Article

Justifying Objective Bayesianism on Predicate Languages

Jürgen Landes * and Jon Williamson

Department of Philosophy, School of European Culture and Languages, University of Kent, Canterbury CT2 7NF, UK; E-Mail: j.williamson@kent.ac.uk

* Author to whom correspondence should be addressed; E-Mail: j.landes@kent.ac.uk;
  Tel.: +44-1227-82-6545.

Abstract: Objective Bayesianism says that the strengths of one's beliefs ought to be probabilities, calibrated to physical probabilities insofar as one has evidence of them, and otherwise sufficiently equivocal. These norms of belief are often explicated using the maximum entropy principle. In this paper we investigate the extent to which one can provide a unified justification of the objective Bayesian norms in the case in which the background language is a first-order predicate language, with a view to applying the resulting formalism to inductive logic. We show that the maximum entropy principle can be motivated largely in terms of minimising worst-case expected loss.

Keywords: objective Bayesianism; g-entropy; predicate language; scoring rule; minimax

1. Introduction

Objective Bayesianism holds that the strengths of one's beliefs should satisfy three norms [1,2]:

- **Probability.** The strengths of one's beliefs should satisfy the axioms of probability: if bel is one's belief function, which assigns a degree of belief to each sentence of one's language, then $bel \in \mathbb{P}$, the set of probability functions defined on the sentences of one's language.

- **Calibration.** The strengths of one's beliefs should fit one's evidence: $bel \in \mathbb{E}$, the set of belief functions compatible with one's evidence. In particular, the strengths of one's beliefs should

be calibrated with physical probabilities, insofar as one has evidence as to what the physical probabilities are: if one's evidence determines just that the physical probability function P^* lies in some non-empty set $\mathbb{P}^*$ of probability functions, then $bel \in \mathbb{E} = \langle \mathbb{P}^* \rangle$, where $\langle \mathbb{P}^* \rangle$ is the convex hull of $\mathbb{P}^*$ [3].

- **Equivocation.** The strengths of one's beliefs should otherwise equivocate sufficiently between the basic possibilities that one can express: bel is some function in $\mathbb{E}$ that is sufficiently *equivocal*. Note that entropy is often used as a measure of the extent to which a probability function equivocates.

These three norms are usually justified in rather different ways. The Probability norm is usually justified as being required if one is to avoid sure loss—the *Dutch book* argument. The Calibration norm needs to hold if one is to avoid loss in the long run when one repeatedly bets on similar events. It has also been argued that the Equivocation norm should hold if one is to minimise worst-case expected loss. See Williamson [1] (Chapter 3) for discussion of these justifications. Unfortunately, these justifications do not cohere particularly well, because the betting set-up and the notion of loss differ in each case—for the Probability norm, the notion of loss is sure single-case loss, where losses may be positive or negative; for the Calibration norm it is almost-sure (*i.e.,* probability 1) long-run loss, positive or negative; for the Equivocation norm, it is worst-case expected loss, where the loss is positive and logarithmic. Furthermore, a justification for the order in which the norms are applied is missing. In particular, the justification of the Equivocation norm presumes that belief is probabilistic; for this justification to work, some argument is needed for the claim that avoiding sure loss should be prioritised over minimising worst-case expected loss; but there is as yet no such argument. The question thus arises as to whether a single, unified justification can be given for the three norms, in order to circumvent the above problems.

Landes and Williamson [4] provided a single, unified justification for the situation in which one's beliefs are defined over propositions, construed as subsets of a finite set Ω of outcomes. It turns out that all three norms must hold if one is to minimise worst-case expected loss: one's belief function should be a probability function in $\mathbb{E} = \langle \mathbb{P}^* \rangle$ that has sufficiently high entropy. This line of argument will be described in Section 2. Landes and Williamson [4] went on to extend this unified justification to the situation in which beliefs are defined over sentences of a propositional language, formed by recursively applying the usual propositional connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ to a finite set of propositional variables.

In this paper we shall show that a similar justification goes through for the situation in which beliefs are defined over sentences of a first-order predicate language, with the use of predicate, constant and variable symbols as well as the quantifiers $\forall, \exists$. In Section 3 we shall formulate the norms of objective Bayesianism in the context of a predicate language. In Section 4 we shall provide a justification for maximising entropy when the language in question is a predicate language without quantifier symbols and when the evidence set is finitely generated. In Section 5, we shall extend this line of argument to predicate languages that contain quantifier symbols. In Section 6 we shall investigate the case of evidence which is not finitely generated. Key concepts and notation are collected in Appendix C for ease of reference.

The key technical results in this paper are Theorem 3, Theorem 6, Theorem 7, and Theorem 8. These results all suppose that the available evidence is finitely generated (in the sense of Definition 5). The

first two jointly show that, on a quantifier-free predicate language, the belief function with the best loss profile is the calibrated probability function which has maximal entropy. Theorem 7 implies that adding new constant or predicate symbols to the language does not change the inferences one draws which are expressible in the original language. Theorem 8 extends Theorem 3 and Theorem 6 to predicate languages with quantifiers. En route to proving Theorem 8, we improve on Gaifman's Unique Extension Theorem [5] (Theorem 1) in Proposition 24.

The case of evidence which cannot be finitely generated is more involved. We consider a case in which no belief function has an optimal loss profile in Proposition 28 and Proposition 30. While there are no functions with the best loss profile in that case, we show in Proposition 29 and Proposition 31 that probability functions in a neighbourhood of the calibrated function with maximal entropy have arbitrarily good loss profiles. We also discuss a case in which the belief function with best loss profile does indeed turn out to be the calibrated probability function which has maximal entropy, see Theorem 9.

2. Beliefs over Propositions

Here we will recap the relevant results of Landes and Williamson [4], to which the reader is referred for further details and motivation. In this section we will be concerned solely with a finite set Ω of possible outcomes. We shall suppose that each member ω of Ω is a state $\pm A_1 \wedge \cdots \wedge \pm A_n$ of a finite propositional language $\mathcal{L} = \{A_1, ..., A_n\}$. A *proposition* F is a subset of Ω. Let Π be the set of all partitions of Ω. We take $\{\emptyset, \Omega\}, \{\Omega\} \in \Pi$. In order to limit the proliferation of partitions, we suppose that the only partition in which $\emptyset$ occurs is $\{\emptyset, \Omega\}$.

Given a belief function $bel : \mathcal{P}\Omega \longrightarrow \mathbb{R}_{\geq 0}$ that is not zero everywhere, we normalise by dividing each degree of belief by $\max_{\pi \in \Pi} \sum_{F \in \pi} bel(F)$ to form a belief function, $B : \mathcal{P}\Omega \longrightarrow [0, 1]$, with degrees of belief in the unit interval. The set of normalised belief functions is

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

On the other hand, the set of probability functions is

$$\mathbb{P} := \{B : \mathcal{P}\Omega \longrightarrow [0, 1] : \sum_{F \in \pi} B(F) = 1 \text{ for all } \pi \in \Pi\} \subset \mathbb{B},$$

where $\subset$ denotes strict subset inclusion. The inclusion is strict since the following normalised belief function B is not in $\mathbb{P}$, $B(\emptyset) = 1$ and $B(F) = 0$ for all $\emptyset \subset F \subseteq \Omega$. Since $\{\Omega\}$ is a partition we have $P(\Omega) = 1$ and since $\{\Omega, \emptyset\}$ is a partition it holds that $P(\emptyset) = 0$ for all $P \in \mathbb{P}$.

Let $L(F, B)$ be the loss incurred by adopting belief function B when proposition F turns out to be true. Arguably, in the absence of knowledge of the true loss function, the loss function L should be taken to be logarithmic, as we shall now see. Consider the following four conditions on a default loss function L:

L1. $L(F, B) = 0$ if $B(F) = 1$.

L2. $L(F, B)$ strictly increases as $B(F)$ decreases from 1 towards 0.

L3. $L(F, B)$ depends only on $B(F)$, not on $B(F')$ for $F' \neq F$.

To express the next condition we need some notation. Suppose $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$: say that $\mathcal{L} = \{A_1, ..., A_n\}$, $\mathcal{L}_1 = \{A_1, ..., A_m\}$, $\mathcal{L}_2 = \{A_{m+1}, ..., A_n\}$ for some $1 \leq m < n$. Then $\omega \in \Omega$ takes the form $\omega_1 \wedge \omega_2$ where $\omega_1 \in \Omega_1$ is a state of $\mathcal{L}_1$, and $\omega_2 \in \Omega_2$ is a state of $\mathcal{L}_2$. Given propositions $F_1 \subseteq \Omega_1$ and $F_2 \subseteq \Omega_2$ we can define $F_1 \times F_2 := \{\omega = \omega_1 \wedge \omega_2 : \omega_1 \in F_1, \omega_2 \in F_2\}$, a proposition of $\mathcal{L}$. Given a fixed belief function B such that $B(\Omega) = 1$, $\mathcal{L}_1$ and $\mathcal{L}_2$ are *independent sublanguages*, written $\mathcal{L}_1 \perp\!\!\!\perp_B \mathcal{L}_2$, if $B(F_1 \times F_2) = B(F_1) \cdot B(F_2)$ for all $F_1 \subseteq \Omega_1$ and $F_2 \subseteq \Omega_2$, where $B(F_1) := B(F_1 \times \Omega_2)$ and $B(F_2) := B(\Omega_1 \times F_2)$. The restriction $B_{|\mathcal{L}_1}$ of B to $\mathcal{L}_1$ is a belief function on $\mathcal{L}_1$ defined by $B_{|\mathcal{L}_1}(F_1) = B(F_1) = B(F_1 \times \Omega_2)$, and similarly for $\mathcal{L}_2$.

L4. Losses are additive when the language is composed of independent sublanguages: if $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ for $\mathcal{L}_1 \perp\!\!\!\perp_B \mathcal{L}_2$, then $L(F_1 \times F_2, B) = L_1(F_1, B_{|\mathcal{L}_1}) + L_2(F_2, B_{|\mathcal{L}_2})$, where L_1, L_2 are loss functions defined on $\mathcal{L}_1, \mathcal{L}_2$ respectively.

Theorem 1. *If a loss function L satisfies L1–4 then $L(F, B) = -k \log B(F)$ for some constant $k > 0$ that does not depend on $\mathcal{L}$.*

When we consider the notion of expected loss, we see that this concept depends on the weight given to the various partitions under consideration. Let $g : \Pi \longrightarrow \mathbb{R}_{\geq 0}$ be a function that assigns a weight to each partition. Then the g-expected loss or g-score of a belief function $B \in \mathbb{B}$ with respect to a probability function $P \in \mathbb{P}$ is defined by

$$S_g^L(P, B) := \sum_{\pi \in \Pi} g(\pi) \sum_{F \in \pi} P(F) L(F, B),$$

for any weighting function g that is *inclusive* in the sense that for any proposition F, some partition π containing F is given positive weight. We adopt the usual convention that $0 \log 0 = 0$. This ensures that $S_g^L(P, B)$ is well-defined. Theorem 1 allows us to focus attention on logarithmic g-score,

$$S_g(P, B) := -\sum_{\pi \in \Pi} g(\pi) \sum_{F \in \pi} P(F) \log B(F). \tag{1}$$

An important property of a scoring rule is that $\arg\inf_{B \in \mathbb{B}} S_g^L(P, B) = \{P\}$ for all $P \in \mathbb{P}$. That is, for fixed $P \in \mathbb{P}$, $S_g^L(P, B)$ is uniquely minimised by $B = P$. This property is known as *strict propriety*.

Proposition 1 (Strict Propriety). *S_g is strictly proper.*

By analogy with the generalised notion of scoring rule, we get a similar generalisation of entropy, g-entropy:

$$H_g(B) := -\sum_{\pi \in \Pi} g(\pi) \sum_{F \in \pi} B(F) \log B(F). \tag{2}$$

The standard entropy function corresponds to the special case in which $g = g_\Omega$, the (non-inclusive) weighting function that gives weight 1 to the partition $\{\{\omega\} : \omega \in \Omega\}$ of states and weight 0 to all other partitions.

It turns out that, if there is such a function, the probability function that minimises worst-case g-score, where the worst case is taken over physical probability functions in the set $\mathbb{E} = \langle \mathbb{P}^* \rangle$, is the probability function in $\mathbb{E}$ that has maximum g-entropy:

Theorem 2. *As noted above, $\mathbb{E}$ is taken to be convex and g inclusive. There is a unique member of* $\arg\sup_{P\in\mathbb{E}} H_g(P)$, *which we shall denote by* $P_g^\dagger$. *Furthermore,*

$$\arg\sup_{P\in\mathbb{E}} H_g(P) = \arg\inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} S_g(P,B) = \{P_g^\dagger\}.$$

Throughout this paper we use $\arg\sup_{P\in\mathbb{E}}$ (and $\arg\inf_{P\in\mathbb{E}}$) to refer to the points in the closure $[\mathbb{E}]$ of $\mathbb{E}$ that achieve the supremum (respectively infimum) *whether or not these points are in* $\mathbb{E}$. (This convention shall also apply mutatis mutandis to suprema and infima over sets of belief functions defined on predicate languages later in this paper.)

The above theorem concerns the minimisation of worst-case g-score. If one replaces the minimisation of worst-case g-score by a more fine-grained criterion (which breaks ties between belief functions with the same worst-case g-score), then an analogue of the above theorem holds: there exists a unique belief function which is best with respect to this criterion and this function is $P_g^\dagger$, which maximises g-score in $[\mathbb{E}]$. When we move to predicate languages we will consider such a refinement in Definition 21.

3. Beliefs over Sentences of a Predicate Language

3.1. Norms

In this section we introduce the norms of objective Bayesianism as they apply to strength of belief in sentences formulated in a predicate language. This framework is presented in more detail in Williamson [1] (Chapter 5). It is this set of norms that we seek to justify in terms of the loss that a belief function exposes one to.

We shall take $\mathcal{L}$ to be a first-order predicate language with finitely many relation symbols $U_1, \ldots, U_s$, countably many constant symbols $t_1, t_2, \ldots$, but no function or equality symbols. We will consider languages with and without the existential quantifier symbol, using the notation $\mathcal{L}^{\not\exists}$ and $\mathcal{L}^{\exists}$ to disambiguate where needed. We shall assume, as is usual in this setting, that each individual in the domain of discourse is picked out by a some constant symbol. The sentences $S\mathcal{L}$ of $\mathcal{L}$ are formed by recursively applying the usual connectives and the existential quantifier, if present. In $\mathcal{L} = \mathcal{L}^{\exists}$, universally quantified sentences may be defined in terms of existentially quantified sentences as usual via $\forall x\theta(x) := \neg\exists x\neg\theta(x)$. Note that $S\mathcal{L}^{\not\exists}$ coincides with the set of quantifier-free sentences of $\mathcal{L}^{\exists}$. We shall also be interested in the finite sublanguages $\mathcal{L}_n$, for $n \geq 1$, which are identical to $\mathcal{L}$ except that they have only finitely many constant symbols $t_1, \ldots, t_n$.

We shall list the atomic sentences of $\mathcal{L}$, *i.e.,* sentences of the form Ut where U is a relation symbol and t is a tuple of constant symbols of the corresponding arity, as $A_1, A_2, \ldots$, ordered in such a way that atomic sentences that can be expressed in $\mathcal{L}_{n+1}$ but not in $\mathcal{L}_n$ occur after the atomic sentences $A_1, \ldots, A_{r_n}$ of $\mathcal{L}_n$, for each n. Ω_n will denote the set of n-*states*, *i.e.,* sentences of the form $\pm A_1 \wedge \cdots \wedge \pm A_{r_n}$. We shall use Greek letters, such as θ, to denote sentences of $\mathcal{L}$, and Roman letters, e.g., F, to denote propositions expressed by such sentences. We shall construe propositions as sets of n-states, $F \subseteq \Omega_n$ for some n (see Section 2).

The norms of objective Bayesianism can then be explicated thus:

Probability. The strengths of one's beliefs should be representable by a probability function, *i.e.,* a function $P : S\mathcal{L} \longrightarrow \mathbb{R}$ that satisfies the properties:

P1. $P(\tau) = 1$ for all tautologies τ.

P2. If $\vDash \neg(\varphi \wedge \psi)$ then $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$.

P3. $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^{m} \theta(t_i)\right)$.

(Clearly P3 is only applicable in the case $\mathcal{L} = \mathcal{L}^{\exists}$.)

Calibration. One's degrees of belief should satisfy constraints imposed by one's evidence. Assuming all evidence is evidence of physical probabilities, P should lie in the set $\mathbb{E}_{\mathcal{L}} = \langle \mathbb{P}^* \rangle$, the convex hull of the set of epistemically possible physical probability functions.

Equivocation. One's degrees of belief should otherwise be sufficiently equivocal. Again, one can explicate this by saying that one's belief function should have sufficiently high entropy. Here P has *higher entropy* than Q if there is some N such that for all $n \geq N$, $H_{\Omega}^n(P) > H_{\Omega}^n(Q)$, where H_{Ω}^n is standard entropy on $\mathcal{L}_n$, $H_{\Omega}^n(P) := -\sum_{\omega \in \Omega_n} P(\omega) \log P(\omega)$.

The key question we attempt to answer here is: can these norms be given a unified justification in terms of avoiding avoidable loss?

3.2. Belief and Probability

A (non-normalised) belief function $bel : S\mathcal{L} \longrightarrow \mathbb{R}_{\geq 0}$ is a function that maps any sentence of the language to a non-negative real number. For technical convenience we shall focus our attention on normalised belief functions, which are defined below.

A (countable) set of mutually exclusive sentences $\pi \subset S\mathcal{L}$ is called *exhaustive* if, for all interpretations $\mathcal{I}$ under which the constants exhaust the universe of $\mathcal{M}$, there exists a sentence $\theta \in \pi$ such that $\mathcal{I} \models \theta$. This means that it is not possible for all $\theta \in \pi$ to be false at the same time. In order to control the number of partitions, we shall assume that the only partitions in which contradictions κ occur are the partitions of the form $\{\tau, \kappa\}$, for some tautology τ. Let $\Pi_{\mathcal{L}}$ denote the set of partitions of $\mathcal{L}$.

Example 1 (**Infinite partitions**). *Even though $\mathcal{L}^{\exists}$ does not contain a symbol for equality and every element of a partition is a sentence of $\mathcal{L}^{\exists}$, which is of finite length, infinite partitions such as the following do exist:*

$$\pi_{\infty} := \{\forall x \neg U_1 x\} \cup \bigcup_{k=1}^{\infty} \{U_1 t_k \wedge \bigwedge_{l=1}^{k-1} \neg U_1 t_l\}.$$

(Here it is presupposed that $\mathcal{L}^{\exists}$ contains a unary predicate symbol U_1.) On the other hand, it turns out that there are no infinite partitions in $\mathcal{L}^{\nexists}$ [6] (§2.5).

We take it that it is a matter of convention on which scale beliefs are measured. For convenience, we want to normalise this scale to the unit interval, $[0, 1]$, so that all belief functions are considered on the same scale.

Definition 1 (**Normalised belief function**). *Let* $M := \sup_{\pi \in \Pi_{\mathcal{L}}} \sum_{\varphi \in \pi} bel(\varphi)$. *Then define the normalisation of* bel *as* $B(\varphi) := \frac{bel(\varphi)}{M}$, *if* $M > 0$. *For a function* f *assigning every* $\varphi \in S\mathcal{L}$ *the same value* $v \in \mathbb{R}_{\geq 0}$ *we write* $f \equiv v$. *We shall consider* bel $\equiv 0$ *as normalised. The set of normalized belief functions on* $S\mathcal{L}$ *then is*

$$\mathbb{B}_{\mathcal{L}} := \{B : S\mathcal{L} \longrightarrow [0,1] : \sum_{\varphi \in \pi} B(\varphi) \leq 1 \text{ for all } \pi \in \Pi_{\mathcal{L}} \text{ and } \sum_{\varphi \in \pi} B(\varphi) = 1 \text{ for some } \pi \in \Pi_{\mathcal{L}}\}$$

$$\cup \{B \equiv 0\}.$$

For the normalisation of bel, B, *it holds that* $B \equiv 0$, *if and only if* $M = +\infty$ *or* bel $\equiv 0$.

We will be particularly interested in the following subset of functions:

$$\mathbb{P}_{\mathcal{L}} := \{P : S\mathcal{L} \longrightarrow [0,1] : \sum_{\varphi \in \pi} P(\varphi) = 1 \text{ for all } \pi \in \Pi_{\mathcal{L}}\}.$$

These are the probability functions:

Proposition 2. $P \in \mathbb{P}_{\mathcal{L}}$, *if and only if* $P : S\mathcal{L} \longrightarrow [0,1]$ *satisfies the axioms of probability:*

P1. $P(\tau) = 1$ *for all tautologies* $\tau \in S\mathcal{L}$.

P2. If $\vDash \neg(\varphi \wedge \psi)$ *then* $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$.

P3. $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^{m} \theta(t_i)\right)$.

Proof. First we shall see that $P \in \mathbb{P}_{\mathcal{L}}$ satisfies the axioms of probability.

P1. For any tautology $\tau \in S\mathcal{L}$ it holds that $P(\tau) = 1$ because $\{\tau\}$ is a partition in $\Pi_{\mathcal{L}}$. $P(\kappa) = 0$ for all contradictions κ because $\{\tau, \kappa\}$ is a partition in $\Pi_{\mathcal{L}}$ and $P(\tau) = 1$.

P2. Suppose that $\varphi, \psi \in S\mathcal{L}$ are such that $\vDash \neg(\varphi \wedge \psi)$. We shall proceed by cases to show that $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$. In the first three cases one of the sentences is a contradiction, in the last two cases there are no contradictions.

(i) $\vDash \varphi$ and $\vDash \neg\psi$, then $\vDash \varphi \vee \psi$. Thus by the above $P(\varphi) = 1$ and $P(\psi) = 0$ and hence $P(\varphi \vee \psi) = 1 = P(\varphi) + P(\psi)$.

(ii) $\vDash \neg\varphi$ and $\vDash \neg\psi$, then $\vDash \neg\varphi \wedge \neg\psi$. Thus $P(\varphi \vee \psi) = 0 = P(\varphi) + P(\psi)$.

(iii) $\nvDash \neg\varphi$, $\nvDash \varphi$, and $\vDash \neg\psi$, then $\{\varphi \vee \psi, \neg\varphi \vee \psi\}$ and $\{\varphi, \neg\varphi \vee \psi\}$ are both partitions in $\Pi_{\mathcal{L}}$. Thus $P(\varphi \vee \psi) + P(\neg\varphi \vee \psi) = 1 = P(\varphi) + P(\neg\varphi \vee \psi)$. Putting these observations together we now find $P(\varphi \vee \psi) = P(\varphi) = P(\varphi) + P(\psi)$.

(iv) $\nvDash \neg\varphi$, $\nvDash \neg\psi$ and $\vDash \varphi \leftrightarrow \neg\psi$, then $\{\varphi, \psi\}$ is a partition and $\varphi \vee \psi$ is a tautology. Hence, $P(\varphi) + P(\psi) = 1$ and $P(\varphi \vee \psi) = 1$. This now yields $P(\varphi) + P(\psi) = P(\varphi \vee \psi)$.

(v) $\nvDash \neg\varphi$, $\nvDash \neg\psi$ and $\nvDash \varphi \leftrightarrow \neg\psi$, then none of the following sentences is a tautology or a contradiction $\varphi, \psi, \varphi \vee \psi, \neg(\varphi \vee \psi)$. Since $\{\varphi, \psi, \neg(\varphi \vee \psi)\}$ and $\{\varphi \vee \psi, \neg(\varphi \vee \psi)\}$ are both partitions in $\Pi_{\mathcal{L}}$ we obtain $P(\varphi) + P(\psi) = 1 - P(\neg(\varphi \vee \psi)) = P(\varphi \vee \psi)$. So $P(\varphi) + P(\psi) = P(\varphi \vee \psi)$.

P3. For the rest of this proof we only have to consider $\mathcal{L} = \mathcal{L}^{\exists}$.

If $\vDash \exists x \theta(x)$, then $P(\exists x \theta(x)) = 1$.

Furthermore, the set $\{\theta_n \ : \ n \in \mathbb{N}\}$ with $\theta_n := \theta(t_n) \wedge \bigwedge_{j=1}^{n-1} \neg\theta(t_j)$ is exhaustive. Note that $\models \bigvee_{i=1}^n \theta(t_i) \leftrightarrow \bigvee_{i=1}^n \theta_i$. P1 and P2 are well-known to imply that logically equivalent sentences are assigned the same probability; see [7] (Proposition 2.1.c). Hence, $P(\bigvee_{i=1}^n \theta(t_i)) = P(\bigvee_{i=1}^n \theta_i)$.

The θ_i are mutually exclusive. We obtain from P2 that $P(\bigvee_{i=1}^n \theta_i) = \sum_{i=1}^n P(\theta_i)$. Next, define a set $\Theta := \{\theta_n \ : \ \theta_n \text{ satisfiable}\}$ which consists of exhaustive, satisfiable and mutually exclusive sentences. Hence Θ is a partition in $\Pi_{\mathcal{L}}$. We finally obtain

$$1 = \sum_{\theta \in \Theta} P(\theta) \leq \lim_{n \longrightarrow \infty} \sum_{i=1}^n P(\theta_n) = \lim_{n \longrightarrow \infty} P(\bigvee_{i=1}^n \theta_i) = \lim_{n \longrightarrow \infty} P(\bigvee_{i=1}^n \theta(t_i)) \leq 1.$$

P1 and P2 are also well-known to imply that if $\models \chi \longrightarrow \psi$ then $P(\chi) \leq P(\psi)$, see [7] (Proposition 2.1.c). Since $\models \bigvee_{i=1}^n \theta(t_i) \longrightarrow \bigvee_{i=1}^{n+1} \theta(t_i)$ we obtain $P(\bigvee_{i=1}^n \theta(t_i)) \leq P(\bigvee_{i=1}^{n+1} \theta(t_i))$. $(P(\bigvee_{i=1}^n \theta_i))_{n \in \mathbb{N}}$ is a (not necessarily strictly) increasing sequence. Then

$$1 = \lim_{n \longrightarrow \infty} P(\bigvee_{i=1}^n \theta(t_i)) = \sup_{n \in \mathbb{N}} P(\bigvee_{i=1}^n \theta(t_i)). \tag{3}$$

The second equality holds also when $1 > \lim_{n \longrightarrow \infty} P(\bigvee_{i=1}^n \theta(t_i))$.

If neither $\models \exists x \theta(x)$ nor $\models \neg\exists x \theta(x)$, then $\{\forall x \neg\theta(x), \exists x \theta(x)\}$ is a partition. We consider two cases.

In the first case the set $\{\forall x \neg\theta(x), \theta(t_1), \theta(t_2) \wedge \neg\theta(t_1), ..., \theta(t_k) \wedge \neg \bigvee_{j=1}^{k-1} \theta(t_j), ...\}$ is not a partition.

For example, this set fails to be a partition for $\theta(x) = \neg U t_2 \wedge U x$: the sentence $\theta(t_2) \wedge \neg\theta(t_1) = \neg U t_2 \wedge U t_2 \wedge \neg(\neg U t_2 \wedge U t_1)$ is a contradiction and hence it cannot be contained in a partition π consisting of infinitely many sentences.

$\neg \bigvee_{i=1}^m \theta(t_i)$ cannot be a contradiction since $\neg\exists\theta(x)$ is satisfiable and $\models \neg\exists\theta(x) \to \neg \bigvee_{i=1}^m \theta(t_i)$. If $\neg \bigvee_{i=1}^m \theta(t_i)$ is a tautology, then all θ_n with $n \leq m$ are contradictions. Hence, for all $m \in \mathbb{N}$ the set $\{\neg \bigvee_{i=1}^m \theta(t_i)\} \cup \{\theta_n \ : \ n \leq m \text{ and } \theta_n \text{ is satisfiable}\}$ is a partition, as is $\{\neg \bigvee_{i=1}^m \theta(t_i), \bigvee_{i=1}^m \theta(t_i)\}$. Furthermore, $\{\forall x \neg\theta(x)\} \cup \{\theta_n \ : \ \theta_n \text{ is satisfiable}\}$ is a partition.

Recalling that $P(\kappa) = 0$ for all contradictions κ we obtain $\sum_{k=1}^m P(\theta_k) = P(\bigvee_{i=1}^m \theta(t_i))$ and

$$P(\exists x \theta(x)) = \lim_{m \longrightarrow \infty} P(\bigvee_{i=1}^m \theta(t_i)).$$

It remains to show that

$$\lim_{m \longrightarrow \infty} P(\bigvee_{i=1}^m \theta(t_i)) = \sup_m P(\bigvee_{i=1}^m \theta(t_i)).$$

This follows as we saw above in (3).

In the second case the set $\{\forall x \neg\theta(x), \theta(t_1), \theta(t_2) \wedge \neg\theta(t_1), ..., \theta(t_k) \wedge \neg \bigvee_{j=1}^{k-1} \theta(t_j), ...\}$ is a partition. Recall that $\{\forall x \neg\theta(x), \exists x \theta(x)\}$ is also a partition. We obtain as in the first case that

$$P(\exists x \theta(x)) = \sum_{k=1}^{\infty} P(\theta(t_k) \wedge \neg \bigvee_{j=1}^{k-1} \theta(t_j)) = \lim_{m \longrightarrow \infty} \sum_{k=1}^m P(\theta_k) = \sup_m P(\bigvee_{i=1}^m \theta(t_i)).$$

For the converse, note that P1–3 imply that P is a probability measure on $S\mathcal{L}$, and so additive over countable partitions (§2 in [8]; §2.5 in [6]). Hence $P \in \mathbb{P}_{\mathcal{L}}$. $\square$

Another key feature of probability functions is that they respect logical equivalence:

Definition 2 (**Respecting logical equivalence**). *For a sublanguage $\mathcal{L}'$ of $\mathcal{L}$ we say that a function $f : S\mathcal{L} \longrightarrow [0,1]$ respects logical equivalence on $\mathcal{L}'$, if and only if for all $\varphi, \psi \in S\mathcal{L}'$ with $\vDash \varphi \leftrightarrow \psi$ it holds that $f(\varphi) = f(\psi)$. For $\mathcal{L}' = \mathcal{L}$ we simply say that f* respects logical equivalence.

Proposition 3. *The probability functions $P \in \mathbb{P}_{\mathcal{L}}$ respect logical equivalence.*

Proof. Suppose $P \in \mathbb{P}_{\mathcal{L}}$ and assume that $\varphi, \psi \in S\mathcal{L}$ are logically equivalent. Observe that $\{\varphi, \neg\varphi\}$ and $\{\psi, \neg\varphi\}$ are partitions in $\Pi_{\mathcal{L}}$. Hence,

$$P(\varphi) + P(\neg\varphi) = 1 = P(\psi) + P(\neg\varphi).$$

But then $P(\varphi) = P(\psi)$.

Thus, the $P \in \mathbb{P}_{\mathcal{L}}$ assign logically equivalent sentences the same probability. □

If a belief function $B : S\mathcal{L} \longrightarrow [0,1]$ respects logical equivalence, it gives sentences which express the same proposition the same degree of belief. Hence, for any $n \in \mathbb{N}$, B induces a function $^{\circ}B$ defined over the propositions $F \subseteq \Omega_n$ (*c.f.*, Section 2). $^{\circ}B$ is defined by:

$$^{\circ}B(F) := B(\bigvee F) = B(\bigvee_{\substack{\omega \in \Omega_n \\ \omega \in F}} \omega).$$

We will use the notation $^{\circ n}B$ to avoid ambiguity in cases where n varies.

The notion of a dominated belief function will prove useful in what follows:

Definition 3 (**Dominated belief function**). $B \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$ is dominated *by a probability function $P \in \mathbb{P}_{\mathcal{L}}$, if and only if for all $\varphi \in S\mathcal{L}$ it holds that $B(\varphi) \leq P(\varphi)$.*

Note that if B is dominated by P, then $B \neq P$, and thus $B(\varphi) < P(\varphi)$ has to hold at least for one sentence φ.

Proposition 4. *There exist $B \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$ which are not dominated.*

Proof. Let U be a relation symbol in $\mathcal{L}$ of arity $a \geq 1$, say. Let Ut_1t be a well-formed formula of $\mathcal{L}_2$, *i.e.,* t is a $a - 1$ tuple with consisting only of t_1 and t_2. Let $O_4 := \{Ut_1t \wedge Ut_2t, Ut_1t \wedge \neg Ut_2t, \neg Ut_1t \wedge Ut_2t, \neg Ut_1t \wedge \neg Ut_2t\}$.

Let $B : S\mathcal{L} \longrightarrow [0,1]$ be such that

$$B(\varphi) := \begin{cases} \frac{1}{100} & \text{iff } \vDash \varphi \leftrightarrow \omega \text{ for some } \omega \in O_4 \\ \frac{50}{100} & \text{iff } \vDash \varphi \leftrightarrow \omega \vee \nu \text{ for different } \omega, \nu \in O_4 \\ \frac{99}{100} & \text{iff } \vDash \varphi \leftrightarrow \neg\omega \text{ for some } \omega \in O_4 \\ 1 & \text{iff } \varphi \text{ is a tautology} \\ 0 & \text{otherwise} \end{cases}$$

Clearly, $B \in \mathbb{B}_{\mathcal{L}}$. We now show that there does *not* exist a $P \in \mathbb{P}_{\mathcal{L}}$ such that $B(\varphi) \leq P(\varphi)$ for all $\varphi \in SL$.

Note that

$$\sum_{\omega \in O_4} B(\neg \omega) = 3 + \frac{96}{100}$$

and that for all $P \in \mathbb{P}_{\mathcal{L}}$ it holds that

$$\sum_{\omega \in O_4} P(\neg \omega) = \sum_{\omega \in O_4} \Big(P(\neg \omega) + P(\omega) \Big) - \sum_{\omega \in O_4} P(\omega) = 4 - 1 = 3. \tag{4}$$

$\square$

Note for later reference that for all $n \geq 3$ and $\omega \in O_4$, $\{\neg \omega\} \cup \{\nu \in \Omega_n \ : \ \nu \vDash \omega\}$ is a partition. So, $\sum_{\substack{\nu \in \Omega_n \\ \nu \vDash \omega}} B(\nu) \leq \frac{1}{100}$ has to hold. Hence, $\sum_{\nu \in \Omega_n} B(\nu) \leq \frac{4}{100}$.

Thus far we have considered partitions of sentences. We shall also need to consider partitions of propositions:

Definition 4 (**Partitions of propositions**). *Let Π_n be the set of partitions on Ω_n. As in Section 2, we take $\{\Omega_n\}$ and $\{\Omega_n, \emptyset\}$ to be partitions and we suppose that there is no further partition containing $\emptyset$. We then define the set of partitions: $\Pi := \bigcup_{n=1}^{\infty} \Pi_n$.*

We use π^n to denote the partition of n-states $\{\{\omega\} \ : \ \omega \in \Omega_n\}$.

Note that $F_1 := \{\omega \in \Omega_1 \ : \ \omega \vDash U_1 t_1\}$ and $F_2 := \{\omega \in \Omega_2 \ : \ \omega \vDash U_1 t_1\}$ are different propositions, where U_1 is a unary predicate symbol. F_1 is a member of $\{F_1, \bar{F}_1\} \in \Pi_1$ and F_2 is a member of $\{F_2, \bar{F}_2\} \in \Pi_2$, but not vice versa. So $\{F_1, \bar{F}_1\}$ and $\{F_2, \bar{F}_2\}$ are *different* partitions, even if these partitions are intuitively equivalent.

3.3. Application to Inductive Logic

We shall be particularly interested in the use of objective Bayesianism over predicate languages to provide semantics for inductive logic.

Inductive logic typically seeks to answer questions of following form [9] (§1.1):

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^?$$

This asks, if premiss sentences $\varphi_1, \ldots, \varphi_k$ of $\mathcal{L}$ have probabilities in sets $X_1, \ldots, X_k \subseteq [0, 1]$ respectively, which probability or set of probabilities should attach to the conclusion sentence ψ?

The answer to this question will depend on the semantics given to the inductive entailment relation $\approx$ [9] (Part I). One natural option is to give the entailment relation objective Bayesian semantics, denoted by $\approx$. Here the premisses are construed as statements about chance, *i.e.,* $P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_k) \in X_k$, and the question concerns rational belief: if one's total evidence is captured by the premisses, to what extent should one believe the conclusion sentence ψ? Applying the norms of objective Bayesianism,

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^Y$$

holds just in case $P(\psi) \in Y$ for every $P \in \mathbb{E}_{\mathcal{L}}$ that has maximal entropy, where

$$\mathbb{E}_{\mathcal{L}} = \langle \varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \rangle := \langle \{ P^* \in \mathbb{P}_{\mathcal{L}} \ : \ P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_k) \in X_k \} \rangle.$$

This application of objective Bayesian epistemology to inductive logic is an example in which $\mathbb{E}_{\mathcal{L}}$ is generated by constraints involving only sentences of some finite sublanguage $\mathcal{L}_n$. We will be particularly interested in the case where $\varphi_1, \ldots, \varphi_k$ are quantifier-free sentences, *i.e.,* sentences of $\mathcal{L}_n^{\not\exists}$ for some n.

Let $\mathbb{P}_{\mathcal{L}_n^{\not\exists}}$ be the set of probability functions on $\mathcal{L}_n^{\not\exists}$, and let

$$\mathbb{E}_n := \{ P_n \in \mathbb{P}_{\mathcal{L}_n^{\not\exists}} : P_n = P_{\restriction n}, \ P \in \mathbb{E}_{\mathcal{L}} \}$$

where $P_{\restriction n}$ is the restriction of P to $S\mathcal{L}_n^{\not\exists}$. Note that,

$$P_{\restriction n}(\theta) := \sum_{\substack{\omega \in \Omega_n \\ \omega \models \theta}} P(\omega)$$

for all $\theta \in S\mathcal{L}_n^{\not\exists}$.

To ease the reading we also let $\mathbb{P}_n := \{ P_n \in \mathbb{P}_{\mathcal{L}_n^{\not\exists}} \}$.

Definition 5 (Finitely generated evidence set). $\mathbb{E}_{\mathcal{L}}$ *is* finitely generated *if it takes the form* $\mathbb{E}_{\mathcal{L}} = \{ P \in \mathbb{P}_{\mathcal{L}} \ : \ P_{\restriction n} \in \mathbb{E}_n \}$ *for some* $n \in \mathbb{N}$*, where* $\mathbb{E}_n \subseteq \mathbb{P}_{\mathcal{L}_n^{\not\exists}}$*. Thus,* $\mathbb{E}_{\mathcal{L}}$ *is generated by constraints involving only some* $\varphi_1, \varphi_2, \ldots \in S\mathcal{L}_n^{\not\exists}$ *and no other sentences.*

From now on, for finitely generated $\mathbb{E}_{\mathcal{L}}$, the letter K is used to denote the smallest number n such that $\mathbb{E}_{\mathcal{L}}$ is generated by constraints on $\mathcal{L}_n^{\not\exists}$.

Note that an evidence set $\mathbb{E}_{\mathcal{L}}$ which is not finitely generated may not be recapturable from $\{ \mathbb{E}_1, \mathbb{E}_2, \ldots \}$. For instance, for

$$\mathbb{E}_{\mathcal{L}} = \{ P \in \mathbb{P}_{\mathcal{L}} \ : \ \lim_{n \longrightarrow \infty} P(\bigwedge_{i=1}^n Ut_i) = 0 \}$$

the following two facts hold simultaneously:

1. $\mathbb{E}_{\mathcal{L}} \subset \mathbb{P}_{\mathcal{L}}$

2. $\mathbb{E}_n = \mathbb{P}_n$ for all $n \in \mathbb{N}$.

4. Quantifier-Free Languages

We would like to develop an analogue of Theorem 2 for beliefs defined over the sentences of a predicate language: we would like to show that belief functions which minimise worst-case expected loss are probability functions in $\mathbb{E}$ that maximise entropy. The main difficulty in moving from the finite domain of propositions to countably many sentences of a predicate language is to ensure that worst-case expected loss is finite where possible, so that these losses can be compared and a belief function can be chosen that minimises worst-case expected loss. For this reason we proceed in two steps. First, in this section, we shall consider the case in which the predicate language has no quantifier symbol, *i.e.,* $\mathcal{L} = \mathcal{L}^{\not\exists}$; comparing worst-case expected loss is more straightforward in this case. Then, in Section 5, we shall examine how far our approach can be extended to handle predicate languages with quantifiers.

