Journal of Philosophical Logic
https://doi.org/10.1007/s10992-022-09680-6

# Determining Maximal Entropy Functions for Objective Bayesian Inductive Logic

**Juergen Landes**[1] · **Soroush Rafiee Rad**[2] · **Jon Williamson**[3]

Received: 23 December 2021 / Accepted: 23 July 2022



## Abstract

According to the objective Bayesian approach to inductive logic, premisses inductively entail a conclusion just when every probability function with maximal entropy, from all those that satisfy the premisses, satisfies the conclusion. When premisses and conclusion are constraints on probabilities of sentences of a first-order predicate language, however, it is by no means obvious how to determine these maximal entropy functions. This paper makes progress on the problem in the following ways. Firstly, we introduce the concept of a limit in entropy and show that, if the set of probability functions satisfying the premisses contains a limit in entropy, then this limit point is unique and is the maximal entropy probability function. Next, we turn to the special case in which the premisses are categorical sentences of the logical language. We show that if the uniform probability function gives the premisses positive probability, then the maximal entropy function can be found by simply conditionalising this uniform prior on the premisses. We generalise our results to demonstrate agreement between the maximal entropy approach and Jeffrey conditionalisation in the case in which there is a single premiss that specifies the probability of a sentence of the language. We show that, after learning such a premiss, certain inferences are preserved, namely inferences to inductive tautologies. Finally, we consider potential pathologies of the approach: we explore the extent to which the maximal entropy approach is invariant under permutations of the constants of the language, and we discuss some cases in which there is no maximal entropy probability function.


**Keywords** Inductive logic · Entropy · Maximum entropy principle · First order logic · Probability logic

✉ Juergen Landes
juergen_landes@yahoo.de

[1] Department of Philosophy “Piero Martinetti”, University of Milan, Milan, Italy

[2] Dutch Institute for Emergent Phenomena (DIEP) and Institute for Logic, Language and Computation (ILLC), Amsterdam, The Netherlands

[3] Philosophy Department and Centre for Reasoning, University of Kent, Canterbury, UK

Published online: 13 October 2022 

## 1 Introduction

Inference under uncertainty remains one of the challenges of our time. While there is widespread agreement that probabilities are well suited to capture uncertainty and that Bayesian and Jeffrey conditionalisation are key principles of rationality, there is significant disagreement about the proper choice of probabilities and their use. One prominent approach to uncertain inference appeals to the Maximum Entropy Principle of Jaynes [15]. This selects a probability function, from all those that agree with the available evidence, that is as equivocal as possible in the sense that it has maximum Shannon entropy [40]. The Maximum Entropy Principle is often employed as part of an objective Bayesian approach to inference [16, 43].

The use of the Maximum Entropy Principle on finite domains is well-understood. A number of axiomatic characterisations highlight some of its most important properties, such as irrelevance of extraneous information, independence in the absence of evidence of dependence, and invariance under uniform refinements of the underlying finite domain [27, 28, 30, 31]. Furthermore, MaxEnt inference is known to agree on finite domains with what one might call 'baseline rationality': Bayesian and Jeffrey conditionalisation turn out to be special cases of MaxEnt inference [41]. While Jeffrey conditionalisation can only deal with a single uncertain premiss at a time, of the form $P(F) = c$, MaxEnt inference can handle multiple uncertain premisses of more complex forms simultaneously. Given a fixed finite domain and premisses of a suitable form, MaxEnt inference introduces an objective relation between premisses and conclusions, independent of the inferring agent. This objectivity facilitates the implementation of MaxEnt inferences in algorithms and automated systems[1].

The application of MaxEnt to infinite domains is much less well understood. Firstly, axiomatic characterisations have yet to be put forward. Second, MaxEnt inference is only known to agree with Jeffrey conditionalisation on certain infinite domains that lack a logical structure [5]. The focus of this paper is to shed some light on the application of MaxEnt to infinite domains—in particular, to its use as semantics for objective Bayesian inductive logic on infinite predicate languages.

There are two different explications of MaxEnt on infinite predicate languages. One, due to Jeff Paris and his co-workers, takes limits of maximum entropy functions on finite sublanguages [2, 29, 35, 37, 38]. The second explication considers maximal entropy probability functions defined on the infinite language as a whole [18, 20, 36, 42, 44]. The limit approach provides a constructive means to determine the probabilities for MaxEnt inference. However, this construction has problems: in some cases, it does not yield an answer at all [29, 35]; in other cases the constructed probabilities fail to satisfy the given premisses [19]. The maximal entropy approach can be used in a wider range of situations [35, 36], but the approach is less constructive and it is less

[1] Note however that inference using MaxEnt can be computationally complex in the worst case—see Paris [27, Chapter 10] and Pearl [34, p. 463], and also Goldman [10], Goldman and Rivest [11], Ormoneit and White [26], Balestrino et al. [1], Chen et al. [6], Landes and Williamson [21, 22]. We will not be concerned with computational complexity in this paper.

clear how to determine maximal entropy probability functions. It has however been conjectured that both approaches agree where the limit approach is well defined [24, 44].

In this paper we study the second of these two approaches: the maximal entropy approach. We first give a method for determining the maximal entropy probability function in many general scenarios, by introducing the concept of a limit in entropy (Theorem 16). Then we show that the maximal entropy approach generalises both Bayesian conditionalisation (Theorem 34) and Jeffrey conditionalisation (Theorem 41). This not only clarifies which probabilities the maximal entropy approach picks out, but also gives a simple way to determine these probabilities and shows that the maximal entropy approach agrees with baseline rationality.

These results expose a surprising fact: where the maximal entropy approach agrees with conditionalisation, the maximal entropy function can be found by conditionalising a particularly simple probability function—the uniform distribution—on particularly simple propositions—namely, quantifier-free propositions. This means that in such cases, inferences in first-order inductive logic from constraints involving quantified propositions can be reduced to what are essentially finite inferences involving quantifier-free premisses. As far as we are aware, the maximal entropy approach is the only viable approach to inductive logic in which inference can be simplified in this way.

We turn next to general features of the maximal entropy approach. We see that certain inferences drawn in the absence of any premisses—inferences to inductive tautologies—are preserved when a premiss is added (Section 7). We show that while the notion of comparative entropy used to define the maximal entropy probability functions can depend on the order of the constant symbols (Proposition 50), this order is rendered irrelevant in all cases in which the maximal entropy approach simplifies to Bayesian or Jeffrey conditionalisation (Theorem 51, Corollary 52). Finally, it becomes clear why the maximal entropy approach fails to provide probabilities in some cases. These cases are those where the premiss has zero prior probability. Updating on events of zero prior probability is notoriously problematic. We investigate the extent of these failures in Section 9, show that they arise in all levels of the arithmetic hierarchy including and above $\Sigma_2$ (Theorem 54), and provide a refinement of the approach to handle these problematic cases.

It is worth noting the relation between this approach and perhaps the most well-known approach to inductive logic, namely that of Rudolf Carnap (see, e.g., [4]). In common with Carnap's approach, we consider the problem of developing an inductive logic involving sentences of a first-order predicate language. However, the maximal entropy approach differs in two key respects. Firstly, our setting is more general, as it considers premiss statements which attach probabilities or sets of probabilities to sentences of the logical language, while Carnap considered only the sentences themselves. Second, our approach is based on the idea of entropy maximisation, while Carnap's approach appeals to Bayesian conditionalisation involving exchangeable prior probability functions. The latter approach is susceptible to serious objections [44, Chapter 4].

## 2 Objective Bayesian Inductive Logic

An important class of probabilistic logics consider entailment relationships of the following form [12]:

$$\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!\!| \;\psi^Y.$$

Here, $\varphi_1, \ldots, \varphi_k, \psi$ are sentences of a logical language $\mathcal{L}$ and $X_1, \ldots, X_k, Y$ are sets of probabilities. This entailment relationship should be interpreted as saying: $\varphi_1, \ldots, \varphi_k$ having probabilities in $X_1, \ldots, X_k$ respectively inductively entails that $\psi$ has probability in $Y$. In Sections 4 and 5 we will be particularly interested in the special case in which the premisses are *categorical*: i.e., $X_1 = \ldots = X_k = \{1\}$. In such a situation we will often omit the superscripts $X_1, \ldots, X_k$.

The objective Bayesian approach to inductive logic interprets probabilities as rational degrees of belief. It takes the premisses on the left-hand side of the entailment relationship to capture all the constraints on rational degrees of belief that are inferred from evidence, and it uses Jaynes' Maximum Entropy Principle to determine a rational belief function with which to calculate the probability of a conclusion statement $\psi$. Thus if $\mathcal{L}$ is a finite propositional language, $X_1, \ldots, X_k$ are closed convex sets of probabilities (i.e., closed intervals), and the premisses are consistent, an entailment relationship holds just when the probability function with maximum entropy, amongst all those that satisfy the premisses, gives a probability in $Y$ to $\psi$ [43, Chapter 7].

This approach has been extended to the case in which $\mathcal{L}$ is a first-order predicate language in the following way. Suppose $\mathcal{L}$ has countably many constant symbols $t_1, t_2, \ldots$ and finitely many relation symbols $U_1, \ldots, U_l$. Let $a_1, a_2, \ldots$ run through the atomic sentences of the form $U_i t_{i_1} \ldots t_{i_k}$ in such a way that those atomic sentences involving only $t_1, \ldots, t_n$ occur before those involving $t_{n+1}$, for each $n$. Consider the finite sub-languages $\mathcal{L}_n$, containing only constant symbols $t_1, \ldots, t_n$.

**Definition 1** ($n$-states) $\Omega_n$ is the set of *$n$-states* of $\mathcal{L}$, i.e., sentences of the form $\pm a_1 \wedge \ldots \wedge \pm a_{r_n}$ involving the atomic sentences $a_1, \ldots, a_{r_n}$ of $\mathcal{L}_n$, which only feature the constants $t_1, \ldots, t_n$.[2] The $n$-states for $\mathcal{L}$ are thus the sentences

$$\bigwedge_{\substack{1\le i\le l \\ 1\le j_1,\ldots,j_{k_i}\le n}} U_i^{\epsilon_{j_1},\ldots,\epsilon_{j_{k_i}}} t_{j_1}\ldots t_{j_{k_i}}$$

where $k_i$ is the arity of $U_i$, $\epsilon_{j_1}, \ldots, \epsilon_{j_{k_i}} \in \{0, 1\}$ and $U_i^1 t_{j_1} \ldots t_{j_{k_i}} = U_i t_{j_1} \ldots t_{j_{k_i}}$ and $U_i^0 t_{j_1} \ldots t_{j_{k_i}} = \neg U_i t_{j_1} \ldots t_{j_{k_i}}$.

Let $S\mathcal{L}$, $S\mathcal{L}_n$ be the sets of sentences of $\mathcal{L}$, $\mathcal{L}_n$ respectively.

**Definition 2** ($N_\varphi$) For a single given sentence $\varphi$ we use $N_\varphi$ to denote the greatest index of the constants appearing in $\varphi$, i.e., the greatest number $n$ such that $t_n$ occurs in $\varphi$. If $\varphi$ has no constants, we adopt the convention that $N_\varphi = 1$.

[2] The $n$-states are sometimes referred to as 'state descriptions'.

**Definition 3** (Probability) A *probability function* $P$ on $\mathcal{L}$ is a function $P : S\mathcal{L} \longrightarrow \mathbb{R}_{\geq 0}$ such that:

P1: If $\tau$ is a tautology, i.e., $\models \tau$, then $P(\tau) = 1$.
P2: If $\theta$ and $\varphi$ are mutually exclusive, i.e., $\models \neg(\theta \wedge \varphi)$, then $P(\theta \vee \varphi) = P(\theta) + P(\varphi)$.
P3: $P(\exists x \theta(x)) = \sup_m P\left(\bigvee_{i=1}^m \theta(t_i)\right)$.

A probability function is determined by the values it gives to the $n$-states—see, e.g., Williamson [44, §2.6.3] and Gaifman [9]. We denote the set of probability functions by $\mathbb{P}$.

Of particular importance will be the *equivocator* function, $P_=$, which gives the same probability to each $n$-state, for each $n$.

**Definition 4** (Equivocator Function) The *equivocator function* is the probability function $P_=$ defined by:

$$P_=(\omega_n) \stackrel{\text{df}}{=} \frac{1}{2^{r_n}} = \frac{1}{|\Omega_n|}$$

for each $n$-state $\omega_n \in \Omega_n$ and each $n \geq 1$.

**Definition 5** (Measure) The *measure* of a sentence $\theta$ is the probability given to it by the equivocator function. In particular, $\theta$ has *positive measure* if and only if $P_=(\theta) > 0$.

