.2. The Canonical Association

In this section we shall see how to canonically associate with every IM an extended SR. We shall give two further examples to illustrate the association.

Definition 5.5 (Canonical Association)**.** *For IM_I define an associated statistical SR S^{aso}_I by:*

$$S^{aso}_I(P, Bel) := \sum_{\omega \in \Omega} P(\omega) \cdot IM_I(\omega, Bel) \tag{16}$$

$$= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subseteq \Omega \\ \omega \in X}} I(X, 1, Bel(X)) + \sum_{\substack{Y \subseteq \Omega \\ \omega \notin Y}} I(Y, 0, Bel(Y)) \Big) \tag{17}$$

$$= \sum_{X \subseteq \Omega} P(X) \cdot I(X, 1, Bel(X)) + P(\bar{X}) \cdot I(X, 0, Bel(X)) \tag{18}$$

$$= \sum_{X \subseteq \Omega} P(X) \cdot \Big(I(X, 1, Bel(X)) + I(\bar{X}, 0, Bel(\bar{X})) \Big) \ . \tag{19}$$

So, letting $L'(X, Bel) := I(X, 1, Bel(X)) + I(\bar{X}, 0, Bel(\bar{X}))$ we see that S_I^{aso} is an extended SR.

For a fixed IM IM_I, $S_I^{aso}(P, Bel)$ is simply the expected inaccuracy of Bel, where expectations are computed with respect to the probability function $P \in \mathbb{P}$.

Theorem 5.6. *IM_I is strictly proper, if and only if S_I^{aso} is strictly $\mathbb{B}$-proper.*

Proof. If IM_I is strictly proper, then for every $\emptyset \subset X \subset \Omega$ and all $P \in \mathbb{P}$

$$P(X) \cdot I(X, 1, Bel(X)) + P(\bar{X}) \cdot I(X, 0, Bel(X))$$

is uniquely minimised by $Bel(X) = P(X)$.

Furthermore, $I(\Omega, 1, Bel(\Omega)) + I(\emptyset, 0, Bel(\emptyset))$ is uniquely minimised by $Bel(\Omega) = 1$ and $Bel(\emptyset) = 0$. Applying (18) we now find that $S_I^{aso}(P, \cdot)$ is uniquely minimised by $Bel = P$.

Now, suppose that S_I^{aso} is strictly $\mathbb{B}$-proper. Then for all $p \in [0, 1]$ and all $P \in \mathbb{P}$ with $P(\omega) = p$ and $P(\omega') = 1 - p$ for different $\omega, \omega' \in \Omega$ we have

$$\begin{aligned} S_I^{aso}(P, Bel) &= \sum_{X \subseteq \Omega} P(X) \cdot I(X, 1, Bel(X)) + P(\bar{X}) \cdot I(X, 0, Bel(X)) \\ &= \sum_{\substack{U \subseteq \Omega \\ \omega, \omega' \in U}} 1 \cdot I(U, 1, Bel(U)) + 0 \cdot I(U, 0, Bel(U)) \\ &\quad + \sum_{\substack{W \subseteq \Omega \\ \omega, \omega' \notin W}} 0 \cdot I(W, 1, Bel(W)) + 1 \cdot I(W, 0, Bel(W)) \\ &\quad + \sum_{\substack{Y \subseteq \Omega \\ \omega \in Y, \omega' \notin Y}} p \cdot I(Y, 1, Bel(Y)) + (1 - p) \cdot I(Y, 0, Bel(Y)) \\ &\quad + \sum_{\substack{Z \subseteq \Omega \\ \omega' \in Z, \omega \notin Z}} (1 - p) \cdot I(Z, 1, Bel(Z)) + p \cdot I(Z, 0, Bel(Z)) \ . \end{aligned}$$

Now observe that every belief function $Bel^+ \in \mathbb{B}$ minimising $S_I^{aso}(P, \cdot)$ minimises each of the four sums above individually, since every sum only depends on degrees of belief no other sum depends on.

By considering the first two sums for $U = \Omega$ and $W = \emptyset$ we find that $I(\Omega, 1, Bel^+(\Omega)) + I(\emptyset, 1, Bel^+(\emptyset))$ is uniquely minimised by $Bel^+(\Omega) = 1$ and $Bel^+(\emptyset) = 0$.

Let us now consider the third sum. Note that any given $Y \subseteq \Omega$ such that $\omega \in Y$ and $\omega' \notin Y$ only appears in this sum once (and it does not appear in any other sum). Thus, $Bel^+(Y) = p = P(Y)$ is the unique minimum of $p \cdot I(Y, 1, \cdot) + (1-p) \cdot I(Y, 0, \cdot)$. By varying $p = P(\omega)$ we obtain that $Bel^+(Y) = P(\omega)$ is the unique minimum of $p \cdot I(Y, 1, \cdot) + (1-p) \cdot I(Y, 0, \cdot)$ for all $p \in [0, 1]$ and all $Y \subseteq \Omega$ with $\omega \in Y$.

Finally, note that the above arguments do not depend on $\omega \in \Omega$. We thus find for all $Y \subseteq \Omega$ that $Bel^+(Y) = p$ is the unique minimum of $p \cdot I(Y, 1, \cdot) + (1-p) \cdot I(Y, 0, \cdot)$ for all $p \in [0, 1]$.

Thus, IM_I is strictly proper. □

From a purely technical point of view, Theorem 5.6 can be most helpful. All one needs to do to check whether a SR S_I^{aso} is strictly $\mathbb{B}$-proper is to check whether the IM IM_I is strictly proper. The latter task can be accomplished simply by checking whether simple sums are uniquely minimised by $Bel(X) = p$ and $Bel(\bar{X}) = 1-p$. Checking strict $\mathbb{B}$-propriety requires one to solve a minimisation problem in $[0, 1]^{|\mathcal{P}\Omega|}$, which is in general a much harder problem.

Furthermore, Theorem 5.6 allows us to easily generate strictly $\mathbb{B}$-proper statistical SRs by association. That means that the class of inaccuracy measures in our sense is a rich class consisting of a great variety of members.

We now give two applications of Theorem 5.6 in which we generate extended strictly $\mathbb{B}$-proper SRs. The logarithmic IM ($I_{\log}(X, 1, x) := -\log(x)$, $I_{\log}(X, 0, x) := -\log(1-x)$) and the spherical IM are well-known to be strictly proper ($I_{sph}(X, 1, x) := 1 + \frac{-x}{\sqrt{x^2 + (1-x)^2}}$, $I_{sph}(X, 0, x) := 1 + \frac{x-1}{\sqrt{x^2+(1-x)^2}}$), see, e.g., [25, Section 8]).

Corollary 5.7. *The following logarithmic SR is strictly $\mathbb{B}$-proper.*

$$S_{\log}^{aso}(P, bel) := \sum_{X \subseteq \Omega} P(X) \cdot \Big(-\log(Bel(X)) - \log(1 - Bel(\bar{X})) \Big)$$

$$= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(- \sum_{\substack{X \subseteq \Omega \\ \omega \in X}} \log(Bel(X)) - \log(1 - Bel(\bar{X})) \Big) \ .$$

As usual in this context, we put $0 \cdot \infty := 0$ and $r \cdot \infty = \infty$ for $r \in (0, 1]$.

By "log" we refer to a logarithm with an arbitrary base $b > 1$ and by "ln" to the natural logarithm, i.e., with base e.

Corollary 5.8. *The following spherical SR is strictly* $\mathbb{B}$*-proper.*

$$S^{aso}_{sph}(P, bel) := \sum_{X \subseteq \Omega} P(X) \cdot$$
$$\Big(2 + \frac{-Bel(X)}{\sqrt{Bel(X)^2 + (1 - Bel(X))^2}} + \frac{Bel(\bar{X}) - 1}{\sqrt{bel(\bar{X})^2 + (1 - Bel(\bar{X}))^2}} \Big) \ .$$

For our running example, Brier Scores, we already considered the canonical association in Definition 5.2. We now note that Proposition 5.3 can alternatively be obtained as a simple corollary from Theorem 5.6

Theorem 5.6 raises one, as of yet, open problem:

> **Open Problem 1:** Does for all strictly $\mathbb{B}$-proper statistical SRs S_L exist an IM IM_I such that
>
> $$S_L(P, Bel) = \sum_{\omega \in \Omega} P(\omega) \cdot IM_I(\omega, Bel) \ ?$$

6. Justifying Probabilism with statistical Scoring Rules

In this section we build on Theorem 4.6 in order to obtain an epistemic justification of *Prob* for rational belief formation.