First, in Section 4.1 we define the notion of a weighting function. This allows us to define and analyse the concept of entropy of a probability function on $\mathcal{L} = \mathcal{L}^{\bar{\exists}}$ in Section 4.2. In Section 4.3 we introduce the idea of the loss profile of a belief function. Finally in Section 4.4 we show that, in various natural scenarios, the belief function that has the best loss profile is the probability function, from all those calibrated with evidence, that has maximal standard entropy.

4.1. Weighting Functions

Definition 6 (**Weighting function**). *A weighting function on $\mathcal{L}_n$, $g_n : \Pi_n \longrightarrow \mathbb{R}_{\geq 0}$, maps partitions $\pi \in \Pi_n$ to non-negative real numbers. A weighting function on $\mathcal{L}$, $g^{\mathcal{L}} : \Pi \longrightarrow \mathbb{R}_{\geq 0}$, is defined over partitions of propositions of all finite sublanguages. A weighting function on $\mathcal{L}$ can be thought of as a family of weighting functions g_n on $\mathcal{L}_n$, where n ranges over the natural numbers. Given a fixed weighting function $g^{\mathcal{L}}$ on $\mathcal{L}$, we shall take $g_n^{\mathcal{L}} := g_{|\Pi_n}$ for each $n \in \mathbb{N}$. A (general) weighting function g is taken to be defined over each predicate language $\mathcal{L} = \mathcal{L}^{\bar{\exists}}$. Different languages $\mathcal{L} = \mathcal{L}^{\bar{\exists}}$, $\mathcal{L}' = \mathcal{L}'^{\bar{\exists}}$ have different sets of relation symbols.*

A weighting function g is *atomic* if for each $\mathcal{L}$ and each n, g_n depends only on the number of atomic propositions in $\mathcal{L}_n$, not on the structure of those atomic propositions. Thus if $\mathcal{L}$ and $\mathcal{L}'$ are such that $\mathcal{L}_n$ and $\mathcal{L}'_m$ have the same number of atomic propositions, then $g_m^{\mathcal{L}} = g_m^{\mathcal{L}'}$. In this paper we shall suppose that all weighting functions are atomic; hence there will be no need to superscript a weighting function on $\mathcal{L}$ or $\mathcal{L}_n$ by the particular language $\mathcal{L}$.

We call g *inclusive*, if and only if it attaches positive weight to each proposition, *i.e.*, if and only if for all n and all $F \subseteq \Omega_n$ it holds that

$$\sum_{\substack{\pi \in \Pi_n \\ F \in \pi}} g(\pi) > 0.$$

As in Section 2, g is *symmetric* if for each n it is invariant under permutations of the states of $\mathcal{L}_n$. It is *refined* if for each n it gives no less weight to a refinement $\pi' \in \Pi_n$ of a partition $\pi \in \Pi_n$ than to π itself.

For example, the *partition weighting* g_{Π} gives weight 1 to each partition, $g_{\Pi}(\pi) = 1$ for all $\pi \in \Pi$. The *proposition weighting* $g_{\mathcal{P}\Omega}$ gives weight 1 to each partition of size 2 and weight 0 to all other partitions; this amounts to giving weight 1 to each proposition. The *standard weighting* g_{Ω} gives weight 1 to the partition π^n of n-states, for each n, and weight 0 to all other partitions. These weighting functions are all symmetric. The partition and proposition weightings are inclusive, but the standard weighting is not. The partition and standard weightings are refined, but the proposition weighting is not.

Definition 7 (**Strongly refined weighting function**). *g is strongly refined if and only if it satisfies the following properties:*

1. *g is refined: in each finite sublanguage, if partition π' is a refinement of partition π, then $g(\pi') \geq g(\pi)$.*

2. *Each finite sublanguage receives the same total weight: for all n, $\sum_{\pi \in \Pi_n} g(\pi)$ is constant.*

3. *A state partition on a richer language should not receive less weight than one one a less rich language: if $m < n$ then $g(\pi^m) \leq g(\pi^n)$*

4. *Non-state-partitions receive finite total weight: the following limit exists (*i.e.*, is finite),*

$$\lim_{k \longrightarrow 0} \sum_{n=1}^{k} \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi).$$

Throughout this paper we will be particularly interested in the following weighting functions:

Definition 8 (**Regular weighting function**). *g is* regular *if it is atomic, inclusive, symmetric and strongly refined.*

4.2. Entropy

Definition 9 (*n-entropy*). *Given a weighting function g and $n \in \mathbb{N}$, we define the n-entropy $H_g^n :$ $\mathbb{P}_{\mathcal{L}} \longrightarrow [0, \infty]$ by:*

$$H_g^n(P) := - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^{\circ}P(F) \log {}^{\circ}P(F). \tag{5}$$

Recall that, for a probability function P (or indeed any belief function that respects logical equivalence) defined on sentences, ${}^{\circ}P$ is the function induced by P over the domain of propositions. Note that by our convention, $-0 \log 0 = 0 = -1 \log 1$. Thus, for all $n \in \mathbb{N}$,

$$g(\{\Omega_n\})P(\Omega_n) \log P(\Omega_n) = 0 = g(\{\Omega_n, \emptyset\})(P(\emptyset) \log P(\emptyset) + P(\Omega_n) \log P(\Omega_n)).$$

In calculating n-entropy we may thus ignore all partitions which contain Ω_n.

Definition 10 (**Standard entropy**). *For the standard weighting g_Ω we denote the corresponding n-entropy by H_Ω^n. We refer to H_Ω^n as* standard entropy *(on $\mathcal{L}_n$). $H_\Omega^n(P)$ is the well-known Shannon Entropy of the n-states of P:*

$$H_\Omega^n(P) = - \sum_{\omega \in \Omega_n} P(\omega) \log P(\omega).$$

For a fixed weighting function g, we say that $P \in \mathbb{P}_{\mathcal{L}}$ has *greater entropy* than $Q \in \mathbb{P}_{\mathcal{L}}$, written $P \gg Q$, if the n-entropy of P eventually dominates that of Q, *i.e.*, if there is some $N \in \mathbb{N}$ such that for all $n \geq N$, $H_g^n(P) > H_g^n(Q)$.

This relation $\gg$ for comparing entropy is preferable to an alternative notion posed in terms of the limiting behaviour of the n-entropy of P and Q, which says that P has greater entropy than Q just when $\lim_{n \longrightarrow \infty} H_g^n(P) > \lim_{n \longrightarrow \infty} H_g^n(Q)$. This is because the limiting behaviour is not fine-grained enough to distinguish greater from lesser entropy: n-entropy will often tend to infinity for both P and Q, and, even where the limiting n-entropy of P and Q are both finite, these limits may be equal even though the entropy of P is intuitively greater than that of Q, insofar as the n-entropy of P eventually dominates that of Q. See Williamson [1] (§5.5) for further discussion of these comparative notions of entropy.

We will be particularly interested in the probability functions in $[\mathbb{E}_{\mathcal{L}}]$ with maximal entropy:

$$\mathrm{maxent}\,\mathbb{E}_{\mathcal{L}} := \{P \in [\mathbb{E}_{\mathcal{L}}] \ : \ \text{there is no } Q \in [\mathbb{E}_{\mathcal{L}}] \text{ such that } Q \gg P\}.$$

We shall also consider entropy maximisers on finite sublanguages. We shall use the notation:

$$\mathbb{P}_n^\dagger := \arg \sup_{P \in \mathbb{E}_n} H_g^n(P).$$

(The members of this set are defined only on the sentences of $\mathcal{L}_n$, not on the sentences of the language $\mathcal{L}$ as a whole.) Note that for convex $\mathbb{E}_{\mathcal{L}}$, $\mathbb{E}_n$ is convex for all $n \in \mathbb{N}$ and that H_g^n is a strictly concave function on $\mathbb{E}_n$ for inclusive g. If g is inclusive, then H_g^n is strictly concave on $\mathbb{P}_n$. Hence $\mathbb{P}_n^\dagger$ contains a unique element, which we will denote by $P_n^\dagger$.

Let us consider the set of limit points of the entropy maximisers on finite sublanguages:

Definition 11 (**Entropy limit**). *A probability function is a limit point of the entropy maximisers on finite sublanguages if it is arbitrarily close to infinitely many such maximisers. We denote the set of such limit points by:*

$$\mathbb{P}^\dagger := \{P \in \mathbb{P}_{\mathcal{L}} \ : \ \forall \epsilon > 0, \exists \text{ infinite } I \subseteq \mathbb{N}, \forall n \in I, \forall \varphi \in S\mathcal{L}_n, |P(\varphi) - P_n^\dagger(\varphi)| < \epsilon\}.$$

Whenever $\mathbb{P}^\dagger$ consists only of a single function we shall denote that function by $P^\dagger$ and refer to $P^\dagger$ as the entropy limit.

One important desideratum for a procedure for choosing a rational belief function, particularly in the context of inductive logic, is language invariance. We shall consider two notions of language invariance: the following notion defined in terms of finite sublanguages, and a second form of language invariance, introduced in Definition 23, which we term infinite-language invariance.

Definition 12 (**Finite-language invariant weighting function**). *A weighting function $g : \Pi \longrightarrow \mathbb{R}_{\geq 0}$ is* finite-language invariant*, if and only if the following holds: for all $\mathbb{E}_{\mathcal{L}}$ finitely generated by constraints on $\mathcal{L}_K$, if $\mathcal{L}_n$ and $\mathcal{L}_m$ are such that $\mathcal{L}_K \subseteq \mathcal{L}_n \subseteq \mathcal{L}_m$, then for all $Q \in \arg \sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^n(P)$ there exists some $R \in \arg \sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^m(P)$ such that $Q_{\restriction n} = R_{\restriction n}$.*

4.2.1. The Standard Entropy Limit

Standard entropy, *i.e.*, entropy with respect to the standard weighting g_Ω, is the subject of a substantial literature. We here collect the features of standard entropy most relevant for our purposes.

Firstly, g_Ω is finite-language invariant; see, e.g., [7]. If $\mathbb{E}_{\mathcal{L}}$ is finitely generated and $g = g_\Omega$, then $\mathbb{P}_n^\dagger$ contains a unique element. Furthermore, there exists a unique function $P \in [\mathbb{E}_{\mathcal{L}}]$ such that for all $n \geq K$ $P_{\restriction n} \in \mathbb{P}_n^\dagger$ holds. This function P is the entropy limit with respect to the standard weighting g_Ω; it will be called the *standard entropy limit* and denoted by $P_\Omega^\dagger$. Henceforth we use $P_\Omega^\dagger$ to denote the standard entropy limit on $\mathcal{L}$, rather than on Ω as in Section 2.

Definition 13 (**Open-minded belief function**). *We say that a belief function $B \in \mathbb{B}_{\mathcal{L}}$ is* open-minded on $\mathcal{L}' \subseteq \mathcal{L}$, *if and only if for all $\varphi \in S\mathcal{L}'$ for which there exists some $P \in [\mathbb{E}_{\mathcal{L}}]$ such that $P(\varphi) > 0$ it holds that $B(\varphi) > 0$. For $\mathcal{L}' = \mathcal{L}$ we say that the belief function $B \in \mathbb{B}_{\mathcal{L}}$ is* open-minded.

The following proposition lists further important properties of $P_\Omega^\dagger$ which we shall make frequent use of in the following two properties—see [7] (p. 95) for a proof of the first property.

Proposition 5. $P_\Omega^\dagger$ *satisfies the following properties:*

- $P_\Omega^\dagger$ *is open-minded.*

- *For a finitely generated* $\mathbb{E}_\mathcal{L}$, *for all* $n \geq K$ *and all* $\nu \in \Omega_n, \omega \in \Omega_K$ *with* $\nu \vDash \omega$ *it holds that* $P_\Omega^\dagger(\nu) = P_\Omega^\dagger(\omega)\frac{|\Omega_K|}{|\Omega_n|}$.

The second property will follow from Proposition 9 and from the fact that g_Ω is language invariant.

Let ν be a consistent conjunction of pairwise different literals such that $\nu \vDash \omega$ for some n-state ω with $n \geq K$. Denoting by $|\nu|, |\omega|$ the number of literals in ν, respectively, ω, it follows from the second property in Proposition 5 that $P_\Omega^\dagger(\nu) = P_\Omega^\dagger(\omega)\frac{2^{|\omega|}}{2^{|\nu|}}$.

4.2.2. General Entropies

The question remains as to how the functions on $\mathcal{L}$ with maximal entropy, *i.e.,* the members of $\mathrm{maxent}\, \mathbb{E}_\mathcal{L}$, relate to the entropy maximisers $P_n^\dagger \in \mathbb{P}_n^\dagger$ on the finite sublanguages $\mathcal{L}_n$. We shall explore this question here.

Proposition 6. $\mathbb{P}^\dagger \subseteq [\mathbb{E}_\mathcal{L}]$.

Proof. Let $P^\dagger \in \mathbb{P}^\dagger$. Thus, for all sentences $\varphi \in S\mathcal{L}$, $P^\dagger(\varphi)$ is the limit of a sequence $(P_n^\dagger)_{n \in I}$ such that $P_n^\dagger \in [\mathbb{E}_n]$ and $I \subseteq \mathbb{N}$ is infinite. Since $[\mathbb{E}_\mathcal{L}]$ and all the $[\mathbb{E}_n]$ are closed, $P^\dagger \in [\mathbb{E}_\mathcal{L}]$. $\square$

Of particular interest is the most equivocal probability function of $\mathbb{P}_\mathcal{L}$, which is called the *equivocator* and denoted by $P_=$. $P_=$ is uniquely defined by the requirement that for all $n \in \mathbb{N}$ it assigns all n-states $\omega \in \Omega_n$ the same probability, $P_=(\omega) = \frac{1}{|\Omega_n|}$. The restriction of $P_=$ to $\mathbb{P}_n$ is denoted by $P_{=\restriction n}$.

In certain cases $\mathbb{P}^\dagger$ will only contain a single limit point $P^\dagger$.

Definition 14. [4] (Definition 16, p. 3573.) *A weighting function* g_n *on* $\mathcal{L}_n$ *is called equivocator-preserving, if and only if*

$$\mathbb{P}_n^\dagger = \{Q_{\restriction n} \;:\; Q \in \arg\sup_{P \in \mathbb{P}_\mathcal{L}} H_g^n(P)\} = \{P_{=\restriction n}\}.$$

g is called equivocator-preserving, if and only if g_n *is equivocator-preserving for all* $n \in \mathbb{N}$.

Proposition 7. *If* $P_= \in [\mathbb{E}_\mathcal{L}]$ *and if* g *is symmetric and inclusive, then* $\mathbb{P}^\dagger = \{P_=\}$.

Proof. By Landes and Williamson [4] (Corollary 6, p. 3574) we have

$$\arg\sup_{P \in \mathbb{E}_\mathcal{L}} H_g^n(P) = \{P \in \mathbb{P}_\mathcal{L} \;:\; P_{\restriction n} = P_{=\restriction n}\}.$$

It follows that

$$\lim_{n \longrightarrow \infty} \arg\sup_{P \in \mathbb{E}_\mathcal{L}} H_g^n(P) = \{P_=\}$$

and hence $\mathbb{P}^\dagger = \{P_=\}$. $\square$

So, if g is symmetric and inclusive, then g is equivocator-preserving. In Appendix B we shall show that there exist non-symmetric g which are equivocator-preserving.

Definition 15 (**State-inclusive weighting function**). *Given $\mathcal{L}$, we call a weighting function $g : \Pi \longrightarrow [0, 1]$ state-inclusive on $\mathcal{L}_n$, if and only if for each state $\omega \in \Omega_n$ there exists a $\pi \in \Pi_n$ such that $\{\omega\} \in \pi$ and $g(\pi) > 0$. A weighting function $g : \Pi \longrightarrow [0, 1]$ is* state-inclusive, *if and only if it is state-inclusive on each $\mathcal{L}_n$. It is* eventually state-inclusive, *if and only if there exists a $J \in \mathbb{N}$ such that for all $n \geq J$, g is state-inclusive on $\mathcal{L}_n$.*

For example, if $g(\pi^n) > 0$ for all $n \in \mathbb{N}$, then g is state-inclusive. Moreover, inclusive implies state-inclusive.

Lemma 1. *If g is state-inclusive on $\mathcal{L}_n$, then H_g^n is strictly concave on $\mathbb{P}_n$.*

Proof. Let $P, Q \in \mathbb{P}_n$ be different and $\lambda \in (0, 1)$. Since for all $\pi \in \Pi_n$ we have $\sum_{F \in \pi} {}^{\circ}P(F) = 1 = \sum_{F \in \pi} {}^{\circ}Q(F)$ we find using the strict concavity of $-x \cdot \log x$ on $[0, 1]$

$$H_g^n(\lambda P + (1 - \lambda)Q) = \sum_{\pi \in \Pi_n} -g(\pi) \sum_{F \in \pi} \left(\lambda {}^{\circ}P(F) + (1 - \lambda){}^{\circ}Q(F) \right) \cdot \log(\lambda {}^{\circ}P(F) + (1 - \lambda){}^{\circ}Q(F))$$

$$\geq \sum_{\pi \in \Pi_n} -g(\pi) \sum_{F \in \pi} \left(\lambda {}^{\circ}P(F) \log(\lambda {}^{\circ}P(F)) \right) + \left((1 - \lambda){}^{\circ}Q(F) \log((1 - \lambda){}^{\circ}Q(F)) \right)$$

$$= H_g^n(\lambda P) + H_g^n((1 - \lambda)Q).$$

The inequality is strict, if and only if there exists some $\pi \in \Pi_n$ with $g(\pi) > 0$ such that there is some $F \in \pi$ with ${}^{\circ}P(F) \neq {}^{\circ}Q(F)$. Since P, Q are different probability functions, there exists some $\omega \in \Omega_n$ such that $P(\omega) \neq Q(\omega)$. Since g is state-inclusive, $g(\pi) > 0$ for some $\pi \in \Pi_n$ with $\{\omega\} \in \pi$. Hence, the inequality is strict. $\square$

Proposition 8. *If $\mathbb{E}_{\mathcal{L}}$ is finitely generated, and g is eventually state-inclusive and language invariant, then $\mathbb{P}^{\dagger}$ consists of a single probability function $P^{\dagger}$ and for all $\varphi \in S\mathcal{L}$ it holds that $\lim_{n \longrightarrow \infty} P_n^{\dagger}(\varphi) = P^{\dagger}(\varphi)$.*

Proof. Recall that $\mathbb{E}_{\mathcal{L}}$ is expressible by constraints in $\mathcal{L}_K$ and let J as in Definition 15. Let $n \geq \max\{J, K\}$.

By the above Lemma 1, H_g^n is strictly concave on $\mathbb{P}_n$. Since $\mathbb{E}_n$ is convex, $\arg\sup_{P \in \mathbb{E}_n} H_g^n(P)$ contains a single element. Hence, $Q, R \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^n(P)$ agree on $S\mathcal{L}_n$.

Since g is language invariant, we have $\arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^m(P) \subseteq \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^l(P)$ for all $n \leq l \leq m$.

For all $\varphi \in S\mathcal{L}$, there exists an $s \in \mathbb{N}$ such that $\varphi \in S\mathcal{L}_s$. Hence, for $l, m \geq \max\{J, K, s\}$ it holds that for $R \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^m(P)$ and $Q \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^l(P)$ that $R(\varphi) = Q(\varphi)$. $\square$

For instance, standard entropy [4] (Equation 80), the substate weighting and other examples generated by Landes and Williamson [4] (Lemma 8) are eventually state-inclusive and language invariant. Note that these weighting functions are not inclusive.

Definition 16. *We say that H_g is strictly concave, if and only if for all $n \in \mathbb{N}$, H_g^n is strictly concave on $\mathbb{P}_n$.*

Proposition 9 (**Equivocation beyond $\mathcal{L}_n$**). *Let $\mathbb{E}_{\mathcal{L}}$ be finitely generated and let g be symmetric. If H_g is strictly concave, then for all $n \geq K$ and all $\nu, \mu \in \Omega_n$ such that there exists an $\omega \in \Omega_K$ with $\nu \vDash \omega$ and $\mu \vDash \omega$ it holds that*

$$P_n^\dagger(\nu) = P_n^\dagger(\mu) = P_n^\dagger(\omega) \cdot \frac{|\Omega_K|}{|\Omega_n|}$$

for all $P_n^\dagger \in \mathbb{P}_n^\dagger$.

We call such $\nu, \mu \in \Omega_n$ *extensions* of $\omega \in \Omega_K$ and say that $P_n^\dagger$ *equivocates beyond* $\mathcal{L}_K$. In particular, $P_n^\dagger$ equivocates beyond $\mathcal{L}_K$ up to $\mathcal{L}_n$.

Proof. Let $n > K$ and let $P \in [\mathbb{E}_{\mathcal{L}}]$ be such that there exist $\nu, \mu \in \Omega_n$ with $P(\nu) \neq P(\mu)$ such that there exists an $\omega \in \Omega_K$ with $\nu \vDash \omega$ and $\mu \vDash \omega$. Assume for contradiction that $P \in \arg\sup_{R \in \mathbb{E}_{\mathcal{L}}} H_g^n(R)$.

Now define a probability function $Q \in \mathbb{P}_{\mathcal{L}}$ by first specifying Q on the n-states. Let

$$Q(\nu) := P(\mu)$$
$$Q(\mu) := P(\nu)$$
$$Q(\eta) := P(\eta) \text{ for all } \eta \in \Omega_n \setminus \{\nu, \mu\}.$$

For a $\lambda \in \Omega_r$ with $r \geq n$ we let $Q(\lambda) := Q(\xi)\frac{|\Omega_n|}{|\Omega_r|}$ where $\xi \in \Omega_r$ is the unique r-state such that $\lambda \vDash \xi$.

By construction, Q and P agree on $S\mathcal{L}_K$. Since $\mathbb{E}_{\mathcal{L}}$ is finitely generated, it follows that $Q \in [\mathbb{E}_{\mathcal{L}}]$. Furthermore, $Q_{\downarrow n}$ can be obtained from $P_{\downarrow n}$ by a renaming of n-states and it holds that $Q_{\downarrow n} \neq P_{\downarrow n}$. Since g_n is symmetric it holds that $H_g^n(P) = H_g^n(Q)$. Since $[\mathbb{E}_n]$ is convex and H_g^n is strictly concave, neither $P_{\downarrow n}$ nor $Q_{\downarrow n}$ can maximise H_g^n over $[\mathbb{E}_n]$.

This contradicts P maximising H_g^n over $[\mathbb{E}_n]$. □

Corollary 1. *Let $\mathbb{E}_{\mathcal{L}}$ be finitely generated. If H_g^n is strictly concave on $\mathbb{P}_n$ for $n \geq K$ and if g is symmetric, then for $n \geq K$ the following maximisation problem*

$$\begin{aligned} \textit{maximise:} \quad & H_g^n(P) \\ \textit{subject to:} \quad & P \in [\mathbb{E}_{\mathcal{L}}] \end{aligned}$$

can be understood as an optimisation problem in the variables $P(\omega)$ with $\omega \in \Omega_K$. In particular, the number of variables does not *grow as n tends to infinity.*

Proof. Follows immediately from the above proposition by noting that $P_n^\dagger \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} H_g^n(P)$ equivocates beyond $\mathcal{L}_K$ up to $\mathcal{L}_n$. □

This corollary shows that in order to compute $P_n^\dagger$ for $n \geq K$ one needs to solve an optimisation problem on Ω_K. If g is not language invariant, then, in general, the objective function of the optimisation problem changes as n changes. So, in general, $(P_n^\dagger)_{\downarrow K}$ varies with n.

Corollary 2. *Under the assumptions of Proposition 9 it holds that for $F \subseteq \Omega_n$ and $\nu, \mu \in \Omega_n$, $^\circ P_n^\dagger(F) = {}^\circ P_n^\dagger(F_{\nu,\mu})$, where $F_{\nu,\mu}$ is the result we obtain by replacing ν by μ and vice versa in F.*

Proof. For an $\eta \in \Omega_n$ denote by $\omega_\eta \in \Omega_K$ the unique K-state such that $\eta \vDash \omega_\eta$. Now simply note that by Proposition 9

$$^\circ P_n^\dagger(F) = \sum_{\substack{\eta \in \Omega_n \\ \eta \in F}} P_n^\dagger(\eta) = \sum_{\substack{\eta \in \Omega_n \\ \eta \in F}} P_n^\dagger(\omega_\eta) \frac{|\Omega_K|}{|\Omega_n|} = \sum_{\substack{\eta \in \Omega_n \\ \eta \in F_{\nu,\mu}}} P_n^\dagger(\omega_\eta) \frac{|\Omega_K|}{|\Omega_n|} = {}^\circ P_n^\dagger(F_{\nu,\mu}).$$

$\square$

Corollary 3. *Let $\mathbb{E}_\mathcal{L}$ be finitely generated. For all $n \geq K$ and all $P \in \mathbb{P}_\mathcal{L}$ equivocating beyond $\mathcal{L}_K$ up to $\mathcal{L}_n$ it holds for all $K \leq k \leq n-1$ that*

$$H_\Omega^{k+1}(P) = H_\Omega^k(P) - \log \frac{|\Omega_k|}{|\Omega_{k+1}|}.$$

If g is symmetric and H_g is strictly concave, then

$$H_\Omega^K(P_k^\dagger) - H_\Omega^K(P_\Omega^\dagger) = H_\Omega^{k+1}(P_{k+1}^\dagger) - H_\Omega^{k+1}(P_\Omega^\dagger).$$

Proof. For $\nu \in \Omega_{k+1}$ let $\omega_\nu \in \Omega_k$ be the unique k state such that $\nu \vDash \omega_\nu$. For $K \leq k \leq n-1$ we now find for a $P \in \mathbb{P}_\mathcal{L}$ equivocating beyond $\mathcal{L}_K$ up to $\mathcal{L}_n$

$$
\begin{aligned}
H_\Omega^{k+1}(P) &= -\sum_{\nu \in \Omega_{k+1}} P(\nu) \log P(\nu) \\
&= -\sum_{\nu \in \Omega_{k+1}} P(\nu) \log\left(P(\omega_\nu) \cdot \frac{|\Omega_k|}{|\Omega_{k+1}|}\right) \\
&= -\log \frac{|\Omega_k|}{|\Omega_{k+1}|} - \sum_{\nu \in \Omega_{k+1}} P(\nu) \log P(\omega_\nu) \\
&= -\log \frac{|\Omega_k|}{|\Omega_{k+1}|} - \sum_{\omega \in \Omega_k} \sum_{\substack{\nu \in \Omega_{k+1} \\ \nu \vDash \omega}} P(\nu) \log P(\omega_\nu) \\
&= -\log \frac{|\Omega_k|}{|\Omega_{k+1}|} - \sum_{\omega \in \Omega_k} \log P(\omega) \cdot \left(\sum_{\substack{\nu \in \Omega_{k+1} \\ \nu \vDash \omega}} P(\nu)\right) \\
&= -\log \frac{|\Omega_k|}{|\Omega_{k+1}|} - \sum_{\omega \in \Omega_k} P(\omega) \log P(\omega) \\
&= -\log \frac{|\Omega_k|}{|\Omega_{k+1}|} + H_\Omega^k(P).
\end{aligned}
$$

The second part of the proof follows directly by observing that $P_\Omega^\dagger$ and $P_n^\dagger$ equivocate beyond $\mathcal{L}_K$ up to $\mathcal{L}_n$ by Proposition 9. $\square$

Corollary 4. *Let $\mathbb{E}_\mathcal{L}$ be finitely generated. For all $n \geq K$ and all $P \in \mathbb{P}_\mathcal{L}$ not equivocating beyond $\mathcal{L}_K$ up to $\mathcal{L}_n$ it holds that $H_\Omega^n(P) < H_\Omega^K(P) - \log \frac{|\Omega_K|}{|\Omega_n|}$.*

Proof. There has to exist at least one $\xi \in \Omega_K$ such that there exist $\nu, \lambda \in \Omega_n$ with $\nu \vDash \xi$ and $\lambda \vDash \xi$ such that $P(\nu) \neq P(\lambda)$. Since P is a probability function it holds that $P(\xi) = \sum_{\substack{\nu \in \Omega_n \\ \nu \vDash \xi}} P(\nu)$. We thus find using the log-sum inequality (see, e.g., Theorem 2.7.1 in [10])

$$
\begin{aligned}
-P(\xi) \log(\frac{|\Omega_K|}{|\Omega_n|}) - P(\xi) \log P(\xi) &= -P(\xi) \log\Big(\frac{|\Omega_K|}{|\Omega_n|} P(\xi)\Big) \\
&= -\sum_{\substack{\nu \in \Omega_n \\ \nu \vDash \xi}} \Big(\frac{|\Omega_K|}{|\Omega_n|} P(\xi)\Big) \log\Big(\frac{|\Omega_K|}{|\Omega_n|} P(\xi)\Big) \\
&> -\sum_{\substack{\nu \in \Omega_n \\ \nu \vDash \xi}} P(\nu) \log P(\nu).
\end{aligned}
$$

If $\xi \in \Omega_K$ is such that for all $\nu, \lambda \in \Omega_n$ with $\nu \vDash \xi$ and $\lambda \vDash \xi$ it holds that $P(\nu) = P(\lambda)$, then the above calculation holds with the exception that the inequality is in fact an equality.

We hence find by summing over all $\omega \in \Omega_K$

$$
\begin{aligned}
H_\Omega^K(P) - \log(\frac{|\Omega_K|}{|\Omega_n|}) &= \sum_{\omega \in \Omega_K} -P(\omega) \Big(\log(\frac{|\Omega_K|}{|\Omega_n|}) + \log P(\omega)\Big) \\
&> -\sum_{\omega \in \Omega_K} \sum_{\substack{\nu \in \Omega_n \\ \nu \vDash \omega}} -P(\nu) \log P(\nu) \\
&= \sum_{\nu \in \Omega_n} -P(\nu) \log P(\nu) \\
&= H_\Omega^n(P).
\end{aligned}
$$

$\square$

Corollary 5. *Let $\mathbb{E}_\mathcal{L}$ be finitely generated. If g is symmetric and if for all $n \geq K$ H_g^n is strictly concave on $\mathbb{P}_n$, then*

$$
\mathbb{P}^\dagger \neq \emptyset.
$$

Proof. By Corollary 1, $P_n^\dagger \in [\mathbb{E}_n]$ is uniquely determined by $P_n^\dagger(\omega)$ for $\omega \in \Omega_K$. That is, we can understand $(P_n^\dagger)_{n \in \mathbb{N}}$ as sequence taking values in $[0,1]^{|\Omega_K|} \subset \mathbb{R}^{|\Omega_K|}$ and $[0,1]^{|\Omega_K|}$ is compact. Hence, the sequence $((P_n^\dagger)_{|K})_{n \in \mathbb{N}}$ has point of accumulation, Q, with $Q \in [\mathbb{E}_K]$. Let $I \subseteq \mathbb{N}$ be infinite such that $\lim_{i \in I, i \longrightarrow \infty} P_{n_i}^\dagger(\omega) = Q(\omega)$ for all $\omega \in \Omega_K$.

Recall that for $n > K$ that $P_n^\dagger$ equivocates under $\mathcal{L}_K$ up to $\mathcal{L}_n$. We now extend Q to a probability function in $[\mathbb{E}_\mathcal{L}]$ by defining it on the n-states $\nu \in \Omega_n$ for $n > K$ as follows: $Q(\nu) := \frac{|\Omega_K|}{|\Omega_n|} \cdot Q(\omega_\nu) = \frac{|\Omega_K|}{|\Omega_n|} \cdot \lim_{i \in I, i \longrightarrow \infty} P_{n_i}^\dagger(\omega_\nu)$. Hence, Q equivocates beyond $\mathcal{L}_K$.

Consider some $\varphi \in S\mathcal{L}$. It follows that there is some $r \geq K$ such that $\varphi \in S\mathcal{L}_r$. For $\nu \in \Omega_r$ denote by ω_ν the unique element of Ω_K such that $\nu \vDash \omega_\nu$.

We thus find

$$
\begin{aligned}
\lim_{\substack{i \longrightarrow \infty \\ i \in I}} P_{n_i}^\dagger(\varphi) &= \lim_{\substack{i \longrightarrow \infty \\ i \in I}} \sum_{\substack{\nu \in \Omega_r \\ \nu \vDash \varphi}} P_{n_i}^\dagger(\nu) \\
&= \sum_{\substack{\nu \in \Omega_r \\ \nu \vDash \varphi}} \lim_{\substack{i \longrightarrow \infty \\ i \in I}} P_{n_i}^\dagger(\nu) \\
&= \sum_{\substack{\nu \in \Omega_r \\ \nu \vDash \varphi}} \frac{|\Omega_K|}{|\Omega_n|} \cdot \lim_{\substack{i \longrightarrow \infty \\ i \in I}} P_{n_i}^\dagger(\omega_\nu) \\
&= \sum_{\substack{\nu \in \Omega_r \\ \nu \vDash \varphi}} \frac{|\Omega_K|}{|\Omega_n|} \cdot Q(\omega_\nu) \\
&= \sum_{\substack{\nu \in \Omega_r \\ \nu \vDash \varphi}} Q(\nu) \\
&= Q(\varphi).
\end{aligned}
$$

$\square$

We now turn our attention to the calibrated functions with maximal entropy, $\mathrm{maxent}\,\mathbb{E}_{\mathcal{L}}$. Our aim is to show that $\mathrm{maxent}\,\mathbb{E}_{\mathcal{L}} = \mathbb{P}^\dagger = \{P_\Omega^\dagger\}$ holds for regular g.

Lemma 2. *If g is regular, then*

$$
\lim_{n \longrightarrow \infty} \log(|\Omega_n|) \cdot \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) = 0.
$$

Proof. Since g is total it is in particular g defined for the language $\mathcal{L}^U$ which only contains a single relation symbol which is unary. When needed, we shall add a superscript ^{U} express that we consider $\mathcal{L}^U$. Now define a sequence $(a_n)_{n \in \mathbb{N}}$ by

$$
a_n := \sum_{\pi \in \Pi_n^U \setminus \{\pi^n\}} g(\pi).
$$

By the Cauchy condensation test [11] (p. 61, Theorem 3.27) for (not necessarily strictly) decreasing sequences we have that

$$
\sum_{n=1}^\infty a_n < \infty \iff \sum_{k=0}^\infty 2^k a_{2^k} < \infty. \tag{6}
$$

Since the series on the left converges by the assumption on finite weights, so does the right, and that implies that $\lim_{k \longrightarrow \infty} 2^k a_{2^k} = 0$.

For $n \in \mathbb{N}$ let $k \in \mathbb{N}$ be such that $2^k \leq n < 2^{k+1}$. Since a_n is (not necessarily strictly) decreasing $a_n \leq a_{2^k}$. Hence,

$$
0 \leq n a_n \leq 2^{k+1} a_n \leq 2^{k+1} a_{2^k} = 2(2^k a_{2^k}).
$$

The right hand side converges to 0 by Cauchy's condensation test (6). Thus,

$$
\begin{aligned}
0 &= \lim_{n \longrightarrow \infty} n \cdot a_n \\
&= \lim_{n \longrightarrow \infty} n \cdot \log_2(2) \cdot a_n \\
&= \lim_{n \longrightarrow \infty} \log_2(2^n) \cdot a_n \\
&= \lim_{n \longrightarrow \infty} \log_2(|\Omega_n^U|) \cdot a_n \\
&= \lim_{n \longrightarrow \infty} \log(|\Omega_n^U|) \cdot a_n \\
&= \lim_{n \longrightarrow \infty} \log(|\Omega_n^U|) \cdot \sum_{\pi \in \Pi_n^U \setminus \{\pi^n\}} g(\pi).
\end{aligned}
$$

Now if $\mathcal{L}$ is some other language in our sense different from $\mathcal{L}^U$, then for all $n \in \mathbb{N}$ there exists an $m_n > n$ such that $|\Omega_n| = |\Omega_n^U|$. This in turn implies the existence of a canonical bijections f_n identifying Π_n with $\Pi_{m_n}^U$ which respect the structure of partitions.

Because g is atomic it follows that for all $\pi \in \Pi_n$ that $g(\pi) = g(f_n(\pi))$ holds. Thus,

$$
a_n = \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) = \sum_{\pi \in \Pi_{m_n}^U \setminus \{\pi^{m_n}\}} g(\pi).
$$

We then observe that the sequence $(\log(|\Omega_n|) \cdot \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi))_{n \in \mathbb{N}}$ is a subsequence of $(\log(|\Omega_n^U|) \cdot \sum_{\pi \in \Pi_n^U \setminus \{\pi^n\}} g(\pi))_{n \in \mathbb{N}}$. Hence,

$$
\begin{aligned}
0 &= \lim_{n \longrightarrow \infty} \log(|\Omega_{m_n}^U|) \cdot \sum_{\pi \in \Pi_{m_n}^U \setminus \{\pi^{m_n}\}} g(\pi) \\
&= \lim_{n \longrightarrow \infty} \log(|\Omega_n|) \cdot \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi).
\end{aligned}
$$

$\square$

Lemma 3. *If g is strongly refined and state-inclusive, then there exist $0 < a \le b < +\infty$ such that for all $n \in \mathbb{N}$, $g(\pi^n) \in [a, b]$.*

Proof. For every $\omega \in \Omega_1$ there exists some $\pi \in \Pi_1$ which contains $\{\omega\}$ with $g(\pi) > 0$. π^1 refines all these partitions (or π^1 is that partition). Hence, $g(\pi^1) > 0$.

Since state partitions on richer languages are assigned more weight it follows that $g(\pi^n) \ge g(\pi^1) > 0$ for all $n \in \mathbb{N}$.

Trivially, $g(\pi^n) \le \sum_{\pi \in \Pi_n} g(\pi)$. The latter is constant for all n. Hence, the sequence $g(\pi^n)$ is bounded from above by $\sum_{\pi \in \Pi_n} g(\pi)$.

We can thus choose a, b as follows $a := g(\pi^1)$ and $b := \sum_{\pi \in \Pi_1} g(\pi)$. $\square$

Following [4] (p. 3556) we define:

Definition 17 (**Spectrum of π**). *The spectrum of a partition π is defined as the multi-set of sizes of the members of π. We write $\sigma(\pi)$ to denote the spectrum of π.*

In other words, if π' can be obtained from π by permuting the states in the members of π, then $\sigma(\pi) = \sigma(\pi')$. If g is symmetric, then $g(\pi)$ only depends on the spectrum of π.

Lemma 4. *If g is symmetric, then for all n and all spectra s*

$$P_= \in \arg\sup_{P \in \mathbb{P}_\mathcal{L}} \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F).$$

Proof. First note that $\sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F)$ is a concave function, since $-x \log x$ is a concave function for $x \in [0, 1]$.

If $P, P' \in \mathbb{P}_\mathcal{L}$ are such that one can be obtained from the other by a permutation of n-states, then for all spectra s

$$\sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F) = \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P'(F) {}^\circ \log P'(F).$$

Hence, for all fixed spectra s $P_{=\restriction n}$ lies inside the contour lines of the function $\sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F)$ in $\mathbb{P}_n$. It follows that

$$P_= \in \arg\sup_{P \in \mathbb{P}_\mathcal{L}} \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F).$$

□

Corollary 6. *If g is symmetric and such that*

$$\lim_{n \longrightarrow \infty} \log |\Omega_n| \sum_{\substack{\pi \in \Pi_n \\ \pi \neq \pi^n}} g(\pi) = 0,$$

then for all $P \in \mathbb{P}_\mathcal{L}$

$$\lim_{n \longrightarrow \infty} \sum_{\substack{\pi \in \Pi_n \\ \pi \neq \pi^n}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F) = 0.$$

Proof. For a fixed spectrum s we have

$$\sup_{P \in \mathbb{E}_\mathcal{L}} \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P(F) = \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} {}^\circ P_=(F) \log {}^\circ P_=(F)$$

$$= \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F \in \pi} \frac{|F|}{|\Omega_n|} \cdot \log \frac{|F|}{|\Omega_n|}$$

$$= \sum_{\substack{\pi \in \Pi_n \\ \sigma(\pi)=s}} -\frac{g(\pi)}{|\Omega_n|} \sum_{F \in \pi} |F| \cdot (\log |F| - \log |\Omega_n|).$$

Thus,

$$
|\sup_{\substack{P\in\mathbb{E}_{\mathcal{L}}}} \sum_{\substack{\pi\in\Pi_n \\ \sigma(\pi)=s}} -g(\pi) \sum_{F\in\pi} {}^{\circ}P(F)\log{}^{\circ}P(F)| \leq \sum_{\substack{\pi\in\Pi_n \\ \sigma(\pi)=s}} \frac{g(\pi)}{|\Omega_n|} \sum_{F\in\pi} |F| \cdot \log|\Omega_n|
$$

$$
= \sum_{\substack{\pi\in\Pi_n \\ \sigma(\pi)=s}} g(\pi) \cdot \log|\Omega_n|.
$$

Summing over all spectra now yields for all $P \in \mathbb{P}_{\mathcal{L}}$

$$
\sum_{\substack{\pi\in\Pi_n \\ \pi\neq\pi^n}} -g(\pi) \sum_{F\in\pi} {}^{\circ}P(F)\log{}^{\circ}P(F) \leq \log|\Omega_n| \sum_{\substack{\pi\in\Pi_n \\ \pi\neq\pi^n}} g(\pi).
$$

The claimed result follows. □

In particular, if g is regular then the above Corollary applies, by Lemma 2.

