Strictly Proper Scoring Rules

Jürgen Landes

May 28, 2014

Abstract
Epistemic scoring rules are the en vogue tool for justifications of the probability norm and further norms of rational belief formation. They are different in kind and application from statistical scoring rules from which they arose. In the first part of the paper I argue that statistical scoring rules, properly understood, are in principle better suited to justify the probability norm than their epistemic brethren. Furthermore, I give a justification of the probability norm applying statistical scoring rules. In the second part of the paper I give a variety of justifications of norms for rational belief formation employing statistical scoring rules. Furthermore, general properties of statistical scoring rules are investigated. Epistemic scoring rules feature as a useful technical tool for constructing statistical scoring rules.

Keywords Scoring rules, probability norm, strict propriety, entropy, principle of indifference, rational belief formation.

Contents

Introduction and Notation 2

1 Introduction 2

2 Notation 4

Part 1 4

3 The Statistical Approach 4
3.1 Statistical Scoring Rules, Applications and Interpretations 4
3.2 Strict Propriety for statistical Scoring Rules 6

4 The Epistemic Approach 8
4.1 The main Ingredients 8
4.2 The Justifications 10
4.3 Against the use of strictly proper epistemic Scoring Rules 12

5 Extended statistical Scoring Rules **13**

6 Connecting epistemic and extended Scoring Rules **15**

7 A Justification of the Probability Norm with statistical Scoring Rules **17**

8 The Bayesian Credo **19**

9 Discussion **21**

Part 2 **23**

10 Maximum Entropy Principles **23**
10.1 The general Arguments 23
10.2 Generalised Entropies 25

11 Justifying the Principle of Indifference **26**

12 Local Scoring Rules **32**
12.1 Locality and strict $\mathbb{P}$-propriety 32
12.2 Locality, strict $\mathbb{B}$-propriety and extended Scoring Rules 34

13 Locality and strictly $\mathbb{B}$-proper extended Scoring Rules **35**
13.1 Penalties 36
13.2 Normalizing Beliefs 38

14 Conclusion **39**

References **39**

Appendix **44**

15 Proof of Proposition 13.1 **44**

16 Rational Belief Formation as Mechanism Design **46**

Introduction and Notation

1. Introduction

Bayesians agree on one basic norm for rational belief formation

Probability Norm: Any rational agent's subjective belief function ought to satisfy the axioms of probability. (PN)

The question arises as to how to justify this norm. Traditionally, axiomatic justifications [6, 38], justifications on logical grounds [23] and Dutch Book Arguments [13, 47] 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 [22].[1]

Recent work in epistemology takes a non-pragmatic approach using epistemic *Scoring Rules* (SRs) to *justify* the probability norm [25, 26, 30, 31, 46]. SRs first appeared in [5] as a tool to *elicit* probabilistic degrees of beliefs from forecasters. Brier's work has been highly influential in the statistical community which developed the notion of a *statistical* SR, these have made their way into the Encyclopedia of Statistics, see [10]. Epistemic SRs differ in form and application from statistical SRs.

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

In Theorem 6.1 we show how strictly proper epistemic SRs give rise to strictly proper statistical SRs in a canonical way. We demonstrate in Theorem 7.1 how so-constructed statistical SRs can be used to justify the PN.

We also briefly consider the consequences of applying a statistical SR which is not strictly proper (Section 8) and obtain unpalatable results in Proposition 8.1 and Proposition 8.2

In the second part of the paper we give a string of results demonstrating the usefulness of strictly proper statistical SRs for rational belief formation beyond justifications of the PN. In more detail, we show how to justify Maximum Entropy Principles (Theorem 10.1 and Theorem 10.2) and a probabilistic Principle of Indifference (Corollary 11.4 and Proposition 11.5).

The logarithmic statistical SR is well-known to be the only local statistical SR which is strictly proper, when applied to belief functions which satisfy the PN. Since we here do not presuppose the PN, we investigate notions of locality applying to statistical SRs for general belief functions (Section 12 and Section 13). We prove an impossibility result for such SRs in Theorem 12.4. Furthermore, we investigate how to weaken the assumption in the impossibility result to obtain strictly proper statistical SRs which are as local as possible in

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

Proposition 13.1 and Proposition 13.2.

Throughout, we use Brier Scores as a running example to illustrate differences and similarities between the epistemic and the statistical approach.

In the appendix we see how to frame the project of rational belief formation employing SRs from a higher, more abstract point of view.

2. Notation

Throughout, we work with a fixed non-empty, finite set Ω, which is interpreted as the set possible worlds or elementary events. The power set $\mathcal{P}\Omega$ may thus be understood as the set of propositions or events. We shall assume throughout that $|\Omega| \geq 2$ and also 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 $X = Y \cup Z$ and $Y \cap Z = \emptyset$, then $P(X) = P(Y) + P(Z)$. We shall use $P(\omega)$ as shorthand for $P(\{\omega\})$. The probability function $P_= \in \mathbb{P}$ defined by $P_=(\omega) := \frac{1}{|\Omega|}$ is called the equivocator.

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

The set of belief functions is the set of functions $bel : \mathcal{P}\Omega \to [0,1]$ and shall be denoted by $\mathbb{B}$. We shall throughout 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*. 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 $\omega \notin X$.

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., base e.

Part 1

3. The Statistical Approach

3.1. Statistical Scoring Rules, Applications and Interpretations

The statistical notion of a SR relies on the following betting scenario: given that a certain elementary event $\omega \in \Omega$ obtains an agent incurs a *loss* which depends on the agent's probabilistic beliefs. Formally, these losses are represented by a loss function $L : \Omega \times \mathbb{P} \to [0, +\infty]$. L is then referred to as a scoring rule. For a guide to the voluminous literature to statistical SRs refer to [19].

As is typical for statistical investigations, the existence of an objective probability function P^* is assumed, which is normally taken to be the distribution of a (or several) random variable(s). The existence of P^* allows the aggregation of losses incurred for different elementary events with respect to P^*. In order to build a general framework one defines a function which aggregates losses with respect to all $P \in \mathbb{P}$:

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

Statisticians virtually always only consider degrees of belief which satisfy the PN. 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 understand $L(\omega, bel)$ as only depending on the first argument, ω, and $\{bel(\omega) \,|\, \omega \in \Omega\}$.

We shall here be interested in *justifying* the PN. Hence, we will have to consider loss functions that also depend on degrees of belief in all non-elementary events $X \subseteq \Omega$. We thus introduce a loss function $L : \Omega \times \mathbb{B} \to [0, +\infty]$ and define the expected loss as

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

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 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.

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

The main areas of application for traditional SRs have been *belief elicitation* and the *assessment of beliefs*. We shall make a few remarks concerning belief elicitation applying SRs. The relevance of these remarks shall become clear in Section 4.3.

In the belief elicitation framework P in (1) is interpreted as the agent's *true* subjective probabilistic belief function, referred to as bel^*. In contrast, bel in (1) is interpreted as the belief function the agent chooses to announce. See [10, p. 211] for a definition in the encyclopedia of statistics and [17] for an overview.

So long as bel^* can be assumed to be in $\mathbb{P}$, the term $\sum_{\omega \in \Omega} bel^*(\omega) \cdot L(\omega, bel)$ can be interpreted as the loss the agent *expects* to incur upon announcing bel.

If bel^* is not a probability function, then there is no widely accepted interpretation of $\sum_{\omega \in \Omega} bel^*(\omega) \cdot L(\omega, bel)$. Thus, belief elicitation using a statistical SR without any further external grounds or assumptions as to why the agent's subjective belief function bel^* satisfies the PN appears to be less than fully satisfactory. For a stark example consider an agent with belief *zero* in every elementary event, i.e. $bel^*(\omega) = 0$ for all $\omega \in \Omega$. For this agent the score $S_L^{stats}(bel^*, bel)$ is *always* zero, independently of bel and independently of L. Thus, any announced belief function bel results in the exact same score. Hence, the agent has no incentive to truthfully announce $bel = bel^*$.

So, for the purposes of belief *elicitation* P may, under the assumption that $bel^* \in \mathbb{P}$, be interpreted as the agent's subjective belief function.

Belief elicitation is at heart an empirical problem, which is normally tackled by employing questionnaires, by conducting interviews and/or by observational studies (of subjects playing [incentive compatible] games). Verily, SRs have made their way into the applied sciences [37, 57]. See [18, Section 3] for a recent philosophical treatment of belief elicitation with statistical SRs.

Let us now consider what happens, if we interpret P as the agent's beliefs bel^* for the purpose of *justifying* of norms of belief formation. Any such justification would then be of the form:

> An agent with private beliefs $bel^* \in \mathbb{B}$ avoiding (some form of) expected loss with respect to bel^* ought to adopt a certain $bel \in \mathbb{B}$ [or some member of a set $B \subseteq \mathbb{B}$].

That is, an agent ought to adopt a belief function $bel \in \mathbb{B}$ which depends on her private beliefs $bel^* \in \mathbb{B}$. A highly circular argument indeed.

There is another major problem with interpreting P as bel^* for the purpose of justifying norms of belief formation. If $bel^*(\omega) = 0$ for all $\omega \in \Omega$, then an agent would be free to adopt any $bel \in \mathbb{B}$.

3.2. Strict Propriety for statistical Scoring Rules

We now turn to the key property for justifications of norms of rational belief formation.

Definition 3.1 (Strict $\mathbb{X}$-propriety). *A statistical SR S_L^{stats} is strictly $\mathbb{X}$-proper with $\mathbb{P} \subseteq \mathbb{X}$, if and only if for all $P \in \mathbb{P}$ the optimisation problem*

$$\begin{aligned} &\textit{minimise} \quad S_L^{stats}(P, bel) \\ &\textit{subject to} \quad bel \in \mathbb{X} \end{aligned}$$

is uniquely solved by $bel = P$.[2]

Following [51], a statistical SR S_L^{stats} is called merely $\mathbb{X}$-proper*, if and only if the following two conditions are satisfied: i) for all $P \in \mathbb{P}$ it holds that $P \in \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel)$. ii) There exists at least one $P \in \mathbb{P}$ and one $bel' \in \mathbb{B} \setminus \{P\}$ such that $\{P, bel'\} \subseteq \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel)$.*

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

[2]Our notion of strict $\mathbb{X}$-propriety differs from that of Γ-strictness of [21]. Γ constraints the domain of P, whereas $\mathbb{X}$ constraints the domain of the belief functions bel.

Note that the pathological SR S_L^{stats} defined as $S_L^{stats}(P, bel) := 0$ for all $P \in \mathbb{P}$ and all $bel \in \mathbb{X}$ is merely $\mathbb{X}$-proper. Unsurprisingly, the class of merely $\mathbb{X}$-proper SRs has received little attention in the literature.

Recall from when we introduced statistical SRs that losses are interpreted pragmatically as losses in a betting game. For our purposes we will interpret the function $L : \Omega \times \mathbb{B} \to [0, \infty]$ as an *inaccuracy measure*. The intended interpretation is that $L(\omega, bel)$ scores the inaccuracy of bel, in case ω obtains. 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 SR $S_L^{stats}(P, bel)$. If $S_L^{stats}(P, bel)$ is strictly $\mathbb{B}$-proper, then $bel = P$ is the unique belief function for which $S_L^{stats}(P, \cdot)$ is minimal. So, $bel = P$ is the unique function which minimises expected inaccuracy. On the other hand, if $S_L^{stats}(P, bel)$ is not strictly $\mathbb{B}$-proper, then there exists a $P \in \mathbb{P}$ and a $bel' \in \mathbb{B} \setminus \{P\}$ such that $bel' \in \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel)$. 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 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.

Ideally, one might think, rational agents aim for beliefs which track the truth rather than tracking objective 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 Bayesians have argued that some version of the Calibration Norm (cf. Section 8) applies to rational agents. If the set of probability functions calibrated to the agent's evidence contains a unique probability function P, then these Calibration Norms entail that adopting $bel = P$ is the unique rational belief function. Thus, proponents of Calibration Norms advocate the tracking of objective probabilities in such cases. We shall take it that these arguments are right and that rational agents aim at tracking objective probabilities.

Having established that minimising inaccuracy can be cached out as tracking objective probabilities, we now turn to motivating the idea that minimising expected inaccuracy is the rational thing to do.

Let us recall that S_L^{stats} is the *expectation* of the inaccuracy measure L where expectations are taken with respect to P. A rational agent forming beliefs on matters relevant to her will use those beliefs for other purposes a great number of times. Typically, formed beliefs may be used to make various decisions. Thus, having formed inaccurate beliefs is, in general, harmful to the agent more than once. In the long run of using once formed beliefs over and over again the sum of incurred inaccuracies tends with probability one to expected inaccuracies. So, a rational agent will aim to minimise *expected* inaccuracy, if she has good reasons to assume that the beliefs she now forms are going to be relevant to her later a great number of times.

So, applying strictly $\mathbb{B}$-proper statistical SRs for the purposes of rational belief formation is in fact a desideratum, if rational agents are assumed to avoid expected inaccuracies. As we have argued above, the avoidance of expected inaccuracies of formed beliefs is very much in line with how rational agents from beliefs.

The most famous and most widely used statistical SR is:

Definition 3.2 (Statistical Brier Score [5]). *The Brier Score S^{stats}_{Brier} takes the following form:*[3]

$$S^{stats}_{Brier}(P, bel) := \sum_{\omega\in\Omega} P(\omega)\cdot\Big((1 - bel(\omega))^2 + \sum_{\nu\in\Omega\setminus\{\omega\}} bel(\nu)^2\Big) \tag{3}$$

$$= 1 + \sum_{\omega\in\Omega} P(\omega)\cdot\Big(-2(bel(\omega)) + \sum_{\nu\in\Omega} bel(\nu)^2\Big) \ . \tag{4}$$

An axiomatic characterization of S^{stats}_{Brier} has been provided in [52].

$S^{stats}_{Brier}(P, bel)$ can visualised as the expectation (with respect to P) of the Euclidean distance on the elementary events of Ω between P and bel. While S^{stats}_{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. Thus, S^{stats}_{Brier} cannot be the SR of choice for rational belief formation approaches that do not presuppose the PN.

Strict $\mathbb{P}$-propriety, in contrast to strict $\mathbb{B}$-propriety, has been argued for by a number of authors in varying contexts. For instance by Selten [52, p. 44] for the purposes of assessing the predictive success of competing probabilistic theories and by Gneiting & Raftery [19] as well as Winkler [58, pp. 4] for the purposes of belief elicitation. Gibbard [18] advocated strict $\mathbb{P}$-propriety for rational belief formation, pre-supposing that degrees of belief are probabilistic.

4. The Epistemic Approach

4.1. The main 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.

The central notion we are here interested in is that of an epistemic SR:

[3] The original definition in [5] 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.