**Definition 6** (Feasible Region) We use $\mathbb{E}$ to refer to the set of probability functions that satisfy the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$, i.e.,

$$\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k\}.$$

Two special cases will be particularly important in this paper. To distinguish the case of a single categorical premiss, $\varphi$, we often write $\mathbb{E}_\varphi$ instead of $\mathbb{E}$. In the case of a single uncertain premiss, $\varphi^X$, we write $\mathbb{E}_{\varphi^X}$. Throughout, we shall assume that the $X$ are intervals or single probability values, and that the feasible region is non-empty, $\mathbb{E} \neq \emptyset$.

**Definition 7** ($n$-entropy) The *$n$-entropy* of a probability function $P$ is defined as

$$H_n(P) \stackrel{\text{df}}{=} -\sum_{\omega \in \Omega_n} P(\omega) \log P(\omega).$$

The $n$-entropies, which only take into account the probabilities on finitely many $n$-states, are then used to define a notion of comparative entropy on the infinite language $\mathcal{L}$ as a whole:

**Definition 8** (Comparative Entropy) We say that the probability function $P \in \mathbb{P}$ *has greater entropy than* $Q \in \mathbb{P}$, if and only if the $n$-entropy of $P$ dominates that of $Q$

for sufficiently large $n$, i.e., if and only if there is an $N \in \mathbb{N}$ such that for all $n \geq N$, $H_n(P) > H_n(Q)$.

The *greater entropy* relation defines a partial order on the probability functions on $\mathcal{L}$. We will focus on the maximal elements in $\mathbb{E}$ of this partial ordering:

**Definition 9** (Maximal Entropy Functions) The set of maximal entropy functions, maxent $\mathbb{E}$, is defined as

$$\text{maxent}\,\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E} \ : \ \text{there is no } Q \in \mathbb{E} \text{ that has greater entropy than } P\}.$$

In the simplest case there are no premisses:

*Example 10* In the absence of any premisses, maxent $\mathbb{E}$ = maxent $\mathbb{P} = \{P_=\}$. To see this note first that in the absence of premisses every probability function is in $\mathbb{E}$, i.e., $\mathbb{E} = \mathbb{P}$. Furthermore, for all $n \in \mathbb{N}$ it holds that $H_n(P)$ is maximal if and only if $P$ agrees with the equivocator $P_=$ on $\Omega_n$. Since every other probability function $Q \in \mathbb{P} \setminus \{P_=\}$ differs from $P_=$ for all large enough $n$, it holds that $H_n(Q) < H_n(P_=)$ for all large enough $n$. Hence, $P_=$ has greater entropy than $Q$ (Definition 8). Since $P_=$ has greater entropy than all other probability functions $Q \neq P_=$ and $P_= \in \mathbb{E}$, we have that maxent $\mathbb{E} = \{P_=\}$ (Definition 9).

In this paper, we invoke the objective Bayesian notion of inductive entailment, denoted by $|\!\!\!\approx^{\circ}$ [44, §5.3]:

**Definition 11** (Objective Bayesian Inductive Entailment) Premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ inductively entail $\psi^Y$, denoted by $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \; |\!\!\!\approx^{\circ} \psi^Y$, iff $P(\psi) \in Y$ for all $P \in$ maxent $\mathbb{E}$.

In the absence of any premisses, we write $|\!\!\!\approx^{\circ} \psi^Y$, which holds if and only if $P_=(\psi) \in Y$, i.e., if and only if the measure of $\psi$ is an element of $Y$ (Example 10).

Note that this definition applies where maxent $\mathbb{E}$ is non-empty. We consider the case in which maxent $\mathbb{E}$ is empty in Section 9.

**Definition 12** We will say that sentence $\psi$ is an *inductive tautology* if $|\!\!\!\approx^{\circ} \psi$, i.e., if it has measure 1. It is an *inductive contradiction* if $|\!\!\!\approx^{\circ} \neg\psi$, i.e., if it has measure 0. It is *inductively consistent* if $\not|\!\!\!\approx^{\circ} \neg\psi$, i.e., if it has positive measure. Sentences $\psi$ and $\theta$ are *inductively equivalent* if $|\!\!\!\approx^{\circ} \psi \leftrightarrow \theta$.

**Proposition 13** *If $\psi$ and $\theta$ are inductively equivalent then:*

$$P_=(\psi) = P_=(\theta)$$

*and as long as $\psi$ has positive measure:*

$$P_=(\cdot|\psi) = P_=(\cdot|\theta).$$

*Proof* That $\psi$ and $\theta$ are inductively equivalent implies that $P_=(\psi \wedge \neg\theta) = P_=(\neg\psi \wedge \theta) = 0$. So,

$$P_=(\psi) = P_=(\psi \wedge \theta) + P_=(\psi \wedge \neg\theta) = P_=(\psi \wedge \theta) + P_=(\neg\psi \wedge \theta) = P_=(\theta).$$

By similar reasoning, $P_=(\varphi \wedge \psi) = P_=(\varphi \wedge \theta)$ for any sentence $\varphi$. Hence,

$$P_=(\varphi|\psi) = \frac{P_=(\varphi \wedge \psi)}{P_=(\psi)} = \frac{P_=(\varphi \wedge \theta)}{P_=(\theta)} = P_=(\varphi|\theta)$$

for any sentence $\varphi$, as required. □

While the objective Bayesian approach provides appropriate semantics for inductive logic, it is often not obvious how to determine the maximal entropy functions in order to ascertain whether a given entailment relationship holds. This is because the definition of maxent $\mathbb{E}$ seems to require a sort through members of $\mathbb{E}$ in order to find those with maximal entropy—a process that would be unfeasible in practice. This paper seeks to address the question of how to determine maximal entropy functions.

Section 3 introduces the concept of a limit in entropy in order to characterise maxent $\mathbb{E}$ in terms of certain limits of $n$-entropy maximisers. This gives a constructive procedure for determining maxent $\mathbb{E}$ when $\mathbb{E}$ contains a limit in entropy.

In Sections 4 and 5 we consider an important special case—that in which the premisses are categorical sentences $\varphi_1, \ldots, \varphi_k$ (without attached probabilities) and where the maximal entropy function can be obtained simply by conditionalising the equivocator function.

## 3 Limits in Entropy

This section adapts the techniques of Landes et al. [24, §5] in order to characterise maxent $\mathbb{E}$ in terms of certain limits of $n$-entropy maximisers. Landes et al. [24] were concerned with a very different question: that of showing that the above objective Bayesian semantics for inductive logic, which appeals to maximal entropy functions, yields the same inferences as those produced by the Barnett-Paris limit approach discussed in Section 1. Nevertheless, the results of Landes et al. [24, §5] can be straightforwardly adapted to the present problem. The proofs of the two results in this section, which are close to those of Landes et al. [24, Proposition 36] and Landes et al. [24, Theorem 39], have been provided in Appendix 1.

We will consider the set of $n$-entropy maximisers for each $n$:

$$\mathbb{H}_n \stackrel{\text{df}}{=} \{P \in \mathbb{E} : H_n(P) \text{ is maximised}\}.$$

We now introduce the key concept of this section:

**Definition 14** (Limit in Entropy) $P \in \mathbb{P}$ is a *limit in entropy* of $\mathbb{P}_1, \mathbb{P}_2, \ldots \subseteq \mathbb{P}$, if there is some sequence $Q_n \in \mathbb{P}_n$ such that $|H_n(Q_n) - H_n(P)| \longrightarrow 0$ as $n \longrightarrow \infty$. $P \in \mathbb{P}$ will be called a *limit in entropy* of $\mathbb{E}$ if it is a limit in entropy of $\mathbb{H}_1, \mathbb{H}_2, \ldots$.

Limits in entropy of $\mathbb{E}$ are of special interest because they are also limit points in terms of the $L_1$ distance,

$$\|P - Q\|_n \stackrel{\text{df}}{=} \sum_{\omega \in \Omega_n} |P(\omega) - Q(\omega)|\,.$$

**Proposition 15** *If $P$ is a limit in entropy of $\mathbb{E}$, then there are functions $Q_n \in \mathbb{H}_n$, for $n \geq 1$, such that $\|Q_n - P\|_n \longrightarrow 0$ as $n \longrightarrow \infty$.*

This property enables us to characterise the set of maximal entropy functions more constructively, in terms of a limit of $n$-entropy maximisers:

**Theorem 16** (Limit in Entropy) *If $\mathbb{E}$ contains a limit in entropy $P$, then*

$$\text{maxent}\,\mathbb{E} = \{P\}.$$

Note that there can be at most one limit in entropy $P$ of $\mathbb{E}$. This is because $\mathbb{E}$ is closed and convex (by the closure and convexity of $X_1, \dots, X_k$) and the $n$-entropy maximiser of a closed, convex set is uniquely determined on $\mathcal{L}_n$. Thus, the $\mathbb{H}_n$ can have at most one $L_1$ limit point.

Theorem 16 provides a simple procedure for showing that a hypothesised function $P$ is in fact a maximal entropy function: show that it is a limit in entropy of $n$-entropy maximisers, and show that it is in $\mathbb{E}$. (Note that this is only a sufficient condition: if $\mathbb{E}$ contains no limit in entropy, then Theorem 16 does not allow us to infer anything about maxent $\mathbb{E}$.) [24, Lemmas 40, 44] provide some tools for demonstrating that a hypothesised function is a limit in entropy of $\mathbb{E}$.

*Example 17* Suppose we have a single premiss $\forall x U x^{\{c\}}$ where $\mathcal{L}$ has a single unary predicate $U$ and $c \in [0, 1]$. (We will often omit the curly braces and write $\varphi^c$ instead of $\varphi^{\{c\}}$ in such cases.) Since there is a single unary predicate in the language, the number $r_n$ of atomic sentences of $\mathcal{L}_n$ is $n$. For $c > 0$ and sufficiently large $n$, the $n$-entropy maximiser gives probability $c$ to the $n$-state $Ut_1 \wedge \dots \wedge Ut_n$, which we abbreviate by $\theta_n$, and divides the remaining probability $1 - c$ amongst all other $n$-states:

$$P^n(\omega_n) = \begin{cases} c & : \quad \omega_n = \theta_n \\ \frac{1-c}{2^n-1} & : \quad \omega_n \models \neg\theta_n. \end{cases}$$

If $c = 0$, on the other hand, $P^n = P_=$ for all $n$.

By the argument of Landes et al. [24, Example 42], the following probability function is a limit in entropy:

$$P(\omega_n) = \begin{cases} c + \frac{1-c}{2^n} & : \quad \omega_n = \theta_n \\ \frac{1-c}{2^n} & : \quad \omega_n \models \neg\theta_n. \end{cases}$$

$P \in \mathbb{E}$ because $P(\forall x U x) = \lim_{n\to\infty} P(\theta_n) = c$. Hence by Theorem 16, maxent $\mathbb{E} = \{P\}$.

*Example 18* Consider a single categorical premiss $U_1 t_1 \vee \exists x \forall y U_2 xy$. In this case, $\mathbb{H}_n$ is the set of probability functions in $\mathbb{E}$ whose restrictions to the sublanguage $\mathcal{L}_n$

match the equivocator function on $\mathcal{L}_n$, $\mathbb{H}_n = \{P \in \mathbb{E} : P_{\restriction\mathcal{L}_n} = P_{=\restriction\mathcal{L}_n}\}$ for all $n$. Thus the equivocator function is the unique limit in entropy of $\mathbb{E}$. However, the equivocator function is not in $\mathbb{E}$, because $P_=(U_1t_1 \vee \exists x\forall y U_2xy) = 1/2 < 1$ (see Example 22), so it cannot be the maximal entropy function. Indeed, as will become apparent later (Theorem 34), $\text{maxent}\,\mathbb{E} = \{P_=(\cdot|U_1t_1)\}$.

## 4 Categorical Premisses and Bayesian Conditionalisation

We now consider an important special case: that in which the premisses are categorical sentences $\varphi_1, \ldots, \varphi_k$ of $\mathcal{L}$, i.e., there are no attached sets of probabilities $X_1, \ldots, X_k$, or equivalently, $X_1 = \ldots = X_k = \{1\}$. Let $\varphi$ be the sentence $\varphi_1 \wedge \ldots \wedge \varphi_k$. In this section and the next, we consider $\mathbb{E} = \mathbb{E}_\varphi \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi) = 1\}$ and we show that there are several cases in which $\text{maxent}\,\mathbb{E}$ can be found simply by conditionalising the equivocator function on $\varphi$.

Our first result directly applies Theorem 16:

**Corollary 19** *If $P_=(\cdot|\varphi)$ is a limit in entropy of $\mathbb{E}_\varphi$, then*

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

Note that the condition that $P_=(\cdot|\varphi)$ is a limit in entropy of $\mathbb{E}_\varphi$ presupposes that the probability function $P_=(\cdot|\varphi)$ is well defined, i.e., that $\varphi$ has positive measure, $P_=(\varphi) > 0$.