Let us consider the application of objective Bayesianism to inductive logic (Section 3.3). It turns out that if g is regular and $\mathbb{E}_{\mathcal{L}}$ is finitely generated then the functions in $[\mathbb{E}_{\mathcal{L}}]$ with maximal entropy coincide with the entropy limits (Definition 11), and moreover there is a unique such function, the standard entropy limit:

Theorem 3. *Let g be symmetric, atomic, state-inclusive and strongly refined, and $\mathbb{E}_{\mathcal{L}}$ be finitely generated. Then*

$$
\text{maxent } \mathbb{E}_{\mathcal{L}} = \mathbb{P}^{\dagger} = \{P_{\Omega}^{\dagger}\}. \tag{7}
$$

Note that if g is also inclusive, then g is regular.

Proof. By Lemma 3 there exist $0 < a \leq b < +\infty$ such that $g(\pi^n) \in [a,b]$ for all $n \in \mathbb{N}$ and by Corollary 6 the combined weight given to all other partitions on Π_n tends to zero, as n increases, fast enough that, for all $P \in \mathbb{P}_{\mathcal{L}}$,

$$
\lim_{n\longrightarrow\infty} \sum_{\substack{\pi\in\Pi_n \\ \pi\neq\pi^n}} -g(\pi) \sum_{F\in\pi} {}^{\circ}P(F)\log{}^{\circ}P(F) = 0.
$$

For $Q \in [\mathbb{E}_{\mathcal{L}}] \setminus \{P_{\Omega}^{\dagger}\}$ there exists a minimal $n \in \mathbb{N}$ with $n \geq K$ such that $(P_{\Omega}^{\dagger})_{\downarrow n} \neq Q_{\downarrow n}$. Since H_{Ω}^n is strictly convex on $\mathbb{E}_n$ and $P_{\Omega}^{\dagger}$ maximises H_{Ω}^n over $[\mathbb{E}_n]$ it holds that $H_{\Omega}^n(P_{\Omega}^{\dagger}) > H_{\Omega}^n(Q)$. Using Corollary 3 and Corollary 4 we obtain $H_{\Omega}^r(P_{\Omega}^{\dagger}) - H_{\Omega}^r(Q) \geq H_{\Omega}^K(P_{\Omega}^{\dagger}) - H_{\Omega}^K(Q)$ for $r \geq n$. Thus,

$$
H_g^r(P_{\Omega}^{\dagger}) - H_g^r(Q) = - g(\pi^r)H_{\Omega}^r(P_{\Omega}^{\dagger}) + \sum_{\substack{\pi\in\Pi_r \\ \pi\neq\pi^r}} -g(\pi) \sum_{F\in\pi} {}^{\circ}P_{\Omega}^{\dagger}(F)\log{}^{\circ}P_{\Omega}^{\dagger}(F)
$$

$$
+ g(\pi^r)H_{\Omega}^r(Q) + \sum_{\substack{\pi\in\Pi_r \\ \pi\neq\pi^r}} g(\pi) \sum_{F\in\pi} {}^{\circ}Q(F)\log{}^{\circ}Q(F)
$$

$$
\geq - g(\pi^r)\Big(H_{\Omega}^n(P_{\Omega}^{\dagger}) - H_{\Omega}^n(Q)\Big)
$$

$$
+ \sum_{\substack{\pi\in\Pi_r \\ \pi\neq\pi^r}} -g(\pi) \sum_{F\in\pi} {}^{\circ}P_{\Omega}^{\dagger}(F)\log{}^{\circ}P_{\Omega}^{\dagger}(F) + \sum_{\substack{\pi\in\Pi_r \\ \pi\neq\pi^r}} g(\pi) \sum_{F\in\pi} {}^{\circ}Q(F)\log{}^{\circ}Q(F).
$$

For large enough r the sums over the $\pi \neq \pi^r$ become negligible. Since $g(\pi^r)$ is bounded there has to exist some $R \in \mathbb{N}$ with $R \geq \max\{K, n\}$ such that for all $r \geq R$ it holds that

$$g(\pi^r)\Big(H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q)\Big) > \sum_{\substack{\pi \in \Pi_r \\ \pi \neq \pi^r}} -g(\pi) \sum_{F \in \pi} {}^\circ P_\Omega^\dagger(F) \log {}^\circ P_\Omega^\dagger(F) + {}^\circ Q(F) \log {}^\circ Q(F).$$

Hence, for all large enough r it holds that $H_g^r(P_\Omega^\dagger) - H_g^r(Q) > 0$.

Thus, $\operatorname{maxent} \mathbb{E}_\mathcal{L} = \{P_\Omega^\dagger\}$.

For the second part of the proof we show that for all $r \in \mathbb{N}$ and all $F \subseteq \Omega_r$ it holds that

$$\lim_{n \longrightarrow \infty} {}^\circ P_n^\dagger(F) - {}^\circ P_\Omega^\dagger(F) = 0. \tag{8}$$

Observe that for all $n \in \mathbb{N}$

$$|H_\Omega^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger)| = |H_\Omega^n(P_n^\dagger) - \frac{1}{g(\pi^n)} H_g^n(P_n^\dagger) + \frac{1}{g(\pi^n)} H_g^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger)|$$

$$\leq \sum_{\substack{\pi \in \Pi_n \\ \pi \neq \pi^n}} -\frac{g(\pi)}{g(\pi^n)} \sum_{F \in \pi} {}^\circ P_n^\dagger(F) \log {}^\circ P_n^\dagger(F) + |\frac{1}{g(\pi^n)} H_g^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger)|.$$

The first sum tends to zero as n goes to infinity by our assumptions on g.

For the second sum observe that for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq \max\{N, K\}$ and all $P \in [\mathbb{E}_\mathcal{L}]$ it holds that $|\frac{1}{g(\pi^n)} H_g^n(P) - H_\Omega^n(P)| < \epsilon$. Hence, $\epsilon > |\sup_{P \in \mathbb{E}_\mathcal{L}} \frac{1}{g(\pi^n)} H_g^n(P) - \sup_{P \in \mathbb{E}_\mathcal{L}} H_\Omega^n(P)| = |\frac{1}{g(\pi^n)} H_g^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger)|$. So,

$$\lim_{n \longrightarrow \infty} H_\Omega^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger) = 0.$$

For all $n \geq K$, $P_n^\dagger$ and $P_\Omega^\dagger$ equivocate under $\mathcal{L}_K$ up to $\mathcal{L}_n$ (Proposition 9). Hence, it holds that $H_\Omega^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger) = H_\Omega^K(P_n^\dagger) - H_\Omega^K(P_\Omega^\dagger)$ (Corollary 3). So,

$$\lim_{n \longrightarrow \infty} H_\Omega^K(P_n^\dagger) - H_\Omega^K(P_\Omega^\dagger) = \lim_{n \longrightarrow \infty} H_\Omega^n(P_n^\dagger) - H_\Omega^n(P_\Omega^\dagger) = 0.$$

H_Ω^K is a strictly concave and continuous function on $\mathbb{P}_K$. Hence, $\lim_{n \longrightarrow \infty} P_n^\dagger(\omega) = P_\Omega^\dagger(\omega)$ for all $\omega \in \Omega_K$. So, $\lim_{n \longrightarrow \infty} (P_n^\dagger)_{\downarrow K} = (P_\Omega^\dagger)_{\downarrow K}$.

For an arbitrary $n \geq K$ and an $F \subseteq \Omega_n$ we find using that $P_\Omega^\dagger$ equivocates beyond $\mathcal{L}_K$

$$\lim_{k \longrightarrow \infty} {}^\circ P_k^\dagger(F) = \lim_{k \longrightarrow \infty} \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} P_k^\dagger(\nu) = \lim_{k \longrightarrow \infty} \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} \frac{|\Omega_K|}{|\Omega_n|} \cdot P_k^\dagger(\omega_\nu)$$

$$= \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} \frac{|\Omega_K|}{|\Omega_n|} \cdot \lim_{k \longrightarrow \infty} P_k^\dagger(\omega_\nu)$$

$$= \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} \frac{|\Omega_K|}{|\Omega_n|} \cdot P_\Omega^\dagger(\omega_\nu)$$

$$= \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} P_\Omega^\dagger(\nu)$$

$$= {}^\circ P_\Omega^\dagger(F).$$

The result for $F \subseteq \Omega_r$ with $r < K$ follows similarly. $\quad\square$

4.3. Loss and Expected Loss

We shall now analyse the notion of the loss incurred by an agent with belief function $B \in \mathbb{B}_\mathcal{L}$. In Section Section 5 we shall be interested how degrees of beliefs in quantified sentences affect losses. The following definition, axioms L1–4, Theorem 4 and Proposition 12 apply within our current, quantifier-free framework, *i.e.,* $\mathcal{L} = \mathcal{L}^{\nexists}$ but they also apply to quantified sentences, *i.e.,* $\mathcal{L} = \mathcal{L}^{\exists}$.

Definition 18 (Independent Sublanguages). *Let $B \in \mathbb{B}_\mathcal{L}$ be a fixed belief function such that $B(\tau) = 1$ for any tautology τ, and $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ where $\mathcal{L}_1$ and $\mathcal{L}_2$ are disjoint: $\mathcal{L}_1$ and $\mathcal{L}_2$ contain the same constants, they do not have a relation symbol in common and the union of the relation symbols in $\mathcal{L}_1$ and $\mathcal{L}_2$ equals $\{U_1, \ldots, U_s\}$, the set of relation symbols in $\mathcal{L}$. We say that $\mathcal{L}_1$ and $\mathcal{L}_2$ are* independent sublanguages*, written $\mathcal{L}_1 \perp\!\!\!\perp_B \mathcal{L}_2$, if and only if $B(\phi_1 \wedge \phi_2) = B(\phi_1) \cdot B(\phi_2)$ for all $\phi_1 \in S\mathcal{L}_1$ and $\phi_2 \in S\mathcal{L}_2$. Let $B_{|\mathcal{L}_1}(\phi_1) := B(\phi_1)$, $B_{|\mathcal{L}_2}(\phi_2) := B(\phi_2)$.*

By analogy with the line of argument of Section 2, we shall suppose that a default loss function $L : S\mathcal{L} \times \mathbb{B}_\mathcal{L} \to (-\infty, \infty]$ satisfies the following requirements. Here $L(\varphi, B)$ is to be interpreted as the loss specific to φ turning out to be true, when one adopts belief function B:

L1. $L(\varphi, B) = 0$, if $B(\varphi) = 1$.

L2. $L(\varphi, B)$ strictly increases as $B(\varphi)$ decreases from 1 towards 0.

L3. $L(\varphi, B)$ only depends on $B(\varphi)$.

L4. Losses are additive when the language is composed of independent sublanguages: if $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ for $\mathcal{L}_1 \perp\!\!\!\perp_B \mathcal{L}_2$, then $L(\phi_1 \wedge \phi_2, B) = L_1(\phi_1, B_{|\mathcal{L}_1}) + L_2(\phi_2, B_{|\mathcal{L}_2})$, where L_1, L_2 are loss functions defined on $\mathcal{L}_1, \mathcal{L}_2$ respectively.

Theorem 4. *If a loss function L on $S\mathcal{L} \times \mathbb{B}_\mathcal{L}$ satisfies L1–4, then $L(\varphi, B) = -k \log B(\varphi)$, where the constant $k > 0$ does not depend on the language $\mathcal{L}$.*

Proof. The proof is exactly analogous to that of Landes and Williamson [4] (Theorem 4), which gives the result in the case in which $\mathcal{L}$ is a finite propositional language. $\square$

Since multiplication by a constant is equivalent to change of base, we can take log to be the natural logarithm. Since we will be interested in the belief functions that minimise loss, rather than in the absolute value of any particular losses, we can take $k = 1$ without loss of generality. Theorem 4 thus allows us to focus on the logarithmic loss function:

$$L^{\log}(\varphi, B) := -\log B(\varphi).$$

Next we define our notion of expected loss. The expectation is taken with respect to a probability function P, and we consider the expectation taken over each partition of propositions. Each partition is weighted by the given weighting function g. Attention is restricted to inclusive weighting functions, so that each belief is evaluated; if the weighting function were not inclusive then degrees of belief in some propositions would fail to contribute to the expectation.

Definition 19 (*n*-**representation**). *A sentence $\theta \in S\mathcal{L}_n$ n-represents a proposition $F \subseteq \Omega_n$, if and only if $F = \{\omega \in \Omega_n : \omega \models \theta\}$. Let $\mathcal{F} \subseteq \mathcal{P}\Omega_n$ be a set of pairwise distinct propositions. We say that $\Theta \subseteq S\mathcal{L}_n$ is a set of n-representatives of $\mathcal{F}$, if and only if each sentence $\theta \in \Theta$ n-represents a unique proposition in $\mathcal{F}$ and each proposition in $\mathcal{F}$ is n-represented by a unique sentence $\theta \in \Theta$.*

A set ρ of n-representatives of $\mathcal{P}\Omega_n$ will be called an n-representation. *We shall use ρF to denote the sentence in ρ which n-represents F. We denote by ϱ_n the set of all n-representations.*

Note that if belief function B respects logical equivalence, then for all $n \in \mathbb{N}$, all $F \subseteq \Omega_n$ and all l-representations ρ with $l \geq n$ it holds that $B(\rho F) = {}^\circ B(F)$. Otherwise there exist an $n \in \mathbb{N}$ a proposition $F \subseteq \Omega_n$ and n-representations ρ, ρ', such that $B(\rho F) \neq B(\rho' F)$.

Definition 20 (*n*-**score**). *Given a loss function L, an inclusive weighting function $g : \Pi \longrightarrow \mathbb{R}_{\geq 0}$, $n \in \mathbb{N}$, and an n-representation $\rho \in \varrho_n$ we define the* representation-relative *n-score $S_{g,\rho}^{L,n} : \mathbb{P}_\mathcal{L} \times \mathbb{B}_\mathcal{L} \longrightarrow [-\infty, \infty]$ by:*

$$S_{g,\rho}^{L,n}(P, B) := \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} P(\rho F) L(\rho F, B).$$

Define the (representation-independent) *n-score $S_g^{L,n} : \mathbb{P}_\mathcal{L} \times \mathbb{B}_\mathcal{L} \longrightarrow [-\infty, \infty]$ by*

$$S_g^{L,n}(P, B) := \sup_{\rho \in \varrho_n} S_{g,\rho}^{L,n}(P, B).$$

(As a technical convenience, we shall consider loss functions and n-scores to be defined more generally, taking arguments $P, B : S\mathcal{L} \longrightarrow [0, 1]$, although we will primarily be concerned with the case above where P is a probability function and B is a belief function.)

In the light of Theorem 4, we will focus exclusively on the logarithmic loss function in this paper:

$$S_{g,\rho}^n(P, B) := - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} P(\rho F) \log B(\rho F),$$

$$S_g^n(P, B) := \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, B).$$

For $P \in \mathbb{P}_\mathcal{L}$ we have that $P(\rho F) = P(\rho' F)$ for all $\rho, \rho' \in \varrho_n$, since P respects logical equivalence. Hence for $P, Q \in \mathbb{P}_\mathcal{L}$ we have

$$\begin{aligned} S_g^n(P, Q) &= \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, Q) \\ &= - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ Q(F) \\ &= S_g({}^\circ P, {}^\circ Q), \end{aligned}$$

where S_g is the propositional scoring rule introduced in Section 2, in the case $\Omega = \Omega_n$. There are also connections with g-entropy H_g^n, defined in (5), and the propositional notion of entropy H_g, defined in Section 2:

$$S_g^n(P, P) = H_g^n(P) = H_g({}^\circ P).$$

If $g = g_\Omega$, we call the resulting function the *standard logarithmic n-score*:

$$S_\Omega^n(P, B) = \sup_{\rho \in \varrho_n} - \sum_{\omega \in \Omega_n} P(\rho\{\omega\}) \log B(\rho\{\omega\})$$

$$= - \sum_{\omega \in \Omega_n} P(\omega) \log B(\omega),$$

where the latter equality applies if B respects logical equivalence.

The question arises as to how S_g^n, the notion of expected loss defined on a finite sublanguage $\mathcal{L}_n$, relates to loss on $\mathcal{L}$, the language as a whole. One particularly natural suggestion is that B has a better overall loss profile than B' if the latter's n-scores eventually dominate those of B or if the worst-case n-score incurred by B' is eventually greater than that of B:

1. If B has lower worst-case expected loss than B' for all sufficiently large n, then B has a better loss profile than B'.

2. If for all $P \in \mathbb{P}_\mathcal{L}$, B has an expected loss which is less than or equal than that of B', and if for some $P \in [\mathbb{E}_\mathcal{L}]$, B has strictly lower expected loss than B' for sufficiently large n, then B has a better loss profile than B'.

We make this precise as follows:

Definition 21 (Better loss profile). B has a better loss profile than B' *if and only if:*

1. *There exists some $N \in \mathbb{N}$ such that for all $n \geq N$, $\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) < \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B')$, or*

2. *$S_g^n(P, B) \leq S_g^n(P, B') < +\infty$ for all $P \in \mathbb{P}_\mathcal{L}$ and all $n \in \mathbb{N}$, and there exist at least one function $Q \in [\mathbb{E}_\mathcal{L}]$ and some $N_Q \in \mathbb{N}$ such that $S_g^n(Q, B) < S_g^n(Q, B')$ for all $n \geq N_Q$.*

We write $B \prec B'$ to denote that B has better loss profile than B'. We will be interested in those belief functions that have the best loss profile, *i.e., the minimal elements of $\prec$, and define:*

$$\mathrm{minloss}\, \mathbb{B}_\mathcal{L} := \{B \in \mathbb{B}_\mathcal{L} \,:\, \text{there is no } B' \in \mathbb{B}_\mathcal{L} \text{ such that } B' \prec B\}. \tag{9}$$

Proposition 10 (Properties of $\prec$). *The binary relation $\prec$ is asymmetric, partial, irreflexive and transitive.*

Proof. Note that if for all $P \in \mathbb{P}_\mathcal{L}$ and all $n \in \mathbb{N}$ it holds that $S_g^n(P, B) \leq S_g^n(P, B')$, then $\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) \leq \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B')$ follows trivially. Hence, conditions 1 and 2 of Definition 21 are consistent, in the sense that the induced relation $\prec$ is asymmetric.

There exist different $B, B' \in \mathbb{B}_\mathcal{L}$ which are not open-minded on $\mathcal{L}_1$ and thus have infinite loss on $\mathcal{L}_n$ for all $n \geq 1$ (*cf.*, Proposition 13). For example, if $B(\tau') = B'(\tau') = 0$ where τ' is a tautology in $S\mathcal{L}_1$, then B and B' have infinite expected loss for all $n \in \mathbb{N}$ and all $P \in \mathbb{P}_\mathcal{L}$. Thus, $\prec$ is only partial.

That $\prec$ is irreflexive follows directly from the definition.

Now consider $B_1, B_2, B_3 \in \mathbb{B}_\mathcal{L}$ such that $B_1 \prec B_2 \prec B_3$. We will consider cases to prove that $B_1 \prec B_3$.

If there exist $N_{1,2}, N_{2,3}$ such that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_1) < \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_2) \text{ for all } n \geq N_{1,2}$$

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_2) < \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_3) \text{ for all } n \geq N_{2,3},$$

then

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_1) < \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_3) \text{ for all } n \geq \max\{N_{1,2}, N_{2,3}\}.$$

Thus, $B_1 \prec B_3$.

Now assume that there exists a number $N_{1,2}$ such that $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_1) < \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_2)$ for all $n \geq N_{1,2}$ and assume that the pair (B_2, B_3) satisfies the second condition of Definition 21. Then, $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_1) < \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B_3)$ for all $n \geq N_{1,2}$. Thus, $B_1 \prec B_3$.

The same argument shows that if the pair (B_1, B_2) satisfies the second condition of Definition 21 and the pair (B_2, B_3) satisfies the first condition, then $B_1 \prec B_3$.

Finally, suppose that the pairs (B_1, B_2) and (B_2, B_3) both satisfy the second condition of Definition 21. Then for all $P \in \mathbb{P}_{\mathcal{L}}$ and all $n \in \mathbb{N}$ it holds that $S_g^n(P, B_1) \leq S_g^n(P, B_3)$. Furthermore, there has to exist a $Q \in [\mathbb{E}_{\mathcal{L}}]$ and an $N_Q \in \mathbb{N}$ such that for all $n \geq N_Q$ it holds that $S_g^n(Q, B_1) < S_g^n(Q, B_2)$. But then $S_g^n(Q, B_1) < S_g^n(Q, B_3)$ for all $n \geq N_Q$. Thus, $B_1 \prec B_3$. $\square$

Since $\prec$ is irreflexive and transitive it cannot contain a cycle.

One main theme of the rest of this paper will be the search for belief functions with the best loss profile. Since the loss function L we are interested in is $-\log B(\varphi)$, and these values monotonically decrease as $B(\varphi)$ increases from 0 to 1, it follows that, ceteris paribus, the belief functions with better loss profiles assign greater degrees of belief to sentences.

It might appear then that the normalisation (see Definition 1) would directly imply that no $B \in \mathbb{B}_{\mathcal{L}} \backslash \mathbb{P}_{\mathcal{L}}$ could have the best loss profile. Intuitively, this might be thought to hold since the belief functions $B \in \mathbb{B}_{\mathcal{L}} \backslash \mathbb{P}_{\mathcal{L}}$ assign smaller degrees of belief than the probability functions $P \in \mathbb{P}_{\mathcal{L}}$. However, Equation (4) shows that some $B \in \mathbb{B}_{\mathcal{L}} \backslash \mathbb{P}_{\mathcal{L}}$ assign *greater* degrees of belief than a probability function $P \in \mathbb{P}_{\mathcal{L}}$ to certain sentences in the following sense: there exists a set of sentences $\Phi \subset S\mathcal{L}$ such that for all $P \in \mathbb{P}_{\mathcal{L}}$ it holds that $\sum_{\varphi \in \Phi} B(\varphi) > \sum_{\varphi \in \Phi} P(\varphi)$.

While Condition 1 of Definition 21 deals with *worst-case* expected loss, Condition 2 deals with *dominance* of expected loss. Now, dominance is often used on its own to justify the Probability norm; see, e.g., de Finetti [12] (Chapter 3) and more recently by Joyce [13,14]. So, one might think that Condition 2 is strong enough on its own to imply the probability norm. However this is not the case:

Proposition 11. *For $\mathbb{E}_{\mathcal{L}} = \mathbb{P}_{\mathcal{L}}$ there exist a weighting function g and a non-probabilistic belief function $B \in \mathbb{B}_{\mathcal{L}} \backslash \mathbb{P}_{\mathcal{L}}$ such that no probability function $P \in \mathbb{P}_{\mathcal{L}}$ has a loss which dominates that of B in the sense of Condition 2.*

Proof. It suffices to show that there exist a weighting g and a $B \in \mathbb{B}_{\mathcal{L}} \backslash \mathbb{P}_{\mathcal{L}}$ such that for all $Q \in \mathbb{P}_{\mathcal{L}}$ there exist a $P \in \mathbb{P}_{\mathcal{L}}$ and infinitely many $n \in \mathbb{N}$ such that $S_g^n(P, B) < S_g^n(P, Q)$.

Consider a $B \in \mathbb{B}_\mathcal{L} \setminus \mathbb{P}_\mathcal{L}$ from Proposition 4 and consider an arbitrary $Q \in \mathbb{P}_\mathcal{L}$. Then there has to exist an $\nu \in O_4$ such that $Q(\nu) \neq B(\nu)$. Next note that $Q(\neg\nu) \neq B(\neg\nu)$ follows. Then, $-\frac{1}{100}\log\frac{1}{100} - \frac{99}{100}\log\frac{99}{100} < -\frac{1}{100}\log Q(\nu) - \frac{99}{100}\log Q(\nu)$ since the logarithmic scoring rule is strictly proper.

So, for $P \in \mathbb{P}_\mathcal{L}$ with $P(\nu) = \frac{1}{100}$ and $g(\{\nu, \neg\nu\}) > 0$ it holds that

$$g(\{\nu, \neg\nu\})\Big(-P(\nu)\log(B(\nu)) - P(\neg\nu)\log(B(\neg\nu))\Big)$$
$$< \ g(\{\nu, \neg\nu\})\Big(-P(\nu)\log(Q(\nu)) - P(\neg\nu)\log(Q(\neg\nu))\Big).$$

Next let $\nu_1 := \neg Ut_1 t \wedge \neg Ut_2 t$, $\nu_2 := Ut_1 t \wedge \neg Ut_2 t$, $\nu_3 := \neg Ut_1 t \wedge Ut_2 t$, and $\nu_4 := Ut_1 t \wedge Ut_2 t$. For $n \geq 4$ let $F_n^i \subset \Omega_n$ be the unique proposition which is equivalent to ν_i, $F_n^i = \{\omega \in \Omega_n \ : \ \omega \models \nu_i\}$.

Now define g_n for $n \geq 4$ as follows:

$$g_n(\{F_n^i, \bar{F}_n^i\}) := 1, \ \text{if} \ n \equiv i \ \mod 4$$
$$g_n(\pi) := 0, \ \text{else.}$$

So, for this B and this g we have found that for all $Q \in \mathbb{P}_\mathcal{L}$ there exist a $P \in \mathbb{P}_\mathcal{L}$ and infinitely many $n \in \mathbb{N}$ (every fourth n) such that

$$S_g^n(P, B) = -\frac{1}{100}\log\frac{1}{100} - \frac{99}{100}\log\frac{99}{100}$$
$$< -\frac{1}{100}\log Q(\nu) - \frac{99}{100}\log Q(\nu)$$
$$= S_g^n(P, Q).$$

$\square$

In general, determining the functions comprising $\mathrm{minloss}\,\mathbb{B}_\mathcal{L}$ is a challenging problem, which we shall tackle in due course. However, there is one general property we can prove directly: assigning zero degree of belief to an epistemically possible sentence is irrational, in the sense that it exposes one to avoidable losses. To see this, first note that:

Proposition 12. *For any $\mathbb{E}_\mathcal{L}$, there exists a probability function $P \in \mathbb{E}_\mathcal{L}$ which is open-minded.*

Proof. The set of consistent sentences in $\mathcal{L}$ is countable. The set

$$\Phi := \{\varphi \in S\mathcal{L} \ : \ \text{there exists a } P \in \mathbb{E}_\mathcal{L} \text{ with } P(\psi) > 0\}$$

is a subset of the set of consistent sentences and is thus countable, too. We can hence enumerate Φ by some countable index set, I, say. Note that $|I| \geq 2$ since $P(\tau) = 1$ for all $P \in \mathbb{P}_\mathcal{L}$ and all tautologies τ.

For all $\varphi \in \Phi$ choose some $P_\varphi \in \mathbb{E}_\mathcal{L}$ such that $P_\varphi(\varphi) > 0$. Next, for all $i \in I$ pick an $\alpha_i \in (0, 1) \subset \mathbb{R}$ such that $\sum_{i \in I} \alpha_i = 1$. Since $|I| \geq 2$ such α_i exist.

We shall now define an open-minded function $P \in \mathbb{E}_\mathcal{L}$ by putting

$$P = \sum_{i \in I} \alpha_i P_{\varphi_i}.$$

Note that P is in $\mathbb{E}_\mathcal{L}$ since it is a convex combination of probability functions in the convex set $\mathbb{E}_\mathcal{L}$.

We next show that P is indeed open-minded. Let $\varphi \in \Phi$ be at the j-th position in the enumeration I of Φ. We now obtain $P(\varphi) \geq \alpha_j P_\varphi(\varphi) > 0$. So, $P(\varphi) > 0$ for all $\varphi \in \Phi$. $\square$

Proposition 13. $B \in \text{minloss} \, \mathbb{B}_{\mathcal{L}}$ *implies that B is open-minded.*

Proof. If B is not open-minded, then there exists a $k \in \mathbb{N}$ and a $\varphi \in S\mathcal{L}_k$ such that $B(\varphi) = 0$ and there exists a $P \in [\mathbb{E}_{\mathcal{L}}]$ such that $P(\varphi) > 0$. Since $\varphi \in S\mathcal{L}_r$ for all $r \geq k$, it holds for all $r \geq k$ that $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^r(P, B) = +\infty$.

By Proposition 12 there exists an open-minded $Q \in [\mathbb{E}_{\mathcal{L}}]$. Thus, $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^r(P, Q) < \infty$ for all r. □

Note that the above proposition does not imply that $\text{minloss} \, \mathbb{B}_{\mathcal{L}}$ is non-empty.

4.4. Minimax Theorems

In this section we shall relate the belief functions that have best loss profile to the probability functions that have maximal g-entropy.

It turns out that an improvement in loss profile is not necessarily accompanied by an increase in entropy (Appendix A). Nevertheless, we shall see that given appropriate conditions on g, there is a close relationship between the belief function that has the *best* loss profile and the probability function which has *maximum* entropy. On a finite sublanguage, the unique belief function with minimum worst-case expected loss is the probability function with maximum entropy (Section 4.4.1). Moreover, on the language $\mathcal{L}$ as whole, if the evidence set $\mathbb{E}_{\mathcal{L}}$ is finitely generated then the unique belief function with the best lost profile (*i.e.,* the belief function that is minimal with respect to $\prec$) is the probability function in $\mathbb{E}_{\mathcal{L}}$ with maximal entropy (Section 4.4.2). However, this is not necessarily so when $\mathbb{E}_{\mathcal{L}}$ is not finitely generated (Section 6.1).

4.4.1. Minimax on Finite Sublanguages

Lemma 5. *For all $n \in \mathbb{N}$, all $P \in \mathbb{P}_{\mathcal{L}}$ and all $B \in \mathbb{B}_{\mathcal{L}}$ respecting logical equivalence on $\mathcal{L}_n$ it holds that $S_g^n(P, B) = S_{g,\rho}^n(P, B)$ for all $\rho \in \varrho_n$.*

Proof. Simply note that $S_{g,\rho}^n(P, B) = -\sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log B(\rho F)$ does not depend on $\rho \in \varrho_n$. □

Lemma 6. *For all inclusive g, for all $n \in \mathbb{N}$ and each belief function*

$$B^\dagger \in \arg \inf_{B \in \mathbb{B}_{\mathcal{L}}} \sup_{P \in \mathbb{E}_{\mathcal{L}}} \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, B),$$

$B^\dagger$ respects logical equivalence on $\mathcal{L}_n$. Furthermore, for all such $B^\dagger$ there exists a partition $\pi \in \Pi_n$ such that $\sum_{F \in \pi} B^\dagger(\rho F) = 1$ for all $\rho \in \varrho_n$.

Proof. Firstly, $B^\dagger$ cannot assign all $\varphi \in S\mathcal{L}_n$ degree of belief 0, since this would an incur an infinite worst-case expected loss; and as we saw in Proposition 13, there are functions which have finite worst-case expected loss.

Assume for contradiction that a $B^\dagger \in \mathbb{B}_{\mathcal{L}}$ does not respect logical equivalence on $\mathcal{L}_n$. Then define a function $B^{\text{inf}} : S\mathcal{L} \longrightarrow [0, 1]$ which respects logical equivalence on $\mathcal{L}_n$ by

$$B^{\text{inf}}(\varphi) := \begin{cases} \inf_{\substack{\psi \in S\mathcal{L}_n \\ \vDash \varphi \leftrightarrow \psi}} B^\dagger(\psi), & \text{if } \varphi \in S\mathcal{L}_n \\ B(\varphi) & \text{otherwise.} \end{cases}$$

The next step in this proof is to show that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\mathrm{inf}}) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\dagger}).$$

In the second part of the proof we shall see that there is a belief function which has a strictly better worst case expected loss than B^{inf}. This then contradicts the assumption that the belief function $B^{\dagger}$ has best worst case expected loss, *i.e.*, $B^{\dagger} \in \arg\inf_{B \in \mathbb{B}_{\mathcal{L}}} \sup_{P \in \mathbb{E}_{\mathcal{L}}} \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, B)$.

Since $B^{\dagger}$ does not respect logical equivalence on $\mathcal{L}_n$, there are logical equivalent $\varphi, \psi \in S\mathcal{L}_n$ such that $B^{\dagger}(\varphi) \neq B^{\dagger}(\psi)$. Thus, $B^{\mathrm{inf}}(\varphi) < \max\{B^{\dagger}(\varphi), B^{\dagger}(\psi)\}$ and hence $B^{\mathrm{inf}}(\varphi) + B^{\mathrm{inf}}(\neg\varphi) < \max\{B^{\dagger}(\varphi), B^{\dagger}(\psi)\} + B^{\dagger}(\neg\varphi) \leq 1$. The last inequality holds since $B^{\dagger} \in \mathbb{B}_{\mathcal{L}}$. So, $B^{\mathrm{inf}}_{\lfloor n} \notin \mathbb{P}_n$.

Recall that we extended the definition of scoring rules allowing the belief function to be any function defined on $S\mathcal{L}$ taking values in $[0,1]$. We shall be careful not to appeal to results that assume a normalised belief function in this situation.

We now find for $P \in \mathbb{P}_{\mathcal{L}}$

$$
\begin{aligned}
S_g^n(P, B^{\dagger}) &= \sup_{\rho \in \varrho_n} S_{g,\rho_n}^n(P, B^{\dagger}) \\
&= \sup_{\rho \in \varrho_n} - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} P(\rho F) \log B^{\dagger}(\rho F) \\
&= - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^{\circ}P(F) \inf_{\rho \in \varrho_n} \log B^{\dagger}(\rho F) \\
&= - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} P(\rho F) \log B^{\mathrm{inf}}(\rho F) \quad \text{for all } \rho \in \varrho_n \\
&= S_{g,\rho}^n(P, B^{\mathrm{inf}}) \quad \text{for all } \rho \in \varrho_n \\
&= \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, B^{\mathrm{inf}}) \\
&= S_g^n(P, B^{\mathrm{inf}}).
\end{aligned}
$$

Hence $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\dagger}) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\mathrm{inf}})$, as claimed above.

Let us now consider cases to derive a contradiciton.

<u>Case i</u> There exists a $\pi \in \Pi_n$ such that $\sum_{F \in \pi} B^{\mathrm{inf}}(\rho F) = 1$.

Since B^{inf} respects logical equivalence this fact is independent of the particular $\rho \in \varrho_n$. Recall that we use the notation ${}^{\circ}B^{\mathrm{inf}} = {}^{\circ n}B^{\mathrm{inf}}$ to denote the function that B^{inf} induces over propositions in Ω_n, defined by ${}^{\circ}B^{\mathrm{inf}}(F) = B^{\mathrm{inf}}(\bigvee F)$.

With this convention we then note that ${}^{\circ}B^{\mathrm{inf}} \in \mathbb{B} \setminus \mathbb{P}$. Let $\mathbb{E}$ be the set of probability functions on Ω_n which are in the canonical one-to-one correspondence with the probability functions on $\mathbb{E}_n$, *i.e.*, $\mathbb{E} := \{{}^{\circ}P : P \in \mathbb{E}_{\mathcal{L}}\}$. We thus find, using Theorem 2 to obtain the strict inequality, that:

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\dagger}) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\mathrm{inf}})$$

$$= \sup_{P \in \mathbb{E}} S_g(^{\circ}P, {}^{\circ}B^{\mathrm{inf}})$$

$$= \sup_{P \in \mathbb{E}} - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^{\circ}P(F) \log {}^{\circ}B^{\mathrm{inf}}(F)$$

$$> \sup_{P \in \mathbb{E}} - \sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^{\circ}P(F) \log {}^{\circ}P_n^{\dagger}(F)$$

$$= \sup_{P \in \mathbb{E}} S_g(^{\circ}P, {}^{\circ}P_n^{\dagger})$$

$$= \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_n^{\dagger}).$$

<u>Case ii</u> For all $\pi \in \Pi_n$ and all $\rho \in \varrho_n$ it holds that $\sum_{F \in \pi} B^{\mathrm{inf}}(\rho F) < 1$.

Since B^{inf} respects logical equivalence on $\mathcal{L}_n$ we may consider the induced function $^{\circ}B^{\mathrm{inf}}$ defined over propositions of Ω_n. Since Π_n is finite, so is the set $\{\sum_{F \in \pi} {}^{\circ}B^{\mathrm{inf}}(F)\}$. Thus, $\sup_{\pi \in \Pi_n} \sum_{F \in \pi} {}^{\circ}B^{\mathrm{inf}}(F) = 1 - \epsilon$ for some $\epsilon \in (0, 1]$.

Let us now define a function $B' : S\mathcal{L} \longrightarrow [0, 1]$. Denote by $\mu \in (0, 1]$ the unique number such that for all $\pi \in \Pi_n$ and all $\rho \in \varrho_n$ it holds that $\sum_{F \in \pi} \mu + B^{\mathrm{inf}}(\rho F) = \sum_{F \in \pi} \mu + {}^{\circ}B^{\mathrm{inf}}(F) \leq 1$ and for at least one $\pi \in \Pi_n$ and one $\rho \in \varrho_n$ we have $\sum_{F \in \pi} \mu + B^{\mathrm{inf}}(\rho F) = \sum_{F \in \pi} \mu + {}^{\circ}B^{\mathrm{inf}}(F) = 1$.

Put $B'(\varphi) := \mu + B^{\mathrm{inf}}(\varphi) > B^{\mathrm{inf}}(\varphi)$ for all $\varphi \in S\mathcal{L}_n$ and $B'(\varphi) := 0$ otherwise. Observe that $B' \in \mathbb{B}_{\mathcal{L}}$ and that $B'(\neg\tau) \geq \mu > 0$ for the tautologies τ of $\mathcal{L}_n$. But then $^{\circ}B' \in \mathbb{B} \setminus \mathbb{P}$. Then for all $\pi \in \Pi_n$ and all $P \in [\mathbb{E}_n]$ we have $-\sum_{F \in \pi} P(\rho F) \log B'(\rho F) < -\sum_{F \in \pi} P(\rho F) \log B^{\mathrm{inf}}(\rho F)$. We now apply Theorem 2 to find the strict inequality below

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\dagger}) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\mathrm{inf}})$$

$$\geq \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B')$$

$$= \sup_{P \in \mathbb{E}} S_g(^{\circ}P, {}^{\circ}B')$$

$$> S_g(^{\circ}P, {}^{\circ}P_n^{\dagger})$$

$$= \sup_{P \in \mathbb{E}} S_g(^{\circ}P, {}^{\circ}P_n^{\dagger})$$

$$= \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_n^{\dagger}).$$

So, in Case i and in Case ii we have found that $P_n^{\dagger}$ has strictly better worst-case expected loss than $B^{\dagger}$ contradicting $B^{\dagger} \in \arg\inf_{B \in \mathbb{B}_{\mathcal{L}}} \sup_{P \in \mathbb{E}_{\mathcal{L}}} \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, B)$.