Definition 4.1 (Epistemic Scoring Rule). *Let L be a function $L : \mathcal{P}\Omega \times \{0,1\} \times [0,1] \to [0,\infty]$. For such a function L, an epistemic SR S_L^{epi} is a map $S_L^{epi} : \Omega \times \mathbb{B} \to [0,\infty]$ such that*

$$S_L^{epi}(\omega, bel) := \sum_{X \subseteq \Omega} L(X, v_\omega(X), bel(X)) \ . \qquad (5)$$

The epistemic score $S_L^{epi}(\omega, bel)$ is interpreted as inaccuracy of the agent's belief function bel at a possible world $\omega \in \Omega$ with respect to the function L. For a proposition $X \subseteq \Omega$ and possible world $\omega \in \Omega$, $L(X, v_\omega, bel(X))$ is construed as the inaccuracy of the belief $bel(X)$ relative to X being true or false at ω.

So, for a given world ω and a given belief function bel, S_L^{epi} sums the inaccuracies over all propositions 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).

The terminology in the literature has not yet converged, the function L has been called an (local) "inaccuracy measure" in [30, 43], whereas Predd et al. call L a SR and refer to S^{epi} as a "penalty function". Groves (private communications) refers to L 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 better understood set-up of Definition 4.1.

The key property for our discussion is:

Definition 4.2 (Strict Propriety). *A SR S_L^{epi} 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 $pL(X,1,x) + (1-p)L(X,0,x)$ is uniquely minimized by $x = p$*
- *$L(\Omega,1,x) + L(\emptyset,0,y)$ is uniquely minimised by $x = 1$ and $y = 0$.*

Some authors do not allow L to depend on X, see for instance [44, 46]. For such a loss function the condition on Ω and $\emptyset$ follows from the first condition. In general, the second condition is required because $P(\emptyset) = 0$ and $P(\Omega) = 1$ for all $P \in \mathbb{P}$.

In particular, if S_L^{epi} is strictly proper, then for all $\omega \in \Omega$ it holds that $S_L^{epi}(\omega, bel)$ is uniquely minimized by $bel = v_\omega$. So, if $\omega^* \in \Omega$ is the actual world, then the strictly least inaccurate function is v_{ω^*}. In this sense, strictly proper epistemic SRs track the actual world.

One intuition behind strict propriety is the following: If p is the chance that the real world ω is a member of X, then $1-p$ is the chance that the real world is in $\bar{X}$. So, for a strictly proper epistemic SR expected inaccuracy is uniquely minimised, iff $bel(X) = p$ and $bel(\bar{X}) = 1-p$. That is, whatever the actual chances are, it is best to hold beliefs matching the chances.

The above intuition has also been framed in terms of subjective degrees of belief rather than chances [44, p. 28]: "An epistemic SR is strictly proper, if and only if a probabilistic agent with credence p in a proposition expects that credence and only that credence to be least inaccurate with respect to L."

Strict propriety as a desideratum for epistemic SRs has been argued for in various contexts [30, Section 3 and 5], see also [15, 18, 20, 36]. We shall not advance further arguments for strict propriety here; in Section 4.3 we shall argue *against* the use of strictly proper epistemic SRs.

A natural condition on epistemic SRs is the following:

DEFINITION 4.3. *An epistemic SR S_L^{epi} is called* continuous*, if and only if L is continuous in $bel(X)$.*

Continuity is here taken in the usual sense: For all $X \subseteq \Omega$, for $i \in \{0,1\}$ and for any sequence $(bel_n(X))_{n\in\mathbb{N}}$ converging to $bel(X) \in [0,1]$ it holds that $\lim_{n\to\infty} L(X,i,bel_n(X)) = L(X,i,bel(X))$. Furthermore, continuity is here understood to be extended to $[0,+\infty]$.

The most popular epistemic SR is:

DEFINITION 4.4 (Epistemic Brier Score). *The epistemic Brier Score is defined as*

$$S_{Brier}^{epi}(\omega, bel) := \sum_{X\subseteq\Omega} (v_\omega(X) - bel(X))^2 \quad . \tag{6}$$

In other words: $S_{Brier}^{epi}(\omega, bel)$ is the distance between v_ω and bel where the distance is Euclidean distance in $\mathbb{R}^{|\mathcal{P}\Omega|}$. S_{Brier}^{epi} is strictly proper and continuous.

Compare the epistemic Brier Score (Definition 4.4) to the statistical Brier Score (Definition 3.2) and note that they differ in various respects. For instance, $S_{Brier}^{epi}(\omega, bel)$ depends on the entire belief function while $S_{Brier}^{stats}(P, bel)$ only depends on beliefs in elementary events. Furthermore, $S_{Brier}^{epi}(\omega, bel)$ is a tuple of real numbers (one number for each $\omega \in \Omega$), whereas $S_{Brier}^{stats}(P, bel)$ is a single real number.

Recently, quadratic inaccuracy measures, such as S_{Brier}^{epi}, 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.

4.2. The Justifications

In justifications of norms of rational belief formation employing epistemic SRs 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 $S_L^{epi}(\omega, bel)$ in different worlds? Surely, one could simply add the inaccuracies up, $\sum_{\omega\in\Omega} S_L^{epi}(\omega, bel)$. But why should one not multiply the inaccuracies, $\prod_{\omega\in\Omega} S_L^{epi}(\omega, bel)$, or consider the sum of the logarithms of the inaccuracies, $\sum_{\omega\in\Omega} \log(S_L^{epi}(\omega, bel))$? Apparently, there is no canonical way to aggregate the inaccuracies $S_L^{epi}(\omega, bel)$ for the possible worlds $\omega \in \Omega$.

The Decision Theoretic Norm (DTN) which springs to mind in this situation is dominance. The first justification of the PN applying dominance was:

THEOREM 4.5 (De Finetti [12]).

- *If* $bel \in \mathbb{B} \setminus \mathbb{P}$, *then there exists some* $P \in \mathbb{P}$ *such that* $S^{epi}_{Brier}(\omega, bel) > S^{epi}_{Brier}(\omega, P)$ *for all* $\omega \in \Omega$.
- *If* $bel \in \mathbb{P}$, *then there is no* $B \in \mathbb{B} \setminus \{bel\}$ *such that* $S^{epi}_{Brier}(\omega, bel) \geq S^{epi}_{Brier}(\omega, B)$ *for all* $\omega \in \Omega$.

De Finetti's result relies on the epistemic Brier Score to measure inaccuracy. Plausibly, there are other epistemic SRs which measure inaccuracy. Recently, the following theorem has been proved:

THEOREM 4.6 (Predd et al. [46]). *If* S^{epi}_L *is a continuous strictly proper epistemic SR, then:*

- *If* $bel \in \mathbb{B} \setminus \mathbb{P}$, *then there exists some* $P \in \mathbb{P}$ *such that* $S^{epi}_L(\omega, bel) > S^{epi}_L(\omega, P)$ *for all* $\omega \in \Omega$.
- *If* $bel \in \mathbb{P}$, *then there is no* $B \in \mathbb{B} \setminus \{bel\}$ *such that* $S^{epi}_L(\omega, bel) \geq S^{epi}_L(\omega, B)$ *for all* $\omega \in \Omega$.

Predd et al. credit Lindley [33] for a precursor of Theorem 4.6. Continuity strikes us as a sensible property an inaccuracy measure should satisfy. As we said above, we shall return to strict propriety in Section 4.3.

We have seen above that the epistemic Brier Score and the statistical Brier Score are fundamentally different creatures. It should therefore come as no surprise that statistical versions of de Finetti's result (Theorem 4.5) or Predd et al.'s result will require extra effort. Only the following result is immediate:

PROPOSITION 4.7. *Let* S^{stats}_L *be a strictly* $\mathbb{B}$*-proper SR. If* $bel \in \mathbb{P}$, *then there is no* $B \in \mathbb{B} \setminus \{bel\}$ *such that* $S^{stats}_L(P, bel) \geq S^{stats}_L(P, B)$ *for all* $P \in \mathbb{P}$.

PROOF. If $bel \in \mathbb{P}$, then $S^{stats}_L(bel, \cdot)$ is uniquely minimized by $bel = bel$. So, for $B \in \mathbb{P} \setminus \{bel\}$ we have $S^{stats}_L(bel, B) > S^{stats}_L(bel, bel)$. ■

The first, and more interesting, direction of Theorem 4.6 shall remain open for the moment:

> **Open Problem 1:** Let S^{stats}_L be a strictly $\mathbb{B}$-proper statistical SR. Under which conditions on S^{stats}_L does $bel \in \mathbb{B} \setminus \mathbb{P}$ imply that there exists some $P \in \mathbb{P}$ such that $S^{stats}_L(Q, bel) > S^{stats}_L(Q, P)$ for all $Q \in \mathbb{P}$?

The two other main justifications of the PN are due to Joyce, see [25] and [26]. Both justificaitons apply dominance as DTN in the same way as de Finetti and Predd et al.

The former justification does not require that the inaccuracy measure $f(\omega, bel)$ 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 convexity property has been objected to in [34, 18] and Gibbard also objected to the normality property. In his 2009 paper Joyce concedes that the objections raised have merit and that it would be best to do without these properties.

The latter justification [26] also does not require that the inaccuracy measure $f(\omega, bel)$ can be written as a sum over the propositions $X \subseteq \Omega$. It is only assumed that the inaccuracy measure f is continuous, finitely valued and satisfies truth-directedness and coherent admissibility. $f(\omega, bel)$ satisfies truth-directedness, iff (if $\omega \in \Omega$ and $bel, bel' \in \mathbb{B}$ are such that for all $X \subseteq \Omega$ it holds that $v_\omega(X) \leq bel(X) < bel'(X)$ or $v_\omega(X) \geq bel(X) > bel'(X)$, then $f(\omega, bel) < f(\omega, bel')$). $f(\omega, bel)$ satisfies coherent admissibility, iff there do not exist a $P \in \mathbb{P}$ and a $bel \in \mathbb{B}$ such that $f(\omega, P) \geq f(\omega, bel)$ for all $\omega \in \Omega$ and $f(\nu, P) > f(\nu, bel)$ for some $\nu \in \Omega$. [This version of coherent admissibility is taken from the erratum to the 2009 paper (p. 280) published on Joyce's website.]

Proponents of logarithmic SRs will object to the finiteness condition. It can also be argued that declaring all probability functions to be a priori admissible (coherent admissibility) singles out probabilistic belief functions as different in kind which is unfortunate for justifications of the PN. We feel that the main draw-back with the 2009 result is that it only applies for every partition of propositions and not for all propositions $X \subseteq \Omega$.

Over the last decade, a number of further results in the vein of de Finetti's theorem have proved. We shall mention [44], where the author proved that a Calibration Norm (cf. Section 8) may be justified in a similar way. A further result in this vein for conditional probabilities may be found in [51].

4.3. Against the use of strictly proper epistemic Scoring Rules

We now return to strict propriety as explained in [44]: "An epistemic SR is strictly proper, if and only if a probabilistic agent with credence p in a proposition expects that credence and only that credence to be least inaccurate with respect to L."[4] In other words, *if* S_L^{epi} is strictly proper and *if* the agent's belief function bel^* is in $\mathbb{P}$, *then* the agent expects bel^* and only bel^* to be least inaccurate.

From a purely technical standpoint, strictly proper epistemic SRs have a highly desirable property for axiomatic justifications of the PN: the second implication in Theorem 4.6 holds trivially.

However, advocating strict propriety on the grounds of Pettigrew's explanation would be arguing for treating probabilistic degrees of belief $bel \in \mathbb{P}$ differently from non-probabilistic degrees of belief $bel \in \mathbb{B} \setminus \mathbb{P}$. It would thus single out probabilistic degrees of belief as different in kind from the belief functions $bel \in \mathbb{B} \setminus \mathbb{P}$ and thereby undermine the intuitive appeal of the justification.

[4] Pettigrew only considers loss functions which do not depend on the proposition X. This technical detail is not relevant here.

Joyce [25, p. 589-590] makes a similar point on the justifications of strict propriety in axiomatic justifications of the PN relying on epistemic SRs.

We also want to remark that Pettigrew's formulation makes reference to an agent concerned with *expected* loss with respect to bel^*. In general, it is not clear why a rational agent aims to avoid expected losses where the expectation is taken with respect the agent's degrees of belief bel^*, unless the agent is (fairly) certain that bel^* is (a good approximation of) P^*. Finally, we remark, if one does not pre-suppose the PN, then there is no sensible way to interpret taking expectations with respect to $bel^* \in \mathbb{B} \setminus \mathbb{P}$ (recall the "stark example" from Section 3.1).

Thus, for our purposes of rational *belief formation* we would need a different motivation for strict propriety from the explanation given by Pettigrew. However, Pettigrew's formulation is apparently the only plausible explanation of strict propriety. Hence, assuming strict propriety for the purposes of rational belief formation is ill-advised.

Predd et al. [46, p. 4786] motivate strict propriety by "Our scoring rule thus encourages sincerity since your interest lies in announcing probabilities that conform to your beliefs." Interpreting their result in these terms, we lay bare the following structure: Because an agent's beliefs bel^* satisfy the PN (or to be more precise: bel^* satisfies $bel^*(X) + bel^*(\bar{X}) = 1$ for all $X \subseteq \Omega$) and because the SR is strictly proper an agent who avoids dominated belief functions will *announce* a belief function $bel \in \mathbb{P}$. That is, Predd et al. avoid the pitfall of giving a circular argument discussed in Section 3.1 by interpreting Theorem 4.6 as result concerning the assessment of forecasters as opposed to the formation of rational beliefs.

5. Extended statistical Scoring Rules

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

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

$$S_{L,ext}^{stats}(P, bel) = \sum_{X \subseteq \Omega} P(X) \cdot L(X, bel) \ , \tag{7}$$

for some loss 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 (7) is over all $X \subseteq \Omega$ and not merely over the $\omega \in \Omega$ as in (1). From now on, we shall mean by an *extended SR* an extended statistical SR.