6.1. The Rationality of Tracking Objective Probabilities

We will assume the existence of objective probabilities and that the set of objective probability functions is $\mathbb{P}$. Whether such probabilities exist in the real world is a metaphysical debate, which we will not enter here. We content ourselves with noting that a number of writers have defended their existence in the real world. While the existence of objective probabilities in the real world is a matter of debate, at least in (statistical) models featuring probability distributions of random variables objective probabilities may safely be assumed to exist.

Ideally, one might think, rational agents aim for beliefs which track the truth rather than tracking probabilities. Determining the truth, if such a thing as the true state of the world exists, has proven to be a rather complicated endeavour. Many have argued that if an agent knows the chances, then the only rational option is to set degrees of belief equal to the chances. We take it here that these arguments are right and that rational agents aim at tracking objective probabilities, at least in situations in which objective probabilities exist.

6.2. The formal Derivation

Lemma 6.1. *Let S_L be a strictly $\mathbb{B}$-proper SR. For all $Bel \in \mathbb{P}$ and all $Bel' \in \mathbb{B} \setminus \{Bel\}$ there exists a $P \in \mathbb{P}$ such that $S_L(P, Bel') > S_L(P, Bel)$.*

Proof. If $Bel \in \mathbb{P}$, then $S_L(Bel, \cdot)$ is uniquely minimized by $Bel = Bel$. So, for $Bel' \in \mathbb{B} \setminus \{Bel\}$ we have $S_L(Bel, Bel') > S_L(Bel, Bel)$. □

Theorem 6.2. *Let S_I^{aso} be strictly proper and let IM_I be continuous.*

- *For all $Bel \in \mathbb{B} \setminus \mathbb{P}$ there exists some $P \in \mathbb{P}$ such that for all $Q \in \mathbb{P}$ $S_I^{aso}(Q, Bel) > S_I^{aso}(Q, P)$.*
- *For all $Bel \in \mathbb{P}$ and all $Bel' \in \mathbb{B} \setminus \{Bel\}$ there exists a $P \in \mathbb{P}$ such that $S_I^{aso}(P, Bel') > S_I^{aso}(P, Bel)$.*

Proof. 1) Let $Bel \in \mathbb{B} \setminus \mathbb{P}$, then by Theorem 4.6 there exists a $P_{Bel} \in \mathbb{P}$ such that for all $\omega \in \Omega$ it holds that $IM_I(v_\omega, Bel) > IM_I(v_\omega, P_{Bel})$. For all $Q \in \mathbb{P}$ there exists some $\omega \in \Omega$ such that $Q(\omega) > 0$. We thus find for all $Q \in \mathbb{P}$ that $S_I^{aso}(Q, Bel) > S_I^{aso}(Q, P_{Bel})$ holds.

2) By Theorem 5.6 S_I^{aso} is strictly $\mathbb{B}$-proper, now apply Lemma 6.1. □

6.3. A brief Discussion

Besides the assumptions that rational agents aim only at accurate beliefs and that inaccuracy may be measured by a statistical SR S_L, the above justification of *Prob* rests on the following: A) The statistical SR S_L is associated with an IM. B) S_I^{aso} is strictly $\mathbb{B}$-proper. C) Continuity of I. D) Dominance as DTN.

In order to make this justification compelling A – D need to be plausible. If rational agents only aim at accurate beliefs, then the statistical SR should be strictly $\mathbb{B}$-proper, as we argued in Section 3.2. If the answer to Open Problem 1 is "yes", then B implies A. If the answer is "no", then we either need to give an argument which singles out the class of statistical SRs which are associated with some IM IM_I or give a proof of Theorem 6.2 that also applies to statistical SRs which are not associated with an IM. Those who consider the class of strictly proper IMs to be the class inaccuracy measures in the epistemic approach seem to be forced to accept that the class of statistical SRs which measure inaccuracy by closeness-to-chances is precisely the class obtained by association.

Continuity is a fairly harmless technical condition. Again, as for A, it might be possible to prove Theorem 6.2 without assuming continuity.

As far as we are aware, no-one has seriously objected to dominance as DTN in this context, when applied to possible worlds. In the setting of this paper, agents aim at tracking objective probabilities and not at tracking worlds. It is thus fitting that dominance applies to objective probabilities in Theorem 6.2.

In Section 4.3 we argued that strict propriety for IMs without presupposing that $Bel \in \mathbb{P}$ is unsatisfactory. For statistical SRs however, strict $\mathbb{B}$-propriety is desirable as a mean to encourage tracking of objective probabilities and thus reduce inaccuracy (Section 3.2). Under the assumption that strict propriety is technically necessary for convincing justifications of *Prob*, the upshot of Section 3.2 is that statistical SRs are *in principle* better suited than IMs for such justifications. Theorem 6.2 demonstrates that it is also *possible* to give a justification of *Prob* in the statistical framework.

The statistical approach has, at least in principle, one further advantage over the epistemic approach. Suppose the $\omega \in \Omega$ are the elementary events of some trial with chance distribution P^*. Given a belief function Bel and a SR S_L we can, at least in principle, approximate $S_L(P^*, Bel)$ by conducting i.i.d. trial runs. Thus, we do not need to have access to P^* to approximate $S_L(P^*, Bel)$. In the epistemic approach one assumes that there is an actual world ω^* among the $\omega \in \Omega$ but one does not know which possible world is the actual world. It is thus not possible, not even in principle, to compute $IM_I(v_{\omega^*}, Bel)$.

Another advantage distinct to the statistical approach is that it canonically lends itself to take the agent's evidence into account, as we shall see in the second part of this paper. The question of whether the classical epistemic framework is able to adequately capture the agent's evidence for justifications of *Prob* is a matter of philosophical debate; see [13, 45]; which we will not enter here.

6.4. Meeting some Objections

One may object that the here presented justification *presupposes* probabilism by assuming the existence objective probability distributions which satisfy Kolmogorov's axioms. We openly acknowledge that we assumed the existence of objective probabilities and that this assumption is key. Note however that the assumption of objective probabilities is an assumption about the "outside world" which is *external* to the agent. We did not presuppose anything about the agent's degrees of belief (other than that they are real numbers in $[0, 1] \subset \mathbb{R}$). Our presupposition thus concerns the agent's environment but not the agent's doxastic state.

We want to make two further points. Firstly, justifications of *Prob* in the framework of Section 4.2 which assume strict propriety presuppose *internal* probabilism, the condition *strict propriety* involves an expectation! Secondly, objective probabilities may well not exist in the real world. However, in (toy) models their existence is guaranteed by the model specifications. The sceptical reader may thus read our proposal as only applying to such toy models. In general, we agree with Jaynes

> In this connection we have to remember that probability theory never solves problems of actual practice, because all such problems are infinitely complicated. We solve only idealizations of the real problem, and the solution is useful to the extent that the idealization is a good one. [23, p. 568]

One may also object that there are further epistemic goods which rational agents ought to care for. It is certainly true that there might be other epistemic goods, or even non-epistemic goods, rational agents ought to care for. In the absence of a convincing account detailing what exactly these goods are, we feel that it is appropriate to ignore these goods and solely focus on inaccuracy minimisation.

The proponent of the classical epistemic framework in Section 4.2 may be drawn to one of the following moves. Firstly, convincing justifications could be given that do not require the IM IM_I to be strictly proper. This move appears very unlikely, but possible, to succeed.

Secondly, one might head down the Joycean path and consider general measures of inaccuracy $f(\omega, Bel)$. This path is, of course, open. The technical challenges one encounters appear to be so substantial, that assumptions need to be made which make the justifications less than fully satisfactory.

Thirdly, an argument may be advanced claiming that the class of appropriate IMs is a proper subclass of the strictly proper IMs. The appeal of such an approach then hinges on the characterisation of this subclass of IMs. Such an argument was put forward in [30, 31]. The class of IMs considered in [30, 31] is so narrow that it does not contain the logarithmic nor the spherical IM. Their justification, improving on de Finetti's result by moderately enlarging the class of IMs, can thus only be a step towards a satisfactory justification of *Prob*. Until such a reasonably large subclass of strictly proper IMs has been discovered, we remain sceptical about this approach.

Part 2

7. Maximum Entropy Principles

The first part of this paper focussed on justifications of *Prob*. A great number of writers invoke further norms to constrain the choice of a belief function more tightly. Typically, such norms are Calibration Norms ([63, Section 3.3]), a Principal Principle ([33, 43, 47]) or the Maximum Entropy Principle (discussed in more detail below) to constrain the choice of a belief function depending on the agent's evidence. Justifications of such approaches are normally given in a two-stage argument. First, *Prob* is justified, then the further norm(s) are justified. This leaves proponents of such approaches with the complicated task of explaining why and how the justification of *Prob* supersedes the justification(s) of the further invoked norm(s).

In this section we give a single justification for *Prob* and Maximum Generalised Entropy Principles at the same time. Since we give a single justification no two-stage justificatory argument is required of the proponent of a combination of *Prob* and a Maximum Generalised Entropy Principle.