Finally, we need to show that for all such belief functions $B^{\dagger}$ there exists a $\pi \in \Pi_n$ such that $\sum_{F \in \pi} {}^{\circ}B^{\dagger}(F) = 1$. Suppose for contradiction that is not the case. Note that $B^{\dagger}$ respects logical equivalence on $\mathcal{L}_n$. Hence, we can define a belief function $B' \in \mathbb{B}_{\mathcal{L}}$ by adding a strictly positive number μ as in Case ii. B' has a worst-case expected loss that is less or equal to the worst-case expected loss of $B^{\dagger}$. Again, we find that $^{\circ}B' \in \mathbb{B} \setminus \mathbb{P}$ and hence B' does not have minimal worst-case expected loss. Clearly then, $B^{\dagger}$ cannot have minimal worst-case expected loss. Contradiction. $\qquad\square$

Theorem 5 (**Finite sublanguage minimax**). *For all inclusive g, all $n \in \mathbb{N}$, all $C \in \arg\inf_{B\in\mathbb{B}_\mathcal{L}} \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,B)$ and all $Q \in \arg\sup_{P\in\mathbb{E}_\mathcal{L}} H_g^n(P)$ it holds that*

$$C_{\downharpoonright n} = Q_{\downharpoonright n} = P_n^\dagger.$$

Proof. From Lemma 6 we know that for every $C \in \arg\inf_{B\in\mathbb{B}_\mathcal{L}} \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,B)$ it holds that $C_{\downharpoonright n}$ respects logical equivalence on $\mathcal{L}_n$ and that $°C := °^nC \in \mathbb{B}$ (since C is normalised). Every probability function in $P \in \mathbb{P}_\mathcal{L}$ respects logical equivalence (Proposition 3).

Thus, $S_g^n(P,C)$ and $S_g^n(P,P)$ collapse to $S_g(°P,°C)$, respectively $S_g(°P,°P)$, the logarithmic scoring rule for propositions (1).

However, for the propositional case we know from Theorem 2 that the unique g-entropy maximiser on $\mathbb{P}$ is the unique worst-case expected loss minimiser on $\mathbb{B}$, $P_g^\dagger = °P_n^\dagger$. That is, $\arg\inf_{B\in\mathbb{B}} \sup_{P\in\mathbb{E}} S_g(P,B) = \arg\sup_{P\in\mathbb{E}} H_g(P) = \{P_g^\dagger\}$.

Thus, for all $F \subseteq \Omega_n$ it holds that $C(\rho F) = P_g^\dagger(F)$ for all $\rho \in \varrho_n$. Hence, $C_{\downharpoonright n} = Q_{\downharpoonright n} = P_n^\dagger$. $\square$

4.4.2. Minimax for Inductive Logic

We shall now consider the language $\mathcal{L}$ as a whole. We shall assume in this section that $\mathbb{E}_\mathcal{L}$ is finitely generated by constraints on $\mathcal{L}_K$. As noted in Section 3.3, this is the scenario that is of key relevance to inductive logic. Our goal is to justify the norms of objective Bayesianism by showing that the belief functions with the best loss profile are the probability functions in $\mathbb{E}_\mathcal{L}$ with maximum entropy.

First we shall see that this is the case if g is language invariant:

Proposition 14 (**Language invariance minimax**). *If g is inclusive and language invariant and if $\mathbb{E}_\mathcal{L}$ is finitely generated, then*

$$\text{minloss}\,\mathbb{B}_\mathcal{L} = \text{maxent}\,\mathbb{E}_\mathcal{L} = \mathbb{P}^\dagger = \{P^\dagger\}.$$

Proof. Note that we have $\mathbb{P}^\dagger = \{P^\dagger\}$ from Proposition 8, in particular $P_n^\dagger = P^\dagger_{\downharpoonright n}$ for all $n \geq K$.

Since g is inclusive, H_g^n is strictly concave on $\mathbb{P}_n$ (Lemma 1). Hence, $P_n^\dagger$ is uniquely determined. By language invariance we obtain $P^\dagger \in \arg\sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,P)$ for all $n \geq K$. Thus, $P^\dagger \in \text{maxent}\,\mathbb{E}_\mathcal{L}$.

For $Q \in [\mathbb{E}_\mathcal{L}] \setminus \{P^\dagger\}$ there has to exist some $N \in \mathbb{N}$ such than $Q_{\downharpoonright n} \neq P^\dagger_{\downharpoonright n}$ for all $n \geq N$. Since H_g^n is a strictly concave function on $\mathbb{P}_n$ and since $P^\dagger$ maximises H_g^n for all $n \geq K$ it follows that $H_g^n(P^\dagger) > H_g^n(Q)$ for all $n \geq \max\{K,N\}$. Thus, $Q \notin \text{maxent}\,\mathbb{E}_\mathcal{L}$.

From Theorem 5 we have that $P_n^\dagger \in \arg\inf_{B\in\mathbb{B}_\mathcal{L}} \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,B)$ for all $n \geq K$. Since $\mathbb{E}_\mathcal{L}$ is finitely generated and g is language invariant we have that $P^\dagger \in \arg\inf_{B\in\mathbb{B}_\mathcal{L}} \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,B)$ for all $n \geq K$. Thus, $P^\dagger \in \text{minloss}\,\mathbb{B}_\mathcal{L}$.

For every $C \in \mathbb{B}_\mathcal{L} \setminus \{P^\dagger\}$ there has to exist an $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $C_{\downharpoonright n} \neq P^\dagger_{\downharpoonright n}$. For all $n \geq \max\{K,N\}$ we now apply Theorem 5 to obtain $\sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,C) > \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P,P^\dagger)$. Hence, $C \notin \text{minloss}\,\mathbb{B}_\mathcal{L}$. $\square$

This result is not entirely satisfactory, because we cannot say anything yet about whether such weighting functions exist. Indeed, it was conjectured in Landes and Williamson [4] (p. 3564) that no inclusive, symmetric and refined weighting function g is language invariant. This conjecture remains open.

Our next result says that, for the standard weighting g_Ω, the *probability function* with the best loss profile is the standard entropy maximiser:

Proposition 15 (**Standard entropy minimax**). *If $\mathbb{E}_\mathcal{L}$ is finitely generated and $g = g_\Omega$, then*

$$\text{minloss}\,\mathbb{P}_\mathcal{L} = \text{maxent}\,\mathbb{E}_\mathcal{L} = \{P_\Omega^\dagger\}.$$

Proof. $\{P_\Omega^\dagger\}$ = maxent $\mathbb{E}_\mathcal{L}$ follows directly, since g_Ω is language-invariant and state-inclusive, Proposition 8.

It is well-known that

$$\arg\inf_{Q\in\mathbb{P}_n}\sup_{P\in\mathbb{E}_n} S_{g_\Omega}(P,Q) = \arg\sup_{P\in\mathbb{E}_n} S_{g_\Omega}(P,P) = \{P_{g_\Omega}^\dagger\},$$

see for instance [15]. Hence,

$$\text{minloss}\,\mathbb{P}_\mathcal{L} = \text{maxent}\,\mathbb{E}_\mathcal{L} = \{P_\Omega^\dagger\}.$$

□

Because it only identifies probability functions with the best loss profile, rather than normalised belief functions with the best loss profile, Proposition 15 provides a justification for only two norms of objective Bayesianism, the Calibration Norm and the Equivocation Norm, under the supposition that $g = g_\Omega$. This is a useful result if there is some independent reason—such as the Dutch book argument—for taking belief functions to be probability functions. But our goal in this paper is to investigate the extent to which the notion of loss profile developed above can be used to justify all three norms at once.

We know that there are weighting functions that are *regular*, *i.e.,* which are atomic, inclusive, symmetric and strongly refined. The plan of the rest this section is to prove the following analogous minimax theorem for regular weighting functions. This says that, for any regular weighting function, the *belief function* with the best loss profile is the probability function in $\mathbb{E}_\mathcal{L}$ which has maximal standard entropy. This theorem thus justifies all three norms at once.

Theorem 6 (**Regularity minimax**). *If g is regular and $\mathbb{E}_\mathcal{L}$ is finitely generated, then*

$$\text{minloss}\,\mathbb{B}_\mathcal{L} = \text{maxent}\,\mathbb{E}_\mathcal{L} = \mathbb{P}^\dagger = \{P_\Omega^\dagger\}.$$

In order to prove this theorem we give a number of lemmata. We shall state these lemmata under more minimal conditions on g. The reader not interested in the details might always replace the stated conditions on g by: "g is regular".

To begin with, we shall consider only belief functions B which respect logical equivalence. (Later we shall relax this restriction.) Hence, $S_{g,\rho}^n(P,B)$ does not depend on ρ and we can ignore the particular representation ρ. This will allow us to focus on propositions.

Lemma 7. *If $n \geq K$, $Q \in \mathbb{P}_\mathcal{L}$ and if $\sup_{P\in\mathbb{E}_\mathcal{L}} S_\Omega^n(P,Q)$ is finite, then it holds that*

$$\sup_{P\in\mathbb{E}_\mathcal{L}} S_\Omega^{n+1}(P,Q) \geq \sup_{P\in\mathbb{E}_\mathcal{L}} S_\Omega^n(P,Q) + \log\frac{|\Omega_{n+1}|}{|\Omega_n|}.$$

Proof. Let $P' \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, Q)$. Then define P'' on Ω_{n+1} by $P''(\nu) := P'(\omega_\nu)\frac{|\Omega_n|}{|\Omega_{n+1}|}$ for all $\nu \in \Omega_{n+1}$ and $\omega_\nu \in \Omega_n$ with $\nu \vDash \omega_\nu$. Now extend P'' arbitrarily to a function in $[\mathbb{E}_{\mathcal{L}}]$. Note that $P''_{|n+1} \in [\mathbb{E}_{n+1}]$ since $\mathbb{E}_{\mathcal{L}}$ is finitely generated and $n \geq K$.

Since $-\log(x)$ is a strictly convex function on $(0, 1]$ and since $Q(\omega) = \sum_{\substack{\nu \in \Omega_{n+1} \\ \nu \vDash \omega}} Q(\nu)$ for all $\omega \in \Omega_n$ it holds for all fixed $\omega \in \Omega_n$ that $\sum_{\substack{\nu \in \Omega_{n+1} \\ \nu \vDash \omega}} -\log Q(\nu) \geq -\frac{|\Omega_{n+1}|}{|\Omega_n|}\log(\frac{|\Omega_n|}{|\Omega_{n+1}|}Q(\omega))$. We now find

$$
\begin{aligned}
\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^{n+1}(P, Q) &\geq S_\Omega^{n+1}(P'', Q) \\
&= -\sum_{\nu \in \Omega_{n+1}} P''(\nu) \log Q(\nu) \\
&= -\sum_{\nu \in \Omega_{n+1}} P'(\omega_\nu)\frac{|\Omega_n|}{|\Omega_{n+1}|} \log Q(\nu) \\
&\geq -\sum_{\omega \in \Omega_n} P'(\omega) \cdot \log \frac{|\Omega_n| \cdot Q(\omega)}{|\Omega_{n+1}|} \\
&= -\log \frac{|\Omega_n|}{|\Omega_{n+1}|} - \sum_{\omega \in \Omega_n} P'(\omega) \cdot \log Q(\omega) \\
&= \log \frac{|\Omega_{n+1}|}{|\Omega_n|} + \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, Q).
\end{aligned}
$$

$\square$

Definition 22 (γ-**weighting**). *To simplify notation we define for $n \in \mathbb{N}$ and $F \subseteq \Omega_n$*

$$
\gamma_n(F) := \sum_{\substack{\pi \in \Pi_n \\ F \in \pi}} g_n(\pi).
$$

If g is symmetric, then $\gamma_n(F)$ only depends on $|F| := |\{\omega \in \Omega_n : \omega \in F\}|$ and we write $\gamma_n(|F|)$.

In particular, since the belief function B is assumed to respect logical equivalence, we can write

$$
\begin{aligned}
S_g^n(P, B) &= \sup_{\rho \in \varrho_n} \sum_{F \subseteq \Omega_n} -\gamma_n(F)P(\rho F) \log B(\rho F) \\
&= \sum_{F \subseteq \Omega_n} -\gamma_n(F)^\circ P(F) \log {}^\circ B(F).
\end{aligned}
$$

Furthermore, we can easily characterise the set of inclusive g. g is inclusive, if and only if for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$ $\gamma_n(F) > 0$.

Lemma 8. *Let g be inclusive and such that there exist $0 < a \leq b < +\infty$ such that $g(\pi^n) \in [a, b]$ for all $n \in \mathbb{N}$ and such that*

$$
\lim_{n \to \infty} \log |\Omega_n| \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) = 0.
$$

Then

$$
Rest_n := \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_\Omega^\dagger) - g(\pi^n) S_\Omega^n(P_\Omega^\dagger, P_\Omega^\dagger) \longrightarrow 0 \text{ as } n \longrightarrow \infty.
$$

Proof. Let us thus first note that

$$S_g^n(P, P_\Omega^\dagger) - g(\pi^n) S_\Omega^n(P, P_\Omega^\dagger) = \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P_\Omega^\dagger(F). \tag{10}$$

Recall that $P_\Omega^\dagger$ is open-minded (Proposition 5). Thus, $P \in [\mathbb{E}_\mathcal{L}]$, $F \subseteq \Omega_n$ and ${}^\circ P(F) > 0$ imply ${}^\circ P_\Omega^\dagger(F) > 0$. Let

$$m := \min\{P_\Omega^\dagger(\omega) \ : \ \omega \in \Omega_K \ \& \ P_\Omega^\dagger(\omega) > 0\} \in (0, 1].$$

Then, for $F \subseteq \Omega_n$ such that ${}^\circ P_\Omega^\dagger(F) > 0$ it holds that

$$\begin{aligned}
{}^\circ P_\Omega^\dagger(F) &\geq \min\{P_\Omega^\dagger(\nu) \ : \ \nu \in \Omega_n \ \& \ P_\Omega^\dagger(\nu) > 0\} \\
&= m \cdot \frac{|\Omega_K|}{|\Omega_n|} \geq \frac{m}{|\Omega_n|},
\end{aligned}$$

since $P_\Omega^\dagger$ equivocates beyond $\mathcal{L}_K$.

Hence, $P \in [\mathbb{E}_\mathcal{L}]$, $F \subseteq \Omega_n$ and ${}^\circ P(F) > 0$ imply that ${}^\circ P_\Omega^\dagger(F) \geq \frac{m}{|\Omega_n|}$. Since $\sum_{F \in \pi} {}^\circ P(F) = 1$ we now find

$$\begin{aligned}
0 &\leq \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) - g(\pi^n) S_\Omega^n(P_\Omega^\dagger, P_\Omega^\dagger) \\
&\leq \sup_{P \in \mathbb{E}_\mathcal{L}} g(\pi^n) S_\Omega^n(P, P_\Omega^\dagger) \\
&\quad + \sup_{P \in \mathbb{E}_\mathcal{L}} \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P_\Omega^\dagger(F) - \sup_{P \in \mathbb{E}_\mathcal{L}} g(\pi^n) S_\Omega^n(P, P_\Omega^\dagger) \\
&= \sup_{P \in \mathbb{E}_\mathcal{L}} \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P_\Omega^\dagger(F) \\
&\leq \sup_{P \in \mathbb{E}_\mathcal{L}} \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log \frac{m}{|\Omega_n|} \\
&= \sup_{P \in \mathbb{E}_\mathcal{L}} \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \log \frac{m}{|\Omega_n|} \sum_{F \in \pi} {}^\circ P(F) \\
&= \log \frac{m}{|\Omega_n|} \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \\
&= \Big(\log(|\Omega_n|) - \log(m) \Big) \cdot \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi)
\end{aligned}$$

To complete the proof, it suffices to note that this sums is eventually positive and converges in $n \in \mathbb{N}$ to zero by our assumption on g and the fact that m is constant. $\square$

Proposition 16. *Let g be inclusive and such that there exist $0 < a \leq b < +\infty$ such that $g(\pi^n) \in [a, b]$ for all $n \in \mathbb{N}$ and such that*

$$\lim_{n \longrightarrow \infty} \log |\Omega_n| \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) = 0.$$

Then for all $B \in \mathbb{B}_\mathcal{L} \setminus \{P_\Omega^\dagger\}$ that respect logical equivalence, $P_\Omega^\dagger \prec B$.

Proof. We shall proceed by considering cases.

<u>Case 1</u> $B \in \mathbb{P}_{\mathcal{L}} \setminus \{P_{\Omega}^{\dagger}\}$.

There exists an $N \geq K$ such that for all $n \geq N$ it holds that $B_{\lfloor n} \neq (P_{\Omega}^{\dagger})_{\lfloor n}$. It is well-known that for all $P \in \mathbb{P}$

$$\arg \inf_{Q \in \mathbb{P}} -\sum_{\omega \in \Omega} P(\omega) \log Q(\omega) = \{P\}. \tag{11}$$

That is, the usual logarithmic scoring rule, when applied to probability functions $P \in \mathbb{P}$ and $Q \in \mathbb{P}$, is *strictly proper*. Savage [16] showed that this scoring rule is not only strictly proper but also unique under the further assumption of locality, which is requirement L3 in our framework. Thus, $S_{\Omega}^{n}(P_{\Omega}^{\dagger}, B) - S_{\Omega}^{n}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) > 0$.

We then find by the first part of Corollary 3 and Lemma 7 for all $n \geq N$ that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{g}^{n}(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{g}^{n}(P, P_{\Omega}^{\dagger})$$

$$= \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{g}^{n}(P, B) - g(\pi^{n}) S_{\Omega}^{n}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) - Rest_{n}$$

$$\geq g(\pi^{n}) \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^{n}(P, B) - g(\pi^{n}) S_{\Omega}^{n}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) - Rest_{n}$$

$$= g(\pi^{n}) \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^{n}(P, B) - g(\pi^{n})(S_{\Omega}^{N}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) + \log \frac{|\Omega_{n}|}{|\Omega_{N}|}) - Rest_{n}$$

$$\geq g(\pi^{n})\Big(\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^{N}(P, B) + \log \frac{|\Omega_{n}|}{|\Omega_{N}|} \Big)$$

$$\quad - g(\pi^{n})(S_{\Omega}^{N}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) + \log \frac{|\Omega_{n}|}{|\Omega_{N}|}) - Rest_{n}$$

$$\geq g(\pi^{n})\Big(S_{\Omega}^{N}(P_{\Omega}^{\dagger}, B) - S_{\Omega}^{N}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) \Big) - Rest_{n}$$

Recall from Lemma 8 that $Rest_{n}$ converges to zero. Furthermore, the sequence $(g(\pi^{n}))_{n \in \mathbb{N}}$ is bounded in $[a, b]$ with $a > 0$. Thus, for all large enough $n \in \mathbb{N}$ it holds that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{g}^{n}(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{g}^{n}(P, P_{\Omega}^{\dagger}) \geq g(\pi^{n})\Big(S_{\Omega}^{N}(P_{\Omega}^{\dagger}, B) - S_{\Omega}^{N}(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) \Big) - Rest_{n}$$

$$> 0.$$

<u>Case 2</u> $B \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$.

<u>Case 2A</u> There exists a $P_{B} \in \mathbb{P}_{\mathcal{L}}$ such that for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_{n}$ it holds that ${}^{\circ}B(F) \leq {}^{\circ}P_{B}(F)$, *i.e.*, P_{B} dominates B.

<u>Case 2Ai</u> $P_{B} = P_{\Omega}^{\dagger}$ and no other $P \in \mathbb{P}_{\mathcal{L}}$ is such that ${}^{\circ}B(F) \leq {}^{\circ}P(F)$ for all n and all $F \subseteq \Omega_{n}$.

Then for all $P \in \mathbb{P}_{\mathcal{L}}$ and all propositions F it holds that

$$\gamma_{n}(F) {}^{\circ}P(F)(-\log {}^{\circ}B(F) + \log {}^{\circ}P_{\Omega}^{\dagger}(F)) \geq 0.$$

Thus, for all $P \in \mathbb{P}_{\mathcal{L}}$ and $n \in \mathbb{N}$ it holds that $S_{g}^{n}(P, B) \geq S_{g}^{n}(P, P_{\Omega}^{\dagger})$.

Since $B \neq P_{\Omega}^{\dagger}$ there exists some $N \in \mathbb{N}$ and a $\emptyset \subset F \subseteq \Omega_{N}$ such that ${}^{\circ}B(F) < {}^{\circ}P_{\Omega}^{\dagger}(F)$. For $n > N$ let $\emptyset \subset F_{n} \subseteq \Omega_{n}$ be such that $F_{n} = \{\omega \in \Omega_{n} : \omega \in F\}$. Hence, for all $n > N$ it holds that

$-\log {}^{\circ}B(F_n) + \log {}^{\circ}P_{\Omega}^{\dagger}(F_n) > 0$. Thus, ${}^{\circ}P_{\Omega}^{\dagger}(F_n)\gamma_n(F_n)(-\log {}^{\circ}B(F_n) + \log {}^{\circ}P_{\Omega}^{\dagger}(F_n)) > 0$. Since g is inclusive ($\gamma_n(F) > 0$ for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$) it holds that $S_g^n(P_{\Omega}^{\dagger}, B) > S_g^n(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger})$ for all $n \geq N$.

Applying the second condition of Definition 21 yields $P_{\Omega}^{\dagger} \prec B$.

<u>Case 2Aii</u> There exists a $P_B \in \mathbb{P}_{\mathcal{L}}$ dominating B such that $P_B \neq P_{\Omega}^{\dagger}$.

Then for all $n \geq K$ and all $P \in \mathbb{E}_{\mathcal{L}}$ it holds that $S_g^n(P, B) - S_g^n(P, P_B) \geq 0$. For all large enough $n \in \mathbb{N}$ it holds by Case 1 that $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger}) > 0$. Thus, we find for all large enough n

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger}) \geq \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger})$$
$$> 0.$$

<u>Cas 2B</u> There does not exist a $P_B \in \mathbb{P}_{\mathcal{L}}$ such that for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$ it holds that ${}^{\circ}B(F) \leq {}^{\circ}P_B(F)$.

For example, the belief functions constructed in Proposition 4 are of this form, *i.e.*, not dominated by a probability function.

Let us assume for contradiction that there exists an infinite set $J := \{j_1, j_2, \dots\} \subseteq \mathbb{N}$ such that $\lim_{i \to \infty} \sum_{\omega \in \Omega_{j_i}} B(\omega) = 1$. Now define a function Q on $S\mathcal{L}$ by requiring that Q respects logical equivalence and that

$$ {}^{\circ}Q(F) := \lim_{i \to \infty} \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \in F}} B(\omega).$$

Next we show $Q \in \mathbb{P}_{\mathcal{L}}$ and ${}^{\circ}B(F) \leq {}^{\circ}Q(F)$ for all F which will allow us to derive the required contradiction.

First note that for all $n \in \mathbb{N}$ it holds that

$$\sum_{\nu \in \Omega_n} Q(\nu) = \lim_{i \to \infty} \sum_{\nu \in \Omega_n} \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \vDash \nu}} B(\omega)$$
$$= \lim_{i \to \infty} \sum_{\omega \in \Omega_{j_i}} B(\omega)$$
$$= 1.$$

Furthermore, we have for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$

$$ {}^{\circ}Q(F) = \lim_{i \to \infty} \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \in F}} B(\omega)$$
$$= \lim_{i \to \infty} \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \vDash \nu}} B(\omega)$$
$$= \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} \lim_{i \to \infty} \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \vDash \nu}} B(\omega)$$
$$= \sum_{\substack{\nu \in \Omega_n \\ \nu \in F}} Q(\nu).$$

So, $Q \in \mathbb{P}_{\mathcal{L}}$.

Now assume that there exists a proposition $F \subseteq \Omega_n$ such that ${}^{\circ}B(F) > {}^{\circ}Q(F)$. Since $Q \in \mathbb{P}_{\mathcal{L}}$ it holds that ${}^{\circ}Q(F) + {}^{\circ}Q(\bar{F}) = 1$. Note that

$$\{\omega \in \Omega_{j_i} : \omega \in F\} \cup \bigcup_{\substack{\omega \in \Omega_{j_i} \\ \omega \in \bar{F}}} \{\omega\}$$

is a partition in Π_{j_i}. Since we assumed that B respects logical equivalence it holds that ${}^{\circ}B(F) = B(\bigvee_{\omega \in \Omega_{j_i} : \omega \in F} \omega)$. Thus,

$$ {}^{\circ}B(F) + \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \in \bar{F}}} B(\omega) \leq 1 $$

has to hold for all large i. We now obtain the required contradiction as follows:

$$ 1 \geq \lim_{i \longrightarrow \infty} \left({}^{\circ}B(F) + \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \in \bar{F}}} B(\omega) \right) $$

$$ = {}^{\circ}B(F) + {}^{\circ}Q(\bar{F}) $$

$$ > {}^{\circ}Q(F) + {}^{\circ}Q(\bar{F}) $$

$$ = 1. $$

Thus, there has to exist an $\alpha > 0$ and an $N \in \mathbb{N}$ with $N \geq K$ such that for all $n \geq N$ it holds that $\sum_{\omega \in \Omega_n} B(\omega) \leq 1 - \alpha$. We have for $n \geq N$ that

$$ \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger}) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B) - g(\pi^n) S_{\Omega}^n(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) - Rest_n $$

$$ \geq g(\pi^n) \left(\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, B) - S_{\Omega}^n(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) \right) - Rest_n $$

$$ \geq g(\pi^n) \left(S_{\Omega}^n(P_{\Omega}^{\dagger}, B) - S_{\Omega}^n(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) \right) - Rest_n. $$

To complete the proof we will now show that there exists some $\beta > 0$, which depends on $\mathbb{E}_{\mathcal{L}}$ and g but does *not* depend on the particular $n \geq N$, such that $S_{\Omega}^n(P_{\Omega}^{\dagger}, B) - S_{\Omega}^n(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) > \beta$. Since $g(\pi^n)$ is bounded, we then obtain that $\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger}) > 0$ for all large enough n.

We need to show that for all large enough n,

$$ -\sum_{\omega \in \Omega_n} P_{\Omega}^{\dagger}(\omega) \log f(\omega) - S_{\Omega}^n(P_{\Omega}^{\dagger}, P_{\Omega}^{\dagger}) \geq \beta > 0 $$

for all functions $f : \Omega_n \longrightarrow [0, 1]$ such that $\sum_{\omega \in \Omega_n} f(\omega) \leq 1 - \alpha$.

Suppose $f' \in \arg\min_f \sum_{\omega \in \Omega_n} -P_{\Omega}^{\dagger}(\omega) \log f(\omega)$. If $P_{\Omega}^{\dagger}(\omega) > 0$ and $f'(\omega) = 0$, then $\sum_{\omega \in \Omega_n} -P_{\Omega}^{\dagger}(\omega) \log f'(\omega) = \infty$. Hence, the minimum cannot obtain for such an f'. On the other hand, if $f'(\omega) > 0$ and $P_{\Omega}^{\dagger}(\omega) = 0$, then there has to exist a $\mu \in \Omega_n \setminus \{\omega\}$ such that $P_{\Omega}^{\dagger}(\mu) > 0$. Then define a function f'' such that $f''(\omega) := 0$, $f''(\mu) := f'(\mu) + f'(\omega) > f'(\mu)$ and $f''(\lambda) := f'(\lambda)$ for all $\lambda \in \Omega_n \setminus \{\omega, \mu\}$. Then $\sum_{\nu \in \Omega_n} -P_{\Omega}^{\dagger}(\nu) \log f'(\nu) > \sum_{\nu \in \Omega_n} -P_{\Omega}^{\dagger}(\nu) \log f''(\nu)$. Again, the minimum cannot obtain for such an f'.

We may thus assume in the following that any f' minimising the above sum satisfies: $P_\Omega^\dagger(\omega) > 0$, if and only if $f'(\omega) > 0$. In particular, the function $f'(\omega) = 0$ for all $\omega \in \Omega_n$ cannot be optimal.

Let $a_f := \sum_{\omega \in \Omega_n} f(\omega) \in (0, 1 - \alpha]$. Then

$$-\sum_{\omega \in \Omega_n} P_\Omega^\dagger(\omega) \log f(\omega) = -\sum_{\omega \in \Omega_n} P_\Omega^\dagger(\omega) \Big(\log \frac{f(\omega)}{a_f} + \log a_f\Big)$$

$$= -\log(a_f) - \sum_{\omega \in \Omega_n} P_\Omega^\dagger(\omega) \log \frac{f(\omega)}{a_f}.$$

By definition, $\sum_{\omega \in \Omega_n} \frac{f(\omega)}{a_f} = 1$. The sum in the above equation is thus standard logarithmic scoring rule on $\mathbb{B}_n$, $S_\Omega^n(P, \frac{f}{a_f})$. For fixed $P \in \mathbb{P}_\mathcal{L}$ the minimum under this scoring rule obtains for a function which agrees with P on the states $\omega \in \Omega_n$.

Thus, for fixed a_f the function f minimising $-\sum_{\omega \in \Omega_n} P_\Omega^\dagger(\omega) \log f(\omega)$ is the a_f multiple of $P_\Omega^\dagger$. In order to minimise $-\sum_{\omega \in \Omega_n} P_\Omega^\dagger(\omega) \log f(\omega)$, $-\log a_f$ has to be minimal. This minimum obtains for $a_f = 1 - \alpha$. We hence find the value of the minimum as

$$\inf_{\substack{f:\Omega_n \longrightarrow [0,1] \\ \sum_{\omega \in \Omega_n} f(\nu) \leq 1-\alpha}} -\sum_{\omega \in \Omega_n} P_\Omega^\dagger(\omega) \log f(\omega) = -\log(1-\alpha) - S_\Omega^n(P_\Omega^\dagger, P_\Omega^\dagger).$$

β may thus be chosen as $\beta = -\log(1-\alpha) > 0$. $\square$

We now drop the assumption that belief functions respect logical equivalence.

Proposition 17. *If g is inclusive and such that there exist $0 < a \leq b < +\infty$ such that $g(\pi^n) \in [a, b]$ for all $n \in \mathbb{N}$ and such that*

$$\lim_{n \longrightarrow \infty} \log |\Omega_n| \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) = 0,$$

then

$$\mathrm{minloss}\,\mathbb{B}_\mathcal{L} = \{P_\Omega^\dagger\}. \tag{12}$$

Proof. We shall consider cases for $B \in \mathbb{B}_\mathcal{L} \setminus \{P_\Omega^\dagger\}$. We will show that $P_\Omega^\dagger \prec B$ holds for all cases. Then $\mathrm{minloss}\,\mathbb{B}_\mathcal{L} = \{P_\Omega^\dagger\}$ follows.

<u>Case 1</u> B respects logical equivalence.

By Proposition 16 we obtain $P_\Omega^\dagger \prec B$.

<u>Case 2</u> B does not respect logical equivalence.

Since B does not respect logical equivalence, there exists a minimal $N \in \mathbb{N}$ such that two different logically equivalent sentences $\varphi, \psi \in S\mathcal{L}_N$ are assigned different degrees of belief, *i.e.*, $B(\varphi) \neq B(\psi)$.

We now inductively define functions $B_n : S\mathcal{L} \longrightarrow [0, 1]$ for $n \geq N$. First, let

$$B_N(\chi) := \begin{cases} \inf\{B(\theta) \,:\, \theta \in S\mathcal{L}_N \,\&\, \vDash \chi \leftrightarrow \theta\} & \text{if } \chi \in S\mathcal{L}_N \\ B(\chi) & \text{if } \chi \in S\mathcal{L} \setminus S\mathcal{L}_N \end{cases}.$$

Now assume $n > N$. For all $\chi \in SL_n$ such that no $\theta \in SL_{n-1}$ is logically equivalent to χ let

$$B_n(\chi) := \inf\{B(\theta) \ : \ \theta \in SL_n \ \& \ \vDash \chi \leftrightarrow \theta\}$$

and otherwise let

$$B_n(\chi) := \begin{cases} B_{n-1}(\theta) & \text{if } \chi \in SL_n \text{ and there exists a } \theta \in SL_{n-1} \text{ with } \vDash \chi \leftrightarrow \theta \\ B(\chi) & \text{if } \chi \in SL \setminus SL_n \end{cases}.$$

Note that B_n is well-defined, B_{n-1} respects logical equivalence on $\mathcal{L}_{n-1}$ and thus $B_{n-1}(\theta)$ does not depend on the particular sentence $\theta \in SL_{n-1}$ which is logically equivalent to χ.

By construction, B_{n+1} agrees with B_n on SL_n.

Finally, let $B^I(\chi) := \lim_{n \to \infty} B_n(\chi)$. Trivially, $B^I\!\restriction_N = B_N\!\restriction_N$.

Since for all $n \geq N$ the B_n respect logical equivalence on $\mathcal{L}_n$, B^I respects logical equivalence on $\mathcal{L}$. Furthermore, B^I agrees with B_n on the sentences of $\mathcal{L}_n$.

Now consider a $\chi \in SL$ and let $k \in \mathbb{N}$ be minimal such that $\chi \in SL_k$ and consider the corresponding proposition $F \subseteq \Omega_k$. For all $n \geq \max\{N, k\}$ we shall show that

$$\inf_{\rho \in \varrho_n} B(\rho F) \leq B^I(\chi).$$

If $\underline{k \leq N}$, then for all $n \geq N$ it holds that $B_n(\chi) = \inf\{B(\theta) \ : \ \theta \in SL_N \ \& \ \vDash \chi \leftrightarrow \theta\} = B_N(\chi)$. Hence, $B^I(\chi) = B_N(\chi)$. For $n \geq N$ there exist $\rho \in \varrho_n$ such that $\rho F = \chi$. Thus, $\inf_{\rho \in \varrho_n} B(\rho F) \leq B_N(\chi) = B^I(\chi)$.

If $\underline{k > N}$, then there are two cases. If no $\theta \in SL_{k-1}$ is logically equivalent to χ, then $B_k(\chi) = \inf\{B(\theta) \ : \ \theta \in SL_k \setminus SL_{k-1} \ \& \ \vDash \chi \leftrightarrow \theta\}$. In which case, we find for all $n \geq k > N$

$$\begin{aligned} \inf_{\rho \in \varrho_n} B(\rho F) &\leq \inf_{\rho \in \varrho_k} B(\rho F) \\ &= \inf\{B(\theta) \ : \ \theta \in SL_k \setminus SL_{k-1} \ \& \ \vDash \chi \leftrightarrow \theta\} \\ &= B^I(\chi). \end{aligned}$$

In the other case there does exist some $\theta \in SL_{k-1}$ which is logically equivalent to χ. Then $B_n(\chi) = B_{k-1}(\theta)$ for all $n \geq k$. So $B^I(\chi) = B_{k-1}(\theta)$. Thus, for all $n \geq \max\{N, k\} \geq k - 1$ it is true that

$$\begin{aligned} \inf_{\rho \in \varrho_n} B(\rho F) &\leq \inf_{\rho \in \varrho_k} B(\rho F) \\ &\leq \inf_{\rho \in \varrho_{\max\{N,k\}}} B(\rho F) \\ &= \inf\{B(\theta) \ : \ \theta \in SL_{k-1} \ \& \ \vDash \chi \leftrightarrow \theta\} \\ &= B^I(\chi). \end{aligned}$$

It thus follows for all $P \in \mathbb{P}_\mathcal{L}$ and all $n \geq N$ that

$$\begin{aligned} S_g^n(P, B) &= \sup_{\rho \in \varrho_n} S_{g,\rho}^n(P, B) \\ &= -\sum_{F \subseteq \Omega_n} \gamma_n(F)^\circ P(F) \inf_{\rho \in \varrho_n} \log B(\rho F) \\ &\geq -\sum_{F \subseteq \Omega_n} \gamma_n(F)^\circ P(F) \log B^I(\rho F) \qquad \text{for all } \rho \in \varrho_n \\ &= S_g^n(P, B^I). \end{aligned} \qquad (13)$$

Let us now note that $B^I(\varphi) < \max\{B(\varphi), B(\psi)\}$. Thus, $B^I(\varphi) + B^I(\neg\varphi) < \max\{B(\varphi), B(\psi)\} + B^I(\neg\varphi)$. Also observe that $B^I(\chi) \le B(\chi)$ for all $\chi \in S\mathcal{L}_N$. Thus, $B^I(\neg\varphi) \le B(\neg\varphi)$. Hence,

$$B^I(\varphi) + B^I(\neg\varphi) < \max\{B(\varphi), B(\psi)\} + B^I(\neg\varphi)$$
$$\le \max\{B(\varphi), B(\psi)\} + B(\neg\varphi)$$
$$\le 1.$$

We infer $B^I(\varphi) + B^I(\neg\varphi) < 1$ and thus $B^I \notin \mathbb{P}_{\mathcal{L}}$.

$\underline{\text{Case 2A}}$ $B^I \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$.

Since B^I respects logical equivalence, we obtain by Proposition 16 that $P_\Omega^\dagger \prec B^I$. Applying (13) we obtain $P_\Omega^\dagger \prec B$.

$\underline{\text{Case 2B}}$ $B^I \notin \mathbb{B}_{\mathcal{L}}$.

We shall now define a function B^J assigning every proposition a value in $[0, 1]$ as follows. Let $\tau \in S\mathcal{L}$ be some tautology. $\{\tau\}$ is a partition. Since $B^I \notin \mathbb{B}_{\mathcal{L}}$ it follows that $B^I(\tau) < 1$. Now put $B^J(\kappa) := 1 - B^I(\tau)$ for all contradictions $\kappa \in S\mathcal{L}$. Clearly, $B^J(\kappa) > 0$. For all satisfiable $\chi \in S\mathcal{L}$ let $B^J(\chi) := B^I(\chi)$.

Note that $B^J \in \mathbb{B}_{\mathcal{L}}$ and since $B^J(\neg\tau) > 0$ it follows that $B^J \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$. Also note that for all $n \in \mathbb{N}$ and all $P \in \mathbb{P}_{\mathcal{L}}$ it holds that $S_g^n(P, B^I) = S_g^n(P, B^J)$ and so

$$S_g^n(P, B) \ge S_g^n(P, B^I) = S_g^n(P, B^J).$$

Since B^J respects logical equivalence we can apply Case 2A to obtain $P_\Omega^\dagger \prec B^J$. But then $P_\Omega^\dagger \prec B$. $\quad\square$

Our main minimax theorem (already stated above on Page 2492) then follows immediately from Proposition 17 by applying Lemma 2 and Theorem 3:

Theorem 6 (Regularity minimax). *If g is regular and $\mathbb{E}_{\mathcal{L}}$ is finitely generated, then*

$$\text{minloss}\,\mathbb{B}_{\mathcal{L}} = \text{maxent}\,\mathbb{E}_{\mathcal{L}} = \mathbb{P}^\dagger = \{P_\Omega^\dagger\}.$$

If $\mathbb{E}_{\mathcal{L}} = \mathbb{P}_{\mathcal{L}}$, then the unique function with greatest entropy is the equivocator (Proposition 7). Thus by Theorem 6,

$$\text{minloss}\,\mathbb{B}_{\mathcal{L}} = \text{maxent}\,\mathbb{P}_{\mathcal{L}} = \{P_\Omega^\dagger\} = \{P_=\}.$$

Recall that $P_=$ assigns all n-states $\omega \in \Omega_n$ the same probability, $P_=(\omega) = \frac{1}{|\Omega_n|}$. So, if the agent does not possess any evidence then all n-states $\omega \in \Omega_n$ are all believed to the same degree. Absence of evidence entails symmetric degrees of belief. In other words, the three norms of objective Bayesianism entail an instance of the Principle of Indifference.

Surprisingly, perhaps, symmetry of the weighting function is not necessary to guarantee this instance of the Principle of Indifference on finite sublanguages—see Appendix B.

4.5. Infinite-Language Invariance

So far, we have been working over a fixed predicate language $\mathcal{L}$ (without quantifiers). One might wonder what would have happened if one had started out with a different such language.

We will investigate this question by considering predicate languages which contain finitely many further relation symbols and/or finitely many further constant symbols than does $\mathcal{L}$.

For all languages we consider here, we shall suppose that the ways the constant symbols are ordered are consistent. Furthermore, we suppose that the order types of the constant symbols are ω, the first infinite ordinal. That is, for $\mathcal{L} \subset \mathcal{L}^1$ let $t_1, t_2, \ldots$ be the constant symbols in $\mathcal{L}$ and let $T^{new} := \{t_1^{new}, \ldots, t_m^{new}\}$ be the set of constant symbols in $\mathcal{L}^1$ which are not in $\mathcal{L}$. Then we require that the constant symbols of $\mathcal{L}^1$ are ordered such that

- for all $n \in \mathbb{N}$, t_n appears before t_{n+1} (consistency),

- for all $t \in T^{new}$ there exists some $n \in \mathbb{N}$ such that t appears before t_n (order type ω).