For our running example the Brier Score we give the following definition:

DEFINITION 5.2 (Extended Brier Score).

$$
\begin{aligned}
S^{stats}_{Brier,ext}(P, bel) &:= \sum_{X \subseteq \Omega} P(X) \cdot \Big((1 - bel(X))^2 + bel(\bar{X})^2 \Big) && (8) \\
&= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subseteq \Omega \\ \omega \in X}} (1 - bel(X))^2 + \sum_{\substack{Y \subseteq \Omega \\ \omega \notin Y}} bel(Y)^2 \Big) && (9) \\
&= \sum_{\omega \in \Omega} P(\omega) \cdot S^{epi}_{Brier}(v_\omega, bel) \ . && (10)
\end{aligned}
$$

PROPOSITION 5.3. *$S^{stats}_{Brier,ext}$ is strictly $\mathbb{P}$-proper.*

PROOF. First note that

$$
\begin{aligned}
2 \cdot S^{stats}_{Brier,ext}(P, bel) = \sum_{X \subseteq \Omega} & P(X) \cdot \Big((1 - bel(X))^2 + bel(\bar{X})^2 \Big) \\
& + P(\bar{X}) \cdot \Big((1 - bel(\bar{X}))^2 + bel(X)^2 \Big) \ .
\end{aligned}
$$

To simplify notation let $x := bel(X)$ and $y := bel(\bar{X})$ and consider the following minimization problem for fixed $P \in \mathbb{P}$ and fixed $X \subseteq \Omega$

$$
\begin{aligned}
\text{minimize} \quad & P(X) \cdot ((1-x)^2 + y^2) + (1 - P(X)) \cdot ((1-y)^2 + x^2) \\
\text{subject to} \quad & x, y \in [0,1] \text{ and } x + y = 1 \ .
\end{aligned}
$$

After substituting $y = 1 - x$ the objective function becomes $P(X) \cdot (2 \cdot (1 - x)^2) + (1 - P(X)) \cdot (2x^2)$ which is equal to $2((P(X) - x)^2 - P(X)^2 + P(X))$. Thus, this optimization problem is uniquely solved by $x = P(X)$. Thus, every summand in $S^{stats}_{Brier,ext}(P, \cdot)$ is uniquely minimized by $bel(X) = P(X)$. Hence, $bel = P$ uniquely minimizes $S^{stats}_{Brier,ext}(P, \cdot)$. ■

In fact, the following stronger statement is true:

PROPOSITION 5.4. *$S^{stats}_{Brier,ext}$ is strictly $\mathbb{B}$-proper.*

PROOF. Consider the following minimization problem for fixed $P \in \mathbb{P}$ and fixed $X \subseteq \Omega$

$$
\begin{aligned}
\text{minimize} \quad & P(X) \cdot ((1-x)^2 + y^2) + (1 - P(X)) \cdot ((1-y)^2 + x^2) \\
\text{subject to} \quad & x, y \in [0,1] \ .
\end{aligned}
$$

Note that the objective function of the minimization 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 $S^{stats}_{Brier,ext}(P, \cdot)$. ■

Interestingly, we can prove a version of de Finetti's Theorem (Theorem 4.5) for statistical SRs:

THEOREM 5.5 (Statistical de Finetti Theorem).

- *If $bel \in \mathbb{B}\backslash\mathbb{P}$, then there exists some $P_{bel} \in \mathbb{P}$ such that $S^{stats}_{Brier,ext}(Q, bel) > S^{stats}_{Brier,ext}(Q, P_{bel})$ for all $Q \in \mathbb{P}$.*
- *If $bel \in \mathbb{P}$, then there is no $P \in \mathbb{P} \setminus \{bel\}$ such that $S^{stats}_{Brier,ext}(Q, bel) \geq S^{stats}_{Brier,ext}(Q, P)$ for all $Q \in \mathbb{P}$.*

PROOF. 1) Let $bel \in \mathbb{B}\backslash\mathbb{P}$, then by Theorem 4.5 there is a $P_{bel} \in \mathbb{P}$ such that for all $\omega \in \Omega$ it holds that $S^{epi}_{Brier}(v_\omega, bel) > S^{epi}_{Brier}(v_\omega, P_{bel})$. Using (10) and that $P_{bel}(\omega) > 0$ for some $\omega \in \Omega$ we have that $S^{stats}_{Brier,ext}(Q, bel) > S^{stats}_{Brier,ext}(Q, P_{bel})$ for all $Q \in \mathbb{P}$.
2) Follows from the Proposition 4.7 using that $S^{stats}_{Brier,ext}$ is strictly $\mathbb{B}$-proper (Proposition 5.4). ∎

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

6. Connecting epistemic and extended Scoring Rules

In this section we shall see how to canonically embed the class of epistemic SRs into the class of extended SRs. We shall give two examples to illustrate the embedding.

Let S^{epi}_L be an epistemic SR. Then we can define an associated extended SR $S^{stats,epi}_{ext}$ as:

$$
\begin{aligned}
S^{stats,epi}_{ext}(P, bel) &:= \sum_{\omega \in \Omega} P(\omega) \cdot S^{epi}_L(\omega, bel) && (11)\\
&= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subseteq \Omega \\ \omega \in X}} L(X, 1, bel(X)) + \sum_{\substack{Y \subseteq \Omega \\ \omega \notin Y}} L(Y, 0, bel(Y)) \Big) && (12)\\
&= \sum_{X \subseteq \Omega} P(X) \cdot L(X, 1, bel(X) + P(\bar{X}) \cdot L(X, 0, bel(X)) && (13)\\
&= \sum_{X \subseteq \Omega} P(X) \cdot \Big(L(X, 1, bel(X)) + L(\bar{X}, 0, bel(\bar{X})) \Big) \ . && (14)
\end{aligned}
$$

Recall that we assumed that L may depend on Ω. Thus, if L depends on X *and* Ω it may also depend on $\bar{X}$. The last equation then shows that we indeed defined an extended SR, since $\Big(L(X, 1, bel(X)) + L(\bar{X}, 0, bel(\bar{X})) \Big)$ is an extended loss function of the form $L_{ext}(X, bel(X), bel(\bar{X}))$. For the following calculations we shall mainly rely on (13).

THEOREM 6.1 (Canonical Embedding). *S^{epi}_L is strictly proper, if and only if $S^{stats,epi}_{ext}$ is strictly $\mathbb{B}$-proper.*

Proof. If S_L^{epi} is strictly proper, then every summand over $\emptyset \subset X \subset \Omega$ in

$$S_{ext}^{stats,epi}(P, bel) = \sum_{\emptyset \subset X \subset \Omega} P(X) \cdot (L(X, 1, bel(X)) + P(\bar{X}) \cdot L(X, 0, bel(X))$$

is uniquely minimised by $bel(X) = P(X)$. Furthermore, $L(\Omega, 1, bel(\Omega)) + L(\emptyset, 0, bel(\emptyset))$ is uniquely minimised by $bel(\Omega) = 1$ and $bel(\emptyset) = 0$. Thus, $S_{ext}^{stats,epi}(P, \cdot)$ is uniquely minimised by $bel = P$.

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

$$\begin{aligned}
S_{ext}^{stats,epi}(P, bel) &= \sum_{X \subseteq \Omega} P(X) \cdot L(X, 1, bel(X)) + P(\bar{X}) \cdot L(X, 0, bel(X)) \\
&= \sum_{\substack{U \subseteq \Omega \\ \omega, \omega' \in U}} 1 \cdot L(U, 1, bel(U)) + 0 \cdot L(U, 0, bel(U)) \\
&\quad + \sum_{\substack{W \subseteq \Omega \\ \omega, \omega' \notin W}} 0 \cdot L(W, 1, bel(W)) + 1 \cdot L(W, 0, bel(W)) \\
&\quad + \sum_{\substack{Y \subseteq \Omega \\ \omega \in Y,\, \omega' \notin Y}} p \cdot L(Y, 1, bel(Y)) + (1 - p) \cdot L(Y, 0, bel(Y)) \\
&\quad + \sum_{\substack{Z \subseteq \Omega \\ \omega' \in Z,\, \omega \notin Z}} (1 - p) \cdot L(Z, 1, bel(Z)) + p \cdot L(Z, 0, bel(Z)) \ .
\end{aligned}$$

Now observe that the belief function bel^+ minimising $S_{ext}^{stats,epi}(P, \cdot)$ minimises each of the four sums above individually, since every sum only depends on beliefs which no other sum depends on.

By considering the first two sums for $U = \Omega$ and $W = \emptyset$ we find that $L(\Omega, 1, bel^+(\Omega)) + L(\emptyset, 1, bel^+(\emptyset))$ 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 $p \cdot L((Y, 1, \cdot) + (1 - p) \cdot L(Y, 0, \cdot)$. By varying $P(\omega) = p$ we obtain that $bel^+(Y) = P(\omega)$ is the unique minimum of $p \cdot L(Y, 1, \cdot) + (1 - p) \cdot L(Y, 0, \cdot)$ 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 $p \cdot L(Y, 1, \cdot) + (1 - p) \cdot L(Y, 0, \cdot)$ for all $p \in [0, 1]$.

Thus, S_L^{epi} is strictly proper. ■

From a purely technical point of view, Theorem 6.1 can be most helpful. All one needs to do to check whether an extended SR is strictly $\mathbb{B}$-proper is to check whether the corresponding epistemic SR is strictly proper. The later task can be accomplished simply by checking whether simple sums are uniquely minimised by $bel(X) = p$ and $bel(\bar{X}) = 1-p$; checking whether an extended SR is strictly $\mathbb{B}$-proper requires one to solve a minimisation problem in $[0,1]^{|\mathcal{P}\Omega|}$. Furthermore, Theorem 6.1 puts us in a position to define strictly $\mathbb{B}$-proper extended SRs by using (14).

For our running example, Brier Scores, we already considered the canonical embedding in Definition 5.2. We now give two applications of Theorem 6.1. The epistemic logarithmic SR is well-known to be strictly proper (see, e.g., [26, Section 8]).

PROPOSITION 6.2. *The following extended SR is strictly $\mathbb{B}$-proper.*

$$\begin{aligned} S^{stats,epi}_{\log,ext}(P,bel) &:= \sum_{X\subseteq\Omega} P(X)\cdot\Big(-\log(bel(X)) - \log(1-bel(\bar{X}))\Big) \\ &= \sum_{\omega\in\Omega} P(\omega)\cdot\Big(-\sum_{\substack{X\subseteq\Omega\\ \omega\in X}} \log(bel(X)) - \log(1-bel(\bar{X}))\Big)\ . \end{aligned}$$

Another popular strictly proper epistemic SR is the spherical epistemic SR (see, e.g., [26, Section 8]). Using Theorem 6.1 we define a spherical extended strictly $\mathbb{B}$-proper SR.

PROPOSITION 6.3. *The following spherical extended SR is strictly $\mathbb{B}$-proper.*

$$\begin{aligned} S^{stats,epi}_{sph,ext}(P,bel) &:= \sum_{X\subseteq\Omega} P(X)\cdot \\ &\Big(1 + \frac{-bel(X)}{\sqrt{bel(X)^2 + (1-bel(X))^2}} + \frac{bel(\bar{X})-1}{\sqrt{bel(\bar{X})^2 + (1-bel(\bar{X}))^2}}\Big)\ . \end{aligned}$$

Theorem 6.1 allow us to generate strictly $\mathbb{B}$-proper extended SRs from strictly proper epistemic SRs. Although, we have argued that strictly proper epistemic SRs are an inadequate tool for rational belief formation without presupposing the PN they can nonetheless be used as a technically convenient tool to generate strictly $\mathbb{B}$-proper extended SRs.

Theorem 6.1 raises one, as of yet, open problem:

Open Problem 2: Is it true that for all strictly $\mathbb{B}$-proper SR S^{stats}_L there exists an epistemic SR $S^{epi}_{L'}$ such that

$$S^{stats}_L(P,bel) = \sum_{\omega\in\Omega} P(\omega)\cdot S^{epi}_{L'}(\omega,bel)\ ?$$

7. A Justification of the Probability Norm with statistical Scoring Rules

By proving the statistical version of de Finetti's theorem (Theorem 5.5) we demonstrated how to transfer a justification of the PN from the epistemic to the statistical approach. We now show how to use the main result from the previous section (Theorem 6.1) to transfer Predd et. al's justification (Theorem 4.6) to the statistical approach.

THEOREM 7.1 (Statistical Predd et al. Theorem). *Let* S_L^{epi} *be strictly proper and continuous.*

- *If* $bel \in \mathbb{B} \setminus \mathbb{P}$*, then there exists some* $P_{bel} \in \mathbb{P}$ *such that* $S_{ext}^{stats,epi}(Q, bel) > S_{ext}^{stats,epi}(Q, P_{bel})$ *for all* $Q \in \mathbb{P}$*.*
- *If* $bel \in \mathbb{P}$*, then there is no* $P \in \mathbb{P} \setminus \{bel\}$ *such that* $S_{ext}^{stats,epi}(Q, bel) \geq S_{ext}^{stats,epi}(Q, P)$ *for all* $Q \in \mathbb{P}$*.*

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 $S_L^{epi}(v_\omega, bel) > S_L^{epi}(v_\omega, P_{bel})$. For all $Q \in \mathbb{P}$ there exists some $\omega \in \Omega$ such that $Q(\omega) > 0$. We thus find that $S_{ext}^{stats,epi}(Q, bel) > S_{ext}^{stats,epi}(Q, P_{bel})$ for all $Q \in \mathbb{P}$.
2) By Theorem 6.1 $S_{ext}^{stats,epi}$ is strictly $\mathbb{B}$-proper. By Proposition 4.7 we now find that for $bel \in \mathbb{P}$ there does not exists a $P \in \mathbb{P} \setminus \{bel\}$ such that $S_{ext}^{stats,epi}(bel, bel) \geq S_{ext}^{stats,epi}(bel, P)$. ■

In the sense in which Theorem 4.6 subsumes Theorem 4.5, Theorem 7.1 subsumes Theorem 5.5.

Theorem 7.1 gives one answer to Open Problem 1 posed on Page 11. If S^{stats} is of the form $S_{ext}^{stats,epi}$ for a strictly proper and continuous epistemic SR S_L^{epi}, then the converse of Proposition 4.7 does hold.

Besides the assumptions that rational agents aim only at accurate beliefs and that inaccuracy may be measured by a SR S_L^{stats}, this justification of the PN rests on the following: A) The statistical SR S_L^{stats} is induced by an epistemic SR. B) S_L^{stats} is strictly $\mathbb{B}$-proper. C) Continuity of the loss function. D) Dominance as DTN.

In order to make this justification compelling A – D need to withstand critiques. If rational agents aim to have accurate beliefs, then B is the obvious condition the statistical SR needs to satisfy to encourage the tracking of objective probabilities (see Section 3.2 and see Propositions 8.1 and 8.2 for what happens for not strictly $\mathbb{B}$-proper SRs). If the answer to Open Problem 2 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 the image of the canonical embedding or give a proof of Theorem 7.1 that also applies for statistical SRs which are not in the image of the embedding. One such argument may be: the set of appropriate inaccuracy measures is (some subset of) the set of epistemic SRs. An appropriate measure of expected inaccuracy is thus a statistical SR in the image of the embedding.

Continuity is a fairly harmless technical condition. Again, as for A, it might be possible to prove Theorem 7.1 without assuming continuity. As far as we are aware, no-one has seriously objected to Dominance as DTN in this context.

Overall, we feel that Theorem 7.1 provides a significantly more convincing justification of the PN than the previous justifications using epistemic SRs.