Exactly as in the first part, we do not presuppose *Prob*, strict $\mathbb{P}$-propriety is hence of little use. The key notion will again be strict $\mathbb{B}$-propriety.

As in the first part of this paper we focus on formal aspects of the justifications and only touch on the question as to when DTNs apply. The DTN we will here use is Worst-Case Expected Loss (WCEL) avoidance. In the formal literature, WCEL has rich history and goes back to the seminal work of Morgenstern and von Neumann. The most obvious toy cases in which WCEL avoidance is an appropriate DTN are two-player single-round games with an adversary playing after Player1 has made her move. Recently, normative arguments for risk sensitivity were advanced in [4]. A maximally risk-sensitive agents adheres to WCEL avoidance.

The justifications we give here apply to interpretations of $P \in \mathbb{E}$ as epistemic subjective probabilities or as objective probabilities.

7.1. The general Arguments

Consider an agent with current evidence which narrows the chance function down to a non-empty and convex set $\emptyset \subset \mathbb{E} \subseteq \mathbb{P}$. $\mathbb{E}$ is called the set of *calibrated* functions. The most prominent objective Bayesian approach then requires an agent to equivocate sufficiently between the basic propositions that the agent can express while adopting a belief function in $\mathbb{E}$, cf. [63].[5] That is, the agent is required to assign the basic propositions the same probabilities as far as this is consistent with the agent's evidence. This norm is then spelled out in terms of the Maximum Entropy Principle:

> **Maximum Entropy Principle** (MaxEnt) A rational agent ought to adopt a probability function $Bel \in \mathbb{E}$ which maximises Shannon Entropy, $H_{\log}$
>
> $$H_{\log}(Bel) := \sum_{\omega\in\Omega} -Bel(\omega)\log(Bel(\omega)) \ . \tag{20}$$

The probability function $P_= \in \mathbb{P}$ defined by $P_=(\omega) := \frac{1}{|\Omega|}$ for all $\omega \in \Omega$ is called the *equivocator*. $P_=$ is the function in $\mathbb{P}$ with greatest entropy. MaxEnt can be understood as requiring an agent to adopt a belief function in $\mathbb{E}$ which is as similar to $P_=$ as possible.

MaxEnt has given rise to a substantial literature on rational belief formation; as examples we mention [1, 8, 23, 29, 40, 41].

Key to MaxEnt is the loss function $L(\omega, Bel) = -\log(Bel(\omega))$ and the logarithmic scoring rule $S_{\log}$

$$S_{\log}(P, Bel) := \sum_{\omega\in\Omega} -P(\omega)\log(Bel(\omega)) \ .$$

We can express Shannon Entropy in terms of this SR, $H_{\log}(P) = S_{\log}(P, P)$.

MaxEnt is well-known to be justified on the following grounds of WCEL avoidance [20, 59]

Theorem 7.1 (Justification of MaxEnt)**.** *If $\emptyset \neq \mathbb{E} \subseteq \mathbb{P}$ is convex and closed, then*

$$\arg\inf_{Bel\in\mathbb{P}} \sup_{P\in\mathbb{E}} S_{\log}(P, Bel) = \arg\sup_{P\in\mathbb{E}} H_{\log}(P) \tag{21}$$

[5] For our purposes, it is not relevant to explain what "*sufficiently* equivocates" amounts to. We shall only be concerned with maximal equivocation.

and there is only one unique such function maximising Shannon Entropy, $P^\dagger$.

So, an agent which aims to minimise $\sup_{P\in\mathbb{E}} S_L(P, Bel)$ by adopting a probabilistic belief function $Bel \in \mathbb{P}$, i.e., avoiding worst-case expected logarithmic loss, has to adopt $P^\dagger$ as her belief function.

We now generalise this well-known justification of MaxEnt to strictly $\mathbb{X}$-proper SRs which satisfy the following minimax equation

$$\inf_{Bel\in\mathbb{X}} \sup_{P\in\mathbb{E}} S_L(P, Bel) = \sup_{P\in\mathbb{E}} \inf_{Bel\in\mathbb{X}} S_L(P, Bel) \ . \tag{22}$$

See [58] for an introduction to such minimax equations which arose from Von Neumann's seminal game theoretical work.

Following [20], we call $H_L(P) := S_L(P, P)$ *generalised entropy*. If the set $\arg\sup_{P\in\mathbb{E}} H_L(P)$ contains a unique function, then this function is denoted by $P^\ddagger$ and called *generalised entropy maximiser*. The following generalises [20, Theorem 6.4] to non-probabilistic belief functions.

Theorem 7.2 (Justification of Generalised Entropy Maximisation)**.** *If $\emptyset \neq \mathbb{E} \subseteq \mathbb{P}$ is convex and closed, S_L strictly $\mathbb{X}$-proper, (22) holds and if $H_L(P)$ is strictly concave on $\mathbb{P}$, then*

$$\arg \inf_{Bel\in\mathbb{X}} \sup_{P\in\mathbb{E}} S_L(P, Bel) = \arg\sup_{P\in\mathbb{E}} H_L(P) =: \{P^\ddagger\} \ . \tag{23}$$

Proof. Let us first use (22) and the fact that S_L is strictly $\mathbb{X}$-proper to obtain

$$\inf_{Bel\in\mathbb{X}} \sup_{P\in\mathbb{E}} S_L(P, Bel) = \sup_{P\in\mathbb{E}} \inf_{Bel\in\mathbb{X}} S_L(P, Bel) \tag{24}$$

$$= \sup_{P\in\mathbb{E}} S_L(P, P) \ . \tag{25}$$

Since $\mathbb{E}$ is convex, closed and non-empty the function $S_L(P, P)$ has a unique supremum in $\mathbb{E}$. That is, the set $\arg\sup_{P\in\mathbb{E}} \sum_{\omega\in\Omega} S_L(P, P)$ consists of a unique probability function which is in $\mathbb{E}$, $P^\ddagger$.

Using $\mathbb{X}$-strict propriety to obtain the strict inequality in (27) we find for all $Bel \in \mathbb{X} \setminus \{P^\ddagger\}$

$$\sup_{P\in\mathbb{E}} S_L(P,Bel) \geq S_L(P^\ddagger,Bel) \tag{26}$$

$$> S_L(P^\ddagger,P^\ddagger) \ . \tag{27}$$

Recall that $\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{E}} S_L(P,Bel)$ equals $S_L(P^\ddagger,P^\ddagger)$. Thus, no $Bel \in \mathbb{X}\setminus\{P^\ddagger\}$ minimises $\sup_{P\in\mathbb{E}} S_L(P,Bel)$. Hence, $P^\ddagger$ is the unique minimiser of $\sup_{P\in\mathbb{E}} S_L(P,Bel)$. □

This means that an agent which aims to minimise $\sup_{P\in\mathbb{E}} S_L(P,Bel)$ by adopting a belief function $Bel \in \mathbb{B}(!)$, i.e., avoiding worst-case expected loss, has to adopt $P^\ddagger$ as her belief function.

So, Theorem 7.2 simultaneously justifies *Prob* and the following principle:

> **Maximum Generalised Entropy Principle** A rational agent ought to adopt the unique probability function in $\mathbb{E}$ which maximises the generalised entropy $H_L(P)$.

The question arises how $P^\ddagger$ changes when the agent receives new information and the set of calibrated functions changes. It is not rational for a WCEL avoiding agent to change her belief, if $\mathbb{E}' \subset \mathbb{E}$ and $P^\ddagger \in \mathbb{E}'$ (see below). This property of unchanged beliefs has been termed *obstinacy*, see for example [40, p. 80].

Corollary 7.3. *Let $\mathbb{E}$ and S_L be as in Theorem 7.2. If $\emptyset \subset \mathbb{E}' \subset \mathbb{E}$ contains $P^\ddagger$, then*

$$\arg\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{E}'} S_L(P,Bel) = \{P^\ddagger\} \ . \tag{28}$$

Note that we do not require that $\mathbb{E}'$ is convex nor that $\mathbb{E}'$ is closed.