*Proof* $P_=(\cdot|\varphi)$ is contained in $\mathbb{E}_\varphi$ because $P_=(\varphi_i|\varphi) = 1$ for each $i = 1, \ldots, k$. Hence, Theorem 16 applies. □

**Corollary 20** *If $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large $n$, then*

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

*Proof* If $P_=(\cdot|\varphi) \in \mathbb{H}_n$ for sufficiently large $n$, then $P_=(\cdot|\varphi)$ is a limit in entropy of $\mathbb{E}_\varphi$. Hence, Corollary 19 applies. □

Corollary 20 is useful because where it applies it provides a particularly simple procedure for determining $\text{maxent}\,\mathbb{E}_\varphi$. Also, it shows that the move to the infinite does not disrupt agreement between the Maximum Entropy Principle and conditionalisation: as long as conditionalising on $\varphi$ maximises $n$-entropy for each sufficiently large $n$, it maximises entropy on the language as a whole. Because of its interest, we provide an alternative, more direct proof of Corollary 20 in Appendix 2.

*Example 21* Suppose we have a single categorical premiss $\varphi = \exists x Ux$, where $\mathcal{L}$ has a single unary predicate symbol $U$. $P_=(\exists x Ux) = P_=(\neg\forall x\neg Ux) = 1 -$

$\lim_{n\to\infty} P_=(\bigwedge_{i=1}^n \neg Ut_i) = 1 - \lim_{n\to\infty} 1/2^n = 1$. So, for all $\psi \in S\mathcal{L}$,

$$\begin{aligned} P_=(\psi|\varphi) &= P_=(\psi|\exists x Ux) = \frac{P_=(\psi \wedge \exists x Ux)}{P_=(\exists x Ux)} \\ &= P_=(\psi \wedge \exists x Ux) + P_=(\psi \wedge \neg\exists x Ux) \\ &= P_=(\psi). \end{aligned}$$

$P_= \in \mathbb{H}_1, \mathbb{H}_2, \ldots$, so Corollary 20 applies and maxent $\mathbb{E}_\varphi = \{P_=\}$.

*Example 22* Suppose we have categorical premisses $Ut_2 \to Vt_3, \forall x \exists y Wxy$, where $\mathcal{L}$ has unary predicate symbols $U$ and $V$ and a binary relation symbol $W$. Now $P_=(Ut_2 \to Vt_3) = 1 - P_=(Ut_2 \wedge \neg Vt_3) = 1 - 1/4 = 0.75$ and

$$\begin{aligned} P_=(\forall x \exists y Wxy) &= 1 - P_=(\exists x \forall y \neg Wxy) \\ &= 1 - \lim_{n\to\infty} P_=\left(\bigvee_{i=1}^n \forall y \neg W t_i y\right) \\ &\geq 1 - \lim_{n\to\infty} \sum_{i=1}^n P_=(\forall y \neg W t_i y) = 1. \end{aligned}$$

So $P_=((Ut_2 \to Vt_3) \wedge \forall x \exists y Wxy) = 0.75$, and $P_=(\cdot|(Ut_2 \to Vt_3) \wedge \forall x \exists y Wxy) = P_=(\cdot|Ut_2 \to Vt_3)$ (Proposition 13). This latter function is in $\mathbb{H}_3, \mathbb{H}_4, \ldots$, so Corollary 20 applies and maxent $\mathbb{E}_\varphi = \{P_=(\cdot|Ut_2 \to Vt_3)\}$.

Finally, we note an important consequence of Corollary 20:

**Theorem 23** *If $\varphi$ is satisfiable and logically equivalent to a quantifier-free sentence, then*

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

*Proof* Suppose $\theta$ is a quantifier-free sentence that is logically equivalent to $\varphi$. Since $\theta$ is logically equivalent to a satisfiable sentence, it is also satisfiable. The equivocator function is 'regular', i.e., it gives every satisfiable, quantifier-free sentence positive probability [32, Chapters 10 and 26]. Hence $P_=(\theta) > 0$ and, since $\varphi$ is logically equivalent to $\theta$, $P_=(\varphi) > 0$. Moreover, the logical equivalence of $\theta$ and $\varphi$ implies that $P_=(\cdot|\varphi) = P_=(\cdot|\theta)$ and $\mathbb{E}_\varphi = \mathbb{E}_\theta$, so maxent $\mathbb{E}_\varphi =$ maxent $\mathbb{E}_\theta$. As we shall show below, $P_=(\cdot|\theta) \in \mathbb{H}_n$ for all $n \geq N_\theta$, where $N_\theta$ is the greatest index of the constant symbols appearing in the quantifier-free sentence $\theta$. The theorem then follows because, by Corollary 20, maxent $\mathbb{E}_\varphi =$ maxent $\mathbb{E}_\theta = \{P_=(\cdot|\theta)\} = \{P_=(\cdot|\varphi)\}$.

It remains to show that $P_=(\cdot|\theta) \in \mathbb{H}_n$ for all $n \geq N_\theta$.

There are two cases: either $P_=(\theta) = 1$ or $P_=(\theta) < 1$.

If $P_=(\theta) = 1$ then $P_= \in \mathbb{E}_\theta = \{P \in \mathbb{P} : P(\theta) = 1\}$ and $P_=(\cdot|\theta) = P_=(\cdot)$. Since $P_{=\restriction\mathcal{L}_n}$ is the unique probability function on $\mathcal{L}_n$ with maximum $n$-entropy and $P_= \in \mathbb{E}_\theta$, $P_= \in \mathbb{H}_n$ for all $n \geq N_\theta$, as required.

So suppose that $P_=(\theta) < 1$ and suppose for contradiction that $P_=(\cdot|\theta) \notin \mathbb{H}_m$ for some $m \geq N_\theta$. Then there is some $R \in \mathbb{E}_\theta$ such that $H_m(R) > H_m(P_=(\cdot|\theta))$. That $R \in \mathbb{E}_\theta$ implies $R(\theta) = 1$ and $R(\cdot|\theta) = R(\cdot)$.

Now define probability function $Q$ by:

$$Q(\cdot) = R(\cdot|\theta)P_=(\theta) + P_=(\cdot|\neg\theta)P_=(\neg\theta).$$

Here $P_=(\cdot|\neg\theta)$ is a well-defined probability function because by assumption, $P_=(\theta) < 1$, i.e., $P_=(\neg\theta) > 0$. Since $Q$ is a convex combination of probability functions, it is a well-defined probability function.

However, we find that $Q$ has greater $m$-entropy than the equivocator function:

$$\begin{aligned}
H_m(Q) &= -\sum_{\omega\in\Omega_m} Q(\omega)\log Q(\omega) \\
&= -\sum_{\omega\in\Omega_m} Q(\omega\wedge\theta)\log Q(\omega\wedge\theta) - \sum_{\omega\in\Omega_m} Q(\omega\wedge\neg\theta)\log Q(\omega\wedge\neg\theta) \\
&= -\sum_{\omega\in\Omega_m} Q(\omega|\theta)Q(\theta)\log\left(Q(\omega|\theta)Q(\theta)\right) - \sum_{\omega\in\Omega_m} Q(\omega|\neg\theta)Q(\neg\theta)\log\left(Q(\omega|\neg\theta)Q(\neg\theta)\right) \\
&= -\sum_{\omega\in\Omega_m} Q(\omega|\theta)Q(\theta)\log Q(\theta) - \sum_{\omega\in\Omega_m} Q(\omega|\neg\theta)Q(\neg\theta)\log Q(\neg\theta) \\
&\quad -\sum_{\omega\in\Omega_m} Q(\omega|\theta)Q(\theta)\log Q(\omega|\theta) - \sum_{\omega\in\Omega_m} Q(\omega|\neg\theta)Q(\neg\theta)\log Q(\omega|\neg\theta) \\
&= -Q(\theta)\log Q(\theta)\sum_{\omega\in\Omega_m} Q(\omega|\theta) - Q(\neg\theta)\log Q(\neg\theta)\sum_{\omega\in\Omega_m} Q(\omega|\neg\theta) \\
&\quad -\sum_{\omega\in\Omega_m} Q(\omega|\theta)Q(\theta)\log Q(\omega|\theta) - \sum_{\omega\in\Omega_m} Q(\omega|\neg\theta)Q(\neg\theta)\log Q(\omega|\neg\theta) \\
&= -Q(\theta)\log Q(\theta) - Q(\neg\theta)\log Q(\neg\theta) \\
&\quad -\sum_{\omega\in\Omega_m} Q(\omega|\theta)Q(\theta)\log Q(\omega|\theta) - \sum_{\omega\in\Omega_m} Q(\omega|\neg\theta)Q(\neg\theta)\log Q(\omega|\neg\theta) \\
&= -P_=(\theta)\log P_=(\theta) - P_=(\neg\theta)\log P_=(\neg\theta) \\
&\quad -P_=(\theta)\sum_{\omega\in\Omega_m} R(\omega|\theta)\log R(\omega|\theta) - P_=(\neg\theta)\sum_{\omega\in\Omega_m} P_=(\omega|\neg\theta)\log P_=(\omega|\neg\theta) \\
&= -P_=(\theta)\log P_=(\theta) - P_=(\neg\theta)\log P_=(\neg\theta) \\
&\quad -P_=(\theta)\sum_{\omega\in\Omega_m} R(\omega)\log R(\omega) - P_=(\neg\theta)\sum_{\omega\in\Omega_m} P_=(\omega|\neg\theta)\log P_=(\omega|\neg\theta) \\
&> -P_=(\theta)\log P_=(\theta) - P_=(\neg\theta)\log P_=(\neg\theta) \\
&\quad -P_=(\theta)\sum_{\omega\in\Omega_m} P_=(\omega|\theta)\log P_=(\omega|\theta) - P_=(\neg\theta)\sum_{\omega\in\Omega_m} P_=(\omega|\neg\theta)\log P_=(\omega|\neg\theta) \\
&= H_m(P_=).
\end{aligned}$$

The sixth equality holds in virtue of the fact that $Q(\cdot|\theta)$ is a probability function and $\Omega_m$ is a partition of sentences, so $\sum_{\omega\in\Omega_m} Q(\omega|\theta) = 1$.

However, that $H_m(Q) > H_m(P_=)$ contradicts the fact that, for each $n$, $P_{=\restriction\mathcal{L}_n}$ is the unique probability function on $\mathcal{L}_n$ that maximises $n$-entropy. Thus $P_=(\cdot|\theta) \in \mathbb{H}_n$ for all $n \geq N_\theta$, as required. □

This result can be thought of as an analogue of [39, Result 1], which demonstrates agreement between the Maximum Entropy Principle and conditionalisation in the case in which the domain is finite. In the next section, we show that this result can be extended to the situation in which $\varphi$ is not logically equivalent to a quantifier-free sentence.

The above proof can be understood as follows. $\varphi$ and $\theta$ being logically equivalent guarantees that (i) $P(\cdot|\varphi) = P(\cdot|\theta)$ and (ii) maxent $\mathbb{E}_\varphi$ = maxent $\mathbb{E}_\theta$, while $\theta$ being quantifier-free ensures that (iii) maxent $\mathbb{E}_\theta = \{P_=(\cdot|\theta)\}$, thanks to Corollary 20. Putting these three facts together, we have that maxent $\mathbb{E}_\varphi$ = maxent $\mathbb{E}_\theta = \{P_=(\cdot|\theta)\} = \{P_=(\cdot|\varphi)\}$ for any $\varphi$ that is logically equivalent to quantifier-free $\theta$. However, (i) does not require full logical equivalence—it is sufficient that $\varphi$ and $\theta$ are inductively equivalent, i.e., $\mid\!\approx^{\circ} \varphi \leftrightarrow \theta$, by Proposition 13. The plan of the next section is to show that maxent $\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}$ for any $\varphi$ that is not an inductive contradiction, by finding some quantifier-free $\theta$ which is inductively equivalent to $\varphi$, and then demonstrating that (ii) holds. (iii) again holds because $\theta$ is quantifier-free, and (i)-(iii) then yield the desired conclusion that maxent $\mathbb{E}_\varphi$ = maxent $\mathbb{E}_\theta = \{P_=(\cdot|\theta)\} = \{P_=(\cdot|\varphi)\}$ for any $\varphi$ that is not an inductive contradiction.

## 5 Bayesian Conditionalisation and Support

This section demonstrates more general agreement between the maximal entropy approach and Bayesian conditionalisation. As above, we consider categorical sentences $\varphi_1, \ldots, \varphi_k$ and abbreviate $\varphi_1 \wedge \ldots \wedge \varphi_k$ by $\varphi$. First we introduce a quantifier-free sentence, the *support* of $\varphi$, which we will show is inductively equivalent to $\varphi$ (see Proposition 28). This will allow us to use the strategy outlined at the end of the last section to show that the maximal entropy approach agrees with Bayesian conditionalisation whenever $\varphi$ has positive measure.