The way the constant symbols of $\mathcal{L}^1$ are ordered can be thought of as inserting the $t \in T^{new}$ into the ordering of the constant symbols of $\mathcal{L}$.

From now on, superscripts are used to refer to such predicate languages, while subscripts continue to refer to their respective finite sublanguages. For example, $\mathcal{L}_n^1$ is the finite sublanguage of $\mathcal{L}^1$ which contains only the first n constants of $\mathcal{L}^1$. For $\mathcal{L} \subset \mathcal{L}^1$, in general, the set of the first n constants of $\mathcal{L}$ may be different from the set of the first n constants of $\mathcal{L}^1$.

Definition 23 (**Infinite-Language Invariance**). *A weighting function g is* infinite-language invariant, *if and only if the following holds: for all $\mathcal{L}$ and for all $\mathbb{E}_\mathcal{L}$ finitely generated by constraints on the finite sublanguage $\mathcal{L}_K$ of $\mathcal{L}$, if $\mathcal{L}^1$ and $\mathcal{L}^2$ are such that $\mathcal{L} \subseteq \mathcal{L}^1 \subseteq \mathcal{L}^2$, then for all $B \in \mathrm{minloss}\, \mathbb{B}_{\mathcal{L}^1}$ there exists a $C \in \mathrm{minloss}\, \mathbb{B}_{\mathcal{L}^2}$ such that $C_{\restriction \mathcal{L}^1} = B$.*

Infinite-language invariance is motivated by the thought that simply adding new constant or predicate symbols to the language $\mathcal{L}$ should not change the inferences which are expressible in the original language $\mathcal{L}$. Note the following qualification: since each element of the domain is picked out by some member of $\mathcal{L}$, one can infer that in $\mathcal{L}'$ formed by adding constants to $\mathcal{L}$, there must be some constants which name the same individual.

We shall now proceed to show that the weighting functions which we focus on in this paper—the regular weighting functions—are infinite-language invariant.

Lemma 9. *If $\mathcal{E}, \mathcal{E}'$ are non-empty and convex sets of the following form*

$$\mathcal{E} \subseteq \{(x_1, \ldots x_n) \in \mathbb{R}^n \; : \; \sum_{i=1}^n x_i = 1 \,\&\, x_i \geq 0\}$$

$$\mathcal{E}' \subseteq \{(y_1, z_1, y_2, z_2, \ldots, y_n, z_n) \in \mathbb{R}^{2n} \; : \; y_i, z_i \geq 0 \quad \& \quad (y_1 + z_1, \ldots, y_n + z_n) \in \mathcal{E}\},$$

then for

$$\{(x_1^\dagger, \ldots, x_n^\dagger)\} = \arg \sup_{(x_1, \ldots, x_n) \in \mathcal{E}} - \sum_{i=1}^n x_i \log x_i$$

$$\{(y_1^\dagger, z_1^\dagger, \ldots, y_n^\dagger, z_n^\dagger)\} = \arg \sup_{(y_1, z_1, \ldots, y_n, z_n) \in \mathcal{E}'} - \sum_{i=1}^n y_i \log y_i + z_i \log z_i$$

it holds that $y_i^\dagger = z_i^\dagger = \frac{x_i^\dagger}{2}$ for all $1 \le i \le n$.

Proof. That the suprema are unique follows from the convexity of the sets $\mathcal{E}, \mathcal{E}'$ and the fact that $H_\Omega^n, H_\Omega^{2n}$ are strictly concave functions on $\mathbb{P}_n$, respectively, $\mathbb{P}_{2n}$.

Recall that $\mathcal{L}^U$ is the language introduced in Lemma 2. $y^\dagger = z^\dagger = \frac{x_i^\dagger}{2}$ is a direct consequence of $P_\Omega^\dagger$ equivocating beyond $\mathcal{L}_k^U$ (Proposition 9). $\square$

Theorem 7. *If g is regular, then g is infinite-language invariant.*

Proof. Let $\mathbb{E}_\mathcal{L}$ be finitely generated by constraints expressible in $\mathcal{L}_K$. Let $\mathcal{L} \subseteq \mathcal{L}^1 \subseteq \mathcal{L}^2$. By Theorem 6 we obtain $\mathrm{minloss}\, \mathbb{B}_{\mathcal{L}^1} = \mathrm{maxent}\, \mathbb{E}_{\mathcal{L}^1} = \{P_\Omega^{\dagger 1}\}$ and $\mathrm{minloss}\, \mathbb{B}_{\mathcal{L}^2} = \mathrm{maxent}\, \mathbb{E}_{\mathcal{L}^2} = \{P_\Omega^{\dagger 2}\}$, where $P_\Omega^{\dagger 1}$ and $P_\Omega^{\dagger 2}$ are the standard entropy limits on $\mathcal{L}^1$, respectively, $\mathcal{L}^2$.

Let $K_2 \in \mathbb{N}$ be minimal such that $\mathcal{L}_K \subseteq \mathcal{L}_{K_2}^2$, *i.e.*, the set of the first K_2 constant symbols of $\mathcal{L}^2$ contains the constant symbols $\{t_1, \ldots, t_K\}$ of $\mathcal{L}$. It suffices to show that for all $n \ge K_2$ and all $\nu \in \Omega_n^1$ it holds that $P_\Omega^{\dagger 1}(\nu) = P_\Omega^{\dagger 2}(\nu)$, where Ω_n^1 is the set of n-states of $\mathcal{L}^1$. Note that the constants in $t_1, \ldots, t_K$ are in $\mathcal{L}_{K_2}^1$.

Since the standard entropy limits is finite-language invariant (Section 4.2.1) it follows for $n \ge K_2$ that $P_\Omega^{\dagger 1}(\nu) = P_{\Omega n}^{\dagger 1}(\nu)$, where $\{P_{\Omega n}^{\dagger 1}\} = \arg \sup_{P \in \mathbb{E}_n^1} S_\Omega^n(P)$, and $P_\Omega^{\dagger 2}(\nu) = P_{\Omega n}^{\dagger 2}(\nu)$, where $\{P_{\Omega n}^{\dagger 2}\} = \arg \sup_{P \in \mathbb{E}_n^2} S_\Omega^n(P)$.

We now obtain from Lemma 9 and Proposition 5 that

$$P_{\Omega n}^{\dagger i}(\nu) = P_\Omega^\dagger(\omega_\nu) \frac{1}{2^{|\nu| - |\omega_\nu|}}$$

where ω_ν is the unique maximal state of $\mathcal{L}$ such that $\nu \vDash \omega_\nu$. Thus, $P_\Omega^{\dagger 1}(\nu) = P_\Omega^{\dagger 2}(\nu)$. $\square$

So, neither adding new redundant names for individuals in the domain to $\mathcal{L}$ nor adding relation symbols which are not constrained by the agent's evidence on $\mathcal{L}$ changes one's rational beliefs in the sentences $\varphi \in S\mathcal{L}$.

Language invariance is an important desideratum for reasoning under uncertainty. We have seen that focussing on regular weighting functions ensures language invariance. We conjecture that, if one imposes the desiderata that g be atomic, inclusive, symmetric, refined and infinite-language invariant, then the standard entropy maximiser will be the belief function with the best loss profile. If this is the case then our results for *regular* weighting functions, which are strongly refined, are symptomatic of a more general phenomenon.

5. Handling Quantifiers

Thus far, we have shown that, on a language $\mathcal{L}^{\not\exists}$ without quantifiers, if the evidence is finitely generated and the weighting function is regular, then the belief function that has the best lost profile is the probability function in $[\mathbb{E}_{\mathcal{L}}]$ that maximises standard entropy. This provides a justification for all the norms of objective Bayesianism on a language without quantifiers.

As we shall see in Section 5.1, that the language is quantifier free was key here: on a language $\mathcal{L}^{\exists}$ with quantifiers, the n-scores become infinite, which makes the comparison of loss profiles impossible. That the evidence is finitely generated is also key: we shall see in Section 6.1 that the minimax result need not hold true if the evidence is not finitely generated.

While the use of scoring rules cannot be readily adapted to a quantified language $\mathcal{L}^{\exists}$, we shall see in Section 5.2 that we can nevertheless justify the norms of objective Bayesianism on $\mathcal{L}^{\exists}$ if we extend our notion of loss profile and add two further desiderata motivated by the application of objective Bayesianism to inductive logic: that inferences should be language invariant, and that, ceteris paribus, universal hypotheses should be afforded substantial credence.

5.1. Limits to the Minimax Approach

Here we explain why the minimax analysis adopted in Section 4 cannot be applied to the case of a language with quantifier symbols. The problem is that n-score becomes infinite, making it impossible to compare the scores of different belief functions.

There are two ways in which n-score becomes infinite. The first is through a failure of super-regularity. A probability function is *super-regular*, if it gives every contingent sentence positive probability. Now, many probability functions that seem eminently rational are not super-regular. For example, if one has no evidence, $\mathbb{E}_{\mathcal{L}} = \mathbb{P}_{\mathcal{L}}$, then it is plausible that one is rationally entitled (even if not rationally compelled) to adopt the equivocator function $P_{=}$, which gives each n-state the same probability, as one's belief function. However, this probability function will give zero probability to a universally quantified sentence such as $\forall x U x$. More generally, if evidence is finitely generated then no inclusive, symmetric entropy maximiser will be super-regular:

Proposition 18. *Let $\mathbb{E}_{\mathcal{L}}$ be finitely generated and let g be symmetric and inclusive. If the sequence $(P_n^{\dagger})_{n \in \mathbb{N}}$ has a point of accumulation $Q \in \mathbb{P}_{\mathcal{L}}$, then Q is not super-regular.*

Proof. Let U be a relation symbol in $\mathcal{L}$ of arity r, say. For all $n \in \mathbb{N}$ let

$$\varphi^n := \bigvee_{\substack{\omega \in \Omega_n \\ \omega \models \bigwedge_{i=1}^{n} U t_i}} \omega,$$

where t_i denotes the tuple of r repetitions of t_i.

If $P_K^{\dagger}(\varphi^K) = 0$, then by the open-mindedness of entropy maximisers $P_n^{\dagger}(\varphi^K) = 0$ for all $n \geq K$. Thus, for all points of accumulation $Q \in \mathbb{P}_{\mathcal{L}}$ it holds that $Q(\varphi^K) = 0$. Hence, Q is not super-regular.

If $P_K^\dagger(\varphi^K) > 0$, then we apply Proposition 9 to find that for all $l \geq n$

$$
\begin{aligned}
P_l^\dagger(\varphi^n) &= P_l^\dagger(\varphi^K)\frac{|\Omega_K^U|}{|\Omega_n^U|} \\
&\leq P_l^\dagger(\varphi^K)2^{K-n} \\
&\leq 2^{K-n},
\end{aligned}
$$

Let Q be a point of accumulation of $(P_n^\dagger)_{n\in\mathbb{N}}$ and let $(P_n^\dagger)_{n_j}$ be a subsequence which converges to Q. Since K is fixed we now find

$$
\begin{aligned}
0 &\leq Q(\forall x U x) \\
&\overset{P3}{=} \lim_{j\longrightarrow\infty} Q(\bigwedge_{i=1}^{n_j} U t_i) \\
&= \lim_{j\longrightarrow\infty} \lim_{m\longrightarrow\infty} P_{n_m}^\dagger(\bigwedge_{i=1}^{n_j} U t_i) \\
&= \lim_{j\longrightarrow\infty} \lim_{m\longrightarrow\infty} P_{n_m}^\dagger(\varphi^{n_j}) \\
&\leq \lim_{j\longrightarrow\infty} 2^{K-j} \\
&= 0.
\end{aligned}
$$

Q is not super-regular. □

Now, a failure of super-regularity is not normally problematic—it is simply a well accepted fact that probability theory forces probability 0 (respectively 1) on many sentences which might be true (respectively false). For example, the strong law of large numbers and the various zero-one laws force extreme probabilities. Moreover, the issue of super-regularity did not arise on $\mathcal{L}^{\overline{\exists}}$, where no contingent sentences are given probability 0 by the entropy maximisers considered above. However, a problem does emerge if we try to apply the scoring rule approach to $\mathcal{L}^{\exists}$, where super-regularity becomes pertinent. If θ is possible yet is given zero belief by belief function B then the logarithmic loss, $-\log B(\theta)$, is infinite if θ turns out to be true. Hence, as long as some epistemically possible physical probability function gives positive probability to θ, belief function B will have infinite score. When scores become infinite, they cannot be readily used to compare belief functions. It is clear, for example, that some non-super-regular belief functions will have better loss profiles than others, but this will not be apparent if we define loss profiles in terms of scores. This problem appears to limit the scope of scoring rules to languages without quantifiers.

One might suggest here that the fact that non-super-regular functions lead to infinite scores merely serves to show that one should adopt a super-regular function as one's belief function. However, there are good grounds for questioning such a conclusion. In particular, consider again the case of a total absence of evidence. As mentioned above, imposing super-regularity rules out the equivocation function $P_=$ as a viable belief function. This means that any super-regular function must, in the total absence of evidence, force a skewed distribution on the n-states, for some n. Thus, one is forced to believe some states to a greater degree than others, despite the fact that one has no evidence to distinguish any such

state from any other. So super-regularity leads to very counter-intuitive consequences and the infinite score problem suggests that the scoring rule approach breaks down on languages with quantifiers.

There is a second way in which the scores become infinite when quantifiers are admitted into the language. When one admits quantifiers into the language, one introduces the possibility of infinite partitions (Example 1) and it is natural, when defining a scoring rule on such a language, to consider scores on these infinite partitions. If a weighting function is inclusive then for any sentence $\theta \in SL^{\exists}$, some partition containing θ will be given positive weight. If it is refined, then any partition that refines this partition will be given positive weight, including any infinite partition which refines this partition. The problem is that, even in the total absence of evidence, every belief function has infinite worst-case expected loss over such a partition:

Proposition 19. *If there exists a partition $\pi_\infty \in \Pi_\mathcal{L}$ consisting of infinitely many sentences such that $g(\pi_\infty) > 0$, then for all $B \in \mathbb{B}_\mathcal{L}$ it holds that*

$$\sup_{P \in \mathbb{P}_\mathcal{L}} - \sum_{\varphi \in \pi} g(\pi_\infty) P(\varphi) \log B(\varphi) = +\infty.$$

Proof. Let $\pi_\infty = \{\varphi_1, \varphi_2, \dots\}$. Let $B \in \mathbb{B}_\mathcal{L}$ be arbitrary but fixed.

If there exists a $\varphi \in \pi_\infty$ such that $B(\varphi) = 0$, then any $P \in \mathbb{P}_\mathcal{L}$ with $P(\varphi) > 0$ satisfies $\sum_{\varphi \in \pi} -g(\pi_\infty) P(\varphi) \log B(\varphi) = +\infty$.

Now assume that $B(\varphi_n) > 0$ for all $n \in \mathbb{N}$.

Since $B \in \mathbb{B}_\mathcal{L}$ it holds that $\sum_{\varphi \in \pi_\infty} B(\varphi) \leq 1$. Thus, there has to exists an infinite set $\mathbb{N}_B \subseteq \mathbb{N} \setminus \{1\}$ such that $n \in \mathbb{N}_B$ implies $0 < B(\varphi_n) < \frac{1}{n} \leq \frac{1}{2}$. Let $\{n_1^B, n_2^B, \dots\}$ be an enumeration of $\mathbb{N}_B$. Let $\{m_2^B, m_3^B, \dots\}$ be an enumeration of an infinite subset of $\mathbb{N}_B$ such that $0 < B(\varphi_{m_k^B}) \leq \frac{1}{e^{(k^2)}} < 1$ and $m_k^B < m_{k+1}^B$ for all $k \in \mathbb{N} \setminus \{1\}$. Since the n_k^B tend to infinity, such a sequence $(m_k^B)_{k \in \mathbb{N} \setminus \{1\}}$ has to exist.

Recall that $\sum_{n \in \mathbb{N}} \frac{1}{n^2} = \frac{\pi^2}{6}$. Let $P \in \mathbb{P}_\mathcal{L}$ be such that for $k \geq 2$ it holds that

$$P(\varphi_{m_k^B}) := \frac{6}{\pi^2} \cdot \frac{1}{k^2}$$

$$P(\varphi_1) := 1 - \sum_{k=2}^{\infty} P(\varphi_{m_k^B}) = \frac{6}{\pi^2}$$

$$P(\varphi_n) := 0 \text{ for all } n \in \mathbb{N} \setminus \{1, m_2^B, m_3^B, \dots\}.$$

We now explain why such a probability function $P \in \mathbb{P}_\mathcal{L}$ exists.

The idea is to define a measure which assigns the set of term structures which are a model of $\varphi_{m_k^B}^B$ the value $\frac{6}{\pi^2} \frac{1}{k^2}$ and assigns value zero to all other term structures which do not model any of the $\varphi_{m_k^B}^B$. The probability of an arbitrary sentence $\chi \in SL$ is then measure assigned to all term structures in which χ holds. One has to be careful of how to set up this measure. Fortunately, the recipe for doing so is well-known.

We follow [7] (pp. 164) and define a term structure $\mathcal{M}$ of $\mathcal{L}$ as a structure with domain $\{t_n : n \in \mathbb{N}\}$ and each constant symbol t_n of $\mathcal{L}$ is interpreted in $\mathcal{M}$ as itself. We use $T\mathcal{L}$ to denote the set of term structures of $\mathcal{L}$.

Now let $\mathcal{P}(T\mathcal{L})$ denote the power set of $T\mathcal{L}$ and put

$$T(\theta) := \{\mathcal{M} \in T\mathcal{L} \;:\; \mathcal{M} \models \theta\}$$
$$R := \{T(\theta) \;:\; \theta \in S\mathcal{L}^{\overrightarrow{7}}\} \subseteq \mathcal{P}(T\mathcal{L}).$$

For a quantified sentence $\theta = \exists x \theta(x)$ let $T(\theta) := \bigcup_{i \in \mathbb{N}} T(\theta(t_i))$, similarly for the universal quantifier $\forall$.

Now let μ^* be any (finitely additive and normalised to one) outer measure on $\mathcal{P}(T\mathcal{L})$ such that $\mu^*(\varphi_{m_k}^B) = \frac{6}{\pi^2}\frac{1}{k^2}$. Particularly simple such outer measures μ^* are measures which for all m_k assign a single particular term structure $\mathcal{M}$ in which $\varphi_{m_k}^B$ holds the value $\frac{6}{\pi^2}\frac{1}{k^2}$.

Next, define R^∞ to be the smallest subset of $\mathcal{P}(T\mathcal{L})$ which contains R and is closed under complements and countable unions. We now define a countably additive measure μ^∞ on R^∞ as follows: $\mu^\infty : R^\infty \to [0,1]$ such that $\mu^\infty(A) = \mu^*(A)$ for all $A \in R^\infty$.

Letting $P(\theta) := \mu^\infty(T(\theta))$ defines a probability function as shown in [7] (pp. 168–171). Furthermore, by construction $\mu^*(\varphi_i) = \frac{6}{i^2\pi^2} = P(\varphi_i)$.

Having demonstrated the existence of the required probability function P, we now show that, for this function P, B incurs an infinite loss. Intuitively, $P(\varphi_n)$ can be obtained from the sequence $(\frac{1}{k^2})_{k \in \mathbb{N}}$ by inserting zeros and normalising by multiplying with $\frac{6}{\pi^2}$. The idea behind this definition is to ensure that for all $k \in \mathbb{N}$ there exists a unique $n \in \mathbb{N}_B$ such that $P(\varphi_n) = \frac{6}{\pi^2} \cdot \frac{1}{k^2}$. Furthermore, for these $n \in \mathbb{N}_B$ it holds that $B(\varphi_n) \leq \frac{1}{e^{(k^2)}}$. For all other $n > 1$ we ensure that $P(\varphi_n)$ vanishes; $P(\varphi_1)$ is defined in such that $\sum_{\varphi \in \pi} P(\varphi) = 1$ holds.

So, when $P(\varphi_n) > 0$ and $n \in \{m_2^B, m_3^B, \dots, \}$ we have

$$\begin{aligned}
-P(\varphi_n)\log B(\varphi_n) &\geq \frac{6}{\pi^2}\frac{1}{k^2}\log e^{(k^2)} \\
&= \frac{6}{\pi^2}\frac{1}{k^2}k^2\log e \\
&= \frac{6}{\pi^2}.
\end{aligned}$$

Finally, we obtain

$$-\sum_{\varphi \in \pi} g(\pi_\infty)P(\varphi)\log B(\varphi) \geq g(\pi_\infty)\sum_{m_2^B, m_3^B, \dots} \frac{6}{\pi^2}$$
$$= +\infty.$$

$\square$

In particular, even the super-regular belief functions have infinite score on any such partition, so one cannot say that any super-regular function has lower overall score than a non-super-regular function. This result, then, casts further doubt on the suggestion that it might be preferable to adopt a super-regular function as one's belief function. Moreover, it clearly suggests that an attempt to extend the minimax approach, which is based on scoring rules, to languages with quantifiers will be fraught with difficulty.

5.2. The Probability Norm

We have argued that there is little scope for straightforwardly extending the minimax analysis to languages with quantifiers because of the problem that scores will quickly become infinite and thus incomparable. So we need another approach, if we are to show that the Probability axioms P1-P3, as well as the Calibration and Equivocation norms, are to apply to languages with quantifiers.

Our plan of attack is as follows. First, as noted in Section 4.5, language invariance is an important desideratum. In particular, one would not want one's degrees of belief on the sentences of a quantifier-free language $\mathcal{L}^{\nexists}$ to change if one were to introduce quantifiers into the language. That is, if evidence determines that one should adopt B_1 as one's belief function on $\mathcal{L}^{\nexists}$ and B_2 as one's belief function on $\mathcal{L}^{\exists}$, where both languages contain the same individuals and relation symbols, then one would want B_1 and B_2 to agree on quantifier-free sentences of $\mathcal{L}$, *i.e.*, one would want that $B_1(\theta) = B_2(\theta)$ for each $\theta \in S\mathcal{L}^{\nexists}$.

Thus far, we have argued that a belief function on $\mathcal{L}^{\nexists}$, given finitely generated $\mathbb{E}$, ought to satisfy the axioms of probability P1 and P2 on $\mathcal{L}^{\nexists}$, as well as the Calibration and Equivocation norms. Given the language invariance desideratum, this implies that the appropriate belief function on $\mathcal{L}^{\exists}$, should, when restricted to quantifier-free sentences, satisfy P1, P2 and the Calibration and Equivocation norms. If we can show that the probability axioms P1-3 should also be satisfied on the language $\mathcal{L}^{\exists}$ as a whole, then degrees of belief in the quantified sentences are uniquely determined by those on the quantifier-free sentences [7] (Theorem 11.2): there is no further role that Calibration or Equivocation can play on the quantified sentences. Thus it suffices to argue for the probability axioms on $\mathcal{L}^{\exists}$. As usual, we restrict attention to evidence sets that are finitely generated in the sense of Definition 5, *i.e.*, $\mathbb{E}_{\mathcal{L}}$ generated by constraints involving sentences of some $\mathcal{L}_K^{\nexists}$ and regular weighting functions g.

In Theorem 4 we showed that the default loss incurred by adopting belief function B when φ is true is such that $L(\varphi, B) = -\log B(\varphi)$, modulo some multiplicative constant. This penalises smaller degrees of belief more than larger degrees of belief. As discussed above, there is little scope for using this to measure the overall expected loss incurred by B on $\mathcal{L}^{\exists}$, and so we cannot directly extend the notion of loss profile developed in Definition 21 to $\mathcal{L}^{\exists}$. However, this default loss function does suggest the following constraint:

(*) Suppose that for all $\theta \in S\mathcal{L}^{\exists}$, $B(\theta) \geq B'(\theta)$, and there is some $\varphi \in S\mathcal{L}^{\exists}$ such that $B(\varphi) > B'(\varphi)$. Then B has a better loss profile than B'.

In other words, if the default loss incurred by B' dominates that incurred by B then B has a better loss profile than B'. We can use (*) to extend our notion of loss profile: the two conditions in Definition 21 apply to quantifier-free sentences in $\mathcal{L}^{\exists}$, and we add the further condition (*) to constrain the quantified sentences. We shall show that the addition of (*) goes some way towards demonstrating P1-3 on $\mathcal{L}^{\exists}$, although we shall have to add a further desideratum in order to complete the derivation.

Definition 24 (**Better loss profile on $\mathcal{L}^{\exists}$**)**.** *B has a better loss profile on $\mathcal{L}^{\exists}$ than B' if and only if:*

1. *$B \prec B'$ (as defined in Definition 21), or*

2. *B dominates B' on $\mathcal{L}^{\exists}$ and there exists some $\varphi \in S\mathcal{L}^{\exists}$ such that $B(\varphi) > B'(\varphi)$.*

We write $B \prec^* B'$ to denote that B has a better loss profile on $\mathcal{L}^{\exists}$ than B'. Clearly, $\prec^*$ is asymmetric. We will be interested in those belief functions on $\mathcal{L}^{\exists}$ that have the best loss profile on $\mathcal{L}^{\exists}$, i.e., the minimal elements of $\prec^*$, and define:

$$\mathrm{minloss}^* \mathbb{B}_{\mathcal{L}} := \{ B \in \mathbb{B}_{\mathcal{L}} \ : \ there\ is\ no\ B' \in \mathbb{B}_{\mathcal{L}}\ such\ that\ B' \prec^* B \}. \tag{14}$$

Note that if B dominates B' on $\mathcal{L}^{\exists}$, then $B \prec B'$ cannot hold. $\prec$ and $\prec^*$ are thus consistent.

Proposition 20. *All $B \in \mathrm{minloss}^* \mathbb{B}_{\mathcal{L}}$ agree with $P_{\Omega}^{\dagger}$ on $\mathcal{L}^{\not\exists}$.*

Proof. Since we assume that g is regular and that $\mathbb{E}_{\mathcal{L}}$ is finitely generated we can apply Theorem 6 to obtain that all all $B \in \mathrm{minloss}\, \mathbb{B}_{\mathcal{L}}$ agree with $P_{\Omega}^{\dagger}$ on $\mathcal{L}^{\not\exists}$.

The claim now follows, since $B \prec B'$ implies $B \prec^* B'$. $\quad\square$

Proposition 21. *If $\mathrm{minloss}\, \mathbb{B}_{\mathcal{L}} = \emptyset$, then $\mathrm{minloss}^* \mathbb{B}_{\mathcal{L}} = \emptyset$.*

Proof. $\prec$ is asymmetric, irreflexive and transitive, Proposition 10; and thus free of cycles. Hence, for all fixed $B' \in \mathbb{B}_{\mathcal{L}}$ there exists some $B \in \mathbb{B}_{\mathcal{L}}$ such that $B \prec B'$. This implies $B \prec^* B'$.

Hence, for all $B' \in \mathbb{B}_{\mathcal{L}}$ there exists some $B \in \mathbb{B}_{\mathcal{L}}$ such that $B \prec^* B'$. We obtain $\mathrm{minloss}^* \mathbb{B}_{\mathcal{L}} = \emptyset$. $\quad\square$

We shall use $B_{\dagger} \in \mathbb{B}_{\mathcal{L}}$ to denote an arbitrary but fixed belief function in $\mathrm{minloss}^* \mathbb{B}_{\mathcal{L}}$. A priori, it is not clear that such a function $B_{\dagger}$ exists.

The rest of this section does not depend on $\mathbb{E}_{\mathcal{L}}$, the weighting function g nor the particular probability function the $B \in \mathrm{minloss}\, \mathbb{B}_{\mathcal{L}}$ agree with on $\mathcal{L}^{\not\exists}$. All that matters is that there exists some probability function $P \in \mathbb{P}_{\mathcal{L}}$ the $B \in \mathrm{minloss}\, \mathbb{B}_{\mathcal{L}}$ agree with on $\mathcal{L}^{\not\exists}$. As we know, this is the case if $\mathbb{E}_{\mathcal{L}}$ is finitely generated and g is regular.

Definition 25. *A sentence $\varphi \in S\mathcal{L}^{\exists}$ is called* contingent, *if and only if φ and $\neg\varphi$ are satisfiable.*

Lemma 10. *For all $\theta, \varphi \in S\mathcal{L}^{\exists}$ such that $\theta \models \varphi$ it holds that $B_{\dagger}(\varphi) \geq B_{\dagger}(\theta)$. In particular, $B_{\dagger}(\psi) = 0$ for all contradictions $\psi \in S\mathcal{L}^{\exists}$ and $B_{\dagger}(\chi) = 1$ for all tautologies $\chi \in S\mathcal{L}^{\exists}$.*

For $\theta, \varphi \in S\mathcal{L}^{\not\exists}$ we have already seen that $B_{\dagger}(\varphi) \geq B_{\dagger}(\theta)$, this followed from $B_{\dagger}$ satisfying P1 and P2 on $\mathcal{L}^{\not\exists}$.

Proof. <u>Case 1</u>. θ is a contradiction.

For a tautology $\tau \in S\mathcal{L}^{\not\exists}$, $\{\tau, \theta\}$ is a partition. Since $B_{\dagger}(\tau) = 1$ and $B_{\dagger}(\tau) + B_{\dagger}(\theta) \leq 1$ it follows that $B_{\dagger}(\theta) = 0$. Hence, $B_{\dagger}(\varphi) \geq 0 = B_{\dagger}(\theta)$.

<u>Case 2</u>. θ is a tautology.

Let $\chi \in S\mathcal{L}^{\exists}$ be a contradiction. We just proved that $B_{\dagger}(\chi) = 0$. The only constraints applying to $B_{\dagger}(\theta)$ are of the form $B_{\dagger}(\theta) + B_{\dagger}(\chi) \leq 1$ where χ is a contradiction and of the form $B_{\dagger}(\theta) \leq 1$. Thus, the only meaningful constraint on $B_{\dagger}(\theta)$ is $B_{\dagger}(\theta) \leq 1$. By (*) we have $B_{\dagger}(\theta) = 1$.

Since θ implies φ, φ has to be a tautology, too. Hence, $B_{\dagger}(\varphi) = 1 = B_{\dagger}(\theta)$.

<u>Case 3</u>. θ is contingent.

If φ is a tautology, then $B_{\dagger}(\varphi) = 1$ by the above and we are done.

Note that φ cannot be a contradiction since θ is satisfiable.

Assume from now on that φ is contingent.

Case 3A $\models \theta \leftrightarrow \varphi$.

For all index sets I and all sentences $\varphi_i \in S\mathcal{L}^\exists$ the following are equivalent

- $\{\varphi\} \cup \bigcup_{i \in I} \{\varphi_i\} \in \Pi_{\mathcal{L}}$,

- $\{\theta\} \cup \bigcup_{i \in I} \{\varphi_i\} \in \Pi_{\mathcal{L}}$.

(*) implies that $B_\dagger(\varphi) = B_\dagger(\theta)$.

Case 3B θ, φ and $\varphi \wedge \neg\theta$ are contingent.

Let I be any countable index set and let $\varphi_i \in S\mathcal{L}^\exists$ for $i \in I$ be contingent such that

$$\{\varphi\} \cup \bigcup_{i \in I} \{\varphi_i\} \in \Pi_{\mathcal{L}}.$$

Then by the consistency of θ and $\varphi \wedge \neg\theta$

$$\{\theta \wedge \varphi\} \cup \{\varphi \wedge \neg\theta\} \cup \bigcup_{i \in I} \{\varphi_i\} \in \Pi_{\mathcal{L}}.$$

And since $\theta \models \varphi$

$$\{\theta\} \cup \{\varphi \wedge \neg\theta\} \cup \bigcup_{i \in I} \{\varphi_i\} \in \Pi_{\mathcal{L}}.$$

From normalisation (Definition 1) we now obtain

$$B_\dagger(\varphi) + \sum_{i \in I} B_\dagger(\varphi_i) \leq 1 \tag{15}$$

$$B_\dagger(\theta) + B_\dagger(\varphi \wedge \neg\theta) + \sum_{i \in I} B(\varphi_i) \leq 1. \tag{16}$$

Note that the equations in (15) are the only constraints which constrain $B_\dagger(\varphi)$. In particular, $B_\dagger(\varphi) = B_\dagger(\theta)$ will not violate any constraint in (15).

The question arises whether $B_\dagger(\varphi) = B_\dagger(\theta)$ imposes any further constraints?

$B_\dagger(\varphi)$ only imposes constraints on the $B_\dagger(\varphi_i)$ for $i \in I$. Let $i \in I$ be fixed and let J be an index set and $(\psi_j)_{j \in J} \in S\mathcal{L}^\exists$ be such that $\{\varphi_i\} \cup \{\varphi\} \cup \bigcup_{j \in J} \{\psi_j\} \in \Pi_{\mathcal{L}}$. Then $\{\varphi_i\} \cup \{\theta\} \cup \{\varphi \wedge \neg\theta\} \cup \bigcup_{j \in J} \{\psi_j\} \in \Pi_{\mathcal{L}}$. Thus, $B_\dagger(\varphi) = B_\dagger(\theta)$ does not impose any further constraint on $B_\dagger(\varphi_i)$ which is not already imposed by $B_\dagger(\theta)$.

By (*) we now find $B_\dagger(\theta) \leq B_\dagger(\varphi)$. $\square$

Corollary 7. *$B_\dagger$ respects logical equivalence on $\mathcal{L}^\exists$.*

Proof. If $\varphi, \theta \in S\mathcal{L}^\exists$ are logically equivalent, then $B_\dagger(\varphi) \leq B_\dagger(\theta) \leq B_\dagger(\varphi)$ and thus $B_\dagger(\varphi) = B_\dagger(\theta)$. $\square$

Corollary 8. *For all $\exists x \theta(x) \in S\mathcal{L}^\exists$ it holds that*

$$\lim_{n \longrightarrow \infty} B_\dagger\left(\bigvee_{i=1}^n \theta(t_i)\right) \leq B_\dagger(\exists x \theta(x)).$$

Proof. First note that $\bigvee_{i=1}^{n}(\theta(t_i))$ implies $\bigvee_{i=1}^{n+1}(\theta(t_i))$. Thus, $B_\dagger(\bigvee_{i=1}^{n}(\theta(t_i)))$ is a (not necessarily strictly) increasing sequence in $[0, 1]$ which has a limit. Finally, note that for all $n \in \mathbb{N}$ $\bigvee_{i=1}^{n}(\theta(t_i))$ implies $\exists x \theta(x)$. Hence, $B_\dagger(\exists x \theta(x))$ has to be greater or equal than the limit. $\square$

Corollary 9 (Superadditivity of $B_\dagger$ on $\mathcal{L}^\exists$). *If* $\models \neg(\theta \wedge \varphi)$, *then* $B_\dagger(\theta) + B_\dagger(\varphi) \leq B_\dagger(\theta \vee \varphi)$.

Proof. If either θ or φ is a contradiction or a tautology, then the Corollary follows trivially.

If $\theta \vee \varphi$ is a tautology, then the corollary follows trivially, too.

It remains to consider the case of contingent $\theta \vee \varphi$. By the above we may assume that θ and φ are contingent. Let I be any countable index set and let $\varphi_i \in S\mathcal{L}^\exists$ for $i \in I$ be satisfiable such that

$$\{\theta\} \cup \{\varphi\} \cup \bigcup_{i \in I}\{\varphi_i\} \in \Pi_\mathcal{L}.$$

Then,

$$\{\theta \vee \varphi\} \cup \bigcup_{i \in I}\{\varphi_i\} \in \Pi_\mathcal{L}.$$

From normalisation (Definition 1) we now obtain

$$B_\dagger(\theta) + B_\dagger(\varphi) + \sum_{i \in I} B_\dagger(\varphi_i) \leq 1$$
$$B_\dagger(\theta \vee \varphi) + \sum_{i \in I} B_\dagger(\varphi_i) \leq 1.$$

The same reasoning a in Lemma 10 about constraints now yields: $B_\dagger(\theta) + B_\dagger(\varphi) \leq B_\dagger(\theta \vee \varphi)$. $\square$

Lemma 11. *For all* $\theta \in S\mathcal{L}^\exists$ *it holds that* $B_\dagger(\theta) + B_\dagger(\neg\theta) = 1$.

In particular, this means that $B_\dagger(\exists x \theta(x)) + B_\dagger(\forall x \neg\theta(x)) = 1$ for all $\exists x \theta(x) \in S\mathcal{L}^\exists$.

Proof. If θ is not contingent, then the lemma holds trivially.

Now assume that θ is contingent and $B_\dagger(\theta) + B_\dagger(\neg\theta) < 1$.

<u>Case 1</u> There exist contingent $(\varphi)_{i \in I}, (\psi)_{j \in J} \in S\mathcal{L}^\exists$ such that

$$\{\theta\} \cup \bigcup_{i \in I}\{\varphi_i\} \in \Pi_\mathcal{L}$$
$$\{\neg\theta\} \cup \bigcup_{j \in J}\{\psi_j\} \in \Pi_\mathcal{L}$$

with

$$B_\dagger(\theta) + \sum_{i \in I} B_\dagger(\varphi_i) = 1$$
$$B_\dagger(\neg\theta) + \sum_{j \in J} B_\dagger(\psi_j) = 1.$$

Note that $\bigcup_{i \in I}\{\varphi_i\} \cup \bigcup_{j \in J}\{\psi_j\} \in \Pi_{\mathcal{L}}$ and thus $\sum_{i \in I} B_{\dagger}(\varphi_i) + \sum_{j \in J} B_{\dagger}(\psi_j) \leq 1$. Adding the above equations we now obtain

$$2 = B_{\dagger}(\theta) + \sum_{i \in I} B_{\dagger}(\varphi_i) + B_{\dagger}(\neg\theta) + \sum_{j \in J} B_{\dagger}(\psi_j)$$

$$\leq B_{\dagger}(\theta) + B_{\dagger}(\neg\theta) + 1.$$

$B_{\dagger}(\theta) + B_{\dagger}(\neg\theta) \geq 1$ follows. Contradiction.

<u>Case 2</u> For all $\pi \in \Pi_{\mathcal{L}}$ with $\theta \in \pi$ and all $\pi' \in \Pi_{\mathcal{L}}$ with $\neg\theta \in \pi'_{\mathcal{L}}$ it holds that $\sum_{\varphi \in \pi} B_{\dagger}(\varphi) < 1$ and $\sum_{\psi \in \pi'} B_{\dagger}(\psi) < 1$.

Applying (*) we obtain a contradiction since $B_{\dagger}(\theta)$ or $B_{\dagger}(\neg\theta)$ could have been set to a greater number.

<u>Case 3</u> For all $\pi \in \Pi_{\mathcal{L}}$ with $\theta \in \pi$ it holds that $\sum_{\psi \in \pi} B_{\dagger}(\psi) < 1$ and there exists a partition $\pi' \in \Pi_{\mathcal{L}}$ with $\neg\theta \in \pi'_{\mathcal{L}}$ such that $\sum_{\varphi \in \pi'} B_{\dagger}(\varphi) = 1$.

Let π' comprise of contingent $(\varphi_i)_{i \in I}$ and $\neg\theta$. For $\pi \in \Pi_{\mathcal{L}}$ with $\theta \in \pi$ we have for all finite $J \subseteq I$ that

$$\bigcup_{j \in J}\{\varphi_j\} \cup \{\theta \wedge \neg \bigvee_{j \in J} \varphi_j\} \cup \{\psi \in \pi \ : \ \psi \neq \theta\} \in \Pi_{\mathcal{L}}.$$

In the same manner as in the proof of Lemma 10 it follows that $B_{\dagger}(\theta) \geq \sum_{j \in J} B_{\dagger}(\varphi_j)$. Since this holds for all finite $J \subseteq I$ and I can be at most countable, it follows that $B_{\dagger}(\theta) \geq \sum_{i \in I} B_{\dagger}(\varphi_i)$.