Proof. First note that

$$\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{E}'} S_L(P,Bel) \leq \inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{E}} S_L(P,Bel) \tag{29}$$

$$= S_L(P^\ddagger,P^\ddagger) \ . \tag{30}$$

For all $Bel \in \mathbb{X}\setminus\{P^\ddagger\}$ we find using strict $\mathbb{X}$-propriety of S_L that

$$\sup_{P \in \mathbb{E}'} S_L(P, Bel) \geq S_L(P^{\ddagger}, Bel) \quad (31)$$

$$> S_L(P^{\ddagger}, P^{\ddagger}) \ . \quad (32)$$

For the belief function $P^{\ddagger}$ we find

$$S_L(P^{\ddagger}, P^{\ddagger}) \leq \sup_{P \in \mathbb{E}'} S_L(P, P^{\ddagger}) \quad (33)$$

$$\leq \sup_{P \in \mathbb{E}} S_L(P, P^{\ddagger}) \quad (34)$$

$$= S_L(P^{\ddagger}, P^{\ddagger}) \ . \quad (35)$$

So, $\sup_{P \in \mathbb{E}'} S_L(P, P^{\ddagger}) = S_L(P^{\ddagger}, P^{\ddagger})$. Hence, $P^{\ddagger}$ uniquely minimises WCEL. □

7.2. Generalised Entropies

Theorem 7.2 gives general conditions under which generalised entropy maximisation is justified with respect to the choice of a *particular* statistical SR. Unsurprisingly, the choice of different SRs, i.e., utility functions, leads to different generalised entropy maximisers. The importance of choosing an appropriate SR has recently been emphasised in [36].

Consider the extended Brier score S^{ext}_{Brier}, the spherical SR S^{aso}_{sph} and $S^{ext}_{llog} := -\frac{|\mathcal{P}\Omega|}{2} + \sum_{Y \subseteq \Omega} Bel(Y) - \sum_{X \subseteq \Omega} P(X) \cdot \ln(Bel(X))$. We now show that all three SRs satisfy satisfy the conditions in Theorem 7.2. We shall not give the rather uninformative calculations but rather state the result of these calculations.

All three SRs are strictly $\mathbb{B}$-proper, see Proposition 5.3, Corollary 5.8 and Proposition 9.1.

Straightforward calculations show that Brier Entropy $H_{Brier}(P)$ and the Spherical Entropy $H_{Sph}(P)$ are strictly concave on $\mathbb{P}$. The entropy of the logarithmic SR is $H_{\mathcal{P}\Omega}(P) := \sum_{X \subseteq \Omega} -P(X) \log(P(X))$ which we shall prove in Section 9.1. This entropy is called *Proposition Entropy* in [29]. Clearly, $H_{\mathcal{P}\Omega}$ is strictly concave on $\mathbb{P}$.

Note that $H_{\mathcal{P}\Omega}$ is different from Shannon Entropy, $H_{\log}$. In $H_{\mathcal{P}\Omega}$ the sum is taken over all events $X \subseteq \Omega$ and not over all elementary events $\omega \in \Omega$. Not only are Proposition Entropy and Shannon Entropy different functions; in general, their respective maximum obtains for different probability functions in $\mathbb{E}$, cf. [29, Figure 1, p. 3536].

That all three entropies considered here are sufficiently regular, satisfying the minimax condition (24), follows for instance from König's result [28, p. 56], see [51] for a discussion of König's result.

These three entropies have different maximisers on rather simple sets $\mathbb{E}$, as can be gleaned from Figure 1 and Figure 2.

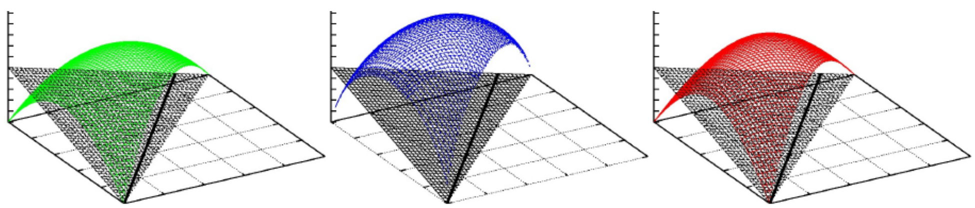

Figure 1: Brier Entropy H_{Brier} (green), Proposition Entropy $H_{\mathcal{P}\Omega}$ (blue) and Spherical Entropy H_{Sph} (red) for $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The black line segment connects $P_1 = (1, 0, 0)$ and $P_2 = (0, \frac{5}{6}, \frac{1}{6})$.

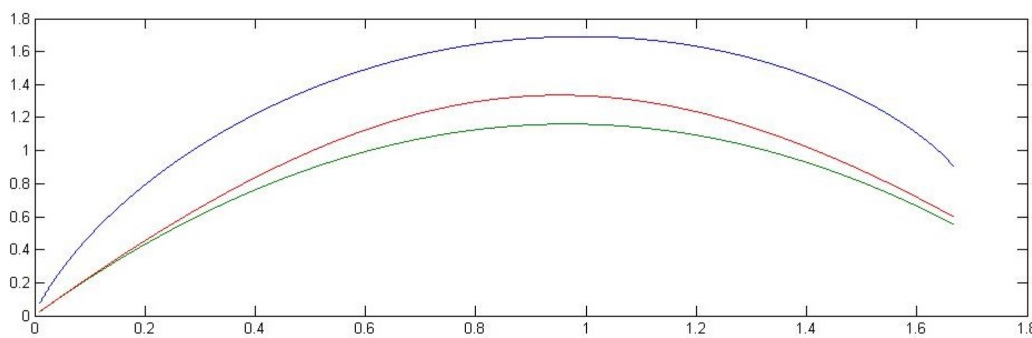

Figure 2: Brier Entropy H_{Brier} (green), Proposition Entropy $H_{\mathcal{P}\Omega}$ (blue) and Spherical Entropy H_{Sph} (red) plotted along the line segment between $P_1 = (1, 0, 0)$ and $P_2 = (0, \frac{5}{6}, \frac{1}{6})$ parametrised as $P_1 + t \cdot (-0.6, 0.5, 0.1)$ for $t \in [0, \frac{10}{6}]$. The Brier Entropy maximiser is $P^{\dagger}_{Brier} = (0.4194, 0.4839, 0.0968)$ $[t = 0.968]$, the Proposition Entropy maximiser is $P^{\dagger}_{\mathcal{P}\Omega} = (0.4054, 0.4955, 0.0991)$ $[t = 0.991]$ and the Spherical Entropy maximiser is $P^{\dagger}_{Sph} = (0.4277, 0.4770, 0.0954)$ $[t = 0.954]$. The absolute value of the Spherical Entropy has been adjusted to fit all curves neatly into the picture.

Theorem 7.2 deals with generalised entropies. The question arises whether we can find a statistical SR to simultaneously justify *Prob* and MaxEnt. Unfortunately, we do not know the answer to this question

Open Problem 2 Does there exist a strictly $\mathbb{B}$-proper statistical SR S_L such that (24) holds and such that for all closed and convex $\emptyset \subset \mathbb{E} \subseteq \mathbb{P}$ it holds that

$$\arg\sup_{P\in\mathbb{E}} S_L(P,P) = \arg\sup_{P\in\mathbb{E}} H_{\log}(P) \ \ ? \tag{36}$$

7.3. Generalised Entropies and the Principle of Indifference

The Principle of Indifference (PoI) has long fascinated philosophers. We here show that maximising generalised entropies entails the PoI for many natural generalised entropies. Recent arguments in its favor can be found in [37, 46, 62].

Definition 7.4. *A SR S_L is called* equivocator neutral, *if and only if for all $\omega,\omega'\in\Omega$ it holds that $L(\omega,P_=)=L(\omega',P_=)$.*

Theorem 7.5 (Generalised Entropies and PoI)**.** *If S_L is equivocator neutral, strictly $\mathbb{X}$-proper with $\mathbb{P}\subseteq\mathbb{X}\subseteq\mathbb{B}$, satisfies (24) and if $H_L(P)$ is strictly concave on $\mathbb{P}$, then*

$$\arg\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{P}} S_L(P,Bel) = \arg\sup_{P\in\mathbb{P}} S_L(P,P) = \{P_=\} \ . \tag{37}$$

So, under complete ignorance, $\mathbb{E}=\mathbb{P}$, the unique rational choice under WCEL avoidance is $Bel=P_=$; this provides a justification of the PoI. For a recent justification of the PoI using IMs we refer the reader to [46].

Proof. From Theorem 7.2 and the fact that $\mathbb{P}$ is convex and closed we obtain

$$\arg\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{P}} S_L(P,Bel) = \arg\sup_{P\in\mathbb{P}} S_L(P,P) \ . \tag{38}$$

Note that since S_L is equivocator neutral, there exists some constant $c\in\mathbb{R}$ such that for all $\omega\in\Omega$ it holds that $L(\omega,P_=)=c$.