**Definition 24** (Support) Let sentence $\varphi^n$ be the disjunction of those $n$-states $\omega$ that are inductively consistent with $\varphi$, i.e., $n$-states $\omega$ such that $\not\mid\!\approx^{\circ} \neg(\omega \wedge \varphi)$. Equivalently, these are the $n$-states $\omega$ such that $\omega \wedge \varphi$ has positive measure. Thus,

$$\varphi^n \stackrel{\text{df}}{=} \bigvee \{\omega \in \Omega_n : P_=(\omega \wedge \varphi) > 0\}.$$

If there are no $n$-states inductively consistent with $\varphi$, we take $\varphi^n$ to be an arbitrary contradiction on $\mathcal{L}_n$.

We call $\varphi^n$ the *inductive support* of $\varphi$ on $\mathcal{L}_n$, or simply the *n-support* of $\varphi$. $\varphi^{N_\varphi}$ will be referred to as the *support* of $\varphi$.[3] We use $|\varphi^n|$ to denote the number of $n$-states in the $n$-support $\varphi^n$, i.e., the number of $n$-states inductively consistent with $\varphi$.

Our main result of this section, Theorem 34, will show that when $\varphi$ has positive measure, the maximal entropy function is the equivocator function conditional on

[3] Recall that $N_\varphi$ is the greatest index of the constants appearing in $\varphi$, or 1 if no constants appear in $\varphi$.

$\varphi$, or, equivalently, the equivocator conditional on the support of $\varphi$. This provides a straightforward way of determining the maximal entropy function in the case in which the premisses are categorical inductive non-contradictions.

We will first prove some technical lemmas to which the main result will appeal. The first lemma invokes the concept of exchangeability:

**Definition 25** (Constant Exchangeability) Let $\theta(x_1, x_2, \ldots, x_l)$ be a formula of $\mathcal{L}$ that does not contain constants. A probability function $P$ on $S\mathcal{L}$ satisfies *constant exchangeability* if and only if for all such $\theta$ and all sets of pairwise distinct constants $t_1, t_2, \ldots, t_l$, and $t'_1, t'_2, \ldots, t'_l$ it holds that

$$P(\theta(t_1, t_2, \ldots, t_l)) = P(\theta(t'_1, t'_2, \ldots, t'_l)).$$

Equivalently, for all $n \in \mathbb{N}$ and all $n$-states $\omega_n, \nu_n \in \Omega_n$, if $\omega_n$ can be obtained from $\nu_n$ by a permutation of the first $n$ constants then $P(\omega_n) = P(\nu_n)$.

Paris and Vencovská [32, Corollary 6.2] show the following: if probability function $P$ on $S\mathcal{L}$ satisfies constant exchangeability and $P(\varphi \wedge \psi) = P(\varphi) \cdot P(\psi)$, whenever $\varphi, \psi$ are quantifier-free sentences of $\mathcal{L}$ that mention no constants in common, then $P(\varphi \wedge \psi) = P(\varphi) \cdot P(\psi)$ for any sentences $\varphi, \psi$ of the language $\mathcal{L}$ which do not mention any constants in common. This has an important consequence:

**Proposition 26** (Zero-one law for constant-free sentences) *Every constant-free sentence has measure 0 or 1.*

*Proof* $P_=$ satisfies constant exchangeability and the assumption of Paris and Vencovská [32, Corollary 6.2] is thus satisfied. Let $\varphi$ be a sentence that does not mention any constant. Then $\varphi, \varphi$ are two sentences that do not mention any constants in common. Since probability functions assign logically equivalent sentences the same probability we now easily find

$$P_=(\varphi) = P_=(\varphi \wedge \varphi) = P(\varphi) \cdot P(\varphi).$$

So, $P_=(\varphi) = P_=(\varphi)^2$. This means that $P_=(\varphi)$ must be either zero or one. □

Hence, every inductively consistent constant-free sentence is an inductive tautology: $P_=(\varphi) > 0$ for constant-free $\varphi$ implies that $P_=(\varphi) = 1$.

We are obliged to Jeff Paris for pointing out the following analogue of Paris and Vencovská [32, Corollary 6.2] and Proposition 28 which follows from it.

**Lemma 27** *Let $\omega_n$ be an $n$-state and suppose that the probability function $P$ on $S\mathcal{L}$ satisfies constant exchangeability and $P(\varphi \wedge \psi|\omega_n) = P(\varphi|\omega_n) \cdot P(\psi|\omega_n)$ for all pairs of quantifier-free sentences $\varphi, \psi$ with shared constants among $\{t_1, \ldots, t_l\}$, $l \leq n$. Then $P(\varphi \wedge \psi|\omega_n) = P(\varphi|\omega_n) \cdot P(\psi|\omega_n)$ for all $\varphi, \psi \in S\mathcal{L}$ whose shared constants are among $\{t_1, \ldots, t_l\}$.*

*Proof* The result follows by a straightforward adaptation of the proof of Paris and Vencovská [32, Corollary 6.2] and proceeds by induction on the quantifier complexity of $\varphi \wedge \psi$ when written in Prenex Normal Form.

The result holds by assumption when $\varphi \wedge \psi$ is quantifier free. For the induction step it is sufficient to consider

$$\exists x_1, \ldots, x_r \theta(x_1, \ldots, x_r, \vec{t}) \wedge \exists x_1, \ldots, x_s \psi(x_1, \ldots, x_s, \vec{t'}) \tag{1}$$

where all constants appearing in both $\vec{t}$ and $\vec{t'}$ are included in $\{t_1, \ldots, t_l\}$. To see that this is sufficient notice that by Eq. 1,

$$\begin{aligned} P(\exists\vec{x}\theta \ \wedge \ \forall\vec{y}\psi) &= P(\exists\vec{x}\theta) - P(\exists\vec{x}\theta \wedge \neg\forall\vec{y}\psi) = P(\exists\vec{x}\theta) - P(\exists\vec{x}\theta \wedge \exists\vec{y}\neg\psi) \\ &= P(\exists\vec{x}\theta) - (P(\exists\vec{x}\theta) \cdot P(\exists\vec{y}\neg\psi)) = P(\exists\vec{x}\theta) - (P(\exists\vec{x}\theta) \cdot (1 - P(\forall\vec{y}\psi))) \\ &= P(\exists\vec{x}\theta) - P(\exists\vec{x}\theta) + P(\exists\vec{x}\theta)P(\forall\vec{y}\psi) \\ &= P(\exists\vec{x}\theta)P(\forall\vec{y}\psi) \end{aligned}$$

and,

$$\begin{aligned} P(\forall\vec{x}\theta \wedge \forall\vec{y}\psi) &= 1 - P(\exists\vec{x}\neg\theta \vee \exists\vec{y}\neg\psi) \\ &= 1 - P(\exists\vec{x}\neg\theta) - P(\exists\vec{y}\neg\psi) + P(\exists\vec{x}\neg\theta \wedge \exists\vec{y}\neg\psi) \\ &= 1 - P(\exists\vec{x}\neg\theta) - P(\exists\vec{y}\neg\psi) + P(\exists\vec{x}\neg\theta) \cdot P(\exists\vec{y}\neg\psi) \\ &= P(\forall\vec{x}\theta) + P(\forall\vec{y}\psi) - 1 + (1 - P(\forall\vec{x}\theta)) \cdot (1 - P(\forall\vec{y}\psi)) \\ &= P(\forall\vec{x}\theta) \cdot P(\forall\vec{y}\psi). \end{aligned}$$

To show (1) let $u_1, u_2, u_3, \ldots$ be distinct constants containing those in $\vec{t}$ and $u'_1, u'_2, u'_3, \ldots$ distinct constants containing those in $\vec{t'}$ such that $\{u_1, u_2, u_3, \ldots\}$ and $\{u'_1, u'_2, u'_3, \ldots\}$ are disjoint except for the constants shared in $\vec{t}$ and $\vec{t'}$.

By Paris and Vencovská [32, Lemma 6.1],

$$\lim_{n\to\infty} P\left(\left(\bigvee_{i_1,\ldots,i_r \leq n} \theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t})\right) \leftrightarrow \exists x_1, \ldots, x_r \theta(x_1, \ldots, x_r, \vec{t}) \,|\omega_n\right) = 1$$

and

$$\lim_{n\to\infty} P\left(\left(\bigvee_{i_1,\ldots,i_s \leq n} \psi(u'_{i_1}, u'_{i_2}, \ldots, u'_{i_s}, \vec{t'})\right) \leftrightarrow \exists x_1, \ldots, x_s \psi(x_1, \ldots, x_s, \vec{t'}) \,|\omega_n\right) = 1.$$

Then for every $\epsilon > 0$ there is $N$ large enough such that for all $n \geq N$

$$P\left(\left(\bigvee_{i_1,\ldots,i_r \leq n} \theta(u_{i_1}, u_{i_2}, \ldots, u_{i_r}, \vec{t})\right) \leftrightarrow \exists x_1, \ldots, x_r \theta(x_1, \ldots, x_r, \vec{t}) \,|\omega_n\right) > 1 - \frac{\epsilon}{4}$$

and

$$P\left(\left(\bigvee_{i_1,\ldots,i_s \leq n} \psi(u'_{i_1}, u'_{i_2}, \ldots, u'_{i_s}, \vec{t'})\right) \leftrightarrow \exists x_1, \ldots, x_s \psi(x_1, \ldots, x_t, \vec{t'}) \,|\omega_n\right) > 1 - \frac{\epsilon}{4}$$

by Paris and Vencovská [32, Lemma 3.7],

$$P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\wedge\exists x_1,\ldots,x_s\psi(x_1,\ldots,x_s,\vec{t'})\,|\omega_n\right)-$$

$$P\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1},u_{i_2},\ldots,u_{i_r},\vec{t})\wedge\bigvee_{i_1,\ldots,i_s\le n}\psi(u'_{i_1},u'_{i_2},\ldots,u'_{i_s},\vec{t'})\,|\omega_n\right)<\frac{\epsilon}{2}.$$

But

$$P\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1},u_{i_2},\ldots,u_{i_r},\vec{t})\wedge\bigvee_{i_1,\ldots,i_s\le n}\psi(u'_{i_1},u'_{i_2},\ldots,u'_{i_s},\vec{t'})\,|\omega_n\right)$$

equals

$$P\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1},u_{i_2},\ldots,u_{i_r},\vec{t})\,|\omega_n\right)\cdot P\left(\bigvee_{i_1,\ldots,i_s\le n}\psi(u'_{i_1},u'_{i_2},\ldots,u'_{i_s},\vec{t'})\,|\omega_n\right)$$

by the induction hypothesis, and taking $n$ large enough we have:

$$P\left(\bigvee_{i_1,\ldots,i_r\le n}\theta(u_{i_1},u_{i_2},\ldots,u_{i_r},\vec{t})\,|\omega_n\right)\cdot P\left(\bigvee_{i_1,\ldots,i_s\le n}\psi(u'_{i_1},u'_{i_2},\ldots,u'_{i_s},\vec{t'})\,|\omega_n\right)-$$

$$P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\,|\omega_n\right)\cdot P\left(\exists x_1,\ldots,x_s\psi(x_1,\ldots,x_s,\vec{t'})\,|\omega_n\right)<\frac{\epsilon}{2}$$

and thus

$$\begin{aligned}&P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\wedge\exists x_1,\ldots,x_s\psi(x_1,\ldots,x_s,\vec{t'})\,|\omega_n\right)\\&\quad-P\left(\exists x_1,\ldots,x_r\theta(x_1,\ldots,x_r,\vec{t})\,|\omega_n\right)\cdot P\left(\exists x_1,\ldots,x_s\psi(x_1,\ldots,x_s,\vec{t'})\,|\omega_n\right)\\&<\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon\end{aligned}$$

which gives the required result. □

**Proposition 28** *$\varphi$ and $\varphi^n$ are inductively equivalent for all $n\ge N_\varphi$.*

*Proof* Consider two sentences $\varphi,\psi\in S\mathcal{L}$. These mention at most the first $N:=N_{\varphi\wedge\psi}$ constants. From Lemma 27 we obtain that for all $n\ge N$ and all $\omega_n\in\Omega_n$,

$$P_=(\varphi\wedge\psi|\omega_n)=P_=(\varphi|\omega_n)\cdot P_=(\psi|\omega_n).$$

Putting $\psi=\varphi$ [32, p. 53] we obtain $P_=(\varphi\wedge\varphi|\omega_n)=P_=(\varphi|\omega_n)=P_=(\varphi|\omega_n)^2$ and so

$$P_=(\varphi|\omega_n)\in\{0,1\}.$$

Using the definition of conditional probability we find

$$P_=(\varphi\wedge\omega_n)=\begin{cases}0 & \text{if and only if } P_=(\varphi|\omega_n)=0\\ P_=(\omega_n) & \text{if and only if } P_=(\varphi|\omega_n)>0.\end{cases}\tag{2}$$