Comparisons of Theorem 7.1 with other justifications of the PN, such as Dutch Book Arguments or axiomatic justifications, are well outside the scope of this paper.

In Section 4.3 we argued that strict propriety for epistemic SRs without presupposing the PN 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 necessary for convincing justifications of the PN, the upshot of Section 3.2 is that statistical SRs are in principle better suited than their epistemic brethren for such justifications. Proving Theorem 7.1 allows us to conclude the following: Not only are statistical SRs better-suited in principle, it is actually possible to give a general justification of the PN in the statistical framework.

8. The Bayesian Credo

Our line of thought so far has been that strict $\mathbb{B}$-propriety of statistical SRs is a desideratum for rational belief formation. Subsequently, we have focussed solely on such SRs. We now investigate some consequences of a applying a statistical SR which is *not* strictly $\mathbb{B}$-proper. We obtain unpalatable results in Propositions 8.1 and Proposition 8.2 for all statistical SRs which are not strictly $\mathbb{B}$-proper.

While the PN is an indispensable ingredient for the Bayesian approach there is another widely advocated norm:

> **Calibration Norm:** A rational agent's belief function $bel \in \mathbb{B}$ ought to satisfy all constraints imposed by her evidence $\mathcal{E}$. (CN)

The CN; also known as the Principal Principle, Straight Rule and Miller's Principle; may be construed in different ways such as: a) *bel* ought to be a member of the set $\mathbb{E}$ of probability functions consistent with $\mathcal{E}$,[5] b) *bel* ought to be a member of the convex hull of the closure of $\mathbb{E}$. We shall refer to the conjunction of the PN and the CN as the *Bayesian Credo*, whatever the particular form of the CN.

Certain brands of Bayesianism advocate adopting a particular belief function $bel \in \mathbb{P}$ consistent with $\mathcal{E}$. If this particular function *bel* is determined by some non-subjective mechanism, the resulting credo is an *objective Bayesian* one. For example, Williamson [55] advocates adopting a calibrated belief function which sufficiently equivocates between the basic propositions that one can express. We shall see how to justify objective Bayesian approaches in Section 10.

Let us now consider a situation where $\mathbb{E} = \{P\}$ and we apply a statistical SR S_L^{stats} which is strictly $\mathbb{B}$-proper. Then it holds that $bel = P$ is the unique expected inaccuracy minimiser and also the unique worst-case expected inaccuracy minimiser, where the worst case is taken with respect to the $P \in \mathbb{E}$. So, for expected inaccuracy minimisation as well as for worst-case expected inaccuracy

[5] We shall always assume that $\mathcal{E}$ is consistent and that $\mathbb{E}$ is not-empty.

minimisation $bel = P$ is the unique best option. Hence, a strictly $\mathbb{B}$-proper statistical SR forces a (worst-case) expected inaccuracy avoiding agent to match her degrees of belief to objective probabilities, if she knows these. In the simple case of $\mathbb{E} = \{P\}$ such an agent will satisfy the Bayesian Credo.

On the other hand we have:

Proposition 8.1 (Violation of the Bayesian Credo). *Let S_L^{stats} be merely $\mathbb{B}$-proper SR. Then there exist a $P \in \mathbb{P}$ and a $bel' \in \mathbb{B} \setminus \{P\}$ such that*

$$bel' \in \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel) \ . \tag{15}$$

If S^{stats} is neither merely strictly $\mathbb{B}$-proper nor strictly $\mathbb{B}$-proper, then there exists a $P \in \mathbb{P}$ such that

$$P \notin \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel) \ . \tag{16}$$

Proof. (15) follows directly from the definition of merely $\mathbb{B}$-proper (Definition 3.1).

For the second part of the proof note that since S_L^{stats} is neither strictly $\mathbb{B}$-proper nor merely $\mathbb{B}$-proper, there exist a $P \in \mathbb{P}$ and a $bel \in \mathbb{B} \setminus \{P\}$ such that $S_L^{stats}(P, bel) < S_L^{stats}(P, P)$. ■

Proposition 8.2 (Violation of the Bayesian Credo for worst cases). *Let S_L^{stats} be merely $\mathbb{B}$-proper SR. Then there exist an $\mathbb{E} \subset \mathbb{P}$ and a $bel' \in \mathbb{B} \setminus \mathbb{E}$ such that*

$$bel' \in \arg\inf_{bel \in \mathbb{B}} \sup_{Q \in \mathbb{E}} S_L^{stats}(Q, bel) \ . \tag{17}$$

If S^{stats} is neither merely strictly $\mathbb{B}$-proper nor strictly $\mathbb{B}$-proper, then there exists an $\mathbb{E} \subset \mathbb{P}$ such that

$$\mathbb{E} \cap \arg\inf_{bel \in \mathbb{B}} \sup_{Q \in \mathbb{E}} S_L^{stats}(Q, bel) = \emptyset \ . \tag{18}$$

Proof. Since S^{stats} is not strictly proper, there exist a $P \in \mathbb{P}$ and a $bel' \in \mathbb{B} \setminus \{P\}$ such that $S^{stats}(P, bel') \leq S^{stats}(P, P)$. Now simply put $\mathbb{E} := \{P\}$. Hence, $\arg\inf_{bel \in \mathbb{B}} \sup_{Q \in \mathbb{E}} S_L^{stats}(Q, bel) = \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel) \ni bel'$.

For the second part of the proof put $\mathbb{E} := \{P\}$ where P is such that there exists some $bel' \in \mathbb{B} \setminus \{P\}$ such that $S_L^{stats}(P, bel') < S_L^{stats}(P, P)$. Hence, $\arg\inf_{bel \in \mathbb{B}} \sup_{Q \in \mathbb{E}} S_L^{stats}(Q, bel) = \arg\inf_{bel \in \mathbb{B}} S_L^{stats}(P, bel)$. ■

It might be worth pointing out that (18) also holds when we allow $Q \in \mathbb{E}$ to take values in the closure of $\mathbb{E}$. This trivial fact follows since $\mathbb{E}$ and the closure of $\mathbb{E}$ are the same set.

Overall, we see that, if S_L^{stats} is not strictly $\mathbb{B}$-proper, then agents avoiding high (worst-case) inaccuracy can rationally to fail to match their degrees of belief to known objective probabilities. If S_L^{stats} is not even merely $\mathbb{B}$-proper, then a particular non-calibrated function has strictly better (worst-case) inaccuracy than the calibrated function.

9. Discussion

In sum: We showed how the epistemic approach to justifying the PN faces a serious difficulty (Section 4.3) and how the statistical approach avoids this difficulty (Section 3.2). In Theorem 7.1 we showed that the statistical approach not only avoids the problem but can also be used to give a more convincing justification of the PN.

The proponent of the use of epistemic SRs in justifications of the PN may be drawn to one of the following moves. Firstly, convincing justifications could be given that do not require the epistemic SRs to be strictly proper. I find this move very unlikely (but possible) to succeed.

Secondly, one might want to defend strict propriety on other grounds than Pettigrew's. However, Pettigrew's interpretation seems to capture the essence of strict propriety. Thus, such a defence will in all likelihood be less than convincing.

The third option to keep the epistemic approach alive is to head down the Joycean path and consider general measures of inaccuracy $f(\omega, bel)$. This can only make the class of inaccuracy measures larger. Any justification of the PN in this framework will thus have to cover more functions and thus be technically more challenging. Such a proof will then also apply to epistemic SRs in our sense. Hence, if strict propriety is indispensable for proofs applying epistemic SRs, then it is so in the Joycean framework.

On the other hand, changing horses from the epistemic to the statistical approach only requires the following two conditions being met. I) One subscribes to some notion of objective probabilities. II) One goes along with aggregating inaccuracies in terms of expected inaccuracies with respect to these objective probabilities.

Whether one could accept I) is well-outside the scope of this paper and shall remain unaddressed here. We shall however address a worry concerning II). It might be feared, that the DTNs acting on the $\omega \in \Omega$ in the epistemic approach cannot be transferred into the statistical framework because taking expectations aggregates inaccuracies and one thus looses relevant information. This worry is unfounded, since one can simply consider these DTNs acting on the set $\{P \in \mathbb{P} \mid \exists \omega \in \Omega : P(\omega) = 1\}$, cf. Section 11.

With the exception of Section 8, the discussion so far has been idealised to a considerable degree by not taking the agent's evidence into account. The charge that the epistemic approach does not properly treat the agent's evidence has already been laid in [14]. One 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 Section 10.

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^{stats} we can, at least in principle, approximate $S_L^{stats}(P^*, bel)$ by conducting i.i.d. trial runs. Thus, we do not need to have access to P^* to approximate $S_L^{stats}(P^*, bel)$. In the epistemic approach one assumes that there in an actual world ω^* among the $\omega \in \Omega$ but one does not know which possible world is the actual world. In practical terms it is thus not possible, even in principle, to compute $S_L^{epi}(v_{\omega^*}, bel)$.

Part 2

While the first part of this paper was devoted to the PN, we shall now turn to other norms of rational belief formation and to strictly $\mathbb{B}$-proper statistical SRs themselves. The logarithmic SR, which stands out as the only strictly $\mathbb{P}$-proper local SR, has received considerable attention in the literature. Subsequently, we will take a keen interest in notions of locality applying to statistical SRs.

To ease the reading, we shall later on drop the subscript $_L$ from S_L^{epi} and S_L^{stats}.

10. Maximum Entropy Principles

Adopting the calibrated probability function with maximal entropy has been vocally advocated by E. T. Jaynes, cf. [24]. Today, entropy maximisation is key to the most popular objective Bayesian approach [55]. In this section we shall show how this approach is justifiable using the statistical SR $S_{\log}^{stats}$ – presupposing the PN. We then note that such justifications also hold when we do not presuppose the PN, as long as we employ a SR which is strictly $\mathbb{B}$-proper and which is sufficiently regular.

In the second part of this section we shall briefly consider the consequences of employing the extended Brier score $S_{Brier,ext}^{stats}$, the extended spherical SR $S_{sph,ext}^{stats,epi}$ and the SR $S_{llog,ext}^{stats}$ in this way.

10.1. The general Arguments

Consider an agent with evidence $\mathcal{E}$ and the thereby induced set of calibrated functions $\mathbb{E} \subseteq \mathbb{P}$. 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. [55].[6] This norm is then spelled out in terms of the Maximum Entropy Principle:

> **Maximum Entropy Principle** A rational agent ought to adopt a probability function $bel \in \mathbb{E}$ which maximises Shannon Entropy (MaxEnt)
>
> $$H_{\log}(bel) := S_{\log}^{stats}(bel, bel) := \sum_{\omega \in \Omega} -bel(\omega) \log(bel(\omega)) \quad . \tag{19}$$

MaxEnt has given rise to a substantial literature on rational belief formation; as examples we mention [2, 7, 21, 24, 29, 39, 40]. MaxEnt is well-known to be justified on the following grounds of worst-case expected loss avoidance:

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

Theorem 10.1 (Justification of MaxEnt). *If $\emptyset \neq \mathbb{E} \subseteq \mathbb{P}$ is convex and closed, then*

$$\arg \inf_{bel \in \mathbb{P}} \sup_{P \in \mathbb{E}} S^{stats}_{\log}(P, bel) = \arg \sup_{P \in \mathbb{E}} H_{\log}(P) \tag{20}$$

and there is only one unique such function maximising Shannon Entropy.

Proof. First observe that the following holds:

$$\inf_{Q \in \mathbb{P}} \sup_{P \in \mathbb{E}} \sum_{\omega \in \Omega} -P(\omega) \log(Q(\omega)) = \sup_{P \in \mathbb{E}} \inf_{Q \in \mathbb{P}} \sum_{\omega \in \Omega} -P(\omega) \log(Q(\omega)) \tag{21}$$

$$= \sup_{P \in \mathbb{E}} \sum_{\omega \in \Omega} -P(\omega) \log(P(\omega)) \tag{22}$$

$$= \sup_{P \in \mathbb{E}} H_{\log}(P) \ . \tag{23}$$

(21) is a typical game-theoretic Mini-Max Theorem in the von Neumann mold, (22) follows directly from strict $\mathbb{P}$-propriety of $S^{stats}_{\log}$, which follows from Theorem 12.2. $H_{\log}(P)$ is a strictly concave function, thus the entropy maximiser is unique, which we shall denote by $P^\dagger$. Then, for $Q \in \mathbb{P} \setminus \{P^\dagger\}$ we find

$$\sup_{P \in \mathbb{E}} \sum_{\omega \in \Omega} -P(\omega) \log(Q(\omega)) \geq \sum_{\omega \in \Omega} -P^\dagger(\omega) \log(Q(\omega)) \tag{24}$$

$$> \sum_{\omega \in \Omega} -P^\dagger(\omega) \log(P^\dagger(\omega)) \ , \tag{25}$$

where the strict inequality follows from strict $\mathbb{P}$-propriety.

Thus, every $Q \in \mathbb{P} \setminus \{P^\dagger\}$ has strictly greater worst-case expected loss than $H_{\log}(P^\dagger)$. However, from (23) we know that there has to be a function $P \in \mathbb{P}$ which has worst-case expected loss equal to $H_{\log}(P^\dagger)$. Hence, $P^\dagger$ minimises worst-case expected loss and this worst-case expected loss equals $H_{\log}(P^\dagger)$. ■

The attentive reader will have noted that the above derivation can be generalised to arbitrary SRs S^{stats} as follows

$$\inf_{bel \in \mathbb{X}} \sup_{P \in \mathbb{E}} S^{stats}(P, bel) = \sup_{P \in \mathbb{E}} \inf_{bel \in \mathbb{X}} S^{stats}(P, bel) \tag{26}$$

$$= \sup_{P \in \mathbb{E}} S^{stats}(P, P) \tag{27}$$

as long as S^{stats} is sufficiently regular (to ensure that (26) holds) and $\mathbb{X}$-strictly proper (so (27) holds). If $\mathbb{E}$ is convex, closed and non-empty and if $S^{stats}(P, P)$ is strictly concave, then $\arg\sup_{P \in \mathbb{E}} \sum_{\omega \in \Omega} S^{stats}(P, P)$ is unique and in $\mathbb{E}$ and shall be denoted by $P^\ddagger$.