Assume for contradiction that there exists some $Q\in\arg\sup_{P\in\mathbb{P}} H_L(P)$ which is different from $P_=$. Since $H_L(P)$ is a strictly concave function on $\mathbb{P}$ the maximum of $H_L(\cdot)$ has to be unique and hence $H_L(Q)>H_L(P)$. We then obtain using (38)

$$\begin{aligned} H_L(P_=) &< H_L(Q) && (39)\\ &= \inf_{Bel \in \mathbb{X}} \sup_{P \in \mathbb{P}} S_L(P, Bel) && (40)\\ &\leq \sup_{P \in \mathbb{P}} S_L(P, P_=) && (41)\\ &= \sup_{P \in \mathbb{P}} \sum_{\omega \in \Omega} P(\omega) L(\omega, P_=) && (42)\\ &= \sup_{P \in \mathbb{P}} \sum_{\omega \in \Omega} P(\omega) c && (43)\\ &= c && (44)\\ &= \sum_{\omega \in \Omega} \frac{1}{|\Omega|} L(\omega, P_=) && (45)\\ &= \sum_{\omega \in \Omega} P_=(\omega) L(\omega, P_=) && (46)\\ &= H_L(P_=) \ . && (47) \end{aligned}$$

Contradiction. Thus, $\{P_=\} = \arg\sup_{P \in \mathbb{P}} H_L(P)$. □

Equivocator neutrality is a very weak symmetry condition on L. Strict $\mathbb{B}$-propriety and satisfying (24) are standing assumptions in this section. Finally, $\arg\sup_{P \in \mathbb{P}} H_L(P)$ containing a unique element would follow from H_L being strictly concave. If S_L is strictly $\mathbb{P}$-proper, then H_L is concave, see [17, p. 361]. Thus, in a large number of cases maximising generalised entropy entails the PoI.

Not only is the equivocator the unique function minimising WCEL under complete ignorance, it is also the unique such function as long as $P_= \in \mathbb{E}$:

Corollary 7.6. *For a SR S_L as in Theorem 7.5 and for all sets $\mathbb{E} \subset \mathbb{P}$ such that $P_= \in \mathbb{E}$ it holds that*

$$\arg \inf_{Bel \in \mathbb{X}} \sup_{P \in \mathbb{E}} S_L(P, Bel) = \arg\sup_{P \in \mathbb{P}} S_L(P, P) = \{P_=\} \ . \qquad (48)$$

Proof. First, let us reason as in Theorem 7.5 to obtain the equality below

$$\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{E}} S_L(P,Bel) \leq \inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{P}} S_L(P,Bel) \tag{49}$$

$$= S_L(P_=,P_=)\ . \tag{50}$$

Using strict propriety we find for all belief functions $Bel \in \mathbb{X}\setminus\{P_=\}$ that

$$\sup_{P\in\mathbb{E}} S_L(P,Bel) \geq S_L(P_=,Bel) \tag{51}$$

$$> S_L(P_=,P_=)\ . \tag{52}$$

So all belief functions different from the equivocator $P_=$ have a strictly sub-optimal WCEL. $P_=$ has the best possible WCEL as we saw in Theorem 7.5. It follows that

$$\arg\inf_{Bel\in\mathbb{X}}\sup_{P\in\mathbb{E}} S_L(P,Bel) = \{P_=\}\ . \tag{53}$$

□

For instance, $S_L = S^{ext}_{Brier}$, $S_L = S^{aso}_{sph}$ and $S_L = S^{ext}_{llog}$ satisfy the assumptions of Theorem 7.5. We hence obtain

Corollary 7.7. *If $S_L = S^{ext}_{Brier}$, $S_L = S^{aso}_{sph}$ or $S_L = S^{ext}_{llog}$ and $P_= \in \mathbb{E}$, then*

$$\arg\inf_{Bel\in\mathbb{B}}\sup_{P\in\mathbb{E}} S_L(P,Bel) = \arg\sup_{P\in\mathbb{P}} S_L(P,P) = \{P_=\}\ . \tag{54}$$

8. Local Scoring Rules

We now turn our attention to strictly $\mathbb{B}$-proper statistical SRs themselves. $S_{\log}$ stands out as the only strictly $\mathbb{P}$-proper local SR and as the heart of MaxEnt. It has hence received considerable attention in the literature. Locality means that if a elementary event $\omega \in \Omega$ obtains, then the loss incurred only depends on $Bel(\omega)$ and not on the entire belief function.

Subsequently, we will take an interest in notions of locality applied to SRs defined on $\mathbb{P}\times\mathbb{B}$. Surprisingly, the most natural way of extending the notion of locality to $\mathbb{P}\times\mathbb{B}$ is incompatible with strict $\mathbb{B}$-propriety.

8.1. Locality and strict $\mathbb{P}$-propriety

Definition 8.1. *A statistical SR $S_L : \mathbb{P} \times \mathbb{P} \to [0, +\infty]$ is called* local*, if and only if $L(\omega, Bel)$ only depends on the belief in ω and not on other beliefs. Abusing the notation in the usual way we write $L(Bel(\omega))$.*

The class of such SRs which are strictly $\mathbb{P}$-proper is rather simple:

Theorem 8.2 (Savage [54])**.** *Up to an affine-linear transformation, the only local and strictly $\mathbb{P}$-proper statistical SR is*

$$S_{\log}(P, Bel) = \sum_{\omega \in \Omega} -P(\omega) \log(Bel(\omega)) \ . \tag{55}$$

Local SRs or logarithmic loss functions have been argued for in a variety of settings. For example, in [66, pp. 16] and [2, p. 72-73] for belief elicitation. See [7, p. 2039-2040] for a discussion on locality and [7, p. 2046] for an axiomatic characterisation of logarithmic SRs in terms of scale-invariance.

Levinstein points out advantages of $S_{\log}$ as a measure of inaccuracy over S_{Brier} applied to probabilistic belief functions, see [32]. We also want to mention that $S_{\log}$ is the only strictly $\mathbb{P}$-proper SR which is consistent with the use of likelihoods or log likelihoods to evaluate assessors, cf. [64, p. 1075]. In [63, p. 64-65], Williamson shows that $S_{\log}$ can be characterised in terms of four natural axioms, one of which is locality. $S_{\log}$ has found applications in a variety of areas, for example in information theory [9, 52], Neyman-Pearson Theory in statistics [15] and the health sciences [26].

Recently, the $IM_{\log}$ has left a positive impression in formal epistemology as a tool to measure a degree of confirmation, see [60].

Let us now consider a local loss function $L : [0, 1] \to [0, +\infty]$ and the corresponding local SR $S_L : \mathbb{P} \times \mathbb{B} \to [0, +\infty]$ (defined on belief functions $Bel \in \mathbb{B}$!)

$$S_L(P, Bel) = \sum_{\omega \in \Omega} P(\omega) \cdot L(Bel(\omega)) \ . \tag{56}$$

Note that only beliefs in elementary events appear in the above expression. Thus, beliefs in non-elementary events will not affect the score $S_L(P, Bel)$. Thus, a DTN applying local statistical SR $S_L(P, Bel)$ can only yield constraints on the agent's beliefs in elementary events; beliefs in non-elementary events are completely unconstrained. So, local SRs are ill-suited for justifications of norms of rational belief formation without presupposing *Prob*.

Thus, we now investigate how to extend the notion of locality, which proved to be technically fruitful when *Prob* was presupposed, without presupposing *Prob*.

8.2. Locality, strict $\mathbb{B}$-propriety and extended Scoring Rules

One obvious way to generalise locality is:

Definition 8.3. *An extended SR is called* ex-local*, if and only if there exists a loss function* $L_{loc} : \mathcal{P}\Omega \times [0,1] \to [0,\infty]$ *such that*

$$S^{ext}_{L_{loc}}(P, Bel) = \sum_{X \subseteq \Omega} P(X) \cdot L_{loc}(X, Bel(X)) \quad (57)$$

$$= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subseteq \Omega \\ \omega \in X}} L_{loc}(X, Bel(X)) \Big) \ . \quad (58)$$

Ex-locality here means that $L(X, Bel)$ is of the form $L_{loc}(X, Bel(X))$, i.e. the loss attributable to event X in isolation of all other events, if X obtains, only depends on X and on $Bel(X)$. Here, we do *not* allow $L(X, Bel(X))$ to depend on further beliefs such as $Bel(\bar{X})$.

This notion of an ex-local extended SR differs from *local* statistical SRs in Savage's sense in two respects. Firstly, the sum is now over all events $X \subseteq \Omega$ and not only over the elementary events $\omega \in \Omega$. Secondly, the loss function L_{loc} may now depend on the event X whereas Savage's loss function only depended on the belief in an elementary event ω and not in the elementary event itself.

If S^{ext}_L is ex-local, then the loss attributable to $Bel(X)$ only enters once into (57). More precisely, the only summand depending on $Bel(X)$ is $P(X) \cdot L_{loc}(X, Bel(X))$. Since P is a probability function, $P(\emptyset) = 0$ holds. Hence, by our convention that $0 \cdot \infty = 0$ we obtain $P(\emptyset) \cdot L_{loc}(\emptyset, Bel(\emptyset)) = 0 \cdot L_{loc}(\emptyset, Bel(\emptyset)) = 0$ for all $P \in \mathbb{P}$. So, $S^{ext}_{L_{loc}}(P, Bel)$ does not depend on $Bel(\emptyset)$.