From $B_{\dagger}(\neg\theta) + \sum_{i \in I} B_{\dagger}(\varphi_i) = \sum_{\varphi \in \pi'} B_{\dagger}(\varphi) = 1$ the required contradiction follows:

$$B_{\dagger}(\theta) + B_{\dagger}(\neg\theta) \geq \sum_{i \in I} B_{\dagger}(\varphi_i) + B_{\dagger}(\neg\theta)$$

$$= 1.$$

$\square$

(*) is not strong enough to uniquely determine constrain $B_{\dagger}$ on $\mathcal{L}^{\exists}$. We invoke the following further desideratum to pin down $B_{\dagger}$: ceteris paribus, prefer belief function B to belief function B' if B gives greater degree of belief to some universally quantified sentence than does B'. One has to be a bit careful about how one formulates such a principle, in order to specify it in such a way that it can be applied consistently. One can appeal to the concept of prenex normal form in order to formulate this desideratum:

($\forall$*) Suppose that neither of B, B' have a better loss profile on $\mathcal{L}^{\exists}$ than the other. Furthermore, suppose there exists a minimal quantifier rank q such that the following hold: For all $\varphi \in S\mathcal{L}^{\exists}$ in prenex normal form with a quantifier rank of $q-1$ or less it holds that $B(\varphi) = B(\varphi')$ and for all universally quantified $\theta \in S\mathcal{L}^{\exists}$ in prenex normal form of quantifier rank q it holds that $B(\theta) \geq B'(\theta)$ and the inequality is strict at least once. Then B is to be preferred to B'.

The motivation behind ($\forall$*) is not in terms of loss. Rather, the motivation stems from the application to inductive logic (see Section 3.3). The use of probability in inductive logic has been roundly criticised for tending to give non-tautological universal laws probability zero, when such laws are widely—and seemingly rationally—believed in science and beyond; see, e.g., Popper [17] (Appendix *vii). Thus there seems good reason to prefer, ceteris paribus, those probability functions which give more credence to

universal hypotheses. (There is a flip-side to ($\forall$*). The more credence one gives to a universal statement $\forall x\theta(x)$, the less credence one must give to $\exists x\neg\theta(x)$. One might motivate the latter policy by appeal to Okham's Razor, which demands scepticism with respect to the existence of entities—particularly new kinds of entity.)

This leaves us with some desiderata that stem from considerations to do with loss, namely the criteria that make up Definition 21—appealing to dominance of loss, dominance of expected loss, and worst-case expected loss—and some desiderata that stem from the application to inductive logic, namely language invariance and ($\forall$*). These desiderata taken together are enough to justify the norms of objective Bayesianism on $\mathcal{L}^{\exists}$, as we shall proceed to show in the remainder of this section.

We shall see first that ($\forall$*) is responsible for ensuring that the degree of belief $B(\forall x\theta(x))$, which is already constrained to $[0, \inf_{n\in\mathbb{N}} \bigwedge_{i=1}^{n} B(\theta(t_i))]$, is equal to the upper bound. On the other hand, $B(\exists x\theta(x))$ comes out to be $\sup_{n\in\mathbb{N}} \bigvee_{i=1}^{n} B(\theta(t_i))$. An arbitrary belief function $B_{\dagger} \in \mathrm{minloss}^* \mathbb{B}_{\mathcal{L}}$ which is also optimal according to ($\forall$*) will be denoted by $B_{\dagger}^{\forall}$.

Proposition 22. *For all universally quantified sentences $\forall x\theta(x) \in \mathcal{L}^{\exists}$ it holds that $B_{\dagger}^{\forall}(\forall x\theta(x)) = \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigwedge_{i=1}^{n} \theta(t_i))$.*

Proof. First note that $\forall x\theta(x) \models \bigwedge_{i=1}^{n} \theta(t_i)$ for all $n \in \mathbb{N}$ and we thus obtain from Lemma 10 that $B_{\dagger}^{\forall}(\forall x\theta(x)) \leq \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigwedge_{i=1}^{n} \theta(t_i))$.

We now prove by an argument on quantifier ranks that

$$B_{\dagger}^{\forall}(\forall x\theta(x)) = \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigwedge_{i=1}^{n} \theta(t_i)).$$

Assume for contradiction that there exists a minimal quantifier rank $q \geq 1$ and a sentence $\forall x\psi(x)$ in prenex normal form of quantifier rank q such that $B_{\dagger}^{\forall}(\forall x\psi(x)) < \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigwedge_{i=1}^{n} \psi(t_i))$.

We now define a function B' which will be preferred to $B_{\dagger}^{\forall}$ which contradicts our standing assumption that no function is preferred to $B_{\dagger}^{\forall}$. Let $B'(\chi) := B_{\dagger}^{\forall}(\chi)$ for all sentences $\chi \in S\mathcal{L}^{\exists}$ which are in prenex normal form and have a quantifier rank of $q - 1$ or less. In particular, $B_{\dagger}^{\forall}$ and B' agree on $\mathcal{L}^{\overline{\exists}}$.

For all $\varphi(x) \in \mathcal{L}^{\exists}$ in prenex normal form of quantifier rank $q - 1$ we let

$$B'(\forall x\varphi(x)) := \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigwedge_{i=1}^{n} \varphi(t_i))$$

and

$$B'(\exists x\neg\varphi(x)) := \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigvee_{i=1}^{n} \neg\varphi(t_i)).$$

Now arbitrarily extend B' to a function in $\mathbb{B}_{\mathcal{L}}$.

Note that $B'(\forall x\psi(x)) > B_{\dagger}^{\forall}(\forall x\psi(x))$ and $B'(\exists x\neg\psi(x)) < B_{\dagger}^{\forall}(\exists x\neg\psi(x))$. So, (*) does not discriminate between $B_{\dagger}^{\forall}$ and B'. Hence, $B_{\dagger}^{\forall}$ and B' are equally preferable according to $\prec^*$.

$B_{\dagger}^{\forall}$ and B' agree on all sentences in prenex normal form of quantifier rank $q-1$. Since $B_{\dagger}^{\forall}(\forall x\varphi(x)) \leq \lim_{n\longrightarrow\infty} B_{\dagger}^{\forall}(\bigwedge_{i=1}^{n} \varphi(t_i))$ has to hold for all $\varphi(x) \in \mathcal{L}^{\exists}$ it follows that for $\varphi(x)$ in prenex normal form of quantifier rank $q-1$ that $B_{\dagger}^{\forall}(\forall x\varphi(x)) \leq B'(\forall x\varphi(x))$ and for $\forall x\varphi(x) = \psi$ the inequality is sharp. ($\forall$*) now implies that B' is preferred to $B_{\dagger}^{\forall}$.

Finally, every sentence of the form $\forall x \theta(x)$ is logically equivalent to a universally quantified sentence $\varphi = \forall x \varphi(x)$ in prenex normal. Note that $\theta(t)$ is logically equivalent to $\varphi(t)$ for all constants t. Hence,

$$
\begin{aligned}
B_\dagger^\forall(\forall x \theta(x)) &= B_\dagger^\forall(\forall x \varphi(x)) \\
&= \lim_{n \longrightarrow \infty} B_\dagger^\forall\Big(\bigwedge_{i=1}^n \varphi(t_i)\Big) \\
&= \lim_{n \longrightarrow \infty} B_\dagger^\forall\Big(\bigwedge_{i=1}^n \theta(t_i)\Big).
\end{aligned}
$$

$\square$

Proposition 23. $B_\dagger^\forall$ *satisfies the axiom P3.*

Proof. Applying Lemma 11, Proposition 22 and applying Lemma 11 a second time we find

$$
\begin{aligned}
B_\dagger^\forall(\exists x \theta(x)) &= 1 - B_\dagger^\forall(\forall x \neg \theta(x)) \\
&= 1 - \lim_{n \longrightarrow \infty} B_\dagger^\forall\Big(\bigwedge_{i=1}^n \neg\theta(t_i)\Big) \\
&= 1 - \lim_{n \longrightarrow \infty} \Big(1 - B_\dagger^\forall\Big(\bigvee_{i=1}^n \theta(t_i)\Big)\Big) \\
&= \lim_{n \longrightarrow \infty} B_\dagger^\forall\Big(\bigvee_{i=1}^n \theta(t_i)\Big).
\end{aligned}
$$

$\square$

The following might be of interest outside the context of this paper since it generalises Gaifman's Theorem, [5] (Theorem 1).

Proposition 24. *If* $f : S\mathcal{L}^\exists \longrightarrow [0,1]$ *satisfies*

- $f(\theta) = 1$ *for all tautologies* $\theta \in S\mathcal{L}^{\bar{\exists}}$ *–[P1 on* $\mathcal{L}^{\bar{\exists}}$*],*

- *for all mutually exclusive* $\theta, \varphi \in S\mathcal{L}^{\bar{\exists}}$ *it holds that* $f(\theta \vee \varphi) = f(\theta) + f(\varphi)$ *– [P2 on* $\mathcal{L}^{\bar{\exists}}$*],*

- $f(\exists x \theta(x)) = \sup_m P(\bigvee_{i=1}^m \theta(t_i))$ *for all* $\exists x \theta(x) \in S\mathcal{L}^\exists$ *and – [P3]*

- f *respects logical equivalence on* $\mathcal{L}^\exists$ *– [P4],*

then f *is a probability function,* i.e., $f \in \mathbb{P}_\mathcal{L}$.

Clearly, P1 on $\mathcal{L}^{\bar{\exists}}$ and P4 jointly imply P1.

Proof. First note that f agrees with some probability function on the quantifier free sentences of $\mathcal{L}$. By Gaifman's Theorem, this probability function is unique on $\mathcal{L}^\exists$; it shall be denoted by P_f.

We now show that $f = P_f$. We need to show that for all $\varphi \in S\mathcal{L}^\exists$ that $f(\varphi) = P_f(\varphi)$.

First, write φ in prenex normal form, φ_{pre}. Note that $f(\varphi) = f(\varphi_{pre})$.

Next, we do a proof by induction on the quantifier-block rank of φ_{pre} to show that $f(\varphi_{pre}) = P_f(\varphi_{pre})$. The quantifier-block rank of φ_{pre} is the number of alternating quantifier blocks in φ_{pre}.

<u>Base case</u> φ_{pre} is of quantifier block rank zero, *i.e.,* φ_{pre} does not contain quantifiers. Then

$$
\begin{aligned}
f(\varphi) &= f(\varphi_{pre}) \\
&= P_f(\varphi_{pre}) \\
&= P_f(\varphi),
\end{aligned}
$$

where the second equation holds since f and P_f agree on all sentences of $\mathcal{L}^{\exists}$. The first and the last equation hold since f and P_f respect logical equivalence on $\mathcal{L}^{\exists}$. This fact will be used without further mention.

<u>Inductive step</u> φ_{pre} is of quantifier block rank $q \geq 1$.

Let us first suppose that $\varphi_{pre} = \exists \bar{x}\chi(\bar{x})$. For $q \geq 2$ the first symbol of χ is a universal quantifier, $\forall$, for $q = 1$, the first symbol of χ is a relation symbol, a negation symbol or an opening bracket. We find for $q = 1$

$$
\begin{aligned}
f(\varphi) &= f(\varphi_{pre}) \\
&= f(\exists \bar{x}\chi(\bar{x})) \\
&\overset{P3}{=} \lim_{n_1 \to \infty} \dots \lim_{n_k \to \infty} f\Big(\bigvee_{i_1=1}^{n_1} \dots \bigvee_{i_k=1}^{n_k} \chi(t_{i_1}, \dots, t_{i_k}) \Big) \\
&= \lim_{n_1 \to \infty} \dots \lim_{n_k \to \infty} P_f\Big(\bigvee_{i_1=1}^{n_1} \dots \bigvee_{i_k=1}^{n_k} \chi(t_{i_1}, \dots, t_{i_k}) \Big) \\
&\overset{P3}{=} P_f(\exists \bar{x}\chi(\bar{x})) \\
&= P_f(\varphi_{pre}) \\
&= P_f(\varphi),
\end{aligned}
$$

where we may substitute P_f for f since χ is quantifier-free and we can thus apply the induction hypothesis.

For $q \geq 2$ $\varphi_{pre} = \exists \bar{x}_1 \forall \bar{x}_2 \dots Q\bar{x}_q \chi(\bar{x}_1, \bar{x}_2, \dots)$, where $Q = \exists$ for odd q and $Q = \forall$ for even q.

First, here is an example of two logically equivalent sentences:

$$
\models \Big(\bigvee_{i=1}^{2} \forall x_1 \exists x_2 U x_1 x_2 t_i \Big) \leftrightarrow \Big(\forall y_1^1 \forall y_2^1 \exists y_1^2 \exists y_2^2 \bigvee_{i=1}^{2} U y_i^1 y_i^2 t_i \Big).
$$

Note that the quantifier block rank on of the sentence on the right of "$\leftrightarrow$" is two. The quantifier block rank has been kept low at the price of larger blocks of quantifiers. Since we are giving a proof by induction on the quantifier block rank, we do not have to worry about paying this price. To denote the larger blocks we will use $\bar{y}$. In general, the greater the number of variables and on the left of an $\bar{x}_i$, the greater the number of variables in $\bar{y}_i$.

Now let us compute

$$
\begin{aligned}
f(\varphi) &= f(\varphi_{pre}) \\
&= f(\exists \bar{x}_1 \forall \bar{x}_2 \ldots Q\bar{x}_q \chi(\bar{x}_1, \bar{x}_2, \ldots \bar{x}_q)) \\
&\overset{P3}{=} \lim_{n_1 \longrightarrow \infty} \ldots \lim_{n_k \longrightarrow \infty} f\Big(\bigvee_{i_1=1}^{n_1} \ldots \bigvee_{i_k=1}^{n_k} \forall \bar{x}_2 \ldots Q\bar{x}_q \chi(t_{i_1}, \ldots, t_{i_k}, \bar{x}_2, \ldots \bar{x}_q) \Big) \\
&= \lim_{n_1 \longrightarrow \infty} \ldots \lim_{n_k \longrightarrow \infty} f\Big(\forall \bar{y}_2 \ldots Q\bar{y}_q \bigvee_{i_1=1}^{n_1} \ldots \bigvee_{i_k=1}^{n_k} \chi(t_{i_1}, \ldots, t_{i_k}, \bar{y}_2, \ldots, \bar{y}_q) \Big) \\
&\overset{IH}{=} \lim_{n_1 \longrightarrow \infty} \ldots \lim_{n_k \longrightarrow \infty} P_f\Big(\forall \bar{y}_2 \ldots Q\bar{y}_q \bigvee_{i_1=1}^{n_1} \ldots \bigvee_{i_k=1}^{n_k} \chi(t_{i_1}, \ldots, t_{i_k}, \bar{y}_2, \ldots, \bar{y}_q) \Big) \\
&= \lim_{n_1 \longrightarrow \infty} \ldots \lim_{n_k \longrightarrow \infty} P_f\Big(\bigvee_{i_1=1}^{n_1} \ldots \bigvee_{i_k=1}^{n_k} \forall \bar{x}_2 \ldots Q\bar{x}_q \chi(t_{i_1}, \ldots, t_{i_k}, \bar{x}_2, \ldots \bar{x}_q) \Big) \\
&\overset{P3}{=} P_f(\exists \bar{x}_1 \forall \bar{x}_2 \ldots Q\bar{x}_q \chi(\bar{x}_1, \bar{x}_2, \ldots \bar{x}_q)) \\
&= P_f(\varphi_{pre}) \\
&= P_f(\varphi).
\end{aligned}
$$

"I H" indicates that we used the induction hypothesis on a sentence of quantifier rank $q - 1$.

The case of $\varphi_{pre} = \forall x \chi(x)$ is analogous, simply replace the disjunctions by conjunctions. $\square$

Theorem 8. *If $\mathbb{E}_{\mathcal{L}}$ is finitely generated and g is regular, then*

$$
\{B_\dagger^\forall\} = \mathrm{maxent}\, \mathbb{E}_{\mathcal{L}} = \{P_\Omega^\dagger\}.
$$

Proof. By Proposition 24 we only need to convince ourselves that $B_\dagger^\forall$ satisfies P1 on $\mathcal{L}^{\bar{\exists}}$, P2 on $\mathcal{L}^{\bar{\exists}}$, P3 and P4 in order to conclude that $B_\dagger^\forall \in \mathbb{P}_{\mathcal{L}}$. Note that we have done so in Theorem 6, Proposition 23 and Corollary 7. So all $B_\dagger^\forall$ are probability functions.

All $B_\dagger^\forall$ agree on $\mathcal{L}^{\bar{\exists}}$ with $P_\Omega^\dagger$. Two different probability functions have to disagree on a quantifier-free sentence (Gaifman's theorem). Hence, $B_\dagger^\forall$ is a unique and equal to $P_\Omega^\dagger$. $\square$

We should point out that ($\forall$*) was only used in Proposition 23. We showed that (*) alone is enough to force that $B_\dagger$ satisfies P1, P2 on $\mathcal{L}^{\bar{\exists}}$, $B_\dagger(\exists x \theta(x)) \geq \lim_{n \longrightarrow \infty} B_\dagger(\bigvee_{i=1}^n \theta(t_i))$ and P4.

In sum, then, by adding invoking two new considerations, (*) and ($\forall$*), one can show that the Probability norm must hold on a predicate language with quantifiers. Since the Calibration and Equivocation norms are already forced on the quantifier-free sentences, and probabilities on these quantifier-free sentences determine those of the quantified sentences, all the norms of objective Bayesianism hold on $\mathcal{L}^\exists$, assuming that the weighting function is regular and the evidence is finitely generated.

6. More Complex Evidence

The question arises as to which functions have an optimal loss profile when $\mathbb{E}_{\mathcal{L}}$ is not finitely generated. In Section 6.2 we shall present a tractable case and show that in that example the function

with maximal standard entropy has the best loss profile. First, in Section 6.1, we shall see that not all examples admit of such an analysis. In particular, we shall analyse an example in some depth in which $\{P^\dagger\} = \mathrm{maxent}\,\mathbb{E}_\mathcal{L}$ but $P^\dagger \notin \mathrm{minloss}\,\mathbb{B}_\mathcal{L}$. Thus, when evidence is not finitely generated, the optimal loss profile may not be achievable by maximising entropy.

6.1. When Losses Cannot Be Minimised

We shall now develop an example in which the minimax theorem fails: $P^\dagger_\Omega \notin \mathrm{minloss}\,\mathbb{B}_\mathcal{L}$, as we shall see in Proposition 27. However, the entropy identity, $\mathbb{P}^\dagger = \{P^\dagger_\Omega\} = \mathrm{maxent}\,\mathbb{E}_\mathcal{L}$, does hold (Proposition 25 and Proposition 26). The connection with optimal loss fails to obtain since $\mathrm{minloss}\,\mathbb{B}_\mathcal{L} = \emptyset$ (Proposition 30). Thus, there is no belief function with an optimal loss profile in this sort of example. Nevertheless, certain equivocal functions $\bar{P}^\dagger_N$ derived from the maximal entropy function come arbitrarily close to having the best loss profile (Proposition 29 and Proposition 31). So, while there is no unique function with the best loss profile, the functions $\bar{P}^\dagger_N$ have a very good loss profile.

In the following discussion we shall focus on the most simple possible language, $\mathcal{L} = \mathcal{L}^U$, which contains only one relation symbol, U, which is unary. We focus on this simple language since the minimax results already fail here and considering more expressive languages does not lead to new insights while creating more notational issues. As a technical convenience, we extend the notion of a loss profile to arbitrary functions $f : S\mathcal{L} \longrightarrow [0, 1]$, not merely normalised belief functions.

The example that we shall consider is generated by the following evidence:

$$E = \{\neg U_1 t_i \to \neg U_1 t_1 \,:\, i = 1, 2, \ldots\}.$$

Let $\omega^n_k \in \Omega_n$ be the k-th n-state of $\mathcal{L} = \mathcal{L}^U$, *i.e.*, $\omega^n_1 := \bigwedge^n_{l=1} \neg U t_l$, $\omega^n_2 := \bigwedge^{n-1}_{l=1} \neg U t_l \wedge U t_n, \ldots$, $\omega^n_{2^n} := \bigwedge^n_{l=1} U t_l$. The set of calibrated probability functions can be characterised in various ways:

$$
\begin{aligned}
\mathbb{E}_\mathcal{L} &= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\neg U_1 t_i \to \neg U_1 t_1) = 1, i = 1, 2, \ldots\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\omega^{n+1}_1) = P(\omega^n_1) \text{ for all } n \geq 1\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\omega^n_i) = 0 \text{ for } 2 \leq i \leq 2^{n-1},\, n = 1, 2, \ldots\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\omega^n_2) = 0 \text{ for all } n \geq 2\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\neg U t_1 \wedge U t_n) = 0 \text{ for all } n \geq 2\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\omega^{n+1}_1 | \omega^n_1) = 1 \text{ for all } n \in \mathbb{N}\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\omega^{n+1}_2 | \omega^n_1) = 0 \text{ for all } n \in \mathbb{N}\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\forall x(U t_1 \vee \neg U x)) = 1\} \\
&= \{P \in \mathbb{P}_\mathcal{L} \,:\, P(\exists x(\neg U t_1 \wedge U x)) = 0\}
\end{aligned}
$$

The last two characterisations employ quantifiers; adding quantifiers to the language enables a finite representation of what is essentially an infinitely generated evidence set. Hence in Definition 5, we specified that an evidence set is finitely generated just if it generated by *quantifier-free* sentences of some finite sublanguage.

We now begin our analysis of this example:

Proposition 25. *If* $g = g_\Omega$ *or if* g *is symmetric and inclusive, then* $\mathbb{P}^\dagger = \{P_\Omega^\dagger\}$ *and* $P_\Omega^\dagger$ *is not open-minded.*

Proof. For all $n \in \mathbb{N}$

$$\mathbb{E}_n = \{P \in \mathbb{P}_{\mathcal{L}_n^{\exists}} \ : \ P(\omega_i^n) = 0 \text{ for all } 2 \leq i \leq 2^{n-1}\}.$$

Then, by Landes and Williamson [4] (Corollary 6, p. 3574) for symmetric and inclusive g

$$P_n^\dagger(\omega_1^n) = P_n^\dagger(\omega_i^n) = \frac{1}{2^{n-1} + 1} \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n$$

and so for all $n \in \mathbb{N}$ and all $1 \leq i \leq 2^{n-1}$

$$\begin{aligned}
\lim_{n \to \infty} P_n^\dagger(\omega_1^n) &= \lim_{n \to \infty} P_n^\dagger(\omega_1^1) \\
&= \lim_{n \to \infty} \frac{1}{2^{n-1} + 1} \\
&= 0.
\end{aligned}$$

For all $n \in \mathbb{N}$ and $i \in \{1, 2^{n-1} + 1, \ldots, 2^n\}$

$$P^\dagger(\omega_i^n) = \frac{1}{2^{n-1}}.$$

The result for $g = g_\Omega$ follows in the same way as above. $\quad\square$

We shall note for later reference that for all $n \geq 2$

$$H_\Omega^n(P_n^\dagger) = -\log \frac{1}{2^{n-1} + 1} > -\log \frac{1}{2^{n-1}} = H_\Omega^n(P_\Omega^\dagger).$$

Proposition 26. *If* $g = g_\Omega$ *or if* g *is regular, then*

$$\operatorname{maxent} \mathbb{E}_\mathcal{L} = \{P_\Omega^\dagger\}.$$

Proof. First note that $[\mathbb{E}_\mathcal{L}] = \mathbb{E}_\mathcal{L}$.

We shall show that for all $Q \in \mathbb{E}_\mathcal{L} \setminus \{P_\Omega^\dagger\}$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ we have $H_\Omega^n(P_\Omega^\dagger) > H_\Omega^n(Q)$ and $H_g^n(P_\Omega^\dagger) > H_g^n(Q)$.

Since $Q \neq P_\Omega^\dagger$ there exists a minimal $k \in \mathbb{N}$ and a k-state $\nu \in \Omega_k$ such that $Q(\nu) > P_\Omega^\dagger(\nu) \geq 0$.

<u>Case 1</u> $\nu = \omega_1^k$.

To simplify notation let $\alpha := P_k^\dagger(\nu) = P_k^\dagger(\omega_1^n) > 0$ for all $n \geq 1$. Let us now define a function $Q' \in \mathbb{E}_\mathcal{L} \setminus \{P_\Omega^\dagger\}$. Note that since we want Q' to be a member of $\mathbb{E}_\mathcal{L}$ we need to let $Q'(\omega_1^n) := Q'(\omega_1^1)$ for all $n \in \mathbb{N}$. Now let for all $n \in \mathbb{N}$

$$\begin{aligned}
Q'(\omega_1^n) &:= \alpha > 0 \\
Q'(\omega_i^n) &:= 0 \text{ for } 2 \leq i \leq 2^{n-1} \\
Q'(\omega_i^n) &:= \frac{1-\alpha}{2^{n-1}} \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n.
\end{aligned}$$

The restriction operator $_{\downarrow n}$ applied to some belief function B continuous to refer to the restriction of B to $\mathcal{L}_n^{\exists}$, rather than to the restriction to $\mathcal{L}_n$.

Note that for all $n \geq 1$

$$\{Q'_{\downarrow n}\} = \arg \sup_{\substack{P \in \mathbb{E}_n \\ P(\omega_1^n) = \alpha}} H_g^n(P)$$

since entropy maximisers assign n-states the same degree of belief whenever possible [4] (Corollary 7, p. 3577). Thus, $H_g^n(Q') \geq H_g^n(Q)$ for all $n \in \mathbb{N}$. Also, $H_\Omega^n(Q') \geq H_\Omega^n(Q)$ for all $n \in \mathbb{N}$.

Let us compute for $n \geq k$

$$\begin{aligned}
H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q') &= -\log(\frac{1}{2^{n-1}}) - [-\alpha \log(\alpha) - (1-\alpha) \log(\frac{1-\alpha}{2^{n-1}})] \\
&= \log(2^{n-1}) + \alpha \log(\alpha) + (1-\alpha)(\log(1-\alpha) - \log(2^{n-1})) \\
&= \log(2^{n-1}) - (1-\alpha) \log(2^{n-1}) + \alpha \log(\alpha) + (1-\alpha) \log(1-\alpha) \\
&= \alpha(n-1) \log(2) + \alpha \log(\alpha) + (1-\alpha) \log(1-\alpha).
\end{aligned}$$

It follows that for all large enough $n \in \mathbb{N}$ that $H_\Omega^n(P_\Omega^\dagger) > H_\Omega^n(Q') \geq H_\Omega^n(Q)$.

For regular g we now find

$$\begin{aligned}
H_g^n(P_\Omega^\dagger) - H_g^n(Q') &= g(\pi^n) \cdot [\alpha(n-1) \log(2) + \alpha \log(\alpha) + (1-\alpha) \log(1-\alpha)] \\
&\quad - \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) \sum_{F \in \pi} {}^\circ P_\Omega^\dagger(F) \log({}^\circ P_\Omega^\dagger(F)) - {}^\circ Q'(F) \log({}^\circ Q'(F)).
\end{aligned}$$

So, as long as $\sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi)$ goes to zero quickly enough it follows that $H_g^n(P_\Omega^\dagger) > H_g^n(Q') \geq H_g^n(Q)$ for large enough n. Corollary 6 shows that this is indeed the case for regular g.

<u>Case 2</u> $\nu \in \{\omega_2^k, \ldots, \omega_{2^{k-1}}^k\}$. Since Q is assumed to be calibrated, $Q \in \mathbb{E}_\mathcal{L}$, this case cannot occur.

<u>Case 3</u> $\nu \in \{\omega_{2^{k-1}+1}^k, \ldots, \omega_{2^k}^k\}$.

<u>Case 3A</u> $Q(\omega_1^k) = 0$.

Then $Q(\omega_1^1) = 0$. But for all $n \in \mathbb{N}$

$$\arg \sup_{\substack{P \in \mathbb{E}_n \\ P(\omega_1^1) = 0}} H_\Omega^n(P) = \{P_{\Omega \downarrow n}^\dagger\} = \arg \sup_{\substack{P \in \mathbb{E}_n \\ P(\omega_1^1) = 0}} H_g^n(P).$$

Since $Q \neq P_\Omega^\dagger$ it follows that there exists some $N \in \mathbb{N}$ such that $Q_{\downarrow n} \neq P_{\Omega \downarrow n}^\dagger$ for all $n \geq N$. But then $H_g^n(P_\Omega^\dagger) > H_g^n(Q)$ and $H_\Omega^n(P_\Omega^\dagger) > H_\Omega^n(Q)$ for all $n \geq N$.

<u>Case B</u> $Q(\omega_1^k) > 0$.

Then $Q(\omega_1^k) > 0 = P_\Omega^\dagger(\omega_1^k)$. Proceed as in Case 1. $\square$

Proposition 27. *If $g = g_\Omega$ or if g is regular, then $P_\Omega^\dagger \notin \mathrm{minloss}\, \mathbb{B}_\mathcal{L}$.*

Proof. We here show that there exists an $R \in \mathbb{E}_\mathcal{L}$ such that for all $n \in \mathbb{N}$ it holds that $S_g^n(R, P^\dagger) = S_\Omega^n(R, P^\dagger) = \infty$ and that there exists an open-minded $Q \in \mathbb{E}_\mathcal{L}$ such that for all $n \in \mathbb{N}$ we have $\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, Q) \leq \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, Q) < \infty$.

Note that the probability function $R \in \mathbb{P}_\mathcal{L}$ with $R(\omega_1^n) := 1$. Then $S_g^n(R, P_\Omega^\dagger) = S_\Omega^n(R, P_\Omega^\dagger) = \infty$ for all $n \in \mathbb{N}$.

We shall now construct an open-minded $Q \in \mathbb{E}_\mathcal{L}$ as advertised. For all $n \in \mathbb{N}$ let

$$Q(\omega_1^n) := \frac{1}{2}$$
$$Q(\omega_i^n) := 0 \text{ for all } 2 \leq i \leq 2^{n-1}$$
$$Q(\omega_i^n) := \frac{1}{2^n} \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n.$$

Thus, Q is open-minded and hence $\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, Q) \leq \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, Q) < +\infty$ for all $n \in \mathbb{N}$. $\qquad \square$

Note that Condition 1 of Definition 21 is solely responsible for the fact that $P_\Omega^\dagger \notin \mathrm{minloss}\,\mathbb{B}_\mathcal{L}$. Condition 2 has played no role here.

So far, we have established that $P^\dagger = P_\Omega^\dagger$ does not have the best loss profile. The question arises whether there exists a belief function $B \in \mathbb{B}_\mathcal{L}$ which is a minimal element of $\prec$, *i.e.*, $B \in \mathrm{minloss}\,\mathbb{B}_\mathcal{L}$.

Proposition 28. *If $g = g_\Omega$, then*

$$\mathrm{minloss}\,\mathbb{B}_\mathcal{L} = \emptyset = \mathrm{minloss}\,\mathbb{P}_\mathcal{L} = \mathrm{minloss}\,\mathbb{E}_\mathcal{L}.$$

Initially, one might suspect that $\mathrm{minloss}\,\mathbb{B}_\mathcal{L} = \emptyset$ would be somehow due to the fact that the S_Ω^n do not take beliefs in all sentences into account. This is not the case. As we will see, $\mathrm{minloss}\,\mathbb{P}_\mathcal{L} = \emptyset = \mathrm{minloss}\,\mathbb{E}_\mathcal{L}$ holds. That is, even when restricting attention to probability functions, whose values on the n-states completely determine degrees of beliefs in all other sentences, we cannot find a function with an optimal loss profile.

Proof. Suppose for contradiction that $Q \in \mathrm{minloss}\,\mathbb{P}_\mathcal{L} \setminus \{P_\Omega^\dagger\}$.

If Q is not open-minded, then there exists an $N \in \mathbb{N}$, an $F \subseteq \Omega_N$ and an $P \in \mathbb{E}_\mathcal{L}$ such that ${}^\circ P(F) > 0$ and ${}^\circ Q(F) = 0$. But then there has to exists some $\omega \in \Omega_N$ with $\omega \in F$ such that $P(\omega) > 0 = Q(\omega)$ since Q and P are probability functions. Thus, for all $n \geq N$ there exists some $\nu \in \Omega_n$ such that $\nu \vDash \omega$ with $P(\nu) > 0 = Q(\nu)$. But then $S_\Omega^n(P, Q) = +\infty$ for all $n \geq N$.

In the proof of Proposition 27 we constructed an open-minded function $Q^+ \in \mathbb{E}_\mathcal{L}$. For Q^+ we have for all n that $\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, Q^+) < +\infty$. So, any $Q \in \mathrm{minloss}\,\mathbb{P}_\mathcal{L}$ has to be open-minded.

<u>Case 1</u> $Q \in \mathrm{minloss}\,\mathbb{P}_\mathcal{L}$ and $Q \notin \mathbb{E}_\mathcal{L}$.

Since $Q \in \mathbb{P}_\mathcal{L} \setminus \mathbb{E}_\mathcal{L}$ there has to exist a minimal $k \geq 2$ such that $Q(\omega_2^k) > 0$.

We next define a probability function $Q' \in \mathbb{E}_\mathcal{L}$ with the following construction for all $n \geq 2$

$$Q'(\omega_i^l) := Q(\omega_i^l) \text{ for all } 1 \leq l \leq k-1 \text{ and all } i$$
$$Q'(\omega_1^n) := Q(\omega_1^k) \text{ for all } n \in \mathbb{N}$$
$$Q'(\omega_i^n) := 0 \text{ for all } n \geq k \text{ and all } 2 \leq i \leq 2^{n-1}$$
$$Q'(\omega_i^n) := Q(\omega_i^n) \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n.$$

It follows that for all $n \geq k$ and all $\omega \in \Omega_n \setminus \{\omega_2^n, \ldots, \omega_{2^{n-1}}^n\}$ and all $P \in \mathbb{E}_n$ such that $P(\omega) > 0$ it holds that $Q(\omega) \leq Q'(\omega)$. For all large enough $n \in \mathbb{N}$ we then find

$$\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, Q) \geq -\log \min_{2^{n-1}+1 \leq i \leq 2^n} \{Q(\omega)\}$$
$$= -\log \min_{2^{n-1}+1 \leq i \leq 2^n} \{Q'(\omega)\}$$
$$= \sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, Q').$$

Hence, there has to exists a $Q' \in \mathbb{E}_{\mathcal{L}} \cap \operatorname{minloss} \mathbb{P}_{\mathcal{L}}$ with $Q' \neq P_\Omega^\dagger$.

<u>Case 2</u> $Q \in \operatorname{minloss} \mathbb{P}_{\mathcal{L}}$ and $Q \in \mathbb{E}_{\mathcal{L}} \setminus \{P_\Omega^\dagger\}$.

Thus, $0 < Q(\omega_1^1) = Q(\omega_1^n)$ for all $n \geq 2$. Let $N \geq 3$ be such that $Q(\omega_1^N) > \min\{Q(\omega_{2^{N-1}+1}^N), \ldots, Q(\omega_{2^N}^N)\}$. For $n \geq N$ let

$$\Omega_n^- := \arg\min\{Q(\omega_{2^{n-1}+1}^n), \ldots, Q(\omega_{2^n}^n)\} \subset \Omega_n.$$

We now find for all fixed $n \geq N$ that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, Q) = -\log Q(\omega_-^n) \text{ for all } \omega_-^n \in \Omega_n^-.$$

We shall now define a function $R \in \mathbb{E}_{\mathcal{L}} \setminus \{P_\Omega^\dagger\}$ by letting for all $n \geq 2$:

$$R(\omega_1^n) := \frac{Q(\omega_1^n)}{2} = \frac{Q(\omega_1^1)}{2}$$
$$R(\omega_i^n) := 0 \text{ for all } 2 \leq i \leq 2^{n-1}$$
$$R(\omega_i^n) := Q(\omega_i^n) + \frac{Q(\omega_1^1)}{2} \frac{2}{|\Omega_n|} > Q(\omega_i^n) \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n.$$

That is, $R = \frac{Q + P_\Omega^\dagger}{2}$.

For large enough $M \in \mathbb{N}$ it holds for all $n \geq M$ that

$$R(\omega_1^n) > \min\{R(\omega_{2^{n-1}+1}^n), \ldots, R(\omega_{2^n}^n)\}.$$

Furthermore, for all $n \geq \max\{M, N\}$ it holds that

$$\arg\min\{Q(\omega_{2^{n-1}+1}^n), \ldots, Q(\omega_{2^n}^n)\} = \arg\min\{R(\omega_{2^{n-1}+1}^n), \ldots, R(\omega_{2^n}^n)\}$$

and hence for all large enough fixed $n \in \mathbb{N}$ and all $\omega_-^n \in \Omega_n^-$

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, R) = -\log R(\omega_-^n) < -\log Q(\omega_-^n) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, Q).$$

Thus, R has a better loss profile than Q. Hence, $Q \notin \operatorname{minloss} \mathbb{P}_{\mathcal{L}}$ and $Q \notin \operatorname{minloss} \mathbb{E}_{\mathcal{L}}$.

Finally, let us consider loss profiles for $B \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$.

<u>Case 3</u> $B \in \operatorname{minloss} \mathbb{B}_{\mathcal{L}}$ and $B \notin \mathbb{P}_{\mathcal{L}}$.

For all $P \in \mathbb{P}_{\mathcal{L}}$, the expression $S_\Omega^n(P, B)$ only depends on the degrees of belief B assigns to sentences which represent an n-state. So, the degree of belief in a sentence $\varphi \in S\mathcal{L}$ which does not n-represent an n-state are ignored by $S_\Omega^n(P, B)$ for all n and all $P \in \mathbb{P}_{\mathcal{L}}$. If B agrees with some probability function $P \in \mathbb{P}_{\mathcal{L}}$ on all sentences of $S\mathcal{L}^{\nexists}$ which n-represent an n-state, then B and P are equally preferable according to $\prec$. As we saw above, for all $P \in \mathbb{P}_{\mathcal{L}}$ there exists some $Q \in \mathbb{P}_{\mathcal{L}}$ with $Q \prec P$. Thus, B cannot be a minimal element of $\prec$.

We can hence assume that for all $P \in \mathbb{P}_{\mathcal{L}}$ there exists some sentence $\varphi \in S\mathcal{L}_n^{\nexists}$ which n-represents an n-state such that $B(\varphi) \neq P(\varphi)$. Since no $P \in \mathbb{P}_{\mathcal{L}}$ is dominated, it follows that $B(\varphi) < P(\varphi)$.

First define a function B_0 as follows:

$$B_0(\varphi) := \inf_{\substack{\rho \in \varrho_n \\ \rho\omega = \varphi}} B(\rho\omega), \text{ if such an } n \in \mathbb{N} \text{ and such a } \rho \in \varrho_n \text{ exist,}$$

$$B_0(\varphi) := 0 \text{ otherwise.}$$

B_0, which does not agree with any probability function on $\mathcal{L}^{\vec{\exists}}$ has been constructed in such a way that B and B_0 are equally preferred according to $\prec$.

Next define a function B^+ by first letting for all fixed $N \in \mathbb{N}$

$$B^+(\varphi) := \sup_{n \geq N} \sum_{\substack{\nu \in \Omega_n \\ \nu \models \omega}} B_0(\nu)$$

for all sentences $\varphi \in S\mathcal{L}_N^{\exists}$ which are logically equivalent to an N-state. Put $B^+(\psi) := 0$ for all other $\psi \in S\mathcal{L}^{\exists}$.

Since B^+ dominates B_0 the loss profile of B^+ cannot be worse than that of B_0. Furthermore, note that for all $N \in \mathbb{N}$, all $\omega \in \Omega_N$ and all $n > N$ it holds that

$$B^+(\omega) \geq \sum_{\substack{\nu \in \Omega_n \\ \nu \models \omega}} B^+(\nu).$$

Let $\alpha := \lim_{n \longrightarrow \infty} \sum_{\omega \in \Omega_n} B^+(\omega)$. For $\alpha = 0$ it follows by the usual reasoning that B^+ cannot have an ideal loss profile. This leads to a contradiction in the usual way.

For $1 \geq \alpha > 0$ define a function B^∞ by first letting for all sentences $\varphi \in S\mathcal{L}^{\exists}$ which are logically equivalent to some n-state ω

$$B^\infty(\varphi) := \frac{1}{\alpha} \lim_{n \longrightarrow \infty} \sum_{\substack{\nu \in \Omega_n \\ \nu \models \varphi}} B^+(\nu).$$

For all other sentences $\varphi \in S\mathcal{L}^{\exists}$ let $B^\infty(\varphi) := 0$.