For $Q \in \mathbb{P} \setminus \{P^\ddagger\}$, we find the following result, using $\mathbb{B}$-strict propriety to obtain the strict inequality (29)

$$\sup_{P \in \mathbb{E}} S^{stats}(P, Q) \geq S^{stats}(P^\ddagger, Q) \tag{28}$$

$$> S^{stats}(P^\ddagger, P^\ddagger) \ . \tag{29}$$

Thus, the worst-case loss minimiser has to be $P^\ddagger$ and the worst-case expected loss for $P^\ddagger$ equals $S^{stats}(P^\ddagger, P^\ddagger)$. Following [21], we call $H(P) := S^{stats}(P, P)$ *generalised entropy*. We have thus proved the following theorem:

THEOREM 10.2 (Justification of Generalised Entropy Maximisation). *If $\emptyset \neq \mathbb{E} \subseteq \mathbb{P}$ is convex and closed, S^{stats} strictly $\mathbb{B}$-proper such that (26) holds, and if $H(P)$ is strictly concave, then*

$$\arg \inf_{bel \in \mathbb{B}} \sup_{P \in \mathbb{E}} S^{stats}(P, bel) = \{P^\ddagger\} = \arg \sup_{P \in \mathbb{E}} H(P) \ . \tag{30}$$

This then justifies the following principle:

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

Note that for convex, closed and non-empty $\mathbb{E}$, the Maximum Entropy Principle and the Generalised Entropy Maximisation Principle both satisfy the Bayesian Credo, because they both advocate adopting a calibrated probability function in $\mathbb{E}$.

10.2. Generalised Entropies

Theorem 10.2 gives general conditions under which generalised entropy maximisation is justified with respect to the choice of a particular statistical SR. From a structural point of view, it would be pleasing if the generalised entropy maximisers for different SRs were the same function. However, this is not the case as we shall now see. In this section, we shall not give the rather uninformative calculations but rather state the result of the calculations.

The following SRs satisfy the conditions in Theorem 10.2: the extended Brier score $S^{stats}_{Brier,ext}$, the extended spherical SR $S^{stats,epi}_{sph,ext}$ and $S^{stats}_{llog,ext} := -\frac{|\mathcal{P}\Omega|}{2} + \sum_{Y \subseteq \Omega} bel(Y) - \sum_{X \subseteq \Omega} P(X) \cdot \ln(bel(X))$. All three SRs are strictly $\mathbb{B}$-proper (see Proposition 5.4, Proposition 6.3 and Proposition 13.1).

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 $H_\Pi(P) := \sum_{X \subseteq \Omega} -P(X) \log(P(X))$ which we shall prove in Section 13.1. This entropy has already appeared in the literature and has been named *proposition entropy* [29]. Clearly, H_Π is strictly convex.

Note H_Π is different from Shannon Entropy. In H_Π the sum is over all events $X \subseteq \Omega$ and not over all elementary events $\omega \in \Omega$. Not only are Partition 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 are sufficiently regular, satisfying the minimax condition (26), follows for instance from König's result [28, p. 56]. The current state of the art of such minimax theorems is reviewed in [48] where König's theorem is also discussed.

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

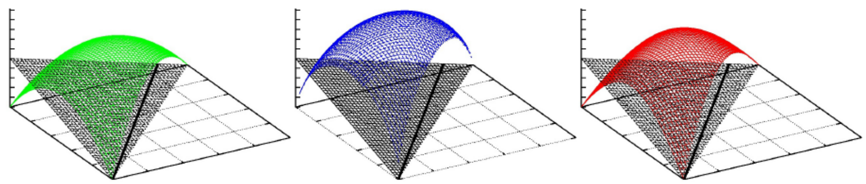

Figure 1. Brier Entropy H_{Brier} (green), Proposition Entropy H_Π (blue) and Spherical Entropy H_{Sph} (red) for $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Let $\mathbb{E}$ be the line segment between $P_1 = (1, 0, 0)$ and $P_2 = (0, \frac{5}{6}, \frac{1}{6})$ (black line segment).

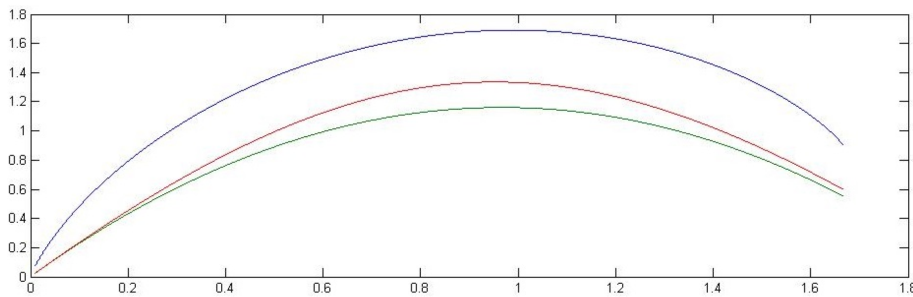

Figure 2. Brier Entropy H_{Brier} (green), Proposition Entropy H_Π (blue) and Spherical Entropy H_{Sph} (red) plotted along the line segment between $P_1 = (1, 0, 0)$ and $P_2 = (0, 5/6, 1/6)$ parametrised as $P_1 + t \cdot (-0.6, 0.5, 0.1)$ for $t \in [0, 10/6]$. The Brier Entropy maximiser is $P^\dagger_{Brier} = (0.4194, 0.4839, 0.0968)$ $[t = 0.968]$, the Proposition Entropy Maximiser is $P^\dagger_\Pi = (0.4054, 0.4955, 0.0991)$ $[t = 0.991]$ 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.

11. Justifying the Principle of Indifference

We shall now show how to use extended SRs to justify a probabilistic version of the Principle of Indifference (PoI). *Probabilistic* here means that our justification singles out the equivocator in the set of *probability* functions as the unique rational function, in case the agent does not possess any evidence. For a justification of the PoI not pre-supposing the PN see [45], this justification however relies on strictly proper epistemic SRs.

Clearly, it would be desirable to generalise our result to belief functions ranging in $\mathbb{B}$ or to follow Pettigrew but give up on the requirement of strict propriety.

Let us now consider an agent without evidence faced with the problem of assigning beliefs to sets of worlds $X_1, \ldots, X_k \subset \Omega$ for $k \geq 2$ such that $|X_1| =$

$|X_2| = \ldots = |X_k|$ and such that all X_i contain some fixed possible world $\omega \in \Omega$. Suppose the agent is required to assign beliefs such that $0 \leq bel(X_1) + ... + bel(X_k) \leq \alpha$ (and of course $0 \leq bel(X_i) \leq 1$), where α is some fixed value $\alpha \in (0, k] \subset \mathbb{R}$. If ω is the actual world, then an *entropic* epistemic SR makes the agent strictly best off for $bel(X_1) = \ldots = bel(X_k) = \frac{\alpha}{k}$.

On the other hand, if $Y_1, Y_2, \ldots, Y_k$ are non-empty sets of worlds of the same size which do not contain the actual world $\omega \in \Omega$, then an entropic epistemic SR would make an agent who assigns at least α strictly best off for $bel(Y_1) = \ldots = bel(Y_k) = \frac{\alpha}{k}$.

This idea of an entropic SR is made precise in the following definition:

Definition 11.1 (Entropic SR). *We call an epistemic SR* symmetric*, if and only if the loss function $L(X, v_\omega(X), bel(X))$ only depends on $v_\omega(X)$ and $bel(X)$. We shall write $L(v_\omega(X), bel(X))$. Such a SR is called an* entropy*, if and only if for all $k \geq 1$, all $\alpha \in (0, k]$, all $\langle x_1, \ldots, x_k\rangle \in [0,1]^k \setminus \langle \frac{\alpha}{k}, \frac{\alpha}{k}, \ldots, \frac{\alpha}{k}\rangle$ with $\sum_{i=1}^k x_i \leq \alpha$ and all $\langle y_1, \ldots, y_k\rangle \in [0,1]^k \setminus \langle \frac{\alpha}{k}, \frac{\alpha}{k}, \ldots, \frac{\alpha}{k}\rangle$ with $\sum_{i=1}^k y_i \geq \alpha$ it holds that*

$$\sum_{i=1}^k L(1, x_i) > \sum_{i=1}^k L(1, \frac{\alpha}{k}) = k \cdot L(1, \frac{\alpha}{k}) \quad \textit{and}$$
$$\sum_{i=1}^k L(0, y_i) > \sum_{i=1}^k L(0, \frac{\alpha}{k}) = k \cdot L(0, \frac{\alpha}{k}) \ .$$

The above technical definition will be required in the proof of Lemma 11.3. We shall now see that for loss functions which are twice continuously differentiable there is a simple condition which allows us to simply read-off whether S_L^{epi} is an entropy or not.[7]

Proposition 11.2. *Let S^{epi} be a symmetric, continuous, finitely-valued epistemic SR such that $L(1, \cdot)$ is strictly decreasing and $L(0, \cdot)$ is strictly increasing. If the first and second derivatives of $L(1, x)$ and $L(0, x)$ exist and are continuous on $(0, 1)$, then the following are equivalent*

- *S^{epi} is an entropy.*
- *$\frac{d^2}{dx^2} L(1, x) > 0$ for $0 < x < 1$ and $\frac{d^2}{dx^2} L(0, x) > 0$ for $0 < x < 1$.*

Proof. The proof is a simple exercise in calculus. We shall briefly sketch it, by noting that the following statements are logically equivalent:

- S^{epi} is an entropy.
- For all $x \in (0, 1)$ and all $\epsilon \in (0, 1)$ such $x + \epsilon \leq 1$ and $x - \epsilon \geq 0$ it holds that $L(1, x + \epsilon) + L(1, x - \epsilon) > 2L(1, x)$ and $L(0, x + \epsilon) + L(0, x - \epsilon) > 2L(0, x)$.

[7] In this entire section, we need not require that L is symmetric and we would still be able to prove our main results. Since this section is already quite heavy on notation and since adding this further complication does not prove to be illuminating, we shall here assume that L does not depend on X.

- $\frac{d^2}{dx^2}L(1,x) > 0$ for $0 < x < 1$ and $\frac{d^2}{dx^2}L(0,x) > 0$ for $0 < x < 1$.
- $L(1,x)$ and $L(0,x)$ are strictly convex on $[0,1]$.

■

In particular, we see that the epistemic Brier Score $S^{epi}_{Brier}(\omega, bel)$ is an entropy.

LEMMA 11.3. *If S^{epi} is strictly proper and an entropy, then*

$$\arg \inf_{bel \in \mathbb{P}} \sup_{P \in \mathbb{P}} S^{stats,epi}_{ext}(P, bel) = \{P_{=}\} \ . \tag{31}$$

PROOF. By definition we have

$$\begin{aligned}
&S^{stats,epi}_{ext}(P, bel) \\
&= L(1, bel(\Omega)) + L(\emptyset, bel(\emptyset)) + \sum_{\emptyset \subset X \subset \Omega} P(X) \cdot (L(1, bel(X)) + L(0, bel(\bar{X}))) \\
&= L(1, bel(\Omega)) + L(\emptyset, bel(\emptyset)) \\
&\quad + \sum_{\emptyset \subset X \subset \Omega} P(X) \cdot L(1, bel(X)) + (1 - P(X)) \cdot L(0, bel(X)) \ .
\end{aligned}$$

Since we assume that $bel \in \mathbb{P}$ we have $bel(\Omega) = 1$ and $bel(\emptyset) = 0$. Thus, only the terms with $\emptyset \subset X \subset \Omega$ depend on the belief function $bel \in \mathbb{P}$. Subsequently, we shall thus ignore the losses for $X = \emptyset$ and $X = \Omega$.

Now, for a fixed $bel \in \mathbb{P}$, maximising

$$\begin{aligned}
&\sum_{\emptyset \subset X \subset \Omega} P(X) \cdot \Big(L(1, bel(X)) + L(0, bel(\bar{X}))\Big) \\
&= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subset \Omega \\ \omega \in X}} L(1, bel(X)) + \sum_{\substack{\emptyset \subset Y \subset \Omega \\ \omega \notin Y}} L(0, bel(Y)) \Big)
\end{aligned} \tag{32}$$

is a linear optimisation problem in the variables $P(\omega)$ with $0 \leq P(\omega) \leq 1$ under the constraint $\sum_{\omega \in \Omega} P(\omega) = 1$. Since the optimisation problem is linear, the optimum obtains at one of the vertices of the feasible region (possibly, there are multiple optima, but at least one optimum obtains at a vertex). In order to compute the worst case expected loss for a fixed $bel \in \mathbb{B}$ it suffices to consider the probability functions which are the vertices of $\mathbb{P}$.

For fixed $\omega \in \Omega$, we put $P_\omega \in \mathbb{P}$ to be the unique function $P \in \mathbb{P}$ such that $P_\omega(\omega) = 1$. In the first part of this paper we denoted such a function by ν_ω. To highlight that this function now represents objective probabilities we write P_ω instead. The set of the vertices of $\mathbb{P}$ equals $\{P_\omega | \omega \in \Omega\}$.

Hence, the maximisation problem for a fixed $bel \in \mathbb{P}$ reduces to finding a/the $\omega \in \Omega$ which maximises the right hand side below

$$\sum_{\emptyset \subset X \subset \Omega} P(X) \cdot \Big(L(1, bel(X)) + L(0, bel(\bar{X}))\Big)$$

$$= \sum_{\substack{X\subset\Omega\\ \omega\in X}} L(1, bel(X)) + L(0, bel(\bar{X})) \ . \tag{33}$$

Now consider a $bel \in \mathbb{P} \setminus \{P_=\}$. Let $\lambda \in \Omega$ be such that $\lambda \in \arg\min_{\omega\in\Omega} bel(\omega)$. Using (33) we find (ignoring losses for $X = \emptyset$ and $X = \Omega$)

$$\sup_{P\in\mathbb{E}} S^{stats,epi}_{ext}(P, bel) \geq S^{stats,epi}_{ext}(P_\lambda, bel) \tag{34}$$

$$= \sum_{\substack{X\subset\Omega\\ \lambda\in X}} L(1, bel(X)) + L(0, bel(\bar{X})) \tag{35}$$

$$= \sum_{n=1}^{|\Omega|-1} \sum_{\substack{\emptyset\subset X\subset\Omega\\ \lambda\in X\\ |X|=n}} L(1, bel(X)) + \sum_{\substack{\emptyset\subset Y\subset\Omega\\ \lambda\notin Y\\ |Y|=|\Omega|-n}} L(0, bel(Y)) \ . \tag{36}$$

Since $bel \in \mathbb{P} \setminus \{P_=\}$ we have $bel(\lambda) < \frac{1}{|\Omega|}$. We shall next show that for all $1 \leq n \leq |\Omega| - 1$ that

$$\sum_{\substack{\emptyset\subset X\subset\Omega\\ \lambda\in X\\ |X|=n}} bel(X) < \sum_{\substack{\emptyset\subset X\subset\Omega\\ \lambda\in X\\ |X|=n}} P_=(X) \ .$$