Hence, a belief function Bel_a which agrees with P on all events $\emptyset \subset X \subseteq \Omega$ and $Bel(\emptyset) = a$ with $a \in (0,1]$ it holds that $S^{ext}_{L_{loc}}(P, P) = S^{ext}_{L_{loc}}(P, Bel_a)$. Thus, no ex-local SR is strictly $\mathbb{B}$-proper.

One might initially think that the incompatibility of ex-locality and strict $\mathbb{B}$-propriety is only due to the fact that for all $P \in \mathbb{P}$ $P(\emptyset) = 0$ holds. However, we shall now see that this is not the case.

Let $\mathbb{B}^- := \{Bel : \mathcal{P}\Omega\backslash\{\emptyset\} \to [0,1]\}$ and define strict $\mathbb{B}^-$-propriety of a SR S_L in the obvious way, i.e., for all $P \in \mathbb{P}$ it holds that $\arg\inf_{Bel\in\mathbb{B}^-} S(P, Bel) = \{P_{\restriction \mathcal{P}\Omega\backslash\{\emptyset\}}\}$. For ease of notation we drop the restriction operator "$\restriction$" from now on.

Theorem 8.4. *There does not exist an ex-local extended strictly $\mathbb{B}^-$-proper SR $S^{ext}_{L_{loc}}$.*

Proof. It is sufficient to show that for all $P \in \mathbb{P}$

$$\arg \inf_{Bel\in\mathbb{B}^-} S^{ext}_{L_{loc}}(P, Bel) = \arg \inf_{Bel\in\mathbb{B}^-} \sum_{X\subseteq\Omega} P(X) \cdot L_{loc}(X, Bel(X)) \tag{59}$$

does not depend on P, since strict $\mathbb{B}^-$-propriety would require that the above minimum obtains uniquely for $Bel = P$.

For a fixed loss function L_{loc} and a fixed event $\emptyset \subset X \subseteq \Omega$ it holds that $\arg\inf_{Bel(X)\in[0,1]} L_{loc}(X, Bel(X))$ only depends on $Bel(X) \in [0, 1]$ and not on P nor on $Bel(Y)$ for $Y \neq X$. Furthermore, $Bel(X)$ may be freely chosen in $[0, 1]$, since Bel does not have to satisfy any further constraints, such as the axioms of probability. Hence, for all $\emptyset \subset X \subseteq \Omega$ the infimum (or infima) of $P(X)L_{loc}(X, Bel(X))$ obtains independently of P.

Thus, $S^{ext}_{L_{loc}}(P, Bel)$ is minimised, if and only if every summand in (59) is minimised. For each summand this minimum obtains independently of P. □

Proposition 8.5. $S^{ext}_{\log}(P, Bel) := \sum_{X\subseteq\Omega} -P(X)\cdot\log(Bel(X))$ *is not strictly $\mathbb{B}^-$-proper.*

Proof. Define a belief function $Bel_1 \in \mathbb{B}$ by $Bel_1(X) := 1$ for all $X \subseteq \Omega$. For all $P \in \mathbb{P}$ and all $X \subseteq \Omega$ it holds that $P(X)\log(Bel_1(X)) = 0$. So, for all $P \in \mathbb{P}$

$$Bel_1 \in \arg \inf_{Bel\in\mathbb{B}} S^{ext}_{\log}(P, Bel) \ . \tag{60}$$

□

Recall from Theorem 8.2 that the logarithmic SR $S_{\log}$ is the only local $\mathbb{P}$-strictly proper statistical SR. Evidently, strict propriety crucially depends on the set of scored belief functions.

The SR considered in Corollary 5.7: $S_{\log}^{aso}(P, Bel) := \sum_{X \subseteq \Omega} P(X) \cdot (-\log(Bel(X)) - \log(1 - Bel(\bar{X})))$ is not ex-local. The loss term depends on $Bel(X)$ *and* $Bel(\bar{X})$. Thus, Proposition 5.7 does not contradict Theorem 8.4.

Note that $S_{L_{loc}}^{ext}(P, Bel)$ does not depend on $Bel(X)$ for all those event $X \subset \Omega$ with $P(X) = 0$. If any genuine measure of inaccuracy has to take into account how $P(X)$ and $Bel(X)$ relate for *all* $X \subseteq \Omega$, then ex-local SRs cannot serve as measures of inaccuracy. In this case, the impossibility theorem only rules out the existence of SRs which are unsuitable for our purposes.

9. Two Notions of Locality

The question we now pose is: how much of the locality condition do we need to give up in order obtain strictly $\mathbb{B}$-proper extended SRs which are local, *in some sense*?

9.1. Penalties

As it turns out, there exists an extended SR employing logarithms which is strictly $\mathbb{B}$-proper.

Proposition 9.1. *The following extended SR is strictly* $\mathbb{B}$*-proper*

$$S_{llog}^{ext}(P, Bel) := \sum_{X \subseteq \Omega} P(X) \cdot \left(-1 + \frac{\sum_{Y \subseteq \Omega} Bel(Y)}{\sum_{Y \subseteq \Omega} P(Y)} - \ln(Bel(X)) \right) \tag{61}$$

$$= -\frac{|\mathcal{P}\Omega|}{2} + \sum_{Y \subseteq \Omega} Bel(Y) - \sum_{X \subseteq \Omega} P(X) \cdot \ln(Bel(X)) \ . \tag{62}$$

This SR is not purely logarithmic since it contains the *penalty term*, $\sum_{Y \subseteq \Omega} Bel(Y)$. This term penalises belief functions for indiscriminately assigning high degrees of belief to all events. In particular it prevents $Bel_1 \in \mathbb{B}$ from being the score minimiser. The penalty term is constant for all $X \subseteq \Omega$, it is thus global.

Proof. Define an IM IM_{llog} by

$$I(X,0,Bel(X)) := Bel(X)$$
$$I(X,1,Bel(X)) := Bel(X) - 1 - \ln(Bel(X)) \ .$$

We now show that IM_{llog} is strictly proper. Clearly, IM_{llog} is never strictly less than zero.

Let $p \in [0,1]$ and $\emptyset \subset X \subset \Omega$ be fixed and let

$$\begin{aligned} f(Bel(X)) :&= p \cdot I(X,1,Bel(X)) + (1-p) \cdot I(X,0,Bel(X)) \\ &= p \cdot Bel(X) - p - p \cdot \ln(Bel(X)) + (1-p) \cdot Bel(X) \\ &= -p - p \cdot \ln(Bel(X)) + Bel(X) \ . \end{aligned}$$

By equating the derivative of $f(Bel(X))$ with zero we find for $p > 0$

$$\frac{d\,f(Bel(X))}{dBel(X)} = -\frac{p}{Bel(X)} + 1 = 0 \ . \tag{63}$$

Trivially, this equation is uniquely solved by $Bel(X) = p > 0$. Considering the second derivative of $f(Bel(X))$ shows that $Bel(X) = p > 0$ is the unique minimum.

For $p = 0$ we recall the usual convention that $0 \ln(Bel(X)) = 0$, even if $Bel(X) = 0$. Hence, $f(Bel(X)) = (1-p) \cdot I(X,0,Bel(X)) = Bel(X)$, which is uniquely minimised by $Bel(X) = p = 0$.

For $X = \emptyset$ and $X = \Omega$ we have

$$I_{llog}(\Omega,1,Bel(X)) + I_{llog}(\emptyset,0,Bel(X)) = Bel(\Omega) - 1 - \ln(Bel(\Omega)) - Bel(\emptyset),$$

which is uniquely minimised by $Bel(\Omega) = 1$ and $Bel(\emptyset) = 0$.