So,

$$
\begin{aligned}
P_{=}(\varphi) &= P_{=}\left(\varphi \wedge \bigvee_{\omega_n \in \Omega_n} \omega_n\right) = P_{=}\left(\varphi \wedge \bigvee_{\substack{\omega_n \in \Omega_n \\ P_{=}(\varphi \wedge \omega_n) > 0}} \omega_n\right) = P_{=}(\varphi \wedge \varphi^n) \\
&= \sum_{\substack{\omega_n \in \Omega_n \\ P_{=}(\varphi \wedge \omega_n) > 0}} P_{=}(\varphi \wedge \omega_n) \\
&\overset{(2)}{=} \sum_{\substack{\omega_n \in \Omega_n \\ P_{=}(\varphi \wedge \omega_n) > 0}} P_{=}(\omega_n) = P_{=}\left(\bigvee_{\substack{\omega_n \in \Omega_n \\ P_{=}(\varphi \wedge \omega_n) > 0}} \omega_n\right) \\
&= P_{=}(\varphi^n). \qquad (3)
\end{aligned}
$$

So,

$$
\begin{aligned}
P_{=}(\neg\varphi \wedge \neg\varphi^n) &= P_{=}(\neg\varphi) + P_{=}(\neg\varphi^n) - P_{=}(\neg\varphi \vee \neg\varphi^n) \\
&= P_{=}(\neg\varphi) + 1 - P_{=}(\varphi^n) - 1 + P_{=}(\varphi \wedge \varphi^n) \\
&= P_{=}(\neg\varphi). \qquad (4)
\end{aligned}
$$

Finally, let us note that

$$
\begin{aligned}
P_{=}\left(\varphi \leftrightarrow \varphi^n\right) &= P_{=}\left(\varphi \wedge \varphi^n\right) + P_{=}\left(\neg\varphi \wedge \neg\varphi^n\right) \\
&\overset{(3)\text{ and }(4)}{=} P_{=}(\varphi) + P_{=}(\neg\varphi) = 1.
\end{aligned}
$$

□

Note that the proportion $\frac{|\varphi^n|}{|\Omega_n|}$ of $n$-states in the $n$-support of a sentence $\varphi$ eventually equals the measure of $\varphi$. This is because $\frac{|\varphi^n|}{|\Omega_n|} = P_{=}(\varphi^n) = P_{=}(\varphi)$ for $n \geq N_\varphi$.

By Proposition 13 we have:

**Corollary 29** *If $\varphi$ has positive measure, then $P_{=}(\cdot|\varphi) = P_{=}(\cdot|\varphi^n)$ for all $n \geq N_\varphi$.*

Moreover,

**Corollary 30** *For all $k \geq 1$, $\models \varphi^{N_\varphi + k} \leftrightarrow \varphi^{N_\varphi}$ and $P_{=}(\cdot|\varphi^{N_\varphi+k}) = P_{=}(\cdot|\varphi^{N_\varphi})$.*

*Proof* By Corollary 29 for all $k \geq 0$, $P_{=}(\cdot|\varphi^{N_\varphi+k}) = P_{=}(\cdot|\varphi)$. This entails $P_{=}(\cdot|\varphi^{N_\varphi+k}) = P_{=}(\cdot|\varphi^{N_\varphi})$ for all $k \geq 1$.

Note that $\varphi^{N_\varphi+k}$ is quantifier-free. Let $\chi$, $\psi$ be quantifier-free and satisfiable, then the probability function $P_{=}(\cdot|\psi)$ is equal to the probability function $P_{=}(\cdot|\chi)$ if and only if $\psi$ and $\chi$ are logically equivalent. Clearly, if $\psi$ and $\chi$ are logically equivalent, then these probability functions are equal. Furthermore, if $\psi$ and $\chi$ are not logically equivalent, then without loss of generality take $\chi$ to be non-tautologous and assume that $\psi$ does not entail $\chi$. Since $\psi$ and $\chi$ are quantifier-free sentences $P_=(\psi|\psi) = 1 > P_=(\chi|\psi)$ follows[4].

Letting $\psi = \varphi^N$ and $\chi = \varphi^{N_\varphi+k}$ we conclude that $\models \varphi^{N_\varphi+k} \leftrightarrow \varphi^{N_\varphi}$. □

Note that every $(N_\varphi + k)$ - state $\omega_{N_\varphi+k}$ extending a state in $\varphi^{N_\varphi}$ is such that $P_=(\omega_{N_\varphi+k} \wedge \varphi) > 0$.

**Corollary 31** *If* $\omega_{N_\varphi+k} \models \varphi^{N_\varphi}$*, then* $\omega_{N_\varphi+k} \models \varphi^{N_\varphi+k}$.

*Proof* We let $N := N_\varphi$. Notice that by Corollary 30, if $\omega_N \in \Omega_N$ appears in $\varphi^N$, then any extension of $\omega_N$ to $\mathcal{L}_m$ (an $m$-state $\omega_m \in \Omega_m$ such that $\omega_m \models \omega_N$ with $m = N + k > N$) will appear in $\varphi^m$. To be more precise, for all $\omega_N \in \Omega_N$ with $\omega_N \models \varphi_N$ and for all $m \geq N$ and $\omega_m \in \Omega_m$, if $\omega_m \models \omega_N$, then $\omega_m \models \varphi^m$. To see this suppose $\omega_N \models \varphi^N$, $\omega'_m \models \omega_N$ but $\omega'_m \not\models \varphi^m$. Then by definition of $P_=$ we have $P_=(\omega_N \,|\, \varphi^N)$, $P_=(\omega'_m \,|\, \omega_N) \neq 0$. Then $0 < P_=(\omega'_m \,|\, \varphi^N) = P_=(\omega'_m \,|\, \varphi^m) = 0$, where the first equality is given by Corollary 30 and second equality is given by the assumption that $\omega'_m \not\models \varphi^m$. □

Consider a sentence $\psi$ with zero measure, $P_=(\psi) = 0$. Intuitively, $\psi$ is only true in few possible worlds.[5] One way to approach this intuition is by exploiting probability axiom P3 according to which the probability of a quantified sentence is the limit of probabilities of quantifier-free sentences. This suggests that—in the limit—only few $n$-states "converge" to $\psi$. So, if $P(\psi) = c > 0$, then $P$ has to assign a joint probability of close to $c$ to *few* $n$-states. That is, for $n$ large enough, there exists set of $n$-states $S_n$, with joint probability of almost $c$, that is arbitrarily small in comparison to the number of all $n$-states. The following result, for which we are obliged to Alena Vencovská, makes this precise.

**Lemma 32** (Concentration of probability on few $n$-states) *Let* $\psi$ *be such that* $P_=(\psi) = 0$ *and* $P(\psi) = c > 0$*, then for any* $\epsilon > 0$ *there exists some* $M \in \mathbb{N}$ *such that for all* $m \geq M$ *there exists a set of* $m$*-states,* $S_m$*, such that*

$$P\left(\bigvee_{\omega_m \in S_m} \omega_m\right) \geq (1-\epsilon)\cdot c \text{ and } \frac{|S_m|}{|\Omega_m|} < \epsilon.$$

[4]The assumption that both sentences are quantifier free is crucial here. For $\chi := \exists x U x$ and $\psi = \forall x(Vx \vee \neg Vx)$ we have $P_=(\psi|\psi) = 1 = P_=(\chi|\psi)$.

[5]More precisely, consider the set of term structures for $\mathcal{L}$ that have a countably infinite domain. Then this means that the proportion of those term structures that satisfy $\psi$ is negligible. But the term structures on a countably infinite domain can be determined as the limiting extensions of terms structures on finite subsets of the domain. This means that for asymptotically large $n$, there are only few term structures with a domain of size $n$ that can be extended to a term structure that satisfies $\psi$. Then dividing the probability mass between the term structures on the full domain in such a way as to assign a probability of $c > 0$ to $\psi$ should inevitably distribute a probability mass close to $c$ between few term structures on a finite subdomain of size $n$ for large $n$.

*Proof* First notice that if the result holds for some $m \in \mathbb{N}$ and a set of $m$-states $S_m$, then it also holds for the set of $m+1$ states $S_{m+1}$ defined as the extensions of $S_m$ to $\mathcal{L}_{m+1}$. Therefore, it is enough to show that result holds for some $m \in \mathbb{N}$.

Let $\mathcal{P} = \{P, P_=\}$. We first show that there exists some $m \in \mathbb{N}$ and a quantifier-free sentence $\chi \in S\mathcal{L}_m$ such that for all $Q \in \mathcal{P}$, $Q(\psi \leftrightarrow \chi) > 1 - \epsilon \cdot c$. (We can think of $\chi$ as a finite approximation of $\psi$.) We proceed by induction on the quantifier complexity; that is we proceed by induction on $n$ for $\psi \in \Sigma_n$ and $\psi \in \Pi_n$.

For the base case, $n = 0$, $\psi$ is quantifier free, and we can simply pick $\chi := \psi$.

For the induction step let $\psi = \forall \vec{x}\xi(x_1, \ldots, x_r) \in \Pi_g$ with $\xi \in \Sigma_{g-1}$ in prenex normal form. The case of $\psi = \exists \vec{x}\xi(\vec{x}) \in \Sigma_g$ is analogous.

By [32, Lemma 3.8] for all probability functions $Q$,

$$Q(\psi) = \lim_{n\to\infty} Q\left(\bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r})\right).$$

Let $n \in \mathbb{N}$ be large enough such that for all $Q \in \mathcal{P}$,

$$\left|Q(\psi) - Q\left(\bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r})\right)\right| < \frac{\epsilon}{2} \cdot c.$$

Now let $Q \in \mathcal{P}$. Notice that $\psi$ logically entails $\bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r})$ and thus

$$\begin{aligned}
&Q\left(\psi \leftrightarrow \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r})\right) = Q\left(\bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r}) \to \psi\right)\\
&= Q\left(\neg \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r})\right) + Q(\psi) - Q\left(\neg \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r}) \wedge \psi\right)\\
&= 1 - Q\left(\bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r})\right) + Q(\psi) - 0\\
&> 1 - \frac{\epsilon}{2} \cdot c.
\end{aligned} \tag{5}$$

By the induction hypothesis, for each $k_1, \ldots k_r \in \{1, \ldots, n\}$ there is a quantifier-free sentence $\lambda_{\vec{k}}(a_1, \ldots, a_{M(\vec{k})}) \in S\mathcal{L}_{M(\vec{k})}$ such that for all $Q \in \mathcal{P}$,

$$Q(\lambda_{\vec{k}} \leftrightarrow \xi(t_{k_1}, \ldots, t_{k_r})) > 1 - \frac{\epsilon}{2n^r} \cdot c. \tag{6}$$

Notice that we have following logical equivalences:

$$
\begin{aligned}
&\neg \left( \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(\vec{t_{k_i}}) \leftrightarrow \bigwedge_{j_1,\ldots,j_r=1}^{n} \lambda_{\vec{j}} \right) \\
&\equiv \left( \bigvee_{k_1,\ldots,k_r=1}^{n} \neg\xi(\vec{t_{k_i}}) \wedge \bigwedge_{j_1,\ldots,j_r=1}^{n} \lambda_{\vec{j}} \right) \vee \left( \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(\vec{t_{k_i}}) \wedge \bigvee_{j_1,\ldots,j_r=1}^{n} \neg\lambda_{\vec{j}} \right) \\
&\equiv \left( \bigvee_{k_1,\ldots,k_r=1}^{n} \left( \neg\xi(\vec{t_{k_i}}) \wedge \bigwedge_{j_1,\ldots,j_r=1}^{n} \lambda_{\vec{j}} \right) \right) \vee \left( \bigvee_{j_1,\ldots,j_r=1}^{n} \left( \neg\lambda_{\vec{j}} \wedge \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(\vec{t_{k_i}}) \right) \right)
\end{aligned}
$$

and that

$$
\begin{aligned}
&\left( \bigvee_{k_1,\ldots,k_r=1}^{n} \left( \neg\xi(\vec{t_{k_i}}) \wedge \bigwedge_{j_1,\ldots,j_r=1}^{n} \lambda_{\vec{j}} \right) \right) \vee \left( \bigvee_{j_1,\ldots,j_r=1}^{n} \left( \neg\lambda_{\vec{j}} \wedge \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(\vec{t_{k_i}}) \right) \right) \models \\
&\left( \bigvee_{k_1,\ldots,k_r=1}^{n} \neg\xi(\vec{t_{k_i}}) \wedge \lambda_{k_1,\ldots,k_r} \right) \vee \left( \bigvee_{j_1,\ldots,j_r=1}^{n} \xi(\vec{t_{k_j}}) \wedge \neg\lambda_{j_1,\ldots,j_r} \right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left( \bigvee_{k_1,\ldots,k_r=1}^{n} \neg\xi(\vec{t_{k_i}}) \wedge \lambda_{k_1,\ldots,k_r} \right) \vee \left( \bigvee_{j_1,\ldots,j_r=1}^{n} \xi(\vec{t_{k_j}}) \wedge \neg\lambda_{j_1,\ldots,j_r} \right) \\
&\equiv \bigvee_{k_1,\ldots,k_r=1}^{n} \neg(\lambda_{k_1,\ldots,k_r} \leftrightarrow \xi(\vec{t_{k_i}}))
\end{aligned}
$$

where we write $\xi(\vec{t_{k_i}})$ for $\xi(t_{k_1}, \ldots, t_{k_r})$.