This will put us in a position to use the assumption that S^{epi} is an entropy. Since bel is a probability function we have for $\{\lambda\} \subseteq X \subset \Omega$ that $bel(X) = bel(\lambda) + bel(X \setminus \{\lambda\})$. Hence,

$$\sum_{\substack{\emptyset\subset X\subset\Omega\\ \lambda\in X\\ |X|=n}} bel(X) = \binom{|\Omega|-1}{n-1} \cdot bel(\lambda) + \sum_{\substack{\emptyset\subseteq X\subset\Omega\setminus\{\lambda\}\\ |X|=n-1}} bel(X) \tag{37}$$

$$= \binom{|\Omega|-1}{n-1} \cdot bel(\lambda) + \sum_{\rho\in\Omega\setminus\{\lambda\}} (bel(\rho) \cdot \sum_{\substack{\emptyset\subseteq X\subseteq\Omega\setminus\{\lambda,\rho\}\\ |X|=n-2}} 1) \tag{38}$$

$$= \binom{|\Omega|-1}{n-1} \cdot bel(\lambda) + \sum_{\rho\in\Omega\setminus\{\lambda\}} \binom{|\Omega|-2}{n-2} \cdot bel(\rho) \tag{39}$$

$$= \binom{|\Omega|-1}{n-1} \cdot bel(\lambda) + \binom{|\Omega|-2}{n-2} \cdot (1 - bel(\lambda)) \tag{40}$$

$$= \binom{|\Omega|-1}{n-1} \cdot bel(\lambda) + \frac{(|\Omega|-2)!}{(n-2)!(|\Omega|-n)!} \cdot \frac{(|\Omega|-1)(n-1)}{(|\Omega|-1)(n-1)} \cdot (1 - bel(\lambda))$$

$$= \binom{|\Omega|-1}{n-1} \cdot bel(\lambda) + \binom{|\Omega|-1}{n-1} \cdot \frac{n-1}{|\Omega|-1} \cdot (1 - bel(\lambda)) \tag{41}$$

$$= \binom{|\Omega|-1}{n-1} \cdot \Big(bel(\lambda) + \frac{n-1}{|\Omega|-1} \cdot (1 - bel(\lambda))\Big) \tag{42}$$

$$= \binom{|\Omega|-1}{n-1} \cdot \Big(\frac{n-1}{|\Omega|-1} + \frac{bel(\lambda)\cdot(|\Omega|-1) - (n-1)\cdot bel(\lambda))}{|\Omega|-1}\Big) \tag{43}$$

$$= \binom{|\Omega|-1}{n-1} \cdot \Big(\frac{n-1}{|\Omega|-1} + \frac{bel(\lambda)\cdot(|\Omega|-n)}{|\Omega|-1}\Big) \tag{44}$$

$$< \binom{|\Omega|-1}{n-1} \cdot \Big(\frac{n-1}{|\Omega|-1} + \frac{|\Omega|-n}{|\Omega|\cdot(|\Omega|-1)}\Big) \tag{45}$$

$$= \binom{|\Omega|-1}{n-1} \cdot \Big(\frac{(n-1)\cdot|\Omega| + |\Omega| - n}{|\Omega|\cdot(|\Omega|-1)}\Big) \tag{46}$$

$$= \binom{|\Omega|-1}{n-1} \cdot \Big(\frac{n\cdot|\Omega| - n}{|\Omega|\cdot(|\Omega|-1)}\Big) \tag{47}$$

$$= \binom{|\Omega|-1}{n-1} \cdot \frac{n}{|\Omega|} \tag{48}$$

$$= |\{\emptyset \subset X \subset \Omega \mid |X| = n \,\&\, \lambda \in X\}| \cdot \frac{n}{|\Omega|} \tag{49}$$

$$= \sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} P_{=}(X) \ . \tag{50}$$

Since S^{epi} is an entropy we now infer that for all $1 \leq n \leq |\Omega| - 1$

$$\sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} L(1, bel(X)) > \sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} L(1, P_{=}(X)) \ . \tag{51}$$

Luckily, there is a shorter route to treat the $L(0, bel(X))$ terms in (36) which enables us to avoid doing the lengthy calculation above again for $L(0, bel(X))$. Since $bel \in \mathbb{P}$ we have for all $P \in \mathbb{P}$ and all $1 \leq n \leq |\Omega| - 1$ that

$$\begin{aligned}
\sum_{\substack{\emptyset\subset Y\subset\Omega \\ \lambda\notin Y \\ |Y|=|\Omega|-n}} bel(Y) + \sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} bel(X) &= \sum_{\substack{\emptyset\subset X\subset\Omega \\ |X|=n}} bel(X) + bel(\bar{X}) \\
&= \sum_{\substack{\emptyset\subset X\subset\Omega \\ |X|=n}} P(X) + P(\bar{X}) \\
&= \sum_{\substack{\emptyset\subset Y\subset\Omega \\ \lambda\notin Y \\ |Y|=|\Omega|-n}} P(Y) + \sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} P(X) \ .
\end{aligned}$$

Thus, for $P = P_{=}$ we find

$$\sum_{\substack{\emptyset\subset Y\subset\Omega \\ \lambda\notin Y \\ |Y|=|\Omega|-n}} bel(Y) = \sum_{\substack{\emptyset\subset Y\subset\Omega \\ \lambda\notin Y \\ |Y|=|\Omega|-n}} P_{=}(Y) + \sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} P_{=}(X) - \sum_{\substack{\emptyset\subset X\subset\Omega \\ \lambda\in X \\ |X|=n}} bel(X) \tag{52}$$

$$\stackrel{(51)}{>} \sum_{\substack{\emptyset \subset Y \subset \Omega \\ \lambda \notin Y \\ |Y| = |\Omega| - n}} P_=(Y) \ . \tag{53}$$

Using that S^{epi} is an entropy we now obtain for all $1 \leq n \leq |\Omega| - 1$ that

$$\sum_{\substack{\emptyset \subset Y \subset \Omega \\ \lambda \notin Y \\ |Y| = |\Omega| - n}} L(0, bel(Y)) > \sum_{\substack{\emptyset \subset Y \subset \Omega \\ \lambda \notin Y \\ |Y| = |\Omega| - n}} L(0, P_=(Y)) \ . \tag{54}$$

Now note that since S^{epi} is symmetric it holds that $S^{stats,epi}_{ext}(P_\rho, P_=) = S^{stats,epi}_{ext}(P_\omega, P_=)$ for all $\rho, \omega \in \Omega$. Thus, by linearity of the maximisation problem $S^{stats,epi}_{ext}(P_\lambda, P_=) = \sup_{P \in \mathbb{E}} S^{stats,epi}_{ext}(P, P_=)$.

Hence, for $bel \in \mathbb{P} \setminus \{P_=\}$

$$\begin{aligned}
\sup_{P \in \mathbb{E}} S^{stats,epi}_{ext}(P, bel) &\stackrel{(34)}{\geq} \sum_{n=1}^{|\Omega|-1} \sum_{\substack{\emptyset \subset X \subset \Omega \\ \lambda \in X \\ |X| = n}} L(1, bel(X)) + \sum_{\substack{\emptyset \subset Y \subset \Omega \\ \lambda \notin Y \\ |Y| = |\Omega| - n}} L(0, bel(Y)) \\
&\stackrel{(50)\&(54)}{>} \sum_{n=1}^{|\Omega|-1} \sum_{\substack{\emptyset \subset X \subset \Omega \\ \lambda \in X \\ |X| = n}} L(1, P_=(X)) + \sum_{\substack{\emptyset \subset Y \subset \Omega \\ \lambda \notin Y \\ |Y| = |\Omega| - n}} L(0, P_=(Y)) && (55) \\
&= S^{stats,epi}_{ext}(P_\lambda, P_=) && (56) \\
&= \sup_{P \in \mathbb{E}} S^{stats,epi}_{ext}(P, P_=) \ . && (57)
\end{aligned}$$

■

An analysis of the above proof unearths that we only considered the P_ω to compute worst-case expected losses. We thus find

COROLLARY 11.4. *[Probabilistic Principle of Indifference] If S^{epi} is strictly proper and an entropy and if $\mathbb{X} \subseteq \mathbb{P}$ is such that $\{P_\omega \mid \omega \in \Omega\} \subseteq \mathbb{X}$, then*

$$\arg \inf_{bel \in \mathbb{P}} \sup_{P \in \mathbb{X}} S^{stats,epi}_{ext}(P, bel) = \{P_=\} \ . \tag{58}$$

To satisfy our curiosity, we now compute the actual worst-case expected loss incurred upon adopting $bel = P_=$. For all P_ω we find

$$\begin{aligned}
S^{stats,epi}_{ext}(P_\omega, P_=) &= \sum_{\substack{X \subseteq \Omega \\ \omega \in X}} L(1, P_=(X)) + L(0, P_=(\bar{X})) \\
&= L(1,1) + L(0,0) + \sum_{\substack{X \subset \Omega \\ \omega \in X}} L(1, \frac{|X|}{|\Omega|}) + L(0, \frac{|\Omega| - |X|}{|\Omega|})
\end{aligned}$$

$$= L(1,1) + L(0,0) + \sum_{n=1}^{|\Omega|-1} \sum_{\substack{X\subset\Omega \\ \omega\in X \\ |X|=n}} L(1, \frac{n}{|\Omega|}) + L(0, \frac{|\Omega|-n}{|\Omega|})$$

$$= L(1,1) + L(0,0) + \sum_{n=1}^{|\Omega|-1} \binom{|\Omega|-1}{n-1} \cdot \Big(L(1, \frac{n}{|\Omega|}) + L(0, \frac{|\Omega|-n}{|\Omega|}) \Big) \ .$$

In particular, this is independent of ω. Thus,

$$\sup_{P\in\mathbb{E}} S^{stats,epi}_{ext}(P, P_{=}) = L(1,1) + L(0,0) + \sum_{n=1}^{|\Omega|-1} \binom{|\Omega|-1}{n-1} \cdot \Big(L(1, \frac{n}{|\Omega|}) + L(0, \frac{|\Omega|-n}{|\Omega|}) \Big) \ .$$

It is well-known that that MaxEnt implies the probabilistic Principle of Indifference. Interestingly, the extended SR induced by the epistemic logarithmic SR also allows for a justification of the probabilistic Principle of Indifference along the same lines.

Proposition 11.5. *[Logarithmic Probabilistic Principle of Indifference] For all $\mathbb{X} \subseteq \mathbb{P}$ such that $\{P \in \mathbb{P} \mid P(\omega) = 1$ for some $\omega \in \Omega\} \subseteq X$ it holds that*

$$\arg \inf_{bel\in\mathbb{P}} \sup_{P\in\mathbb{X}} S^{stats,epi}_{\log,ext}(P, bel) = \{P_{=}\} \ . \tag{59}$$

Proof. We need to show that $S^{stats,epi}_{\log,ext}$ is an entropy. This follows directly since $-\log(x)$ and $-\log(1-x)$ are strictly convex on $[0,1]$. ■

12. Local Scoring Rules

An important property of statistical SRs for belief functions in $\mathbb{P}$ is locality. For our purposes however, local strictly $\mathbb{P}$-proper statistical SRs are of little use, since they only take beliefs in elementary events into account. Beliefs in non-elementary events are *not* scored. After briefly reviewing the pertinent notions in the first part of this section, we shall study in the second part of this section how to extend the notion of locality to extended SRs. We shall see that the most natural way of extending the notion of locality is incompatible with strict $\mathbb{B}$-propriety.

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

Definition 12.1. *A statistical SR $S^{stats}_{L} : \mathbb{P} \times \mathbb{P} \to [0, +\infty]$ 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 shall write $L(bel(\omega))$.*

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

THEOREM 12.2 (Savage 1971). *Up to an affine-linear transformation, the only local and strictly $\mathbb{P}$-proper statistical SR is*

$$S_{\log}^{stats}(P, bel) := \sum_{\omega \in \Omega} -P(\omega) \log(bel(\omega)) \ . \tag{60}$$

Earlier versions of this theorem using stronger assumptions have appeared in [1, 35, 53]. Note that Savage does not require the SR to be continuous. See [3] for a version of this result for continuous probability densities.

Locality of statistical SRs has been argued for in a variety of settings. For example in [58, pp. 16] and [4, p. 72-73] for belief elicitation. The argument given is along the following lines; if an elementary event $\omega \in \Omega$ is guaranteed to obtain, then the loss incurred ought to only depend on the announced belief in ω.

We also want to mention that this logarithmic SR is the only strictly $\mathbb{P}$-proper statistical SR which is consistent with the use of likelihoods or log likelihoods to evaluate assessors, cf. [56, p. 1075].

Williamson defends in [55] $S_{\log}^{stats}$ as the only SR with a loss function $L : \Omega \times \mathbb{P} \longrightarrow [0, \infty]$ which satisfies the following four axioms:

L1 If $bel(\omega) = 1$, then $L(\omega, bel) = 1$.

L2 If $bel(\omega) > bel'(\omega)$, then $L(\omega, bel) < L(\omega, bel')$.

L3 $L(\omega, bel)$ is local, i.e. $L(\omega, bel)$ is a function of the form $L(bel(\omega))$.

L4 Losses are additive over independent sublanguages: For $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$ with $bel(\omega_1 \wedge \omega_2) = bel(\omega_1) \cdot bel(\omega_2)$ it holds that $L(\omega_1 \wedge \omega_2, bel) = L(\omega_1, bel) + L(\omega_2, bel)$.

See [29, p. 3538] for further motivation of these axioms.

The notion of locality in infinite continuous sample spaces has been extended to allow the loss function to also depend on the derivatives of $bel(X)$, see [41]. The same authors transfer their extended notion of locality to finite discrete sample spaces in the companion paper [11].

The statistical logarithmic SR $S_{\log}^{stats}$ has found applications in a variety of areas, for example in information theory [8, 49], Neyman-Pearson Theory in statistics [16] and the health sciences [27].

Recently, the epistemic logarithmic SR has been argued for by van Enk on the grounds that it yields a better measure of confirmation than the epistemic Brier Score [54, p. 108]. Another advantage of the logarithmic epistemic SR over the epistemic Brier Score came to light in [32], under a certain rule of conditionalisation (L & P conditionalisation to be exact) the epistemic Brier Score makes you believe in ghosts while the epistemic logarithmic SR does not.

Let us now consider a general local loss function $L : [0, 1] \rightarrow [0, +\infty]$ and the corresponding local SR $S_L^{stats} : \mathbb{P} \times \mathbb{B} \rightarrow [0, +\infty]$

$$S_L^{stats}(P, bel) = \sum_{\omega \in \Omega} P(\omega) \cdot L(bel(\omega)) \ . \tag{61}$$

Note that only beliefs in elementary events appear in the above expression. Thus, beliefs in non-elementary events will not affect the score $S_L^{stats}(P, bel)$. Thus, a DTN applying local statistical SR $S_L^{stats}(P, bel)$ 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 the PN. We thus now investigate how to extend the notion of locality, which proved to be technically fruitful when the PN was presupposed, when the PN is not presupposed.