Observe that for all $k \in \mathbb{N}$ and all $\omega \in \Omega_k$

$$B^\infty(\omega) = \frac{1}{\alpha} \lim_{n \longrightarrow \infty} \sum_{\substack{\nu \in \Omega_n \\ \nu \models \omega}} B^+(\nu)$$

$$= \frac{1}{\alpha} \lim_{n \longrightarrow \infty} \sum_{\substack{\lambda \in \Omega_{k+1} \\ \lambda \models \omega}} \sum_{\substack{\nu \in \Omega_n \\ \nu \models \lambda}} B^+(\nu)$$

$$= \sum_{\substack{\lambda \in \Omega_{k+1} \\ \lambda \models \omega}} B^\infty(\lambda).$$

Finally, we note that B^∞ agrees with some $P \in \mathbb{P}_\mathcal{L}$ on all sentences in $S\mathcal{L}^{\exists}$ which represent a state. Then B cannot have a better loss profile than P. As we saw in Case1 and Case2, for all $P \in \mathbb{P}_\mathcal{L}$ there exists a $Q \in \mathbb{P}_\mathcal{L}$ which has a strictly better loss profile than P. This contradicts $B \in \mathrm{minloss}\,\mathbb{B}_\mathcal{L}$. $\square$

Denote by $\bar{P}_N^\dagger$ the unique probability function in $\mathbb{E}_\mathcal{L}$ satisfying for all $n \in \mathbb{N}$

$$\bar{P}_N^\dagger(\omega_1^n) = P_N^\dagger(\omega_1^n) = P_N^\dagger(\omega_1^1) = \frac{1}{2^{N-1} + 1}$$

$$\bar{P}_N^\dagger(\omega_i^n) = 0 \text{ for all } 2 \leq i \leq 2^{n-1}$$

$$\bar{P}_N^\dagger(\omega_i^n) = \left(1 - \frac{1}{2^{N-1} + 1}\right) \cdot \frac{2}{|\Omega_n|} = \frac{1}{\frac{|\Omega_N|}{2} + 1} \frac{|\Omega_N|}{|\Omega_n|} \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n.$$

That is, $\bar{P}_N^\dagger$ agrees with $\bar{P}_N^\dagger$ on $\mathcal{L}_N$ and equivocates beyond $\mathcal{L}_N$ as much as possible while satisfying $\bar{P}_N^\dagger \in \mathbb{E}_\mathcal{L}$.

Proposition 29. *For all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$*

$$\sup_{P \in \mathbb{E}_\mathcal{L}} S^n_\Omega(P, \bar{P}^\dagger_N) - \sup_{P \in \mathbb{E}_\mathcal{L}} S^n_\Omega(P, P^\dagger_n) \leq \epsilon.$$

Proof. For all large enough $N \in \mathbb{N}$ and even larger $n \in \mathbb{N}$ we find

$$0 \leq \sup_{P \in \mathbb{E}_\mathcal{L}} S^n_\Omega(P, \bar{P}^\dagger_N) - \sup_{P \in \mathbb{E}_\mathcal{L}} S^n_\Omega(P, P^\dagger_n)$$

$$= -\log\Big(\frac{\frac{|\Omega_N|}{2}}{\frac{|\Omega_N|}{2} + 1} \frac{2}{|\Omega_n|} \Big) + \log\Big(\frac{1}{\frac{2^n}{2} + 1} \Big)$$

$$= -\log\Big(\frac{2^{N-1}}{2^{N-1} + 1} \Big) + \log\Big(\frac{2^n}{2} \Big) + \log\big(\frac{1}{2^{n-1} + 1} \big)$$

$$= -\log\Big(\frac{2^{N-1}}{2^{N-1} + 1} \Big) + \log\Big(\frac{2^{n-1}}{2^{n-1} + 1} \Big).$$

For $\epsilon > 0$ let $N > 2$ be such that $0 < -\log \frac{2^{N-1}}{2^{N-1}+1} < \epsilon$. Then for all $n \geq N$ it holds that $0 > \log \frac{2^{n-1}}{2^{n-1}+1} > \log \frac{2^{N-1}}{2^{N-1}+1}$. For $n \geq N$ large enough we now obtain

$$0 \leq \sup_{P \in \mathbb{E}_\mathcal{L}} S^n_\Omega(P, \bar{P}^\dagger_N) - \sup_{P \in \mathbb{E}_\mathcal{L}} S^n_\Omega(P, P^\dagger_n)$$

$$= -\log\Big(\frac{2^{N-1}}{2^{N-1} + 1} \Big) + \log\Big(\frac{2^{n-1}}{2^{n-1} + 1} \Big)$$

$$< \epsilon.$$

□

Having considered loss for $g = g_\Omega$ we now investigate loss for regular g.

Proposition 30. *If g is regular, then* $\mathrm{minloss}\,\mathbb{B}_\mathcal{L} = \emptyset$.

Proof. We will show that $\prec$ has no minimal element. Suppose for contradiction that $B \in \mathbb{B}_\mathcal{L}$ is such a minimal element.

Define a function $B' : S\mathcal{L} \longrightarrow [0, 1]$ by

$$B'(\varphi) := 0 \text{ for all } \varphi \in S\mathcal{L} \text{ for which there exists an } n \in \mathbb{N} \text{ with } \bigvee_{i=2}^{2^{n-1}} \omega^n_i \vDash \varphi$$

$$B'(\psi) := B(\psi) \text{ else.}$$

B' and B are equally preferable according to $\prec$ since $P(\varphi) = 0$ for all $P \in \mathbb{E}_\mathcal{L}$ and all such φ.

For all $\varphi \in S\mathcal{L}^\exists$ let n_φ be the minimal n such that $\varphi \in S\mathcal{L}^\exists_{n_\varphi}$. Now define a function B^inf by first letting

$$B^\mathrm{inf}(\varphi) = \inf_{\substack{\psi \in S\mathcal{L}^\exists_{n_\varphi} \\ \vDash \varphi \leftrightarrow \psi}} B'(\psi).$$

Put $B^\mathrm{inf}(\varphi) := B'(\varphi)$ for all other $\varphi \in S\mathcal{L}^\exists$. For all $\varphi \in S\mathcal{L}$ it holds that $B^\mathrm{inf}(\varphi) \leq B(\varphi)$. Furthermore, B^inf is equally preferable to B' according to $\prec$. We now consider cases to show that there is a function with a strictly better loss profile than B^inf, which contradicts our assumption that $B \in \mathrm{minloss}\,\mathbb{B}_\mathcal{L}$.

<u>Case A</u> There exists some $N \in \mathbb{N}$ such that for all $n \geq N$, B^{inf} and $P_\Omega^\dagger$ agree on all n-states.

Since $B \neq P_\Omega^\dagger$ it holds that $B^{\text{inf}} \neq P_\Omega^\dagger$ and hence $B_{\frac{1}{2}} := \frac{B^{\text{inf}} + P_\Omega^\dagger}{2} \neq P_\Omega^\dagger$. Thus, for all $n \geq N$ $B_{\frac{1}{2}}$ and $P_\Omega^\dagger$ agree on all n-states. But then for all $n \geq N$ all $F \subseteq \Omega_n$ and all $\rho \in \varrho_n$ $B^{\text{inf}}(\rho F) \leq B_{\frac{1}{2}}(\rho F)$. Hence, for all $P \in \mathbb{P}_\mathcal{L}$ it holds that $S_g^n(P, B_{\frac{1}{2}}) \leq S_g^n(P, B^{\text{inf}})$.

From the above we have that for all $n \geq N$ there exists an $F \subseteq \Omega_n$ such that $F \setminus \{\omega_2^N, \ldots, \omega_{2^{N-1}}^N\} \neq \emptyset$ and such that $B^{\text{inf}}(\rho F) < B_{\frac{1}{2}}(\rho F)$ for some ρ. Thus, there exists some $P \in \mathbb{E}_\mathcal{L}$ with $^\circ P(F) > 0$. Then $S_g^n(P, B_{\frac{1}{2}}) < S_g^n(P, B^{\text{inf}})$ for this $P \in \mathbb{P}_\mathcal{L}$ and all $n \geq N$.

Thus, $B_{\frac{1}{2}} \prec B^{\text{inf}}$ by Condition 2 of Definition 21.

<u>Case B</u> There exist infinitely many $n \in \mathbb{N}$ where B^{inf} and $P_\Omega^\dagger$ agree on all n-states and infinitely many $n \in \mathbb{N}$ many where they do not agree on all n-states.

Since $P_\Omega^\dagger$ is a probability function it follows that for all $n \in \mathbb{N}$, all $F \subseteq \Omega_n$ and all $\rho \in \varrho_n$ $B^{\text{inf}}(\rho F) \leq P_\Omega^\dagger(\rho F)$ has to hold. Now proceed as in Case A.

<u>Case C</u> The number of $n \in \mathbb{N}$ for which B^{inf} and $P_\Omega^\dagger$ agree on all n-states is finite (possibly zero).

<u>Case C1</u> There exists an infinite set $J \subseteq \mathbb{N}$, $J = \{j_1, j_2, \ldots\}$, such that $\lim_{i \to \infty} \sum_{\omega \in \Omega_{j_i}} B^{\text{inf}}(\omega) = 1$. If $P_\Omega^\dagger$ dominates B^{inf}, we are done.

If $P_\Omega^\dagger$ does not dominate B^{inf}, then define a function $B_1 \in \mathbb{P}_\mathcal{L}$ by letting for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$

$$^\circ B_1(F) := \lim_{i \to \infty} \sum_{\substack{\omega \in \Omega_{j_i} \\ \omega \in F}} B^{\text{inf}}(\omega)$$

and requiring that B_1 satisfies logical equivalence on $\mathcal{L}^{\nexists}$. For all $\varphi \in S\mathcal{L}^{\exists} \setminus S\mathcal{L}^{\nexists}$ use Gaifman's condition to ensure that B_1 is a probability function.

Since we assumed that $P_\Omega^\dagger$ does not dominate B^{inf} $B_1 \neq P_\Omega^\dagger$ holds. Furthermore, B_1 dominates B^{inf}. So, the loss profile of $B_1 \in \mathbb{P}_\mathcal{L}$ is at least equally good as that of B.

We complete this proof by showing that $\mathbb{P}_\mathcal{L} \cap \text{minloss}\, \mathbb{B}_\mathcal{L} = \emptyset$.

Now suppose for contradiction that there exists a function $Q \in \mathbb{P}_\mathcal{L} \cap \text{minloss}\, \mathbb{B}_\mathcal{L}$ such that $Q(\omega_2^n) > 0$ for some $n \geq 2$, *i.e.*, $Q \notin \mathbb{E}_\mathcal{L}$. It needs to hold that $Q(\omega_1^n) > 0$ for all $n \in \mathbb{N}$ (open-mindedness).

Let $k \geq 2$ be minimal such that $Q(\omega_2^k) > 0$. Now define a function $R \in \mathbb{P}_\mathcal{L}$ by letting for all $n > k$

$$R(\omega_i^k) := \frac{Q(\omega_i^k) + P_\Omega^\dagger(\omega_i^k)}{2} \text{ for all } 1 \leq i \leq 2^k$$

$$R(\omega_1^n) := R(\omega_1^k) = \frac{Q(\omega_1^k)}{2} = \frac{Q(\omega_1^k) + P_\Omega^\dagger(\omega_1^k)}{2} \text{ for all } n > k$$

$$R(\omega_{2^{n-k}+1}^n) := \frac{Q(\omega_2^k) + P_\Omega^\dagger(\omega_2^k)}{2} = \frac{Q(\omega_2^k)}{2} > 0$$

$$R(\omega_i^n) := \frac{Q(\nu) + P_\Omega^\dagger(\nu)}{2} \text{ for all } 2^{n-1} + 1 \leq i \leq 2^n \text{ where } \nu \in \Omega_k \text{ with } \omega_i^n \vDash \nu$$

$$R(\omega_i^n) := 0 \text{ otherwise.}$$

That is, R is the arithmetic mean of Q and $P_\Omega^\dagger$ on $\mathcal{L}_k$. Beyond $\mathcal{L}_k$, R equivocates under the k-states which imply Ut_1. For such n-states $R(\omega_i^n) = \frac{Q(\nu) + \frac{1}{2^{n-1}}}{2}$ holds. Beyond $\mathcal{L}_k$, there are only two n-states which imply $\neg Ut_1$ which are assigned non-zero probability, w_1^n and $w_{2^{n-k}+1}^n$.

We now show that R has a strictly better loss profile than Q what contradicts $Q \in \text{minloss}\, \mathbb{B}_\mathcal{L}$.

Let $\nu_k^- \in \arg\min_{\omega \in \{\omega_{2^{k-1}+1}^k, \ldots, \omega_{2^k}^k\}} Q(\omega)$. Trivially, $0 < Q(\nu_k^-) < \frac{1}{2^{k-1}}$. Next note that for all $n \geq k$ which are large enough it holds that

$$\min_{\omega \in \{\omega_1^n, \omega_{2^{n-1}+1}^n, \ldots, \omega_{2^n}^n\}} R(\omega) = \frac{\frac{1}{2^{k-1}} + Q(\nu_k^-)}{2} \cdot \frac{|\Omega_k|}{|\Omega_n|}$$

and that

$$\min_{\omega \in \{\omega_{2^{n-1}+1}^n, \ldots, \omega_{2^n}^n\}} Q(\omega) \leq Q(\nu_k^-) \cdot \frac{|\Omega_k|}{|\Omega_n|}.$$

We now find for all large enough $n > k$ that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, Q) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, R) \geq g(\pi^n) \log\Big(Q(\nu_k^-) \cdot \frac{|\Omega_k|}{|\Omega_n|} \Big) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, R)$$

$$\geq g(\pi^n)\Big(-\log\Big(Q(\nu_k^-) \cdot \frac{|\Omega_k|}{|\Omega_n|} \Big) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, R) \Big)$$

$$- \sup_{P \in \mathbb{E}_{\mathcal{L}}} - \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ R(F).$$

Whenever ${}^\circ P(F) > 0$ with $F \subseteq \Omega_n$, then ${}^\circ R(F)$ is bounded from below by $\frac{1}{2^n}$. Hence, the last term in the above sum converges to zero, since g is regular.

We now obtain the contradiction as follows: there exists some $\epsilon > 0$ such that for all large enough $n \geq k$ it holds that

$$-\log\Big(Q(\nu_k^-) \cdot \frac{|\Omega_k|}{|\Omega_n|} \Big) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, R)$$

$$= -\log\Big(Q(\nu_k^-) \cdot \frac{|\Omega_k|}{|\Omega_n|} \Big) + \log\Big(\frac{\frac{1}{2^{k-1}} + Q(\nu_k^-)}{2} \cdot \frac{|\Omega_k|}{|\Omega_n|} \Big)$$

$$= \log\Big(\frac{\frac{1}{2^{k-1}} + Q(\nu_k^-)}{2} \Big) - \log Q(\nu_k^-)$$

$$\geq \epsilon.$$

We have thus shown that if $\mathbb{P}_{\mathcal{L}} \cap \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}} \neq \emptyset$, then there exists some $Q \in \mathbb{E}_{\mathcal{L}} \cap \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$.

<u>Case C1A</u> $Q(\omega_1^1) = 0$. Then Q has infinite worst-case expected loss for all $n \in \mathbb{N}$ and we are done.

<u>Case C1B</u> $Q(\omega_1^1) > 0$.

By open-mindedness, $Q(\omega_1^1) < 1$ has to hold.

For all $n \in \mathbb{N}$ let $\omega_-^n \in \arg\min_{\omega \in \{\omega_{2^{n-1}+1}^n, \ldots, \omega_{2^n}^n\}} Q(\omega)$. From $Q \in \mathbb{E}_{\mathcal{L}}$ we now obtain that for all large enough n there exists a probability function $R \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, Q)$ such that $R(\omega_-^n) = 1$.

Next, define a probability function $Q' \in \mathbb{E}_{\mathcal{L}}$ where $Q'(\omega_1^n) := Q'(\omega_1^1) := Q(\omega_1^1)$ and Q' equivocates over Ut_1, $Q'(\omega_i^n) := Q(\omega_2^1)\frac{|\Omega_1|}{|\Omega_n|}$ for all $n \in \mathbb{N}$ and for all $2^{n-1} + 1 \leq i \leq 2^n$. Assume for contradiction that $Q \neq Q'$.

We next show that $Q' \prec Q$. This contradicts $Q \in \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$. To this end let us note that for all large enough n

$$\sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,Q') \leq \sup_{P\in\mathbb{E}_{\mathcal{L}}} g(\pi^n) S_\Omega^n(P,Q') + \sup_{P\in\mathbb{E}_{\mathcal{L}}} - \sum_{\pi\in\Pi_n\setminus\{\pi^n\}} g(\pi) \sum_{F\in\pi} {}^\circ\!P(F) \log {}^\circ\!Q'(F)$$

$$\leq -g(\pi^n) \log Q'(\omega_{2^n}^n) + \sup_{P\in\mathbb{E}_{\mathcal{L}}} - \sum_{\pi\in\Pi_n\setminus\{\pi^n\}} g(\pi) \sum_{F\in\pi} {}^\circ\!P(F) \log\, Q'(\omega_{2^n}^n)$$

$$= -g(\pi^n) \log Q'(\omega_{2^n}^n) - \log\!\Big(Q'(\omega_2^1)\frac{|\Omega_1|}{|\Omega_n|}\Big) \cdot \sum_{\pi\in\Pi_n\setminus\{\pi^n\}} g(\pi).$$

Since whenever ${}^\circ\!P(F) > 0$, then ${}^\circ\!Q'(F)$ is bounded from below by $Q'(\omega_{2^n}^n)$.

Thus, for all large enough n we have

$$0 \leq \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,Q') - g(\pi^n) \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_\Omega^n(P,Q')$$

$$\leq -\log\!\Big(Q'(\omega_2^1)\frac{|\Omega_1|}{|\Omega_n|}\Big) \cdot \sum_{\pi\in\Pi_n\setminus\{\pi^n\}} g(\pi).$$

g is regular, hence, this last term converges to zero. We thus obtain

$$\lim_{n\longrightarrow\infty} \sup_{P\in\mathbb{E}_{\mathcal{L}}} g(\pi^n) S_\Omega^n(P,Q') - \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,Q') = 0. \tag{17}$$

Since $Q \neq Q'$, $Q, Q' \in \mathbb{E}_{\mathcal{L}}$ and $Q(\omega_1^1) = Q'(\omega_1^1)$, there has to exist some minimal $k \in \mathbb{N}$ and a minimal $i \geq 2^{k-1}+1$ such that $Q(\omega_i^k) < Q'(\omega_i^k)$. We now find for all large enough n that

$$\sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,Q) - g(\pi^n) \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_\Omega^n(P,Q') \geq g(\pi^n)\cdot\Big(\sup_{P\in\mathbb{E}_{\mathcal{L}}} S_\Omega^n(P,Q) - \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_\Omega^n(P,Q')\Big)$$

$$\geq g(\pi^n)\cdot\Big(-\log Q(\omega_-^n) + \log Q'(\omega_{2^n}^n)\Big)$$

$$\geq g(\pi^n)\cdot\Big(-\log\!\Big(Q(\omega_i^k)\frac{|\Omega_k|}{|\Omega_n|}\Big) + \log\!\Big(Q'(\omega_{2^i}^k)\frac{|\Omega_k|}{|\Omega_n|}\Big)\Big)$$

$$\geq g(\pi^n)\cdot\Big(-\log Q(\omega_i^k) + \log Q'(\omega_i^k)\Big)$$

$$> 0.$$

Recall that there exists $0 < a \leq b$ such that for all $n \in \mathbb{N}$ $a \leq g(\pi^n) \leq b$ holds. Hence, there exists some constant $c > 0$ such that $g(\pi^n)(-\log Q(\omega_i^k) + \log Q'(\omega_i^k)) \geq c > 0$. From (17) we conclude that for all large enough n

$$\sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,Q) - \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,Q') > 0$$

holds. Thus, $Q' \prec Q$. So, $Q \notin \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$.

To complete the proof of Case C1B we show that there exists some $N \in \mathbb{N}$ such that $\bar{P}_N^\dagger$ has a strictly better loss profile than Q'.

Let $N \in \mathbb{N}$ be such that $\bar{P}_N^\dagger(\omega_1^1) < Q'(\omega_1^1)$. Analogous to the above it holds that

$$\lim_{n\longrightarrow\infty} \sup_{P\in\mathbb{E}_{\mathcal{L}}} S_g^n(P,\bar{P}_N^\dagger) - \sup_{P\in\mathbb{E}_{\mathcal{L}}} g(\pi^n) S_\Omega^n(P,\bar{P}_N^\dagger) = 0. \tag{18}$$

It hence suffices to show that there exists some $\epsilon > 0$ such that for large enough $N \in \mathbb{N}$ and all $n \geq N$

$$g(\pi^n) \cdot \left(\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, Q') - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, \bar{P}_N^{\dagger}) \right) > \epsilon.$$

We now recall that $Q'(\omega_1^1) > \bar{P}_N^{\dagger}(\omega_1^1)$. The required inequality follows for large enough $n \in \mathbb{N}$

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, Q') - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, \bar{P}_N^{\dagger}) = -\log(\frac{1 - Q'(\omega_1^1)}{2^{n-1}}) + \log(\frac{1 - \bar{P}_N^{\dagger}(\omega_1^1)}{2^{n-1}})$$

$$> \epsilon.$$

Hence, $\bar{P}_N^{\dagger} \prec Q'$.

<u>Case C2</u> There exist an $\alpha > 0$ and an minimal N_1 such that for all $n \geq N_1$ $\sum_{\omega \in \Omega_n} B^{\inf}(\omega) \leq 1 - \alpha$ holds.

We may assume that $B^{\inf}$ is open-minded on $\mathcal{L}^{\nexists}$. Thus there has to exist some minimal $N \geq N_1$ such that $0 < P_n^{\dagger}(\omega_1^1) < B^{\inf}(\omega_1^1)$ for all $n \geq N$. For all large enough $n \geq N$ we now find

$$\frac{1}{g(\pi^n)} \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\inf}) \geq \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, B^{\inf})$$

$$= \max_{\omega \in \Omega_n \setminus \{\omega_{2^{n-1}+1}^n, \ldots, \omega_{2^n}^n\}} -\log B^{\inf}(\omega)$$

$$= -\log(\min_{\omega \in \Omega_n \setminus \{\omega_{2^{n-1}+1}^n, \ldots, \omega_{2^n}^n\}} B^{\inf}(\omega))$$

$$\geq -\log \frac{1 - \alpha - B^{\inf}(\omega_1^1)}{2^{n-1}}.$$

Using (18) we find for all large enough $n \in \mathbb{N}$

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B^{\inf}) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, \bar{P}_N^{\dagger}) \geq g(\pi^n) \cdot \left(-\log \frac{1 - \alpha - B^{\inf}(\omega_1^1)}{2^{n-1}} + \log \frac{1 - P_N^{\dagger}(\omega_1^1)}{2^{n-1}} \right)$$

$$+ \left(\log(|\Omega_n|) - \log(|\Omega_N|) - \log(P_N^{\dagger}(\omega_{2^N}^N)) \right) \cdot \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi)$$

$$> 0.$$

$\square$

Proposition 31. *For all regular g and all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$*

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, \bar{P}_N^{\dagger}) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_n^{\dagger}) \leq \epsilon.$$

Proof. Let $\epsilon > 0$ be fixed. By (18) it suffices to show that there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that

$$0 \leq \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, \bar{P}_N^{\dagger}) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_n^{\dagger})$$

$$\leq g(\pi^n) \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, \bar{P}_N^{\dagger}) - g(\pi^n) \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, P_n^{\dagger})$$

$$\leq \epsilon.$$

Now simply note that we have proved this already in Proposition 29. $\square$

Hence, for all $\epsilon > 0$ there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ and all $Q \in \mathbb{B}_{\mathcal{L}}$

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, \bar{P}_N^\dagger) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, Q) \leq \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, \bar{P}_N^\dagger) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_n^\dagger)$$

$$\leq \epsilon.$$

Although, $\bar{P}_N^\dagger$ is not a minimal element of $\prec$, the losses incurred by adopting any other $B \in \mathbb{B}_{\mathcal{L}}$ can only be marginally better, eventually.

Thus, for fixed k and $\delta > 0$ there exists an $N \in \mathbb{N}$ such that for all $\varphi \in S\mathcal{L}_k$ $|\bar{P}_N^\dagger(\varphi) - P_\Omega^\dagger(\varphi)| < \delta$. Hence, belief functions with an arbitrarily good loss can be found within an (Euclidean) neighbourhood of $P_\Omega^\dagger$.

Since the $\bar{P}_N^\dagger$ are probability functions, there does not exist a $B \in \mathbb{B}_{\mathcal{L}}$ which dominates $\bar{P}_N^\dagger$ on $\mathcal{L}^\exists$ or on $\mathcal{L}^{\not\exists}$. Furthermore, the $\bar{P}_N^\dagger$ are optimal according to $(\forall*)$. The $\bar{P}_N^\dagger$ thus are *almost* optimal in all the senses we here considered.

In essence, the phenomenon of $\mathrm{minloss}\,\mathbb{B}_{\mathcal{L}} = \emptyset$ arises from $\bar{P}_{n+1}^\dagger$ having a strictly better loss profile than $\bar{P}_n^\dagger$ but the limit of the sequence $(\bar{P}_n^\dagger)_{n \in \mathbb{N}}$ is $P_\Omega^\dagger$, which is not open-minded. This phenomenon is reminiscent of $\min\{x \in \mathbb{R} \; : \; 0 < x < 1\} = \emptyset$, where it is possible to get ever closer to zero but it is impossible to reach it.

6.2. When Losses Can Be Minimised

The analysis of Section 6.1, shows that there can be no general minimax theorem which covers any evidence that is not finitely generated. On the other hand, we shall see in this section that for certain natural cases evidence which cannot be finitely generated, minimax theorems do obtain.

Let $\mathcal{L}$ contain only one m-ary relation symbol, U, and $c \in [0, 1]$. Let $\nu_1^n :=$ $\bigwedge_{1 \leq i_1, \ldots, i_m \leq n} \neg U t_{i_1} t_{i_2} \ldots t_{i_m} \in \Omega_n$ and let $\nu_2^n, \ldots, \nu_{|\Omega_n|}^n$ be an enumeration of the remaining n-states. We shall consider the following example:

$$\mathbb{E}_{\mathcal{L}} = \{P \in \mathbb{P}_{\mathcal{L}} \; : \; \lim_{n \longrightarrow \infty} P(\nu_1^n) = c\}$$

$$= \{P \in \mathbb{P}_{\mathcal{L}} \; : \; P(\forall x_1 x_2 \ldots x_m \neg U x_1 x_2 \ldots x_m) = c\}.$$

Slightly less general versions of $\mathbb{E}_{\mathcal{L}}$ have attracted recent interest in the literature [18] (Example 3, p. 95), [19] (Example 3.5, p. 172) and [1] (Example 5.7, p. 99). We here consider relations symbols U of arbitrary arity, while previously U was taken to be unary.

First of all, if $c = 0$ and g is symmetric and inclusive, then $P_= \in \mathbb{E}_{\mathcal{L}}$ and we immediately obtain that $P_= = P_\Omega^\dagger$ and $\{P_=\} = \mathrm{maxent}\,\mathbb{E}_{\mathcal{L}} = \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$.

We shall assume from now on that $c > 0$.

Proposition 32. *For symmetric and inclusive g it holds that $\mathbb{P}^\dagger = \{P^\dagger\} = \{P_\Omega^\dagger\}$ and $P_\Omega^\dagger(\nu_1^n) = c + \frac{1-c}{|\Omega_n|}$ and $P_\Omega^\dagger(\nu_i^n) = \frac{1-c}{|\Omega_n|}$ for all $n \in \mathbb{N}$ and all $1 \leq i \leq |\Omega_n|$.*

Proof. For all $n \geq 2$ and symmetric and inclusive g_n it holds that $P_n^\dagger(\nu_2^n) = P_n^\dagger(\nu_{2+i}^n)$ for all $1 \leq i \leq |\Omega_n| - 2$ by [4] (Corollary 7, p. 3577). Thus, there exists some $\lambda_n \geq 0$ such that $P_n^\dagger(\nu_1^n) = \lambda_n$ and $P_n^\dagger(\nu_k^n) = \frac{1-\lambda_n}{|\Omega_n|-1}$ for all $2 \leq k \leq |\Omega_n|$.

For all $n \in \mathbb{N}$, now define a function $P_1 \in [\mathbb{E}_n]$ by $P_1(\nu_1^n) := 1$. Then, define a convex combination of the equivocator on $\mathbb{E}_n$ and P_1 by $P_{\lambda_n} := \lambda_n P_1 + (1 - \lambda_n) P_{=\lfloor n}$. Recall that g_n is equivocator-preserving (Proposition 7) and that H_g^n is strictly concave on $\mathbb{P}_n$ (Lemma 1). Thus, $H_g^n(P_{\lambda_n}) > H_g^n(P_{\lambda_n'})$ for all $0 \le \lambda_n < \lambda_n' \le 1$.

On the one hand g-entropy strictly increases with decreasing λ_n on the other hand $P_n^\dagger \in [\mathbb{E}_{\mathcal{L}}]$ imposes the constraint $P_n^\dagger(\nu_1^n) \ge c$. Let $N \in \mathbb{N}$ be minimal with $|\Omega_N| > \frac{1}{c}$. Then for all $n \ge N$ it holds that $P_n^\dagger(\nu_1^n) = c$ and $P_n^\dagger(\nu_2^n) = P_n^\dagger(\nu_{2+i}^n) = \frac{1-c}{|\Omega_n|-1}$ for all $1 \le i \le |\Omega_n| - 2$.

For all $r \ge N$ it follows that

$$P_r^\dagger(\nu_2^N) = P_r^\dagger(\nu_2^r)\frac{|\Omega_r|}{|\Omega_N|} = \frac{1-c}{|\Omega_r|-1}\frac{|\Omega_r|}{|\Omega_N|}$$
$$P_r^\dagger(\nu_1^N) = 1 - (|\Omega_N| - 1) \cdot P_r^\dagger(\nu_2^N).$$

Thus, for all $r \ge N$ we find

$$P^\dagger(\nu_2^r) = \lim_{n \to \infty} P_n^\dagger(\nu_2^r) = \frac{1-c}{|\Omega_r|}$$
$$P^\dagger(\nu_1^r) = \lim_{n \to \infty} P_n^\dagger(\nu_1^r)$$
$$= 1 - (|\Omega_r| - 1) P^\dagger(\nu_2^r)$$
$$= \frac{|\Omega_r| - (|\Omega_r| - 1)(1 - c)}{|\Omega_r|}$$
$$= c + \frac{1-c}{|\Omega_r|}$$
$$= c + P^\dagger(\nu_2^r).$$

Thus, for all $n \in \mathbb{N}$ $P^\dagger(\nu_1^n) = c + \frac{1-c}{|\Omega_n|}$ and $P^\dagger(\nu_2^n) = \frac{1-c}{|\Omega_n|}$.

We now show that $P^\dagger$ is indeed a probability function. We need to show that $\sum_{\substack{\nu \in \Omega_{n+1} \\ \nu \models \omega}} P^\dagger(\nu) = P^\dagger(\omega)$ for all $n \in \mathbb{N}$ and all $\omega \in \Omega_n$:

$$P^\dagger(\nu_i^n) = \frac{1-c}{|\Omega_n|} = \frac{|\Omega_{n+1}|}{|\Omega_n|}\frac{1-c}{|\Omega_{n+1}|} = \frac{|\Omega_{n+1}|}{|\Omega_n|}P^\dagger(\nu_i^{n+1}) \text{ for all } 2 \le i \le |\Omega_n|$$
$$P^\dagger(\nu_1^n) = c + \frac{1-c}{|\Omega_n|}$$
$$= c + \frac{1-c}{|\Omega_{n+1}|} + (\frac{|\Omega_{n+1}|}{|\Omega_n|} - 1)\frac{1-c}{|\Omega_{n+1}|}$$
$$= P^\dagger(\nu_1^{n+1}) + (\frac{|\Omega_{n+1}|}{|\Omega_n|} - 1)P^\dagger(\nu_2^{n+1}).$$

Finally, observe that $\mathbb{P}_n^\dagger = \arg\sup_{P \in \mathbb{E}_n} H_\Omega^n(P)$. Hence, $\mathbb{P}^\dagger = \{P_\Omega^\dagger\}$. $\square$

Proposition 33. *If $g = g_\Omega$ or if g is regular, then* $\mathrm{maxent}\,\mathbb{E}_{\mathcal{L}} = \{P_\Omega^\dagger\}$.

Proof. Let $Q \in \mathbb{E}_{\mathcal{L}} \setminus \{P_\Omega^\dagger\}$. For regular g, it suffices to show that there exists an $N \in \mathbb{N}$ such that for all $n \ge N$ $H_g^n(Q) < H_g^n(P_\Omega^\dagger)$ holds.

Since $Q \ne P_\Omega^\dagger$ there has to exist a minimal $N \in \mathbb{N}$ and an N-state $\omega' \in \Omega_N \setminus \{\nu_1^N\}$ such that $Q(\omega') \ne P_\Omega^\dagger(\omega')$.

Now define a function $Q' : S\mathcal{L} \longrightarrow [0,1]$ by requiring that Q' respects logical equivalence, Q and Q' agree on $S\mathcal{L}_N$,

- $Q'(\nu') := \frac{|\Omega_N|}{|\Omega_n|} Q(\omega')$ for all $n > N$ all $\nu' \in \Omega_n$ with $\nu' \vDash \omega'$,

- $Q'(\nu_1^n) := Q(\nu_1^n)$ for all $n > N$ and

- $Q'(\nu) := \frac{1 - Q(\nu_1^n) - Q(\omega')}{|\Omega_n| - 1 - \frac{|\Omega_n|}{|\Omega_N|}}$ for all $n > N$ and all $\nu \in \Omega_n \setminus \{\nu_1^n\}$ with $\nu \nvDash \omega'$.

In general, Q' is not a probability function because

$$Q'(\nu_1^n) < \sum_{\substack{\omega \in \Omega_{n+1} \\ \omega \vDash \nu_1^n}} Q'(\omega).$$

Note that for all $n \geq N$ $H_\Omega^n(Q) \leq H_\Omega^n(Q')$ holds.

We now show that for all large enough n $H_\Omega^n(Q') < H_\Omega^n(P_\Omega^\dagger)$ holds. Let us first compute

$$
\begin{aligned}
-H_\Omega^n(Q') =\ & Q(\nu_1^n) \log(Q(\nu_1^n)) + \left(\frac{|\Omega_n|}{|\Omega_N|} \frac{|\Omega_N|}{|\Omega_n|} Q(\omega') \right) \log \frac{Q(\omega') \cdot |\Omega_N|}{|\Omega_n|} \\
& + (1 - Q(\nu_1^n) - Q(\omega')) \log \frac{1 - Q(\nu_1^n) - Q(\omega')}{|\Omega_n| - 1 - \frac{|\Omega_n|}{|\Omega_N|}} \\
=\ & Q(\nu_1^n) \log(Q(\nu_1^n)) + Q(\omega') \cdot \left(\log(Q(\omega')) + \log\left(\frac{|\Omega_N|}{|\Omega_n|} \right) \right) \\
& + (1 - Q(\nu_1^n) - Q(\omega')) \cdot \left(\log\left(\frac{1 - Q(\nu_1^n) - Q(\omega')}{|\Omega_N| - \frac{|\Omega_N|}{|\Omega_n|} - 1} \right) + \log\left(\frac{|\Omega_N|}{|\Omega_n|} \right) \right) \\
=\ & Q(\nu_1^n) \log(Q(\nu_1^n)) + Q(\omega') \log(Q(\omega')) + (1 - Q(\nu_1^n)) \log\left(\frac{|\Omega_N|}{|\Omega_n|} \right) \\
& + (1 - Q(\nu_1^n) - Q(\omega')) \cdot \log\left(\frac{1 - Q(\nu_1^n) - Q(\omega')}{|\Omega_N| - \frac{|\Omega_N|}{|\Omega_n|} - 1} \right).
\end{aligned}
$$

Since

$$H_\Omega^n(P_\Omega^\dagger) = -\left(c + \frac{1-c}{|\Omega_n|} \right) \log\left(c + \frac{1-c}{|\Omega_n|} \right) - (|\Omega_n| - 1) \frac{1-c}{|\Omega_n|} \log\left(\frac{1-c}{|\Omega_n|} \right)$$

we now find with $\lim_{n\longrightarrow\infty} Q'(\nu_1^n) = c$ that

$$\lim_{n\longrightarrow\infty} H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q') = -c\log(c) + \lim_{n\longrightarrow\infty}\left(-(|\Omega_n| - 1)\frac{1-c}{|\Omega_n|}\log(\frac{1-c}{|\Omega_n|})\right.$$

$$+ Q(\nu_1^n)\log(Q(\nu_1^n)) + Q(\omega')\log(Q(\omega')) + (1 - Q(\nu_1^n))(\log(\frac{|\Omega_N|}{|\Omega_n|}))$$

$$\left. + (1 - Q(\nu_1^n) - Q(\omega')) \cdot (\log(\frac{1 - Q(\nu_1^n) - Q(\omega')}{|\Omega_N| - \frac{|\Omega_N|}{|\Omega_n|} - 1})))\right)$$

$$= \cancel{-c\log(c)} - (1-c)\log(1-c) + \lim_{n\longrightarrow\infty}\left(\frac{(1-c)(|\Omega_n| - 1)}{|\Omega_n|}\log(|\Omega_n|)\right.$$

$$\cancel{+c\log(c)} + Q(\omega')\log(Q(\omega')) + (1 - Q(\nu_1^n))(\log(\frac{|\Omega_N|}{|\Omega_n|})))\Big)$$

$$+ (1 - c - Q(\omega')) \cdot (\log(\frac{1 - c - Q(\omega')}{|\Omega_N| - 1}))$$

$$= -(1-c)\log(1-c) + Q(\omega')\log(Q(\omega'))$$

$$+ (1 - c - Q(\omega')) \cdot (\log(\frac{1 - c - Q(\omega')}{|\Omega_N| - 1}))$$

$$+ \lim_{n\longrightarrow\infty}\left(\frac{(1-c)(|\Omega_n| - 1)}{|\Omega_n|}\log(|\Omega_n|)\right.$$

$$\left. + (1 - Q(\nu_1^n))(\log(|\Omega_N|) - \log(|\Omega_n|))\right)$$

$$= -(1-c)\log(\frac{1-c}{|\Omega_N|}) + Q(\omega')\log(Q(\omega'))$$

$$+ (1 - c - Q(\omega')) \cdot (\log(\frac{1 - c - Q(\omega')}{|\Omega_N| - 1}))$$

$$+ \lim_{n\longrightarrow\infty}\frac{(1-c)(|\Omega_n| - 1)}{|\Omega_n|}\log(|\Omega_n|) - (1 - Q(\nu_1^n))\log(|\Omega_n|)$$

$$\overset{Q(\nu_n^1)\geq c}{\geq} -(1-c)\log(\frac{1-c}{|\Omega_N|}) + Q(\omega')\log(Q(\omega'))$$

$$+ (1 - c - Q(\omega')) \cdot (\log(\frac{1 - c - Q(\omega')}{|\Omega_N| - 1}))$$

$$+ \lim_{n\longrightarrow\infty}\frac{(1-c)(|\Omega_n| - 1)}{|\Omega_n|}\log(|\Omega_n|) - (1 - c)\log(|\Omega_n|)$$

$$= -(1-c)\log(\frac{1-c}{|\Omega_N|}) + Q(\omega')\log(Q(\omega'))$$

$$+ (1 - c - Q(\omega')) \cdot (\log(\frac{1 - c - Q(\omega')}{|\Omega_N| - 1}))$$

$$+ (1-c)\lim_{n\longrightarrow\infty}(\frac{|\Omega_n| - 1}{|\Omega_n|} - 1)\log(|\Omega_n|)$$

$$= -(1-c)\log(\frac{1-c}{|\Omega_N|})$$

$$+ Q(\omega')\log(Q(\omega')) + (1 - c - Q(\omega')) \cdot (\log(\frac{1 - c - Q(\omega')}{|\Omega_N| - 1})).$$

Since $Q(\omega') \neq \frac{1-c}{|\Omega_N|}$ there exists some $\epsilon > 0$ such that for all large enough n

$$H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q') > \epsilon > 0.$$

This establishes the result for $g = g_\Omega$.