We next show that S^{ext}_{llog} is strictly $\mathbb{B}$-proper. We do so by showing that it is associated with IM_{llog} and hence strictly $\mathbb{B}$-proper by Theorem 5.6.

$$\sum_{\omega \in \Omega} P(\omega) \cdot IM_{llog}(\omega, Bel)$$

$$
\begin{aligned}
&= \sum_{\omega\in\Omega} P(\omega)\cdot\Big(\sum_{\substack{X\subseteq\Omega\\ \omega\in X}} I(X,1,Bel(X)) + \sum_{\substack{Y\subseteq\Omega\\ \omega\notin Y}} I(Y,0,Bel(Y))\Big)\\
&= \sum_{\omega\in\Omega} P(\omega)\cdot\Big(\sum_{\substack{X\subseteq\Omega\\ \omega\in X}} Bel(X) - 1 - \ln(Bel(X)) + \sum_{\substack{Y\subseteq\Omega\\ \omega\notin Y}} Bel(Y)\Big)\\
&= \sum_{\omega\in\Omega} P(\omega)\cdot\Big(\sum_{Z\subseteq\Omega} Bel(Z) + \sum_{\substack{X\subseteq\Omega\\ \omega\in X}} -1 - \ln(Bel(X))\Big)\\
&= \sum_{Z\subseteq\Omega} Bel(Z) + \sum_{\omega\in\Omega} P(\omega)\cdot\Big(\sum_{\substack{X\subseteq\Omega\\ \omega\in X}} -1 - \ln(Bel(X))\Big)\\
&= \sum_{Z\subseteq\Omega} Bel(Z) + \sum_{X\subseteq\Omega} P(X)\cdot\Bigg(-1 - \ln(Bel(X))\Bigg)\\
&= \sum_{X\subseteq\Omega} P(X)\cdot\sum_{Z\subseteq\Omega}\frac{Bel(Z)}{\sum_{Y\subseteq\Omega} P(Y)} + \sum_{X\subseteq\Omega} P(X)\cdot\Bigg(-1 - \ln(Bel(X))\Bigg)\\
&= \sum_{X\subseteq\Omega} P(X)\cdot\Bigg(\frac{\sum_{Z\subseteq\Omega} Bel(Z)}{\sum_{Y\subseteq\Omega} P(Y)} - 1 - \ln(Bel(X))\Bigg)\\
&= S^{ext}_{llog}(P,Bel)\ .
\end{aligned}
$$

□

S^{ext}_{llog} contains a local term, $\ln(Bel(X))$, and a global term, $\sum_{Y\subseteq\Omega} Bel(Y)$. The constant term, $-\frac{|\mathcal{P}\Omega|}{2}$, has been added for the following cosmetic reason. For $Bel \in \mathbb{P}$ we have

$$
\begin{aligned}
S^{ext}_{llog}(P,Bel) &= -\sum_{X\subseteq\Omega} P(X)\cdot\ln(Bel(X)) \qquad (64)\\
&= S^{ext}_{\log}(P,Bel)\ . \qquad (65)
\end{aligned}
$$

So, for $Bel \in \mathbb{P}$ we recapture the SR considered in Proposition 8.5 (for the natural logarithm) and we note that

$$
S^{ext}_{llog}(P,P) = -\sum_{X\subseteq\Omega} P(X)\cdot\ln(P(X))\ .
$$

At first glance, S^{ext}_{llog} appears to be an extended strictly $\mathbb{B}$-proper SR which is not associated to an IM. If this were the case, then we would have solved Open Problem 1 (Section 5.2) in the negative. However, we saw in the above proof that S^{ext}_{llog} is indeed associated with the strictly proper IM_{llog}. We have thus not solved Open Problem 1.

Finally, let us remark that proving strict $\mathbb{B}$-propriety of S^{ext}_{llog} directly is a rather complicated endeavour. The above proof is a nice illustration of the technical helpfulness of Theorem 5.6 to which we alluded to in Section 5.2.

9.2. Normalising Beliefs

In Proposition 9.1 we saw how one can use a penalty term to construct a strictly $\mathbb{B}$-proper logarithmic SR. In [29] the authors showed that the penalty term can be dropped, if the belief functions are normalised, that is the belief functions considered are in some set $\mathbb{B}_{norm} \supset \mathbb{P}$.

We shall now quickly summarise the relevant points in [29]: Denote by π a set of non-empty mutually exclusive, jointly exhaustive proper subsets of Ω, i.e., a partition. Denote by Π the union of $\{\Omega, \emptyset\}$, $\{\Omega\}$ and the set of these partitions. Then define

$$\mathbb{B}_{norm} := \{B : \mathcal{P}\Omega \to [0,1] \mid \sum_{F\in\pi} B(F) = 1 \text{ for some } \pi \in \Pi \\ \text{and } \sum_{F\in\pi} B(F) \leq 1 \text{ for all } \pi \in \Pi\} \ .$$

For a given a weighting function $g : \Pi \to \mathbb{R}_{\geq 0}$ such that for all $\emptyset \subseteq X \subseteq \Omega$ it holds that $\sum_{\substack{\pi\in\Pi \\ X\in\pi}} g(\pi) > 0$, a SR is defined on $\mathbb{P} \times \mathbb{B}_{norm}$ by:

$$S^{ext}_{normlog,g}(P,B) := -\sum_{\pi\in\Pi} g(\pi) \sum_{X\in\pi} P(X) \cdot \log(B(X)) \tag{66}$$

$$= \sum_{X\subseteq\Omega} P(X) \cdot \Big(\sum_{\substack{\pi\in\Pi \\ X\in\pi}} g(\pi)\Big) \cdot \log(B(X)) \ . \tag{67}$$

Proposition 9.2. *[29, Corollary 3, p. 3542] $S^{ext}_{normlog,g}(P,B)$ is strictly $\mathbb{B}_{norm}$-proper for all such g.*

Note that since $\mathbb{P} \subset \mathbb{B}_{norm}$, strict $\mathbb{B}_{norm}$-propriety is well defined in the sense of Definition 3.2.

The above proposition does not contradict Theorem 8.4, since we here consider normalised belief functions in $\mathbb{B}_{norm}$ while Theorem 8.4 concerns belief functions in $\mathbb{B}$.

The SRs S^{ext}_{llog} and $S^{ext}_{normlog,g}$ rely on the same idea: The main culprit in the impossibility Theorem 8.4 is that in (59) there is no interaction between the degrees of belief in different events. Normalising beliefs re-introduces such an interaction. The main structural difference between the two SRs is how normalisation is achieved. The former SR, S^{ext}_{llog}, introduces a penalty (i.e. normalisation) term into the SR, for the latter SR, $S^{ext}_{normlog,g}$, one presupposes normalised belief functions.

10. Conclusion

In the first part of this paper we saw how to use statistical SRs to justify *Prob*. In this second part we demonstrated the usefulness of statistical SRs for justifications of further norms of rational belief formation. In particular, we saw how an agent's evidence can be naturally taken into account by applying WCEL avoidance as DTN.

Logarithmic SRs occupy a prominent place in the literature as protagonists in Savage's theorem and objective Bayesianism. We hence set out to investigate how to construct statistical logarithmic SRs which are strictly $\mathbb{B}$-proper. We found three such logarithmic SRs (Proposition 5.7, Proposition 9.1 and Proposition 9.2).

Ideas from the epistemic and the statistical approach have been influential in the development of this paper. Looking into the future, pulling strands from both approaches together appears to have the potential to be beneficial for both approaches. Generally speaking, extending Richard Pettigrew's Epistemic Utility Theory Programme [42, 48] to statistical SRs appears to be a research avenue holding great promise. We thus hope for many more exciting entries to be added to Table 1.

Unfortunately, we did not answer all the questions we raised. Hopefully, future work will solve the problems left open in this paper.

Acknowledgements. I would like to thank the anonymous referees for their help and XX for helpful comments. I am also grateful to the UK Arts and Humanities Research Council for funding this research.

Decision Theoretic Norm	Inaccuracy Measures	Scoring Rules
Dominance w.r.t. $\omega \in \Omega$	[11], [49], [24], [25],[43], [44]	[55] [56]
Dominance w.r.t. $P \in \mathbb{P}$		Corollary 5.4, Theorem 6.2
Expected Loss w.r.t. Bel^*		Belief Elicitation
Worst-Case Loss w.r.t. $\omega \in \Omega$	[46]	
WCEL w.r.t. $P \in \mathbb{E}$		Theorems 7.1, 7.2, 7.5 [20], [29]

Table 1: Combinations of IMs and SRs with DTNs

References

[1] Owen Barnett and Jeff B. Paris. Maximum Entropy Inference with Quantified Knowledge. *Logic Journal of IGPL*, 16(1):85–98, 2008.

[2] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. Wiley, 2 edition, 2000.

[3] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1):1–3, 1950.

[4] Lara Buchak. Risk and Tradeoffs. *Erkenntnis*, 79(6 (Supplement)):1091–1117, 2014.

[5] Michael Caie. Rational Probabilistic Incoherence. *Philosophical Review*, 122(4):527–575, 2013.

[6] R. T. Cox. Probability, Frequency and Reasonable Expectation. *American Journal of Physics*, 14(1):1–13, 1946.

[7] Imre Csiszár. Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems. *The Annals of Statistics*, 19(4):2032–2066, 1991.

[8] Imre Csiszàr. Axiomatic Characterizations of Information Measures. *Entropy*, 10(3):261–273, 2008.