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

One obvious way to generalise locality is:

DEFINITION 12.3. *An extended SR is called* ex-local*, iff there exists a loss function* $L_{loc} : \mathcal{P}\Omega \times [0,1] \to [0,\infty]$ *such that*

$$S^{stats}_{L_{loc},ext}(P, bel) = \sum_{X \subseteq \Omega} P(X) \cdot L_{loc}(X, bel(X)) \tag{62}$$

$$= \sum_{\omega \in \Omega} P(\omega) \cdot \Big(\sum_{\substack{X \subseteq \Omega \\ \omega \in X}} L_{loc}(X, bel(X)) \Big) \ . \tag{63}$$

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 $bel(X)$.

This notion of an ex-local extended SR generalises *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 the elementary event itself.

If $S^{stats}_{L,ext}$ is ex-local, then the loss attributable to $bel(X)$ only enters once into (62). 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, $S^{stats}_{L_{loc},ext}(P, bel)$ does not depend on $bel(\emptyset)$. 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 somehow due to $P(\emptyset) = 0$ for all $P \in \mathbb{P}$. However, we shall now prove that this is not the case.

Let $\mathbb{B}^- := \{bel : \mathcal{P}\Omega \setminus \{\emptyset\} \to [0,1]\}$ and define strict $\mathbb{B}^-$-propriety of a SR S in the obvious way, i.e., for all $P \in \mathbb{P}$ it holds that $\arg\inf_{bel \in \mathbb{B}^-} S(P, bel) = \{P\}$.

THEOREM 12.4. *There does not exists an ex-local extended strictly* $\mathbb{B}^-$*-proper SR* $S^{stats}_{L_{loc},ext}$.

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

$$\arg\min_{bel \in \mathbb{B}^-} S^{stats}_{L_{loc},ext}(P, bel) = \arg\min_{bel \in \mathbb{B}^-} \sum_{X \subseteq \Omega} P(X) \cdot L_{loc}(X, bel(X)) \tag{64}$$

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

For a fixed loss function L_{loc} and a fixed event $\emptyset \subset X \subseteq \Omega$ it holds that $\arg\min_{bel(X) \in [0,1]} L_{loc}(X, bel(X))$ 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 of $L_{loc}(X, bel(X))$ obtains independently of P.

Thus, $S^{stats}_{L,ext}(P, bel)$ is minimised, if and only if every summand in (64) is minimised. For each summand this minimum obtains independently of P. ∎

Corollary 12.5. *$S_{ext,log}(P, bel) := \sum_{X \subseteq \Omega} -P(X) \cdot \log(bel(X))$ is not strictly $\mathbb{B}^-$-proper.*

Proof. If $bel(X) = 1$ for all $\emptyset \subset X \subseteq \Omega$, then $S_{ext,log}(P, bel) = 0$. Since $S_{ext,log}$ is a map with range $[0, \infty]$, it follows that $bel(X) = 1$ for all $\emptyset \subset X \subseteq \Omega$ minimises $S_{ext,log}(P, \cdot)$ for all $P \in \mathbb{P}$. Clearly, $S_{ext,log}$ cannot be strictly $\mathbb{B}^-$-proper. ∎

Recall from Theorem 12.2 that the logarithmic SR $S^{stats}_{\log}$ 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 Proposition 6.2: $S^{stats,epi}_{\log,ext}(P, bel) := \sum_{X \subseteq \Omega} P(X) \cdot \Big(-\log(bel(X)) - \log(1 - bel(\bar{X}))\Big)$ is not ex-local. The loss term depends on $bel(X)$ and $bel(\bar{X})$. Thus, Proposition 6.2 does not contradict Theorem 12.4.

Note that $S^{stats}_{L_{loc},ext}(P_\omega, bel) = \sum_{X \subseteq \Omega, \omega \in X} L_{loc}(X, bel(X))$. That is, only beliefs in events containing ω are scored while beliefs in events which do not contain ω are entirely ignored. Clearly, any genuine measure of inaccuracy will have to take into account how $P(X)$ and $bel(X)$ relate for *all* $X \subseteq \Omega$. Thus, ex-local SRs cannot serve as measures of inaccuracy. Hence, the impossibility theorem only rules out the existence of SRs in which are unsuitable for justifications of norms of rational belief formation.

13. Locality and strictly $\mathbb{B}$-proper extended Scoring Rules

We saw in Theorem 12.4 that there are no ex-local strictly $\mathbb{B}$-proper extended SRs. The question arises how much of the locality condition we need to give up in order obtain strictly $\mathbb{B}$-proper extended SRs which are local, *in some sense*.

13.1. Penalties

As it turns out, there exists a logarithmic extended SR which is strictly $\mathbb{B}$-proper. However, the SR is not purely logarithmic since it contains a *penalty term*. In essence, this penalty term ($\sum_{Y\subseteq\Omega} bel(Y)$) prevents $bel \in \mathbb{B}$ defined as $bel(X) = 1$ for all $X \subseteq \Omega$ from being the score minimiser, since it inflicts a heavy penalty.

PROPOSITION 13.1. *The following extended SR is strictly* $\mathbb{B}$*-proper*

$$S^{stats}_{llog,ext}(P, bel) := \sum_{X\subseteq\Omega} P(X) \cdot \left(-1 + \frac{\sum_{Y\subseteq\Omega} bel(Y)}{\sum_{Y\subseteq\Omega} P(Y)} - \ln(bel(X))\right) \tag{65}$$

$$= -\frac{|\mathcal{P}\Omega|}{2} + \sum_{Y\subseteq\Omega} bel(Y) - \sum_{X\subseteq\Omega} P(X) \cdot \ln(bel(X)) \ . \tag{66}$$

PROOF. A direct, but rather long and technical, proof has been exiled and banned to the Appendix, cf. Section 15. ■

Recall that for $P \in \mathbb{P}$ we have $\sum_{Y\subseteq\Omega} P(Y) = \sigma$. Hence, for $bel \in \mathbb{P}$ we have

$$S^{stats}_{llog,ext}(P, bel) = -\sum_{X\subseteq\Omega} P(X) \cdot \ln(bel(X)) \tag{67}$$

$$= S_{ext,log}(P, bel) \ . \tag{68}$$

So, for $bel \in \mathbb{P}$ we recapture the SR considered in Corollary 12.5. Note that $S^{stats}_{llog,ext}(P, P) = -\sum_{X\subseteq\Omega} P(X) \cdot \ln(P(X))$.

$S^{stats}_{llog,ext}$ is not ex-local, since $L(X, bel)$ depends on $bel(X)$ *and* also on $\sum_{Y\subseteq\Omega} bel(Y)$. However, the loss term only depends on the belief in event X and the sum of beliefs taken over all $Y \subseteq \Omega$. The non-local term is constant for all $X \subseteq \Omega$. Calling such an extended SR *semi-local* we pose an interesting open problem:

> **Open Problem 3:** Is $S^{stats}_{llog,ext}$ the only extended semi-local strictly $\mathbb{B}$-proper SR (unique up to multiplication and addition of a constant)?

While Proposition 13.1 raises the above question it apparently allows us to answer the Open Problem 2 left open in Section 6 in the negative. $S^{stats}_{llog,ext}$ is strictly $\mathbb{B}$-proper but it does not appear to be an expectation of an epistemic SR à la (11). Furthermore, the factor in (65) multiplied by $P(X)$ depends not only on $bel(X)$ and $bel(\bar{X})$ but on all $bel(Y)$ for $\emptyset \subseteq Y \subseteq \Omega$, whereas in (14) the factor only depends on $bel(X)$ and $bel(\bar{X})$. Alternatively, the summand $\sum_{Y\subseteq\Omega} bel(Y)$ in (66) depends on the entire function bel, i.e. it is not independent of $bel(Y)$ for $X \neq Y \neq \bar{X}$. However, this only appears to solve the problem left open, but it does not as we shall now see.

Let $L(X,0,bel(X)) := bel(X)$ and $L(X,1,bel(X)) := bel(X) - 1 - \ln(bel(X))$ for the thereby induced epistemic SR S^{epi}_{llog} we find

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

It follows immediately from Theorem 6.1 that S^{epi}_{llog} is strictly proper. To ease the mind of a sceptical reader we shall now prove strict $\mathbb{B}$-propriety of $S^{stats}_{llog,ext}$ directly.

Proof. Let

$$\begin{aligned}
f(bel(X)) :&= p\cdot L(X,1,bel(X)) + (1-p)\cdot L(X,0,bel(X)) &(69)\\
&= p\cdot bel(X) - p - p\cdot\ln(bel(X)) + (1-p)\cdot bel(X) &(70)\\
&= -p - p\cdot\ln(bel(X)) + bel(X)\ . &(71)
\end{aligned}$$

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

$$\frac{d\,f(bel(X))}{dbel(X)} = -\frac{p}{bel(X)} + 1 = 0\ . \qquad (72)$$

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 have $f(bel(X)) = (1-p)\cdot L(X,0,bel(X)) = bel(X)$ which is uniquely minimised by $bel(X) = p = 0$. ■

So, $S^{stats}_{llog,ext}$ is in fact induced by a strictly proper epistemic SR and $S^{stats}_{llog,ext}$ is strictly $\mathbb{B}$-proper. We have thus not solved Open Problem 2.

Finally, let us remark that we now have a direct proof that S^{epi}_{llog} is strictly proper. Thus, we can use Theorem 6.1 to infer that $S^{stats}_{llog,ext}$ is strictly $\mathbb{B}$-proper. Thus, Proposition 13.1 follows from Theorem 6.1 and the fact that S^{epi}_{llog} is strictly proper. The technical proof of Proposition 13.1 in the appendix is thus not essential for our purposes. This then nicely illustrates the technical helpfulness of Theorem 6.1 to which we alluded to in Section 6.

13.2. Normalizing Beliefs

In Proposition 13.1 we saw how one can use a penalty term to ensure that only $bel \in \mathbb{P}$ can minimise a logarithmic SR $S^{stats}_{llog,ext}(P, bel)$. For $bel \in \mathbb{P}$ it holds that $\sum_{F\in\pi} bel(F) = 1$ for all partitions π of Ω. In [29] the authors showed that the penalty term can be dropped, if 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 subsets of Ω, which is henceforth called a partition. Denote by Π the union of $\{\Omega, \emptyset\}$, $\{\Omega\}$ and the set of these partitions. Then define

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

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:

$$S^{stats}_{normlog,ext,g}(P, B) := -\sum_{\pi\in\Pi} g(\pi) \sum_{X\in\pi} P(X) \cdot \log(B(X)) \tag{73}$$

$$= \sum_{X\subseteq\Omega} P(X) \cdot \Big(\sum_{\substack{\pi\in\Pi\\ X\in\pi}} g(\pi)\Big) \cdot \log(B(X)) \ . \tag{74}$$

PROPOSITION 13.2. *[29, Corollary 3, p. 3542] $S^{stats}_{normlog,ext,g}(P, B)$ 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 3.1.

The above proposition does not contradict Theorem 12.4, since we here consider normalised belief functions in $\mathbb{B}_{norm}$ while Theorem 12.4 concerns belief functions in $\mathbb{B}$.

The SRs $S^{stats}_{llog,ext}$ and $S^{stats}_{normlog,ext,g}$ rely on the same idea: The main culprit in the impossibility Theorem 12.4 is that in (64) there is no interaction between the beliefs 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 ($S^{stats}_{llog,ext}$) introduces a penalty (i.e. normalisation) term into the SR, for the latter SR ($S^{stats}_{normlog,ext,g}$) one pre-supposes normalised belief functions.

14. Conclusion

In the first part of this paper we saw how to use statistical SRs to justify the PN. In this second part we demonstrated the usefulness of statistical SRs for further norms of rational belief formation. In particular, we saw how an agent's evidence $\mathcal{E}$ can be naturally taken into account by applying worst-case expected loss avoidance as DTN. This seems to us like a clear advantage of the statistical approach over the epistemic approach.

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 6.2, Proposition 13.1 and Proposition 13.2).

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 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.

Decision Theoretic Norm	Applications of Epistemic Scoring Rules	Applications of Statistical Scoring Rules
Dominance w.r.t. $\omega \in \Omega$ Dominance w.r.t. $P \in \mathbb{P}$	[12], [46], [25], [26],[43], [44]	[50] [51] Proposition 4.7, Theorem 5.5, Theorem 7.1
Expected Loss w.r.t. bel^*		Belief Elicitation
Worst-Case Loss w.r.t. $\omega \in \Omega$	[45]	
Worst-Case Expected Loss w.r.t. $P \in \mathbb{E}$		Theorem 10.1, Theorem 10.2 Corollary 11.4, Proposition 11.5, [21] [29]

Table 1. Applications of SRs to rational belief formation

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 XX for helpful comments and YY for supporting this research.

References

[1] Aczél, János, and J. Pfanzagl, 'Remarks on the measurement of subjective probability and information', *Metrika*, 11 (1967), 91–105.

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

[3] Bernardo, José M., 'Expected Information as Expected Utility', *The Annals of Statistics*, 7 (1979), 3, 686–690.

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

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

[6] Cox, R. T., 'Probability, Frequency and Reasonable Expectation', *American Journal of Physics*, 14 (1946), 1, 1–13.

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

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

[9] Dash, Rajdeep K., Nicholas R. Jennings, and David C. Parkes, 'Computational-Mechanism Design: A Call to Arms', *IEEE Intelligent Systems*, 18 (2003), 40–47.

[10] Dawid, A. Philip, 'Probability forecasting', in Samuel Kotz, and Norman Lloyd Johnson, (eds.), *Encyclopedia of Statistical Sciences*, vol. 7, Wiley, 1986, pp. 210–218.

[11] Dawid, A. Philip, Steffen Lauritzen, and Matthew Parry, 'Proper local scoring rules on discrete sample spaces', *Annals of Statistics*, 40 (2012), 1, 593–608.

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

[13] de Finetti, Bruno, 'Foresight: Its logical laws, its subjective sources', in Henry Ely Kyburg, and Howard Edward Smokler, (eds.), *Studies in Subjective Probability*, 2 edn., Krieger, 1980, pp. 53–118.

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

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

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

[17] Garthwaite, Paul H, Joseph B Kadane, and Anthony O'Hagan, 'Statistical Methods for Eliciting Probability Distributions', *Journal of the American Statistical Association*, 100 (2005), 470, 680–701.

[18] Gibbard, Allan, 'Rational Credence and the Value of Truth', in Tamar Szabó Gendler, and John Hawthorne, (eds.), *Oxford Studies in Epistemology: Volume 2*, chap. 6, Oxford University Press, 2007, pp. 143–164.

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

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

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

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

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

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

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

[26] Joyce, James M., 'Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief', in Franz Huber, and Christoph Schmidt-Petri, (eds.), *Degrees of Belief*, vol. 342 of *Synthese Library*, Springer, 2009, pp. 263–297.