We now turn to regular g.

$$H_g^n(P_\Omega^\dagger) - H_g^n(Q) \geq H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q) - \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) \sum_{F \in \pi} {}^\circ Q(F) \log {}^\circ Q(F)$$

$$\geq H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q') - \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi) \sum_{F \in \pi} {}^\circ Q(F) \log {}^\circ Q(F).$$

The last sum goes to zero since g is regular, Corollary 6 . Eventually, $H_\Omega^n(P_\Omega^\dagger) - H_\Omega^n(Q')$ is greater some $\epsilon > 0$ as we established in the first part of the proof. Thus, for all large enough $n \in \mathbb{N}$ and all $Q \in \mathbb{E}_\mathcal{L} \setminus \{P^\dagger\}$ we have

$$H_g^n(P_\Omega^\dagger) - H_g^n(Q) > 0.$$

$\square$

Lemma 12. *The following three conditions are equivalent for all large enough $n \in \mathbb{N}$ and inclusive and symmetric g*

- $P' \in \arg\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger)$

- $P'(\nu_1^n) = c$

- $P' \in \arg\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, P_\Omega^\dagger)$.

Proof. Note that for all $P \in \mathbb{P}_\mathcal{L}$

$$S_g^n(P, P_\Omega^\dagger) = -\sum_{\pi \in \Pi_n} g(\pi) \sum_{F \in \pi} {}^\circ P(F) \log {}^\circ P_\Omega^\dagger(F)$$

$$= -\sum_{\nu \in \Omega_n} P(\nu) \sum_{\substack{F \subseteq \Omega_n \\ \nu \in F}} \gamma_n(F) \log {}^\circ P_\Omega^\dagger(F)$$

$$= -P(\nu_1)\left(\gamma_n(\nu_1) \log P_\Omega^\dagger(\nu_1) + \sum_{\substack{F \subseteq \Omega_n \\ \nu_1 \in F \\ |F| \geq 2}} \gamma_n(F) \log {}^\circ P_\Omega^\dagger(F)\right)$$

$$+ \sum_{i=2}^{|\Omega_n|} -P(\nu_i)\left(\gamma_n(\nu_i) \log P_\Omega^\dagger(\nu_i) + \sum_{\substack{F \subseteq \Omega_n \\ \nu_i \in F \\ |F| \geq 2}} \gamma_n(F) \log {}^\circ P_\Omega^\dagger(F)\right).$$

The term between the last set of brackets $\left(\right)$ does not depend on i. So, $S_g^n(P, P_\Omega^\dagger)$ only depends on $P(\nu_1)$ but not on how P distributes probabilities among the other n-states.

For large enough $N \in \mathbb{N}$ it holds that $P_\Omega^\dagger(\nu_1) > P_\Omega^\dagger(\nu_2) = P_\Omega^\dagger(\nu_i)$ for all $3 \leq i \leq |\Omega_n|$.

Since g is symmetric, $\gamma_n(F)$ is only a function of the size of F, $|F|$, it follows that every $P' \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_\Omega^\dagger)$ assigns as little probability as possible to ν_1. Since we require that $P \in \mathbb{E}_{\mathcal{L}}$ it follows that $P'(\nu_1) = c$.

The result for S_Ω^n follows as above by noting that for $g = g_\Omega$ it holds that $\gamma_n(\nu) = 1$ for all n-states $\nu \in \Omega_n$ and $\gamma_n(F) = 0$ otherwise. $\quad\square$

Adapting Joyce's notion of truth-directedness [14] we define:

Definition 26 (**Chance-directed scoring rule**). *A function $F^f : [0,1] \times [0,1] \longrightarrow [0,+\infty]$ of the form $F^f(x,y) = x \cdot f(y) + (1-x) \cdot f(1-y)$ is called* chance-directed, *if and only if for all $x \in [0,1]$, all $0 \leq \lambda < 1$ and all $y \in [0,1] \setminus \{x\}$*

$$
\begin{aligned}
F^f(x,y) &= x \cdot f(y) + (1-x) \cdot f(1-y) \\
&> x \cdot f((1-\lambda)x + \lambda y) + (1-x) \cdot f(1 - (1-\lambda)x - \lambda y) \\
&= F^f(x, (1-\lambda)x + \lambda y)
\end{aligned}
$$

holds. For a scoring rule F^f this formalises the idea that beliefs which are closer to the chances on two mutually exclusive and exhaustive events are strictly better scored.

In particular, $F^f(x,y) = -x \log y - (1-x) \log(1-y)$ is chance-directed. The score improves by simultaneously moving y closer to x and $1-y$ closer to $1-x$.

Proposition 34. *If g is regular, then all $B \in \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$ agree with $P_\Omega^\dagger$ on $\mathcal{L}^{\bar{\exists}}$.*

Proof. If $c = 1$, then $|\mathbb{E}_{\mathcal{L}}| = 1$ and $\mathrm{maxent}\,\mathbb{E}_{\mathcal{L}} = \{P_\Omega^\dagger\}$ follows trivially. By Theorem 5 we have that for every function $B' \in \arg\inf_{B \in \mathbb{B}_{\mathcal{L}}} \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B)$ it holds that $B'_{\lfloor n} = P_{\Omega \lfloor n}^\dagger$. Thus, all $B \in \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$ agree with $P_\Omega^\dagger$ on $\mathcal{L}^{\bar{\exists}}$.

We now focus on $0 < c < 1$.

From the above lemma we obtain

$$
\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^n(P, P_\Omega^\dagger) = -c \log\left(c + \frac{1-c}{|\Omega_n|}\right) - (1-c) \log\left(\frac{1-c}{|\Omega_n|}\right).
$$

We now follow the structure of the proof of Proposition 16 for fixed $0 < c < 1$. Let $B \in \mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$.

<u>Case1</u> $B \in \mathbb{P}_{\mathcal{L}} \setminus \{P_\Omega^\dagger\}$.

<u>Case1A</u> $B \in [\mathbb{E}_{\mathcal{L}}] \setminus \{P_\Omega^\dagger\}$.

If there exists an $n \in \mathbb{N}$ such that $B(\nu_1^n) > P_\Omega^\dagger(\nu_1^n)$, then $\sum_{\nu \in \Omega_n \setminus \{\nu_1^n\}} B(\nu) < \sum_{\nu \in \Omega_n \setminus \{\nu_1^n\}} P_\Omega^\dagger(\nu)$. If there exists an $m \in \mathbb{N}$ such that $B(\nu_1^m) < P_\Omega^\dagger(\nu_1^m)$, then there has to exist some $k > m$ such that

$$
\sum_{\substack{\nu \in \Omega^k \setminus \{\nu_1^k\} \\ \nu \models \nu_1^{k-1}}} B(\nu) < \sum_{\substack{\nu \in \Omega^k \setminus \{\nu_1^k\} \\ \nu \models \nu_1^{k-1}}} P_\Omega^\dagger(\nu).
$$

Since $B \neq P_\Omega^\dagger$ either such an $n \in \mathbb{N}$ or such a $k \in \mathbb{N}$ has to exist, possibly both exist. Overall, there has to exist some $N \in \mathbb{N}$, a $\nu^N \in \Omega_N \setminus \{\nu_1^N\}$ and an $\epsilon > 0$ such that $B(\nu^N) + \epsilon = P_\Omega^\dagger(\nu^N)$.

For large enough $n \in \mathbb{N}$, depending on B, $\mathcal{L}$ and c, it holds that

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, B) \geq -c \log B(\nu_1^n) - (1-c) \log\left(B(\nu^N)\frac{|\Omega_N|}{|\Omega_n|}\right)$$

$$> -c \log B(\nu_1^n) - (1-c) \log\left((B(\nu^N) + \frac{\epsilon}{2})\frac{|\Omega_N|}{|\Omega_n|}\right)$$

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, P_{\Omega}^{\dagger}) = -c \log\left(c + \frac{1-c}{|\Omega_n|}\right) - (1-c) \log\left(P_{\Omega}^{\dagger}(\nu^N)\frac{|\Omega_N|}{|\Omega_n|}\right).$$

Since we may assume that $B(\nu_1^n)$ converges in n to c ($B \in \mathbb{E}_{\mathcal{L}}$) we now find

$$\lim_{n \longrightarrow \infty} \frac{\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, P_{\Omega}^{\dagger})}{1-c}$$

$$\geq \lim_{n \longrightarrow \infty} -\log\left(B(\nu^N)\frac{|\Omega_N|}{|\Omega_n|}\right) + \log\left(P_{\Omega}^{\dagger}(\nu^N)\frac{|\Omega_N|}{|\Omega_n|}\right)$$

$$> -\log(B(\nu^N) + \frac{\epsilon}{2}) + \log P_{\Omega}^{\dagger}(\nu^N)$$

$$> 0.$$

Whether this limit exists or not, we have thus established that for large enough $n \in \mathbb{N}$ there exists a lower bound of the sequence

$$(\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, P_{\Omega}^{\dagger}))_{n \in \mathbb{N}}$$

which is strictly positive, since we take $N \in \mathbb{N}$ to be fixed here.

For all fixed $n \in \mathbb{N}$ let $P_n' \in \mathbb{E}_{\mathcal{L}}$ be such that $P_n'(\omega_1^n) := c$ and $P_n'(\omega_2^n) := 1 - c$. Note that $P_n' \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger})$ for all large enough n and $P_n' \in \arg\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_{\Omega}^n(P, P_{\Omega}^{\dagger})$ for all large enough n, Lemma 12.

To simplify notation let $R_n := \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^{\circ}P_n'(F) \log {}^{\circ}P_{\Omega}^{\dagger}(F)$. With this notation we have for all large enough $n \in \mathbb{N}$

$$0 \leq R_n = \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^{\circ}P_n'(F) \log {}^{\circ}P_{\Omega}^{\dagger}(F)$$

$$\leq \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \sum_{F \in \pi} {}^{\circ}P_n'(F) \log \frac{1-c}{|\Omega_n|}$$

$$= \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} -g(\pi) \log \frac{1-c}{|\Omega_n|}$$

$$= \left(\log(|\Omega_n|) - \log(1-c)\right) \cdot \sum_{\pi \in \Pi_n \setminus \{\pi^n\}} g(\pi).$$

By our standing assumption on g (regularity), we obtain that R_n converges to zero. We now find

$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, P_{\Omega}^{\dagger}) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, B) - g(\pi^n) S_{\Omega}^n(P_n', P_{\Omega}^{\dagger}) - R_n$$

$$\geq g(\pi^n)\left(S_{\Omega}^n(P_n', B) - S_{\Omega}^n(P_n', P_{\Omega}^{\dagger})\right) - R_n.$$

Because $g(\pi^n)$ is bounded and R_n converges to zero, we obtain for all large enough $n \in \mathbb{N}$ that

$$\sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P, B) - \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) > 0.$$

<u>Case1B</u> $B \notin [\mathbb{P}_\mathcal{L}] \setminus [\mathbb{E}_\mathcal{L}]$.

<u>Case1Bi</u> $\lim_{n\to\infty} B(\nu_1^n) > c$.

Let us first note that this limit has to exist, because $B(\nu_1^n)$ is a (not necessarily strictly) decreasing sequence bounded from below by c. Let $b_1 := \lim_{n\to\infty} B(\nu_1^n) > c$.

Note that there has to exist some $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $B(\nu_1^n) > P_\Omega^\dagger(\nu_1^n)$. For all $n \geq N$ there has to exist some $\nu \in \Omega_n \setminus \{\nu_1^n\}$ such that $B(\nu) < P_\Omega^\dagger(\nu)$. Then, for all $n \geq N$

$$\frac{1}{g(\pi^n)} \cdot \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P, B) \geq -c\log B(\nu_1^n) - (1-c)\log \frac{1 - B(\nu_1^n)}{|\Omega_n| - 1}$$

$$= -c\log B(\nu_1^n) - (1-c)\Big(\log(1 - B(\nu_1^n)) + \log\frac{1}{|\Omega_n| - 1}\Big)$$

$$> -c\log(c + \frac{b_1 - c}{2}) - (1-c)\log(1 - c - \frac{b_1 - c}{2}) + (1-c)\log\frac{1}{|\Omega_n| - 1}$$

$$= c\log(c + \frac{b_1 - c}{2}) - (1-c)\log \frac{1 - c - \frac{b_1-c}{2}}{|\Omega_n| - 1}$$

$$= -c\log(\frac{b_1 + c}{2}) - (1-c)\log \frac{1 - \frac{b_1+c}{2}}{|\Omega_n| - 1},$$

where the strict inequality follows from chance-directedness. We now find

$$\lim_{n\to\infty} \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P, B) - \sup_{P\in\mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger)$$

$$> \lim_{n\to\infty} g(\pi^n)\Big(-c\log(\frac{b_1 + c}{2}) - (1-c)\log\Big(\frac{1 - \frac{b_1+c}{2}}{|\Omega_n| - 1}\Big)$$

$$+ c\log\Big(c + \frac{1 - c}{|\Omega_n|}\Big) + (1-c)\log\Big(\frac{1 - c}{|\Omega_n|}\Big)\Big) - R_n$$

$$= \lim_{n\to\infty} g(\pi^n)\Big(-c\log(\frac{b_1 + c}{2}) - (1-c)\log\Big(\frac{1 - \frac{b_1+c}{2}}{|\Omega_n| - 1} \cdot |\Omega_n|\Big)$$

$$+ c\log\Big(c + \frac{1 - c}{|\Omega_n|}\Big) + (1-c)\log(1 - c)\Big)$$

$$= \lim_{n\to\infty} g(\pi^n)\Big(-c\log(\frac{b_1 + c}{2}) - (1-c)\log(1 - \frac{b_1 + c}{2})$$

$$+ c\log\Big(c + \frac{1 - c}{|\Omega_n|}\Big) + (1-c)\log(1 - c)\Big)$$

$$= \Big(\lim_{n\to\infty} g(\pi^n)\Big) \cdot \Big(-c\log(\frac{b_1 + c}{2}) - (1-c)\log(1 - \frac{b_1 + c}{2})$$

$$+ c\log(c) + (1-c)\log(1 - c)\Big)$$

$$> 0,$$

where the last line follows from the fact that the standard logarithmic scoring rules is strictly proper, *i.e.*, Equation (11) holds.

<u>Case1Bii</u> $\lim_{n \to \infty} B(\nu_1^n) < c$.

Let $b_2 := \lim_{n \to \infty} B(\nu_1^n) < c$, b_2 exists for the same reasons b_1 exists. Note that there has to exist some $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $B(\nu_1^n) < b_2 + \frac{c - b_2}{2} < c < P_\Omega^\dagger(\nu_1^n)$. Using chance-directedness we find for all $n \geq N$

$$
\begin{aligned}
\frac{1}{g(\pi^n)} \cdot \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) &\geq -c \log B(\nu_1^n) - (1 - c) \log \frac{1 - B(\nu_1^n)}{|\Omega_n| - 1} \\
&> -c \log(c + \frac{b_2 - c}{2}) - (1 - c) \log \frac{1 - c - \frac{b_2 - c}{2}}{|\Omega_n| - 1} \\
&= -c \log(\frac{b_2 + c}{2}) - (1 - c) \log \frac{1 - \frac{b_2 + c}{2}}{|\Omega_n| - 1}.
\end{aligned}
$$

Now proceed as in Case1Bi.

<u>Case2</u> $B \in \mathbb{B}_\mathcal{L} \setminus \mathbb{P}_\mathcal{L}$ and B respects logical equivalence on $\mathcal{L}^{\overline{\exists}}$.

<u>Case2A</u> There exists a $P_B \in \mathbb{P}_\mathcal{L}$ such that for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$ it holds that $^\circ B(F) \leq {}^\circ P_B(F)$. Since $B \notin \mathbb{P}_\mathcal{L}$ there has to exists an $N \in \mathbb{N}$ and an $F' \in \Omega_N$ such that $^\circ B(F') < {}^\circ P_B(F')$.

<u>Case2Ai</u> $P_B = P_\Omega^\dagger$ and no other $P \in \mathbb{P}_\mathcal{L}$ is such that $^\circ B(F) \leq {}^\circ P(F)$ for all n and all $F \subseteq \Omega_n$.

Follows as does Case2Ai in Proposition 16.

<u>Case2Aii</u> There exists a $P_B \in \mathbb{P}_\mathcal{L}$ such that $P_B \neq P_\Omega^\dagger$.

Then for all $n \geq N$ and all $P \in [\mathbb{E}_\mathcal{L}]$ it holds that $S_g^n(P, B) - S_g^n(P, P_B) \geq 0$. For all large enough $n \in \mathbb{N}$ it holds by Case1 that $\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_B) - \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) > 0$. Thus,

$$
\begin{aligned}
\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) &\geq \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_B) - \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) \\
&> 0.
\end{aligned}
$$

<u>Case2B</u> There does not exist a $P_B \in \mathbb{P}_\mathcal{L}$ such that for all $n \in \mathbb{N}$ and all $F \subseteq \Omega_n$ it holds that $^\circ B(F) \leq {}^\circ P_B(F)$.

As in Case2B in Proposition 16 we obtain that there has to exist an $\alpha > 0$ and a $N \in \mathbb{N}$ such that for all $n \geq N$ it holds that $\sum_{\omega \in \Omega_n} B(\omega) \leq 1 - \alpha$.

We have for $n \geq N$ that

$$
\begin{aligned}
\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) &= \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) - g(\pi^n) S_\Omega^n(P_n', P_\Omega^\dagger) - R_n \\
&\geq g(\pi^n) \Big(S_\Omega^n(P_n', B) - S_\Omega^n(P_n', P_\Omega^\dagger) \Big) - R_n.
\end{aligned}
$$

To complete the proof we will now show that there exists some $\beta > 0$, which depends on $\mathbb{E}_\mathcal{L}$ and g but does *not* depend on the particular $n \geq N$, such that $S_\Omega^n(P_n', B) - S_\Omega^n(P_n', P_\Omega^\dagger) > \beta > 0$. Since $g(\pi^n)$ is bounded, we then obtain that $\sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, B) - \sup_{P \in \mathbb{E}_\mathcal{L}} S_g^n(P, P_\Omega^\dagger) > 0$ for all large enough $n \in \mathbb{N}$.

We show that for all large enough $n \in \mathbb{N}$ that

$$
-\sum_{\omega \in \Omega_n} P_n'(\omega) \log f(\omega) - c \log\Big(c + \frac{1 - c}{|\Omega_n|}\Big) - (1 - c) \log\Big(\frac{1 - c}{|\Omega_n|}\Big) \geq \beta
$$

for all functions $f : \Omega_n \longrightarrow [0, 1]$ such that $\sum_{\omega \in \Omega_n} f(\omega) \leq 1 - \alpha$.

The minimum obtains, if and only if $f(\omega) = (1-\alpha)P'_n(\omega)$ for all $\omega \in \Omega_n$ as we saw in Proposition 16. Thus, the minimum obtains for $f(\nu_1^n) = (1-\alpha)(c+\frac{1-c}{|\Omega_n|})$ and $f(\nu_i^n) = (1-\alpha)\frac{1-c}{|\Omega_n|-1}$ for all other $\nu_i^n \in \Omega_n$. Let us now compute

$$
\begin{aligned}
-\sum_{\omega \in \Omega_n} P'_n(\omega) \log f(\omega) &= -c \log((1-\alpha)(c+\frac{1-c}{|\Omega_n|})) - (1-c)\log\Big(\frac{(1-c)(1-\alpha)}{|\Omega_n|-1}\Big) \\
&= -c\Big(\log(c+\frac{1-c}{|\Omega_n|}) + \log(1-\alpha)\Big) - (1-c)\Big(\log(\frac{1-c}{|\Omega_n|-1}) + \log(1-\alpha)\Big) \\
&= -c\log(c+\frac{1-c}{|\Omega_n|}) - (1-c)\log(\frac{1-c}{|\Omega_n|-1}) - \log(1-\alpha).
\end{aligned}
$$

For n approaching infinity we find

$$
\lim_{n \longrightarrow \infty} -\sum_{\omega \in \Omega_n} P'_n(\omega) \log f(\omega) + c \log\Big(c+\frac{1-c}{|\Omega_n|}\Big) + (1-c)\log\Big(\frac{1-c}{|\Omega_n|}\Big)
$$

$$
= -\log(1-\alpha)
$$

which is strictly greater some $\beta > 0$ as required.

<u>Case3</u> $B \in \mathbb{B}_{\mathcal{L}} \setminus \mathbb{P}_{\mathcal{L}}$ and B does not respect logical equivalence on $\mathcal{L}^{\exists}$.

Simply proceed as in Case2 in Theorem 17. $\quad\square$

Theorem 9.

$$
\operatorname{maxent} \mathbb{E}_{\mathcal{L}} = \{P_\Omega^\dagger\} = \{B_\dagger^\forall\}.
$$

Proof. Since all $B \in \operatorname{minloss} \mathbb{B}_{\mathcal{L}}$ agree with $P_\Omega^\dagger$ on $\mathcal{L}^{\exists}$, all $B_\dagger \in \operatorname{minloss}^* \mathbb{B}_{\mathcal{L}}$ agree with $\{P_\Omega^\dagger\}$ on $\mathcal{L}^{\exists}$; as we noted in Proposition 20.

Recall that Theorem 8 does not depend on the particular probability function, as we stated on Page 2508. We can thus apply Theorem 8 to infer that

$$
\operatorname{maxent} \mathbb{E}_{\mathcal{L}} = \{P_\Omega^\dagger\} = \{B_\dagger^\forall\}.
$$

$\square$

7. Conclusion

In this paper we have set out to provide a unified justification of the three norms of objective Bayesianism in the setting in which the underlying language is a first-order predicate language. We have seen that an approach based on scoring rules can be used to justify the norms on sentences without quantifiers: if the evidence is finitely generated, then the belief function with the best loss profile is a probability function in the set of those calibrated with evidence which has maximum standard entropy, as long as the scoring rule used in the definition of loss profile is defined in terms of a regular weighting function. One can extend this line of argument to handle sentences with quantifiers if one extends the notion of loss profile and imposes two extra desiderata: (i) language invariance and (ii) that one should not give universal hypotheses less credence than the maximum forced by the evidence.

Finally, we saw that this line of justification also applies in some cases in which evidence is not finitely generated. However, we investigated another case in which the justification does not apply because the evidence is such that there is no belief function with the best loss profile. The most one can ask in such a situation is for a belief function that has a *sufficiently* good loss profile. We saw that in this case one can use standard entropy maximisers to determine belief functions which are arbitrarily close to optimal.

We would identify two main questions for further research. First, it remains an open question as to whether, when the evidence is not finitely generated, a construction appealing to standard entropy maximisers always leads to belief functions that are arbitrarily close to optimal. Second, it would be interesting to investigate the extent to which one can relax the condition that a weighting function should be regular. We speculated that it may be the case that language invariance can be used in place of the condition that the weighting function be strongly refined, but we have little evidence, at this stage, to warrant apportioning a high degree of belief to this claim.

Acknowledgments

Author Contributions

Both authors conceived the idea, did the analysis and wrote the paper. Both authors have read and approved the final manuscript.

Appendix

A. Non-maximal entropies and non-minimal losses

In Section Section 4.4 we gave a number of minimax theorems for finitely generated evidence. As we saw in Section Section 6.1 the case of evidence which is not finitely generated is more complex. Entropy limits incur, in certain cases, infinite worst case expected loss.

While the minimax theorems relate entropy maximisers (respectively entropy limits) to loss minimisers (respectively belief functions with the best loss profile), these theorems do not tell us much about the general relation between entropy and loss. In particular, the minimax theorems leave open the question as to whether an improvement in loss profile is always accompanied by greater entropy. In this section we will show that this is not the case, by appealing to an example involving a set of calibrated probability functions $\mathbb{E}_{\mathcal{L}} \subset \mathbb{P}_{\mathcal{L}}$ which is finitely generated and two probability functions, $Q, R \in \mathbb{E}_{\mathcal{L}}$, such that Q has a better loss profile than R but has lower entropy than R.

In contrast to Section 6.1, our functions Q and R are open-minded. So, all losses we consider are finite. The fact that R has greater entropy than Q but also incurs a greater loss is thus not due to taking logarithms of zero.

For the sake of simplicity, we shall consider $\mathcal{L} = \mathcal{L}^U$.

Proposition 35. *There exist regular weightings g, a finitely generated set $\mathbb{E}_{\mathcal{L}} \subset \mathbb{P}_{\mathcal{L}}$ and probability functions $Q, R \in \mathbb{E}_{\mathcal{L}}$ such that for all $n \in \mathbb{N}$*

$$H_g^n(Q) < H_g^n(R)$$
$$\sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, Q) < \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^n(P, R).$$

The standard weighting g_Ω is another such weighting.

Thus, Q has a better loss profile than R, $Q \prec R$, but Q also has lower entropy than R, $R \gg Q$.

Proof. Let

$$\mathbb{E}_{\mathcal{L}} = \{P \in \mathbb{P}_{\mathcal{L}} \ : \ P(\omega_2^2) = P(\omega_3^2) = 0 \,\&\, P(\omega_2^1) \geq 0.495\}.$$

We now define R and Q as follows for $n \geq 3$:

$$R(\omega_1^1) := 0.505 =: R(\omega_1^2) \text{ and } R(\omega_2^1) := 0.495 =: R(\omega_4^2)$$

$$R(\omega_i^n) := 0.505 \cdot \frac{4}{|\Omega_n|} \text{ for all } 1 \leq i \leq \frac{1}{4}|\Omega_n|$$

$$R(\omega_i^n) := 0 \text{ for all } \frac{1}{4}|\Omega_n| + 1 \leq i \leq \frac{3}{4}|\Omega_n|$$

$$R(\omega_i^n) := 0.495 \cdot \frac{4}{|\Omega_n|} \text{ for all } \frac{3}{4}|\Omega_n| + 1 \leq i \leq |\Omega_n|$$

$$Q(\omega_1^1) := 0.490 =: R(\omega_1^2) \text{ and } R(\omega_2^1) := 0.510 =: R(\omega_4^2)$$

$$Q(\omega_i^n) := 0.490 \cdot \frac{4}{|\Omega_n|} \text{ for all } 1 \leq i \leq \frac{1}{4}|\Omega_n|$$

$$Q(\omega_i^n) := 0 \text{ for all } \frac{1}{4}|\Omega_n| + 1 \leq i \leq \frac{3}{4}|\Omega_n|$$

$$Q(\omega_i^n) := 0.510 \cdot \frac{4}{|\Omega_n|} \text{ for all } \frac{3}{4}|\Omega_n| + 1 \leq i \leq |\Omega_n|.$$

That is, Q and R equivocate beyond $\mathcal{L}_2$.

We find for $n = 1$

$$
\begin{aligned}
g(\pi^1) \cdot H_\Omega^1(Q) = H_g^1(Q) &= -g(\pi^1)(0.49 \log(0.49) + 0.51 \log(0.51)) \\
&\approx 0.6929 \cdot g(\pi^1) \\
&< 0.6931 \cdot g(\pi^1) \\
&\approx -g(\pi^1)(0.505 \log(0.505) + 0.495 \log(0.495)) \\
&= H_g^1(R) = g(\pi^1) \cdot H_\Omega^1(R)
\end{aligned}
$$

and

$$
\begin{aligned}
g(\pi^1) \cdot \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^1(P, Q) = \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^1(P, Q) &= -g(\pi^1)(0.495 \log(0.51) + 0.505 \log(0.49)) \\
&\approx 0.6935 \cdot g(\pi^1) \\
&< 0.7032 \cdot g(\pi^1) \\
&\approx -g(\pi^1) \log(0.495) \\
&= \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_g^1(P, R) = g(\pi^1) \cdot \sup_{P \in \mathbb{E}_{\mathcal{L}}} S_\Omega^1(P, R).
\end{aligned}
$$

Having established the result for $n = 1$ we shall now to the general case for $n \geq 2$.

For $n = 2$ note that

$$H_\Omega^2(Q) = H_\Omega^1(Q) < H_\Omega^1(R) = H_\Omega^1(R)$$

$$\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^2(P, Q) = \sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^1(P, Q) < \sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^1(P, R) = \sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^2(P, R).$$

For $n \geq 3$ we have

$$H_\Omega^n(Q) = -(0.49 \cdot \frac{4}{|\Omega_n|}) \frac{|\Omega_n|}{4} \log(0.49 \cdot \frac{4}{|\Omega_n|}) - (0.51 \cdot \frac{4}{|\Omega_n|}) \frac{|\Omega_n|}{4} \log(0.51 \cdot \frac{4}{|\Omega_n|})$$

$$= -0.49 \log(0.49) - 0.51 \log(0.51) + \log \frac{|\Omega_n|}{4}$$

$$= H_\Omega^1(Q) + \log \frac{|\Omega_n|}{4}$$

and in the same way we find

$$H_\Omega^n(R) = H_\Omega^1(R) + \log \frac{|\Omega_n|}{4}$$

Furthermore,

$$\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, Q) = -0.505 \log(0.49 \cdot \frac{4}{|\Omega_n|}) - 0.495 \log(0.51 \cdot \frac{4}{|\Omega_n|})$$

$$= -0.49 \log(0.49) - 0.51 \log(0.51) + \log \frac{|\Omega_n|}{4}$$

$$= \sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^1(P, Q) + \log \frac{|\Omega_n|}{4}$$

$$\sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^n(P, R) = -\log(0.495 \cdot \frac{4}{|\Omega_n|})$$

$$= \sup_{P \in \mathbb{E}_\mathcal{L}} S_\Omega^1(P, R) + \log \frac{|\Omega_n|}{4}.$$

This establishes the result for $g = g_\Omega$.

The result follows for such general weightings g which converge quickly enough to the standard weighting g_Ω so that all further terms are negligible. $\square$

B. Symmetry and equivocator preservation

Recall Definition 14: g is called *equivocator-preserving*, if and only if g_n is equivocator-preserving for all $n \in \mathbb{N}$, *i.e.,* if and only if

$$\mathbb{P}_n^\dagger = \{Q_{\downarrow n} \; : \; Q \in \arg \sup_{P \in \mathbb{P}_\mathcal{L}} H_g^n(P)\} = \{P_{=\downarrow n}\}.$$

So, if g is equivocator-preserving and if $P_= \in [\mathbb{E}_\mathcal{L}]$, then $P_=$ maximises $\sup_{P \in \mathbb{E}_\mathcal{L}} H_g^n(P)$ and thus $\mathbb{P}^\dagger = \{P_=\} = \text{maxent} \, \mathbb{E}_\mathcal{L} = \text{minloss} \, \mathbb{B}_\mathcal{L}$. We know from Proposition 7 that inclusive and symmetric g are equivocator-preserving.

Interestingly, we shall see that there exist non-symmetric g_n which are equivocator-preserving. This answers the question posed at the bottom of Landes and Williamson [4] (p. 3574) in the negative.

Proposition 36 (**Non-symmetric equivocator preservation**). *For all $n \in \mathbb{N}$ such that $|\Omega_n| \geq 4$ there exist inclusive, equivocator-preserving and non-symmetric weighting functions g_n. The set of such weighting functions g_n is convex.*

Proof. By Landes and Williamson [4] (Lemma 9 p. 3573) it holds that g_n is inclusive and equivocator-preserving, if and only if

$$y(\omega) := \sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} \sum_{\substack{\pi \in \Pi_k \\ F \in \pi}} -g_n(\pi)(1 - \log|\Omega_n| + \log|F|) = c$$

for some constant, c.

Note that we can simply this expression as follows

$$\sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} \sum_{\substack{\pi \in \Pi_n \\ F \in \pi}} -g_n(\pi)(1 - \log|\Omega_n| + \log|F|)$$

$$= (1 - \log|\Omega_n|) \sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} \sum_{\substack{\pi \in \Pi_n \\ F \in \pi}} -g_n(\pi) + \sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} \sum_{\substack{\pi \in \Pi_n \\ F \in \pi}} -g_n(\pi) \log|F|$$

$$= (1 - \log|\Omega_n|) \sum_{\pi \in \Pi_n} -g_n(\pi) + \sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} \log(|F|) \cdot \sum_{\substack{\pi \in \Pi_n \\ F \in \pi}} -g_n(\pi)$$

$$= (1 - \log|\Omega_n|) \sum_{\pi \in \Pi_n} -g_n(\pi) + \sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} -\gamma_n(F) \log|F|.$$

The first sum does not depend on ω. Thus, $y(\omega)$ is constant, if and only if $z(\omega) := \sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} -\gamma_n(F) \log|F|$ is constant.

Let us now define an inclusive and non-symmetric weighting function g'_n which satisfies this condition. Let k be such that $\Omega_n = \{\omega_1, \dots \omega_{2^k}\}$ and put

$$g'_n(\{\omega_1, \dots, \omega_{2^{k-1}}\}, \{\omega_{2^{k-1}+1}, \dots, \omega_{2^k}\}) := \frac{1}{2}$$

$$g'_n(\pi) := 1 \text{ for all other } \pi \in \Pi_n.$$

Clearly, g'_n is inclusive ($g_n(\pi) > 0$ for all $\pi \in \Pi_n$), non-symmetric (there are two partitions π, π' such that the classes of π and π' have the same number of elements but $g'_n(\pi) \neq g'_n(\pi')$) and $z(\omega)$ is constant (since $\sum_{\substack{F \subseteq \Omega_n \\ \omega \in F}} \log(|F|) \cdot \sum_{\substack{\pi \in \Pi_k \\ F \in \pi}} -g_n(\pi)$ is invariant under permutations of n-states).

Addressing the second part of the proof: For inclusive and equivocator-preserving g_n it holds that H_g^n is a strictly concave function on $\mathbb{P}_n$ and $\sup_{P \in \mathbb{P}_n} H_g^n(P)$ always obtains for $P_{=\lfloor n}$. $\mathbb{P}_n$ is convex. Hence, the unique maximum of every convex combination of such g_n obtains for $P_{=\lfloor n}$. $\square$

In general, computing a function which maximises $H_g^n(P)$ for $P \in [\mathbb{E}_\mathcal{L}]$ is a non-trivial computational problem, even for $g = g_\Omega$. The only widely shared intuition is that $P_=$ ought to be the function in $\mathbb{P}_\mathcal{L}$ which has greatest entropy. Imposing symmetry is sufficient—but, as we have just seen, not necessary—to ensure that this constraint is satisfied. Imposing symmetry has further structural consequences such as: if $\mathbb{E}_\mathcal{L}$ is invariant under renaming of states, then so is $P_n^\dagger$; see Landes and Williamson [4] (Appendix B.3) for details.

C. Key notation

Here we summarise key notation, for ease of reference.

Symbol	Reference	Meaning
$\langle\ \rangle$	Page 2460	Convex hull
$[\ ]$	Page 2463	Closure
$\mathcal{L}^{\exists}$	Page 2463	Predicate language with quantifiers
$\mathcal{L}^{\nexists}$	Page 2463	Predicate language without quantifiers
$\mathbb{B}_{\mathcal{L}}$	Definition 1	(Normalised) belief functions on sentences $S\mathcal{L}$
$\mathbb{P}_{\mathcal{L}}$	Page 2465	Probability functions on $S\mathcal{L}$
$\mathbb{P}_n$	Page 2469	Probability functions on $S\mathcal{L}_n$
Ω_n	Page 2463	n-states
$P_=$	Page 2473	Equivocator function in $\mathbb{P}_{\mathcal{L}}$, $P_=(\omega) = 1/|\Omega_n|$ for each $\omega \in \Omega_n$
$\Pi_{\mathcal{L}}$	Page 2464	Partitions of sentences $S\mathcal{L}$
Π_n	Definition 4	Partitions of propositions $\mathcal{P}\Omega_n$
Π	Definition 4	All partitions of propositions, $\bigcup_{n=1}^{\infty} \Pi_n$
π^n	Page 2470	$\{\{\omega\} : \omega \in \Omega_n\}$, the finest partition of Ω_n
$\mathbb{E}_{\mathcal{L}}$	Page 2464	Calibrated belief functions on $S\mathcal{L}$
$\mathbb{E}_n$	Page 2469	Restrictions of these functions to $S\mathcal{L}_n^{\nexists}$
$^{\circ}B$	Page 2467	Belief function on propositions induced by B defined on sentences
g	Definition 6	Weighting function
g_{Ω}	Page 2470	Standard weighting function
H_g^n	Definition 9	n-entropy
H_{Ω}^n	Definition 10	Standard (Shannon) entropy
$\mathrm{maxent}\,\mathbb{E}_{\mathcal{L}}$	Page 2472	Calibrated functions on $\mathcal{L}$ with maximal entropy
$\mathbb{P}_n^{\dagger}$	Page 2472	Calibrated functions on $\mathcal{L}_n$ with maximum n-entropy
$P_n^{\dagger}$	Page 2472	Unique such function
$\mathbb{P}^{\dagger}$	Definition 11	Limit points of maximum n-entropy functions
$P^{\dagger}$	Page 2472	Unique such entropy limit
$P_{\Omega}^{\dagger}$	Page 2472	Standard entropy limit
ϱ_n	Definition 19	n-representations
S_g^n	Page 2484	Logarithmic n-score of a belief function wrt a probability function
$S_{g,\rho}^n$	Page 2484	Representation-relative n-score
$\mathrm{minloss}\,\mathbb{B}_{\mathcal{L}}$	Definition 21	Belief functions on $\mathcal{L}^{\nexists}$ with the best loss profile
$\mathrm{minloss}^*\,\mathbb{B}_{\mathcal{L}}$	Definition 24	Belief functions on $\mathcal{L}^{\exists}$ with the best loss profile

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Williamson, J. *In defence of objective Bayesianism*; Oxford University Press: Oxford, UK, 2010.

2. There are several alternatives to the objective Bayesian account of strength of belief, including subjective Bayesianism, imprecise probability, the theory of Dempster-Shafer belief functions and related theories. Here we only have the space to motivate objective Bayesianism, not to assess these other views.

3. Taking the convex hull may mean that a calibrated belief function does not satisfy the known constraints on physical probability. For example, if θ is known to be a statement about the past then it is known that its physical probability is 0 or 1; bel is not constrained to be 0 or 1, however, unless it is also known whether or not θ is true. Similarly, it may be known that two propositions are probabilistically independent with respect to physical probability; this need not imply that they are probabilistically independent with respect to epistemic probability. See Williamson [1] (pp. 44–45) for further discussion of this point.

4. Landes, J.; Williamson, J. Objective Bayesianism and the Maximum Entropy Principle. *Entropy* **2013**, *15*, 3528–3591.

5. Gaifman, H. Concerning Measures in First Order Calculi. *Isr. J. Math.* **1964**, *2*, 1–18.

6. Williamson, J. *Lectures on Inductive Logic*; Oxford University Press: Oxford, UK, 2015.

7. Paris, J.B. *The Uncertain Reasoner's Companion*; Cambridge University Press: Cambridge, UK, 1994.

8. Williamson, J. Probability logic. In *Handbook of the Logic of Argument and Inference: the Turn toward the Practical*; Gabbay, D., Johnson, R., Ohlbach, H.J., Woods, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2002; pp. 397–424.

9. Haenni, R.; Romeijn, J.W.; Wheeler, G.; Williamson, J. *Probabilistic Logics and Probabilistic Networks*; Synthese Library, Springer: Dordrecht, The Netherlands, 2011.

10. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley and Sons: New York, NY, USA, 1991.

11. Rudin, W. *Principles of Mathematical Analysis*, 3 ed.; McGraw-Hill: New York, USA, 1973.

12. de Finetti, B. *Theory of Probability*; Wiley: London, UK, 1974.

13. Joyce, J.M. A Nonpragmatic Vindication of Probabilism. *Philos. Sci.* **1998**, *65*, 575–603.

14. Joyce, J.M. Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In *Degrees of Belief*; Huber, F., Schmidt-Petri, C., Eds.; Synthese Library 342; Springer: New York, NY, USA, 2009.

15. Grünwald, P.; Dawid, A.P. Game Theory, Maximum Entropy, Minimum Discrepancy, and Robust Bayesian Decision Theory. *Ann. Stat.* **2004**, *32*, 1367–1433.

16. Savage, L.J. Elicitation of Personal Probabilities and Expectations. *J. Am. Stat. Assoc.* **1971**, *66*, 783–801.

17. Popper, K.R. *The Logic of Scientific Discovery*; Routledge: London, UK, 1999.

18. Barnett, O.; Paris, J.B. Maximum Entropy Inference with Quantified Knowledge. *Logic J. IGPL* **2008**, *16*, 85–98.

19. Williamson, J. Objective Bayesian Probabilistic Logic. *J. Algorithm.* **2008**, *63*, 167–183.