[9] Daryl J. Daley and David Vere-Jones. Scoring Probability Forecasts for Point Processes: The Entropy Score and Information Gain. *Journal of Applied Probability*, 41:297–312, 2004.

[10] Alexander Philip Dawid. Probability forecasting. In Samuel Kotz and Norman Lloyd Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 7, pages 210–218. Wiley, 1986.

[11] Bruno de Finetti. *Theory of Probability*. Wiley, 1974.

[12] Bruno de Finetti. Foresight: Its logical laws, its subjective sources. In Henry Ely Kyburg and Howard Edward Smokler, editors, *Studies in Subjective Probability*, pages 53–118. Krieger, 2 edition, 1980.

[13] Kenny Easwaran and Branden Fitelson. An "Evidentialist" Worry About Joyce's Argument for Probabilism. *Dialectica*, 66(3):425–433, 2012.

[14] Don Fallis. Attitudes toward Epistemic Risk and the Value of Experiments. *Studia Logica*, 86(2):215–246, 2007.

[15] Andrey Feuerverger and Sheikh Rahman. Some aspects of probability forecasting. *Communications in Statistics - Theory and Methods*, 21(6):1615–1632, 1992.

[16] Allan Gibbard. Rational Credence and the Value of Truth. In Tamar Szabó Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology: Volume 2*, chapter 6, pages 143–164. Oxford University Press, 2007.

[17] Tilmann Gneiting and Adrian E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

[18] Hilary Greaves. Epistemic Decision Theory. *Mind*, 122(488):915–952, 2013.

[19] Hilary Greaves and David Wallace. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility. *Mind*, 115(459):607–632, 2006.

[20] Peter D. Grünwald and A.Philip Dawid. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics*, 32(4):1367–1433, 2004.

[21] Alan Hájek. Arguments for - or against - Probabilism? *British Journal for the Philosophy of Science*, 59(4):793–819, 2008.

[22] Colin Howson. Probability and logic. *Journal of Applied Logic*, 1(3-4):151–165, 2003.

[23] Edwin T Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003.

[24] James M. Joyce. A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65(4):575–603, 1998.

[25] James M. Joyce. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In Franz Huber and Christoph Schmidt-Petri, editors, *Degrees of Belief*, volume 342 of *Synthese Library*, pages 263–297. Springer, 2009.

[26] Leonhard Knorr-Held and Evi Rainer. Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. *Biostatistics*, 2(1):109–129, 2001.

[27] Jason Konek and Ben Levinstein. The Foundations of Epistemic Decision Theory.

[28] Heinz König. A general minimax theorem based on connectedness. *Archiv der Mathematik*, 59:55–64, 1992.

[29] Jürgen Landes and Jon Williamson. Objective Bayesianism and the maximum entropy principle. *Entropy*, 15(9):3528–3591, 2013.

[30] Hannes Leitgeb and Richard Pettigrew. An Objective Justification of Bayesianism I: Measuring Inaccuracy. *Philosophy of Science*, 77(2):201–235, 2010.

[31] Hannes Leitgeb and Richard Pettigrew. An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy. *Philosophy of Science*, 77(2):236–272, 2010.

[32] Benjamin Anders Levinstein. Leitgeb and Pettigrew on Accuracy and Updating. *Philosophy of Science*, 79(3):413–424, 2012.

[33] David Lewis. A Subjectivist's Guide to Objective Chance. In Richard C. Jeffrey, editor, *Studies in Inductive Logic and Probability*, volume 2, chapter 13, pages 263–293. Berkeley University Press, 1980.

[34] Dennis V. Lindley. Scoring rules and the inevitability of probability. *International Statistical Review / Revue Internationale de Statistique*, 50(1):1–11, 1982.

[35] Patrick Maher. Joyce's Argument for Probabilism. *Philosophy of Science*, 69(1):pp. 73–81, 2002.

[36] Edgar C. Merkle and Mark Steyvers. Choosing a Strictly Proper Scoring Rule. *Decision Analysis*, 10(4):292–304, 2013.

[37] Greg Novack. A Defense of the Principle of Indifference. *Journal of Philosophical Logic*, 39(6):655–678, 2010.

[38] Graham Oddie. Conditionalization, Cogency, and Cognitive Value. *The British Journal for the Philosophy of Science*, 48(4):533–541, 1997.

[39] Theo Offerman, Joep Sonnemans, Gijs Van De Kuilen, and Peter P. Wakker. A Truth Serum for Non-Bayesians: Correcting Proper Scoring Rules for Risk Attitudes. *The Review of Economic Studies*, 76(4):1461–1489, 2009.

[40] Jeff B. Paris. Common Sense and Maximum Entropy. *Synthese*, 117:75–93, 1998.

[41] Jeff B. Paris. *The Uncertain Reasoner's Companion: A Mathematical Perspective*, volume 39 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2 edition, 2006.

[42] Richard Pettigrew. Epistemic Utility Arguments for Probabilism. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Stanford University, winter 2011 edition, 2011.

[43] Richard Pettigrew. Accuracy, Chance, and the Principal Principle. *The Philosophical Review*, 121(2):241–275, 2012.

[44] Richard Pettigrew. A New Epistemic Utility Argument for the Principal Principle. *Episteme*, 10:19–35, 2 2013.

[45] Richard Pettigrew. Accuracy and Evidence. *Dialectica*, 67(4):579–596, 2013.

[46] Richard Pettigrew. Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research*, n/a:n/a, 2014. Article first published online: 24 MAR 2014.

[47] Richard Pettigrew. What Chance-Credence Norms Should Not Be. *Noûs*, 49(1):177–196, 2015.

[48] Richard Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, forthcoming.

[49] J.B. Predd, R. Seiringer, E.H. Lieb, D.N. Osherson, H.V. Poor, and S.R. Kulkarni. Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792, 2009.

[50] F.P. Ramsey. Truth and probability. *Histoy of Economic Thought Chapters*, pages 156–198, 1926.

[51] Biagio Ricceri. Recent Advances in Minimax Theory and Applications. In Altannar Chinchuluun, PanosM. Pardalos, Athanasios Migdalas, and Leonidas Pitsoulis, editors, *Pareto Optimality, Game Theory And Equilibria*, volume 17 of *Optimization and Its Applications*, pages 23–52. Springer, 2008.

[52] Mark S. Roulston and Leonard A. Smith. Evaluating Probabilistic Forecasts Using Information Theory. *Monthly Weather Review*, 130(6):1653–1660, 2002.

[53] Leonard Jimmie Savage. *The Foundations of Statistics*. Dover Publications, 1954.

[54] Leonard Jimmie Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[55] Mark J. Schervish. A General Method for Comparing Probability Assessors. *The Annals of Statistics*, 17(4):1856–1879, 1989.

[56] Mark J. Schervish, Teddy Seidenfeld, and Joseph B. Kadane. Proper Scoring Rules, Dominated Forecasts, and Coherence. *Decision Analysis*, 6(4):202–221, 2009.

[57] Reinhard Selten. Axiomatic characterization of the quadratic scoring rule. *Experimental Economics*, 1:43–62, 1998.

[58] Stephen Simons. Minimax Theorems and Their Proofs. In Ding-Zhu Du and PanosM. Pardalos, editors, *Minimax and Applications*, volume 4 of *Nonconvex Optimization and Its Applications*, pages 1–23. Springer, 1995.

[59] F. Topsøe. Information theoretical optimization techniques. *Kybernetika*, 15:1–27, 1979.

[60] Steven J. van Enk. Bayesian Measures of Confirmation from Scoring Rules. *Philosophy of Science*, 81(1):101–113, 2014.

[61] Jonathan Weisberg. You,ve Come a Long Way, Bayesians. *Journal of Philosophical Logic*, pages 1–18, 2015. early view.

[62] Roger White. Evidential Symmetry and Mushy Credence. In T. Szabo Gendler and J. Hawthorne, editors, *Oxford Studies in Epistemology*, volume 3, pages 161–186. Oxford University Press, 2009.

[63] Jon Williamson. *In Defence of Objective Bayesianism*. Oxford University Press, 2010.

[64] Robert L. Winkler. Scoring Rules and the Evaluation of Probability Assessors. *Journal of the American Statistical Association*, 64(327):1073–1078, 1969.

[65] Robert L. Winkler, Victor Richmond R. Jose, James J. Cochran, Louis A. Cox, Pinar Keskinocak, Jeffrey P. Kharoufeh, and J. Cole Smith. Scoring Rules. In *Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., 2010.

[66] Robert L. Winkler, Javier Muñoz, José Cervera, José Bernardo, Gail Blattenberger, Joseph Kadane, Dennis Lindley, Allan Murphy, Robert Oliver, and David Ríos-Insua. Scoring Rules and the Evaluation of Probabilities. *TEST*, 5:1–60, 1996.