Then

$$
\begin{aligned}
&Q\left( \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r}) \leftrightarrow \bigwedge_{j_1,\ldots,j_r=1}^{n} \lambda_{\vec{j}} \right) \\
&= 1 - Q\left( \neg \left( \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1}, \ldots, t_{k_r}) \leftrightarrow \bigwedge_{j_1,\ldots,j_r=1}^{n} \lambda_{\vec{j}} \right) \right) \\
&\geq 1 - Q\left( \bigvee_{k_1,\ldots,k_r=1}^{n} \neg(\lambda_{k_1,\ldots,k_r} \leftrightarrow \xi(t_{k_1}, \ldots, t_{k_r})) \right) \\
&\overset{(6)}{>} 1 - n^r \frac{\epsilon}{2n^r} \cdot c \\
&= 1 - \frac{\epsilon}{2} \cdot c.
\end{aligned} \tag{7}
$$

Let $\Xi = \bigwedge_{k_1,\ldots,k_r=1}^{n} \xi(t_{k_1},\ldots,t_{k_r})$, and $\Lambda = \bigwedge_{k_1,\ldots,k_r=1}^{n} \lambda_{\vec{k}}$. Then by Eqs. 5 and 7 we have

$$Q(\psi \leftrightarrow \Xi) > 1 - \frac{\epsilon}{2} \cdot c,$$

and

$$Q(\Xi \leftrightarrow \Lambda) > 1 - \frac{\epsilon}{2} \cdot c.$$

And we have

$$\begin{aligned} Q\ (\psi \leftrightarrow \Lambda) &= Q(\psi \wedge \Lambda) + Q(\neg\psi \wedge \neg\Lambda) \\ &= Q(\psi \wedge \Lambda \wedge \Xi) + Q(\psi \wedge \Lambda \wedge \neg\Xi) + Q(\neg\psi \wedge \neg\Lambda \wedge \Xi) + Q(\neg\psi \wedge \neg\Lambda \wedge \neg\Xi) \\ &\geq Q(\psi \wedge \Lambda \wedge \Xi) - Q(\psi \wedge \Lambda \wedge \neg\Xi) - Q(\neg\psi \wedge \neg\Lambda \wedge \Xi) + Q(\neg\psi \wedge \neg\Lambda \wedge \neg\Xi). \end{aligned}$$

Noticing that

$$Q(\psi \wedge \Lambda \wedge \Xi) = Q(\psi \wedge \Xi) - Q(\psi \wedge \neg\Lambda \wedge \Xi)$$

and

$$Q(\neg\psi \wedge \neg\Lambda \wedge \neg\Xi) = Q(\neg\psi \wedge \neg\Xi) - Q(\neg\psi \wedge \Lambda \wedge \neg\Xi)$$

we get

$$Q\ (\psi \leftrightarrow \Lambda) \geq Q(\psi \leftrightarrow \Xi) - Q(\neg(\Lambda \leftrightarrow \Xi)) > 1 - \epsilon \cdot c. \tag{8}$$

Since Eq. 8 holds for all $Q \in \mathcal{P} = \{P_=, P\}$ and $P_=(\psi) = 0$,

$$\begin{aligned} 1 - \epsilon \cdot c \ < \ P_=(\psi \leftrightarrow \Lambda) &= P_=(\psi \wedge \Lambda) + P_=(\neg\psi \wedge \neg\Lambda) \\ &= P_=(\neg\psi \wedge \neg\Lambda) = 1 - P_=(\psi \vee \Lambda) \leq 1 - P_=(\Lambda) \end{aligned}$$

and thus $P_=(\Lambda) < \epsilon \cdot c \leq \epsilon$.

Now let $m = \max\{M(\vec{k}) \,|\, \vec{k} \in \{1,\ldots,n\}^r\}$, then $\Lambda \in S\mathcal{L}_m$ and since $\Lambda$ is quantifier-free, there is a set of $m$-states $S_m$, such that

$$\models \Lambda \leftrightarrow \bigvee_{\omega_m \in S_m} \omega_m$$

and we have $\frac{|S_m|}{|\Omega_m|} = P_=(\Lambda) < \epsilon$. Note that $S_m$ is the set of $m$-states entailing $\Lambda$[6].

[6]One might think that the following statement can play a similar role to that played by $S_m$. Let $\varphi_0^n \stackrel{\text{df}}{=} \bigvee\{\omega \in \Omega_n : P_=(\omega \wedge \varphi) = 0, \not\models \neg(\omega \wedge \varphi)\}$, i.e., the disjunction of $n$-states deductively but not inductively consistent with $\varphi$. (If there are no such states, take $\varphi_0^n$ to be an arbitrary contradiction on $\mathcal{L}_n$.)

Now suppose that $\varphi$ has measure zero and that $P(\varphi) = c$. Since $\varphi$ has measure zero, $\varphi^n$ is a contradiction on $\mathcal{L}_n$. Hence,

$$c = P(\varphi) = P(\varphi \wedge \varphi^n) + P(\varphi \wedge \varphi_0^n) = P(\varphi \wedge \varphi_0^n),$$

so $P(\varphi_0^n) \geq c$. $P$ must concentrate probability at least $c$ on $\varphi_0^n$.

Thus the question arises as to whether $P_=(\varphi_0^n) \longrightarrow 0$ as $n \longrightarrow \infty$. This would imply that $\frac{|\varphi_0^n|}{|\Omega_n|} = P_=(\varphi_0^n) \longrightarrow 0$ as $n \longrightarrow \infty$, in which case $\varphi_0^n$ would represent an increasingly negligible number of states.

However, it turns out that while this last condition holds true for some measure-zero $\varphi$, e.g., $\forall x U_1 x$, it does not hold true for all such sentences (see Section 9). For example, in the case of $\exists x \forall y U_2 xy$, which also has zero measure, $P_=(\varphi_0^n) = 1$ for all $n$.

Furthermore,

$$P\left(\bigvee_{\omega_m \in S_m} \omega_m\right) = P(\Lambda) \geq P(\Lambda \wedge \psi) = P(\Lambda \wedge \psi) + P(\neg\Lambda \wedge \neg\psi) - P(\neg\Lambda \wedge \neg\psi)$$
$$= P(\psi \leftrightarrow \Lambda) - P(\neg\Lambda \wedge \neg\psi)$$
$$> 1 - \epsilon \cdot c - P(\neg\Lambda \wedge \neg\psi)$$
$$\geq 1 - \epsilon \cdot c - P(\neg\psi)$$
$$= P(\psi) - \epsilon \cdot c = c - \epsilon \cdot c = c \cdot (1 - \epsilon).$$

□

The next Lemma shows that any maximal entropy function must assign probability one to the support $\varphi^{N_\varphi}$ of $\varphi$ (and thus to the $n$-support $\varphi^n$ for $n \geq N_\varphi$). Note that this lemma does not prove the existence of a maximal entropy function.

**Lemma 33** *Let $\varphi \in S\mathcal{L}$ with $P_=(\varphi) \in (0, 1]$. If $P \in \mathbb{E}$ with $P(\varphi^n) < 1$ for some $n \geq N_\varphi$, then $P_=(\cdot \mid \varphi^n)$ has greater entropy than $P$.*

*Proof* Let $N := N_\varphi$. If $P_=(\varphi) = 1$, then $P_=(\cdot|\varphi^N) = P_= \in \mathbb{E}$. It suffices to recall that the equivocator has greater entropy than all other probability functions (Example 10).

Now consider $0 < P_=(\varphi) < 1$.

Since $\varphi^N$ and $\varphi^n$ are logically equivalent for $n \geq N$ (Corollary 30) and since probability functions respect logical equivalence, $P(\varphi^N) < 1$ follows from the assumption that $P(\varphi^n) < 1$.

So, let $P$ be such that $P(\varphi) = 1$ and $P(\varphi^N) < 1$, then $P(\varphi \wedge \neg\varphi^N) = c > 0$ for some $1 \geq c > 0$. Let $\psi := \varphi \wedge \neg\varphi^N$ and notice that by definition of $\varphi^N$, $P_=(\psi) = 0$. Let $\epsilon > 0$ and take $M$ and $S_M$ as given by Lemma 32 and let $K_M$ be the set of $M$-states in $\Omega_M \setminus S_M$ such that $P_=(\varphi \wedge \omega_M) > 0$ for $M \geq N$. Corollary 30 shows that $|K_M| = |\varphi^N| \frac{|\Omega_M|}{|\Omega_N|}$, since all $M$-states $\omega_M \in \Omega_M$ extending an $N$-state in $\varphi^N$ are such that $P_=(\varphi \wedge \omega_M) > 0$. Let $b_M = P(\bigvee_{S_M} \omega_M) \geq (1-\epsilon)c > 0$ and notice that since $P(\varphi) = 1$ we have $P(\bigvee_{K_M} \omega_M) = 1 - b_M$.

Then by convexity

$$H_M(P) \leq -b_M \log\left(\frac{b_M}{|S_M|}\right) - (1 - b_M)\log\left(\frac{1 - b_M}{|K_M|}\right)$$
$$= b_M \log(|S_M|) - b_M \log(b_M) + (1 - b_M)\log(|K_M|) - (1 - b_M)\log(1 - b_M).$$

The $M$-entropy of $P_=(\cdot|\varphi^N)$ is

$$H_M(P_=(\cdot|\varphi^N)) = -\sum_{\omega_M \models \varphi^N} \frac{1}{|K_M|}\log\left(\frac{1}{|K_M|}\right) = \log(|K_M|)$$
$$= \log\left(|\varphi^N| \cdot \frac{|\Omega_M|}{|\Omega_N|}\right). \quad (9)$$

We thus note

$$\frac{H_M(P) - H_M(P_=(\cdot|\varphi^N))}{\log(|K_M|)}$$
$$\leq b_M \log(|S_M| - |K_M|) - \frac{b_M \log(b_M) + (1-b_M)\log(1-b_M)}{\log(|K_M|)} + (1-b_M) - 1.$$

Now consider the three summands in turn. Since $\frac{|S_M|}{|K_M|} = \frac{|S_M|\cdot|\Omega_N|}{|\Omega_M|\cdot|\varphi^N|}$ becomes arbitrarily small by Lemma 32 and $1 \geq b_M > 0$, the first term is eventually less than zero. The second term goes to zero, since $K_M$ increases without bounds. Finally, $b_M \geq (1-\epsilon)c > 0$. This means that for all large enough $M$ it holds that $H_M(P) - H_M(P_=(\cdot|\varphi^N)) < 0$ and hence $H_M(P) < H_M(P_=(\cdot|\varphi^N))$. This entails that $P_=(\cdot|\varphi^N)$ has greater entropy than $P$. Thus, $P \notin \text{maxent}\,\mathbb{E}_\varphi$ since $P_=(\cdot|\varphi^N) \in \mathbb{E}$ (Corollary 29).

In particular, we note for later use that the sequence $f_n(P) := H_n(P_=(\cdot|\varphi^N)) - H_n(P)$ is bounded from below by $\frac{b_M}{2} \geq \frac{(1-\epsilon)\cdot c}{2} > 0$ for all large enough $n$. □

We are now in a position to present the main result of this section:

**Theorem 34** (Agreement with Bayesian Conditionalisation) *For all* $\varphi \in S\mathcal{L}$ *with* $P_=(\varphi) \in (0, 1]$ *and all* $n \geq N_\varphi$,

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\} = \{P_=(\cdot|\varphi^n)\} = \{P_=(\cdot|\varphi^{N_\varphi})\}.$$

*Proof* We let $N := N_\varphi$. We prove that $P_=(\cdot|\varphi^N)$ has greater entropy than every other probability function $P \in \mathbb{E}$.

By Corollary 29, $P_=(\cdot\,|\,\varphi^N) = P_=(\cdot\,|\,\varphi^n) = P_=(\cdot\,|\,\varphi)$ for all $n \geq N$. This establishes the two last equalities in the statement of the theorem.

Consider first the case of $P_=(\varphi) = 1$. In this case, the equivocator $P_=$ is in $\mathbb{E}$, and, since it is the probability function in $\mathbb{P}$ with maximal entropy, it is the unique member of $\text{maxent}\,\mathbb{E}_\varphi$. Moreover, since $P_=(\varphi) = 1$, $P_= = P_=(\cdot|\varphi)$.