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

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

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

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

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

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

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

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

[35] McCarthy, John, 'Measures of the value of information', *Proceedings of the National Academy of Sciences*, 42 (1956), 9, 654–655.

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

[37] Offerman, Theo, 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 (2009), 4, 1461–1489.

[38] Paris, Jeff B., *The Uncertain Reasoner's Companion*, Cambridge Univerity Press, 1994.

[39] Paris, Jeff B., 'Common Sense and Maximum Entropy', *Synthese*, 117 (1998), 75–93.

[40] Paris, Jeff B., and Alena Vencovská, 'A note on the inevitability of maximum entropy', *International Journal of Approximate Reasoning*, 4 (1990), 3, 183–223.

[41] Parry, Matthew, A. Philip Dawid, and Steffen Lauritzen, 'Proper local scoring rules', *Annals of Statistics*, 1 (2012), 561–592.

[42] Pettigrew, Richard, 'Epistemic Utility Arguments for Probabilism', in Edward N. Zalta, (ed.), *Stanford Encyclopedia of Philosophy*, winter 2011 edn., Stanford University, 2011.

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

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

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

[46] Predd, J.B., 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 (2009), 10, 4786–4792.

[47] Ramsey, F.P., 'Truth and probability', *Histoy of Economic Thought Chapters*, (1926), 156–198.

[48] Ricceri, Biagio, 'Recent Advances in Minimax Theory and Applications', in Altannar Chinchuluun, PanosM. Pardalos, Athanasios Migdalas, and Leonidas Pitsoulis, (eds.), *Pareto Optimality, Game Theory And Equilibria*, vol. 17 of *Optimization and Its Applications*, Springer, 2008, pp. 23–52.

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

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

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

[52] Selten, Reinhard, 'Axiomatic characterization of the quadratic scoring rule', *Experimental Economics*, 1 (1998), 43–62.

[53] Shuford, Emir H., Arthur Albert, and H. Edward Massengill, 'Admissible probability measurement procedures', *Psychometrika*, 31 (1966), 2, 125–145.

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

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

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

[57] Winkler, Robert L., 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.

[58] Winkler, Robert L., 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 (1996), 1–60.

Appendix

15. Proof of Proposition 13.1

PROPOSITION 13.1. *The following scoring rule is strictly* $\mathbb{B}$*-proper*

$$S^{stats}_{llog,ext}(P, bel) := \sum_{X\subseteq\Omega} P(X)\cdot\left(-1+\frac{\sum_{Y\subseteq\Omega} bel(Y)}{\sum_{Y\subseteq\Omega} P(Y)} - \ln(bel(X))\right)$$
$$= -\frac{|\mathcal{P}\Omega|}{2} + \sum_{Y\subseteq\Omega} bel(Y) - \sum_{X\subseteq\Omega} P(X)\cdot\ln(bel(X)) \ .$$

PROOF. First note that for every $P\in\mathbb{P}$ that $P(X)+P(\bar{X})=1$, thus $\sum_{X\subseteq\Omega} P(X) = \frac{|\mathcal{P}\Omega|}{2}$. This then explains why the two expression given for the SR are equal.

Suppose that there exists an $X\subseteq\Omega$ such that $bel(X)=0$ and $P(X)>0$, then $S^{stats}_{llog,ext}(P, bel) = +\infty$. Since, $S^{stats}_{llog,ext}(P, P) < +\infty$ it follows that such a belief function bel cannot minimise $S^{stats}_{llog,ext}(P,\cdot)$.

On the other hand suppose that there exists an $X\subseteq\Omega$ such that $bel(X)>0$ and $P(X)=0$. Then define a belief function $bel'\in\mathbb{B}$ by letting $bel'(X):=0$ and $bel'(Y):=bel(Y)$ for all other $Y\subseteq\Omega$. Then,

$$\begin{aligned}
&S^{stats}_{llog,ext}(P, bel) - S^{stats}_{llog,ext}(P, bel') \\
&\quad = \sum_{Z\subseteq\Omega} bel(Z) - \sum_{Z\subseteq\Omega} bel'(Z) - \sum_{Z\subseteq\Omega} P(Z)\cdot\Big(\ln(bel(Z)) - \ln(bel'(Z))\Big) \\
&\quad \geq bel(X) \\
&\quad > 0 \ .
\end{aligned}$$

Thus, bel' has a better score than bel. Overall, it follows that any function bel minimizing $S^{stats}_{llog,ext}(P, bel)$ has to satisfy that $bel(X)>0$, if and only if $P(X)>0$.

For $\lambda\in\mathbb{R}_{>0}$ let $\mathbb{B}_\lambda := \{bel\in\mathbb{B} \mid \sum_{X\subseteq\Omega} bel(X)=\lambda\}$. Now consider a fixed $P\in\mathbb{P}$ and let $\mathcal{P}\Omega^+ := \{X\subseteq\Omega \mid P(X)>0\}$. We will first show that $bel(X) := \lambda\frac{2}{|\mathcal{P}\Omega|}P(X)$ is the unique belief function in $\arg\inf_{bel\in\mathbb{B}_\lambda} S^{stats}_{llog,ext}(P, bel)$.

We need to consider the following minimisation problem for fixed but arbitrary $P\in\mathbb{P}$

$$\begin{aligned}
\text{minimise} \quad & S^{stats}_{llog,ext}(P, bel) \\
\text{subject to} \quad & bel\in\mathbb{B}_\lambda \\
& bel(X)>0, \text{ if and only if } X\in\mathcal{P}\Omega^+ \ .
\end{aligned}$$

Due to the fact that we fixed λ, this is a strictly convex minimisation problem. Thus, it has a unique solution in the closure of the convex set $\mathbb{B}_\lambda$. By the above, $bel(X)=0$ for $X\in\mathcal{P}\Omega^+$ cannot minimise score. Thus, the derivatives $\frac{\partial}{\partial bel(X)}$ for $X\in\mathcal{P}\Omega^+$ of $S^{stats}_{llog,ext}(P, bel)$ are well-defined where the minimum obtains.

We may thus apply the Lagrange Multiplier Method (LMM) to solve the minimisation problem. Define the Lagrange function of the problem as

$$
\begin{aligned}
Lag(bel) &:= S^{stats}_{llog,ext}(P, bel) + \mu(\sum_{X\subseteq\Omega} bel(X) - \lambda) \\
&= \sum_{X\subseteq\Omega} bel(X) - \frac{|\mathcal{P}\Omega|}{2} - \sum_{X\subseteq\Omega} P(X)\cdot\ln(bel(X)) + \mu(\sum_{X\subseteq\Omega} bel(X) - \lambda)
\end{aligned}
$$

where μ is the Lagrange multiplier. As it will turn out, we do not need further Lagrange multipliers for the constraints $0 < bel(X) \le 1$ for all $X \in \mathcal{P}\Omega$.

Taking derivatives with respect to the independent variables $bel(X)$ for $X \in \mathcal{P}\Omega^+$ and equating with zero we obtain the following set of equations

$$
\frac{\partial}{\partial bel(X)} Lag = 1 - \frac{P(X)}{bel(X)} + \mu = 0 \text{ for all } X \in \mathcal{P}\Omega^+ \ . \tag{75}
$$

Only if $\frac{P(X)}{bel(X)}$ does not depend on X can this set of equations be solved by choosing a single value for μ.

Recall that we have convinced ourselves that the optimization problem has a unique solution. Furthermore, the LMM will find all minima of the convex minimisation problem on the convex set $\mathbb{B}_\lambda$. Thus, there has to be at least on $\mu \in \mathbb{R}$ which solves (75).

Hence, bel has to be a multiple of P. Given that $\sum_{X\subseteq\Omega} P(X) = \frac{\mathcal{P}\Omega}{2}$ and $\sum_{X\subseteq\Omega} bel(X) = \lambda$ the claim for $\mathbb{B}_\lambda$ follows.

Let us now consider varying the parameter λ. If $\lambda = 0$, then $S^{stats}(P, bel)$ equals infinity. So, not assigning any positive belief to any proposition does not lead to a finite score.

For $\lambda > 0$ define $S(\lambda) = S^{stats}_{llog,ext}(P, \lambda P)$. Taking the derivative with respect to λ and equating with zero we obtain

$$
\begin{aligned}
\frac{d}{d\lambda} S(\lambda) &= \frac{d}{d\lambda}\Big(\lambda\frac{|\mathcal{P}\Omega|}{2} - \frac{|\mathcal{P}\Omega|}{2} - \sum_{X\subseteq\Omega} P(X)\cdot\ln(\lambda\cdot P(X))\Big) && (76)\\
&= \frac{d}{d\lambda}\Big(\lambda\frac{|\mathcal{P}\Omega|}{2} - \frac{|\mathcal{P}\Omega|}{2} - \sum_{X\subseteq\Omega} P(X)\cdot(\ln(\lambda) + \ln(P(X)))\Big) && (77)\\
&= \frac{|\mathcal{P}\Omega|}{2} - \sum_{X\subseteq\Omega} P(X)\frac{1}{\lambda} && (78)\\
&= \frac{|\mathcal{P}\Omega|}{2} - \frac{|\mathcal{P}\Omega|}{2}\frac{1}{\lambda} && (79)\\
&= \frac{|\mathcal{P}\Omega|}{2}\cdot(1 - \frac{1}{\lambda}) && (80)\\
&= 0 \ . && (81)
\end{aligned}
$$

Thus, for $\lambda = 1$ the unique minimum of $S(\lambda)$ obtains. Hence $bel = P$ is the unique minimum of $S^{stats}_{llog,ext}(P, bel)$ with bel ranging in $\mathbb{B}$. ■

If the SR did not employ the natural logarithm but rather use some arbitrary base b for the logarithm, then the above proof holds with the single exception that in (78) we obtain $\frac{1}{\lambda \log b}$ rather than $\frac{1}{\lambda}$. So, the belief function with the lowest score for a given fixed P is $\frac{P}{\log b}$. It follows that $S^{stats}_{llog,ext,b}(P, bel) = \sum_{Y\subseteq\Omega} \frac{bel(Y)}{\log b} - \frac{|\mathcal{P}\Omega|}{2} - \sum_{X\subseteq\Omega} P(X)\cdot\log_b(bel(X))$ is strictly $\mathbb{B}$-proper.

16. Rational Belief Formation as Mechanism Design

Let us now take a step back and consider the project of justifying norms of rational belief formation via SRs from a higher, more abstract point of view. We will see that, from a purely technical point of view, this project can be framed in more general terms. We shall begin by giving two examples.[8]

1) Say, you, a dear reader of these lines, are a proponent of the PN and you are looking for a justification of it in terms of epistemic SRs. What you are looking for is an epistemic SR S and a DTN D, where D employs S as a dis-utility function. You want S and D to be such that an agent acting according to D will adopt some $bel \in \mathbb{P}$ and such that adopting any $bel \in \mathbb{P}$ is acting in accordance with D while adopting some $bel \in \mathbb{B}\setminus\mathbb{P}$ would not be in accordance with D. Theorem 4.5 shows that $S = S^{epi}_{Brier}$ and dominance as DTN are what you have been looking for.

2) Let us now assume that you are convinced that adopting the calibrated $bel^\dagger \in \mathbb{E}$ which maximises Shannon Entropy is (for the application you have in mind) the most appropriate belief function. You are wondering for which dis-utility function S^{stats} and which DTN D a rational agent acting in accordance with D will adopt $bel^\dagger$. Assuming the PN, your choice of $S^{stats}_{\log}$ (see (19)) and worst-case loss avoidance as DTN will do (cf. Theorem 10.1).

Both these examples have the following common structure. You are looking for a triple: a norm of rational belief formation N, a SR S and a DTN D such that an agent acting in accordance with D and minimising dis-utility with respect to S will adopt some bel which is consistent with N and such that every bel consistent with N minimises dis-utility with respect to S in accordance with D.

This structure has already been laid bare by the approach termed *Epistemic Utility Theory* (EUT), cf. [42]. In EUT, an epistemic SR is interpreted as a measure of dis-utility. Norms governing the actions of agents, such as dominance, are then framed in epistemic utility terms. For instance, an agent avoiding epistemic dis-utility dominated belief functions with respect to an epistemic SR S^{epi}_L will adopt a/the belief function which is not dominated with respect to S^{epi}_L.

Taking another step back a deeper structure can be unearthed.

The reasoner aiming to prove theorems (such as Theorem 4.5 and Theorem 10.1) is looking for a DTN D and a SR S such that every rational agent acting in accordance with D will conform to a norm N. A rational agent faced

[8]The remarks in this section also apply to frameworks in which degrees of belief are measured by other means, by intervals or fuzzy numbers, say.

with the problem of minimising score in accordance with D is thus facing a minimisation problem (with respect to S) in the variables $bel(X)$ for all $X \subseteq \Omega$. From a purely formal point of view, the agent is playing a single-player game. The rules of the game are: assign every $X \subseteq \Omega$ a unique value $bel(X) \in [0, 1]$. The agent's aim in the game is to minimise score measured by S in accordance with D.

Speaking in these terms, the reasoner's quest for proofs of theorems can be understood as the search for a single player game. In such a game the player has to assign every $X \subseteq \Omega$ some value in $[0, 1]$. Depending on the assignment, i.e. bel, and a (dis-)utility function S the agent will be awarded some (dis-)utility. Assuming that the player acts in accordance with a DTN D, the only rational assignments $bel : \mathcal{P}\Omega \to [0, 1]$ are those assignments which are consistent with some norm N and every assignment consistent with N is in accordance with D. Briefly put, the reasoner is designing a game.

The study of (agents playing) games is a very well-trodden path in the literature. A rather recent development is the study of the *design of games*. In this approach one focuses on the design of games such that rational players playing such a game will act in a way which the *game designer* considers desirable. Consider the prisoner's dilemma as a famous example, but this time consider it from the perspective of the minister of justice aiming to incentivise the prisoners to tell the truth. Whether the minister of justice has been successful in incentivising prisoners to tell the truth is still a matter of lively philosophical debate. In the prisoner's dilemma the rules of the game, the dis-utility function and the minister's desired outcome are evident. On the other hand, it is unclear to which notion of rationality the imprisoned agents (ought to) subscribe to. From the minister's point of view, the design of the game has only been moderately successful because (apparently) not all rational agents tell the truth.

The design of such games, or in technical terms: the *design of such mechanisms*, has become a sub-field of game theory and is known under the name Mechanism Design (MD). In general, MD concerns the design of games such that the rules of the game incentivise rational players to act in ways the designer of the game considers desirable; see [9] for an introduction to MD. Thus, EUT can be seen to be a sub-field of MD. Game-theoretic machinery may thus, in the future, enrich EUT.