Now consider $0 < P_=(\varphi) < 1$.

Let $P$ be a probability function with $P(\varphi) = 1$. If $P(\varphi^N) < 1$ then $P_=(\cdot\,|\,\varphi^N)$ has greater entropy than $P$ by Lemma 33.

If $P(\varphi^N) = 1$, on the other hand, but $P \neq P_=(\cdot|\varphi^N)$, then $P_=(\cdot|\varphi^N)$ has greater entropy than $P$ because $\varphi^N$ is quantifier-free and so by Theorem 23, $\text{maxent}\,\mathbb{E}_{\varphi^N} = \{P_=(\cdot|\varphi^N)\}$.

So, $P_=(\cdot|\varphi^N)$ has greater entropy than every other probability function $P \in \mathbb{E}$. □

*Example 35* For the premiss sentence $\varphi = (\exists x \forall y Uxy \wedge Ut_1t_1) \vee (\forall x \exists y \neg Uxy \wedge \neg Ut_1t_1)$,

$$\text{maxent}\,\mathbb{E}_\varphi = \{P_=(\cdot|\neg Ut_1t_1)\}.$$

*Proof* There is only one constant mentioned in $\varphi$, $t_1$. So, $N_\varphi = 1$. We here consider the simple case of the language containing only the relation symbol $U$. The general case follows from the fact that entropy maximisation is language invariant [27, Chapter 6].

There are two 1-states, $Ut_1t_1$ and $\neg Ut_1t_1$. $\varphi \wedge Ut_1t_1$ is logically equivalent to $\exists x\forall y Uxy \wedge Ut_1t_1$ and $P_=(\exists x\forall y Uxy \wedge Ut_1t_1) \leq P_=(\exists x\forall y Uxy) = 0$. $\varphi \wedge \neg Ut_1t_1$ is logically equivalent to $\forall x\exists y\neg Uxy \wedge \neg Ut_1t_1$. For this sentence it holds that $P_=(\forall x\exists y\neg Uxy \wedge \neg Ut_1t_1) = 0.5$. $\varphi_1$ is the disjunction of all 1-states $\omega_1 \in \Omega_1$ such that $P_=(\varphi \wedge \omega_1) > 0$. Thus, $\varphi^{N_\varphi} = \varphi_1 = \neg Ut_1t_1$. That maxent $\mathbb{E}_\varphi = \{P_=(\cdot|\neg Ut_1t_1)\}$ follows from Theorem 34. □

The following observation shows that the maximal entropy function not only has greatest entropy in the sense defined above, but also in a cumulative sense.

**Corollary 36** *If $\varphi$ has positive measure, then for all $P \in \mathbb{E}_\varphi \setminus \{P_=(\cdot|\varphi)\}$,*

$$\lim_{n\to\infty} \sum_{i=1}^{n} (H_i(P_=(\cdot|\varphi)) - H_i(P)) = \infty.$$

*Proof* The proof shows a slightly stronger property: for all $P \in \mathbb{E}_\varphi \setminus \{P_=(\cdot|\varphi)\}$ the sequence $f_n(P) := H_n(P_=(\cdot|\varphi)) - H_n(P)$ is such that there exists some $M \geq N_\varphi$ such that $f_n(P)$ is strictly positive and never decreasing for all $n \geq M$.

Let us first consider the case in which $P(\varphi^N) < 1$. The claim of this corollary follows directly from the final observation in the proof of Lemma 33.

The second and final case is when $P(\varphi^{N_\varphi}) = 1$. Since $P_=(\cdot|\varphi) \neq P$ there has to exist some $M \geq N_\varphi$ such that for all $m \geq M$ the probability functions $P$ and $P_=(\cdot|\varphi)$ disagree on the $m$-states. This is because, as noted after Definition 3, probability functions on first-order predicate languages are determined by their values on the $n$-states, for each $n$.

Since both functions assign non-zero probability to, at most, the $m$-states extending those in $\varphi^N$, and $P_=(\cdot|\varphi)$ is maximally equivocal on this set of $M$-states, it follows that $H_M(P_=(\cdot|\varphi)) > H_M(P)$.

For all $m \geq M$ we let $\omega_{m\downharpoonright M}$ denote the $M$-state entailed by $\omega_m$. We now find

$$\begin{aligned} H_m(P) &= -\sum_{\omega_m\in\Omega_m} P(\omega_m)\log(P(\omega_m)) \\ &\leq -\sum_{\omega_m\in\Omega_m} P(\omega_{m\downharpoonright M})\frac{|\Omega_M|}{|\Omega_m|}\cdot\log\left(P(\omega_{m\downharpoonright M})\cdot\frac{|\Omega_M|}{|\Omega_m|}\right) \\ &= -\sum_{\omega_M\in\Omega_M} P(\omega_{m\downharpoonright M})\log\left(P(\omega_{m\downharpoonright M})\cdot\frac{|\Omega_M|}{|\Omega_m|}\right) \\ &= H_M(P)+\log\left(\frac{|\Omega_m|}{|\Omega_M|}\right) \end{aligned}$$

$$H_m(P_=(\cdot|\varphi)) \stackrel{(9)}{=} \log\left(|\varphi^N|\cdot\frac{|\Omega_m|}{|\Omega_N|}\right)$$

$$= \log(|\varphi^N|) + \log\left(\frac{|\Omega_m| \cdot |\Omega_M|}{|\Omega_N| \cdot |\Omega_M|}\right)$$
$$= H_M(P_=(\cdot|\varphi)) + \log\left(\frac{|\Omega_m|}{|\Omega_M|}\right).$$

where $\omega_{m\restriction\omega_M}$ is the restriction of $\omega_m$ to $\mathcal{L}_M$, i.e $\omega_{m\restriction\omega_M}$ is the $M$-state induced by $\omega_m$. It thus easily follows that $H_m(P_=(\cdot|\varphi)) - H_m(P) \geq H_M(P_=(\cdot|\varphi)) - H_M(P)$ for all $m \geq M$. In turn, this implies that

$$\lim_{n\to\infty} \sum_{M=1}^{n} H_i(P_=(\cdot|\varphi)) - H_i(P) \geq \lim_{n\to\infty} (n-M) \cdot (H_M(P_=(\cdot|\varphi)) - H_M(P)).$$

Since the last difference is strictly positive, this limit is $+\infty$. The Corollary follows trivially by adding the first $M-1$ bounded terms to the above limit. □

Given a finite set of premisses of the form $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ we showed in Theorem 15 how a maximal entropy function can be characterised in terms of a limit in entropy. In case of a single categorical premiss, $\varphi$, if $P_=(\cdot \mid \varphi)$ is a limit in entropy then it is the unique maximal entropy function (Corollary 19). In particular, this is the case when $\varphi$ is equivalent to a quantifier-free sentence (Theorem 23). Theorem 34 shows that for any inductively consistent premiss $\varphi$, there exists a unique maximal entropy function, which can be determined by conditionalising the equivocator on the support of $\varphi$, the quantifier-free sentence $\varphi^{N_\varphi}$ expressible in the sublanguage $\mathcal{L}_{N_\varphi}$. For example, for $\varphi = U_1 t_1 \vee \exists x \forall y U_2 xy$ every 1-state is consistent with $\varphi$. However, only the 1-states entailing $U_1 t_1$ are in the support of $\varphi$. These 1-states have the feature that almost all their extensions contribute to the probability of $P_=(\varphi)$ via probability axiom P3. What is more, Theorem 34 shows that the maximal entropy probability function equivocates between the $N_\varphi$-states, and also between their extensions. That is, the unique maximal entropy probability function divides the full probability measure equally between these $N_\varphi$-states and similarly between their extensions to any $\mathcal{L}_n$ with $n \geq N_\varphi$.

Given Theorem 34, conditionalising the equivocator function is a simple method for determining the maximal entropy probabilities in objective Bayesian inductive logic. Although this approach to inductive logic is Bayesian, conditionalisation is not taken here as a principle that is constitutive or core to the Bayesian method, but rather as an inference tool that is appropriate in certain specific circumstances. Indeed, conditionalisation has been criticised as being problematic outside a circumscribed range of circumstances [14, 43]. The fact that it agrees with the maximal entropy approach can be taken to justify the use of conditionalisation on learning $\varphi$, in the circumstances in which $\varphi$ has positive measure and is 'simple' in the sense that it only imposes the constraint $P(\varphi) = 1$ [44, Definition 5.14].

## 6 Jeffrey Conditionalisation

In this section, we generalise our results for conditionalisation from the case in which the premiss is a categorical sentence $\varphi$ to the case in which the premiss is a sentence of the language with a specific probability attached, $\varphi^c$, with $c \in (0, 1)$. Thus in this section, $\mathbb{E} = \mathbb{E}_{\varphi^c} \stackrel{\text{df}}{=} \{P \in \mathbb{P} : P(\varphi) = c\}$.

**Definition 37** (Jeffrey Conditionalisation of the Equivocator) Where $P_=(\varphi) \in (0, 1)$ we can define the Jeffrey conditionalisation of the equivocator function:

$$P_{\varphi^c}(\cdot) \stackrel{\text{df}}{=} c \cdot P_=(\cdot|\varphi) + (1 - c) \cdot P_=(\cdot|\neg\varphi).$$

First, we have a straightforward generalisation of Corollary 19:

**Proposition 38** *If $P_{\varphi^c}$ is a limit in entropy of $\mathbb{E}_{\varphi^c}$, then*

$$\text{maxent}\, \mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\}.$$

*Proof* $P_{\varphi^c}$ is contained in $\mathbb{E}_{\varphi^c}$ because $P_{\varphi^c}(\varphi) = c \cdot 1 + (1 - c) \cdot 0 = c$. Hence, Theorem 16 applies. □

We also have an analogue of Corollary 20:

**Proposition 39** *If $\mathbb{H}_n$ contains $P_{\varphi^c}$ for sufficiently large $n$, then*

$$\text{maxent}\, \mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\}.$$

*Proof* If $P_{\varphi^c} \in \mathbb{H}_n$ for sufficiently large $n$, then $P_{\varphi^c}$ is a limit in entropy of $\mathbb{E}_{\varphi^c}$. Hence, Proposition 38 applies. □

Thus (cf. Theorem 23), if $\varphi$ is logically equivalent to a quantifier-free sentence and $P_{\varphi^c}$ is well defined ($1 > P_=(\varphi) > 0$), then $\text{maxent}\, \mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\}$. Interestingly, as we show shortly, this holds true even when $\varphi$ is not logically equivalent to a quantifier-free sentence. First we make the following observation:

**Proposition 40** $\neg\varphi^n = (\neg\varphi)^n$.

*Proof* Recall from Eq. 2 that for all $n \geq N_\varphi$ it is true that

$$P_=(\varphi \wedge \omega_n) = \begin{cases} 0, & \text{if and only if } P_=(\varphi|\omega_n) = 0 \\ P_=(\omega_n), & \text{if and only if } P_=(\varphi|\omega_n) > 0 \end{cases}$$

$$P_=(\neg\varphi \wedge \omega_n) = \begin{cases} 0, & \text{if and only if } P_=(\neg\varphi|\omega_n) = 0 \\ P_=(\omega_n), & \text{if and only if } P_=(\neg\varphi|\omega_n) > 0. \end{cases}$$

Since $0 < P_=(\omega_n) = P_=(\varphi \wedge \omega_n) + P_=(\neg\varphi \wedge \omega_n)$ it follows that for every fixed $n$-state $\omega_n \in \Omega_n$ either $P_=(\varphi \wedge \omega_n) > 0$ or $P_=(\neg\varphi \wedge \omega_n) > 0$ is true but not both. Since $\varphi^n$ is the disjunction of such $\omega_n$, in particular $\varphi^n$ is quantifier-free, we have

$$\neg\varphi^n = (\neg\varphi)^n$$

and $\langle\varphi^n, (\neg\varphi)^n\rangle$ is a partition. □

We are now in a position to provide the main result of this section.

**Theorem 41** (Agreement with Jeffrey Conditionalisation) *For all* $c \in (0, 1)$ *and all* $\varphi \in S\mathcal{L}$ *such that* $P_=(\varphi) \in (0, 1)$*, the maximal entropy function for the premiss* $\varphi^c$ *is obtained by Jeffrey updating the equivocator function:*

$$\text{maxent}\,\mathbb{E}_{\varphi^c} = \{P_{\varphi^c}\} = \{c \cdot P_=(\cdot|\varphi^{N_\varphi}) + (1-c) \cdot P_=(\cdot|\neg\varphi^{N_\varphi})\}.$$

Recall that Theorem 34 covers the boundary cases of $c = 0$ and $c = 1$, in which the maximal entropy function is determined by Bayesian conditionalisation.

*Proof* The main idea in the proof comes from the intuition that it is always beneficial in terms of entropy to take the probability mass from those $n$-states that have few extensions to $m$-states that simulate $\varphi$ (in the sense in which the states in the set $S_m$, introduced in Lemma 32, simulate $\varphi$), as $m$ increases to infinity, and divide it (equally) between the extensions of those $n$-states for which almost all extensions to an $m$-state simulate $\varphi$ as $m$ increases to infinity.

Let $N := N_\varphi$ and note that by Theorem 34,

$$\begin{aligned} & c \cdot P_=(\cdot|\varphi^N) + (1-c) \cdot P_=(\cdot|\neg\varphi^N) \\ &= c \cdot P_=(\cdot|\varphi) + (1-c) \cdot P_=(\cdot|\neg\varphi) \\ &= P_{\varphi^c} \in \mathbb{E}_{\varphi^c}. \end{aligned}$$

Williams [41, p. 136] shows that this probability function has maximum $N$-entropy in $\mathbb{E}_{(\varphi^N)^c}$. Moreover, $P_{\varphi^c}$ is equivocal beyond $N$ in the following sense: it assigns all $n$-states extending $\varphi^N$ the same probability and it also assigns assigns all $n$-states extending $(\neg\varphi)^N = \neg\varphi^N$ (Proposition 40) the same probability. Hence, by Williamson [44, Theorem 5.13], $\text{maxent}\,\mathbb{E}_{(\varphi^N)^c} = \{P_{\varphi^c}\}$. Thus $P_{\varphi^c}$ has greater entropy than any function $Q \in \mathbb{E}_{\varphi^c}$ such that $Q(\varphi^N) = c$.

Now consider some other $Q \in \mathbb{E}_{\varphi^c}$ with $Q(\varphi^N) \neq c$. We show that $Q \notin \text{maxent}\,\mathbb{E}_{\varphi^c}$ by proving that $P_{\varphi^c}$ has greater entropy than $Q$. Without loss of generality we assume that $\alpha := Q(\varphi^N) < c$, so $1 - \alpha = Q(\neg\varphi^N) > 1 - c$. Then there has to exist some state $\nu_N \vDash \neg\varphi^N$ (recall that this means that $P_=(\nu_N \wedge \varphi) = 0$) such that $Q(\nu_N \wedge \varphi) > 0$.

The $n$-entropy of $P_{\varphi^c}$ is given by:

$$\begin{aligned} H_n(P_{\varphi^c}) &= H_n(c \cdot P_=(\cdot|\varphi) + (1-c) \cdot P_=(\cdot|\neg\varphi)) \\ &= -\sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \vDash \varphi^N}} \frac{c \cdot |\Omega_N|}{|\varphi^N| \cdot |\Omega_n|} \cdot \log\left(\frac{c \cdot |\Omega_N|}{|\varphi^N| \cdot |\Omega_n|}\right) \\ &\quad - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \vDash \neg\varphi^N}} \frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N| \cdot |\Omega_n|} \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N| \cdot |\Omega_n|}\right) \end{aligned}$$

$$
\begin{aligned}
&= -c \cdot \log\left(\frac{c \cdot |\Omega_N|}{|\varphi^N| \cdot |\Omega_n|}\right) - (1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N| \cdot |\Omega_n|}\right) \\
&= -c \cdot \log\left(\frac{c \cdot |\Omega_N|}{|\varphi^N|}\right) - (1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N|}\right) \\
&\quad + c \cdot \log(|\Omega_n|) + (1-c) \cdot \log(|\Omega_n|) \\
&= -c \cdot \log\left(\frac{c \cdot |\Omega_N|}{|\varphi^N|}\right) - (1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N|}\right) \\
&\quad + \log(|\Omega_n|).
\end{aligned}
$$

Note that this expression is of the form $\log(|\Omega_n|)$ plus some constant.

We now calculate an upper bound on the $n$-entropy of $Q \in \mathbb{E}_{\varphi^c}$ for all large enough $n \geq N$. Since $\alpha = Q(\varphi^N) < c$ and $n$-entropy is a sum, which can be split into a sum over the $n$-states which imply $\varphi^N$ and into a sum over the $n$-states which do not entail $\varphi^N$ (those $n$-states entailing $\neg\varphi^N$), $Q$ is maximally equivocal on $\Omega_n$ if $Q$ equivocates beyond $\varphi^N$. That is, all $n$-states entailing $\varphi^N$ are assigned the same probability $\frac{\alpha}{|\varphi^N|} \cdot \frac{|\Omega_N|}{|\Omega_n|}$.

The remaining probability mass of $1 - \alpha$ has then to be assigned to the $n$-states entailing $\neg\varphi^N$. Since $Q$ has to assign $\varphi$ probability $c$, there have to exist sets of $n$-states $S_n$ (extending $\neg\varphi^N$), which are jointly assigned a probability of $c - \alpha$. Furthermore, probability of $1 - c$ needs to be assigned to $\neg\varphi$. Firstly, note that for large enough $n$, the $n$-entropy of $Q$ is bounded from above by interpreting these two constraints as constraints on different sets of $n$-states. Furthermore, the $n$-entropy of $Q$ is bounded from above by assuming that all states in $S_n$ have equally many $n+1$-states extending it in $S_{n+1}$. Finally, the $n$-entropy is maximised by equivocating among all these $n$-states.

We hence have for large enough $n \geq N$:

$$
\begin{aligned}
H_n(Q) \leq & - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \varphi^N}} \frac{\alpha \cdot |\Omega_N|}{|\varphi^N| \cdot |\Omega_n|} \cdot \log\left(\frac{\alpha \cdot |\Omega_N|}{|\varphi^N| \cdot |\Omega_n|}\right) \\
& - \sum_{\substack{\omega_n \in \Omega_n \\ \omega_n \models \neg\varphi^N}} \frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N| \cdot |\Omega_n|} \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N| \cdot |\Omega_n|}\right) \\
& - \sum_{S_n \subset \Omega_n} \frac{c-\alpha}{|S_n|} \cdot \log\left(\frac{c-\alpha}{|S_n|}\right) \\
= & -\alpha \cdot \log\left(\frac{\alpha \cdot |\Omega_N|}{|\varphi^N| \cdot |\Omega_n|}\right) \\
& -(1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N| \cdot |\Omega_n|}\right) \\
& -(c-\alpha) \cdot \log\left(\frac{c-\alpha}{|S_n|}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= -\alpha \cdot \log\left(\frac{\alpha \cdot |\Omega_N|}{|\varphi^N|}\right) - (1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N|}\right) \\
&\quad -(c-\alpha) \cdot \log(c-\alpha) \\
&\quad +(1-[c-\alpha]) \log(|\Omega_n|) + (c-\alpha) \cdot \log(|S_n|) \\
&< -\alpha \cdot \log\left(\frac{\alpha \cdot |\Omega_N|}{|\varphi^N|}\right) - (1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N|}\right) \\
&\quad -(c-\alpha) \cdot \log(c-\alpha) \\
&\quad +(1-[c-\alpha]) \cdot \log(|\Omega_n|) + (c-\alpha) \cdot \log\left(\frac{|\Omega_n|}{2}\right) \\
&= -\alpha \cdot \log\left(\frac{\alpha \cdot |\Omega_N|}{|\varphi^N|}\right) - (1-c) \cdot \log\left(\frac{(1-c) \cdot |\Omega_N|}{|\neg\varphi^N|}\right) \\
&\quad -(c-\alpha) \cdot \log(c-\alpha) \\
&\quad +\left(1-\frac{c-\alpha}{2}\right) \cdot \log(|\Omega_n|).
\end{aligned}
$$

The inequality follows from Lemma 32 noting that $\frac{|S_n|}{|\Omega_n|}$ is a null-sequence.

By comparing the factors of $\log(|\Omega_n|)$ (recall that $c > \alpha$) we find for all large $n$ that

$$H_n(Q) < H_n(P_{\varphi^c}).$$

Hence, $c \cdot P_=(\cdot|\varphi) + (1-c) \cdot P_=(\cdot|\neg\varphi)$ has greater entropy than $Q$. This completes the proof. □

**Corollary 42** (Generalisation to a sentence with an interval attached) *For all intervals $\emptyset \neq X \subset [0,1]$ and all sentences $\varphi \in S\mathcal{L}$ such that $P_=(\varphi) \in (0,1)$ it holds that $c \cdot P_=(\cdot|\varphi^N) + (1-c) \cdot P_=(\cdot|\neg\varphi^N)$ has greater entropy than every other function in $\mathbb{E}_{\varphi^X}$ where $c := \arg\min_{x \in X} |x - P_=(\varphi)|$. Given the premiss $\varphi^X$, the maximal entropy function is obtained by Jeffrey conditionalisation of the equivocator on $\varphi^c$ where $c$ is closest to the measure of $\varphi$. Hence,*

$$\text{maxent}\, \mathbb{E}_{\varphi^X} = \{c \cdot P_=(\cdot|\varphi^N) + (1-c) \cdot P_=(\cdot|\neg\varphi^N)\}.$$

*Proof* If $P_=(\varphi) \in X$, then $c = \arg\min_{x \in X} |x - P_=(\varphi)| = P_=(\varphi) = P_=(\varphi^N)$. Hence, for all sentences $\psi \in S\mathcal{L}$

$$
\begin{aligned}
P_{\varphi^c}(\psi) &= P_=(\varphi^N) \cdot P_=(\psi|\varphi^N) + P_=(\neg\varphi^N) \cdot P_=(\psi|\neg\varphi^N) \\
&= P_=(\psi \wedge \varphi^N) + P_=(\psi \wedge \neg\varphi^N) \\
&= P_=(\psi).
\end{aligned}
$$

Since $P_= \in \mathbb{E}_{\varphi^X}$ the result follows.

If $P_=(\varphi) \notin X$, then for all $P \in \mathbb{E}_{\varphi^X}$ it holds that $x := P(\varphi) \neq P_=(\varphi)$. By the proof of Theorem 41 we see that $x \cdot P_=(\cdot|\varphi^N) + (1-x) \cdot P_=(\cdot|\neg\varphi^N)$ has greatest entropy among all functions with $x = P(\varphi)$. Hence,

$$\text{maxent}\, \mathbb{E}_{\varphi^X} \subseteq \{x \cdot P_=(\cdot|\varphi^N) + (1-x) \cdot P_=(\cdot|\neg\varphi^N) \;:\; x \in X\}.$$

Letting $P_x := x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)$, we now compute the $n$-entropies for all these probability functions for $n \geq N$ to be equal to

$$\begin{aligned} H_N(P_x) &= H_N(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) \\ &= -\sum_{\omega_N \models \varphi^N} x \cdot \frac{1}{|\varphi^N|} \log\left(x \cdot \frac{1}{|\varphi^N|}\right) \\ &\quad - \sum_{\omega_N \models \neg\varphi^N} (1-x) \cdot \frac{1}{|\neg\varphi^N|} \log\left((1-x) \cdot \frac{1}{|\neg\varphi^N|}\right) \\ &= -x \cdot \log\left(\frac{x}{|\varphi^N|}\right) - (1-x) \cdot \log\left(\frac{1-x}{|\neg\varphi^N|}\right) \end{aligned}$$

and

$$\begin{aligned} H_n(P_x) &= H_n(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) \\ &= -x \cdot \log\left(\frac{x}{|\varphi^N| \cdot \frac{|\Omega_n|}{|\Omega_N|}}\right) - (1-x) \cdot \log\left(\frac{1-x}{|\neg\varphi^N| \cdot \frac{|\Omega_n|}{|\Omega_N|}}\right) \\ &= H_N(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) \\ &\quad -x \cdot \log\left(\frac{|\Omega_N|}{|\Omega_n|}\right) - (1-x) \cdot \log\left(\frac{|\Omega_N|}{|\Omega_n|}\right) \\ &= H_N(x \cdot P_=(\cdot|\varphi) + (1-x) \cdot P_=(\cdot|\neg\varphi)) + \log(|\Omega_n|) - \log(|\Omega_N|) \\ &= H_N(P_x) + \log(|\Omega_n|) - \log(|\Omega_N|). \end{aligned}$$

It hence holds for all $x, y \in [0, 1]$ and all $n > N$ that

$$H_n(P_x) > H_n(P_y), \text{ if and only if } H_N(P_x) > H_N(P_y). \tag{10}$$

Let us next note that $P_{P_=(\varphi^N)} = P_=$. Furthermore, every $P_x$ is a convex combination of $P_=(\cdot|\varphi)$ and of $P_=(\cdot|\neg\varphi)$. Along this line from $P_=(\cdot|\varphi)$ to $P_=(\cdot|\neg\varphi)$ $N$-entropy is maximised by $P_{P_=(\varphi^N)} = P_=$ since it is the equivocator (on $\Omega_N$). Since the $P_x$ (on $\Omega_N$) all are part of a line segment and $H_N$ is strictly concave, it follows that $N$-entropy is uniquely maximised by the equivocator and strictly decreases the further one moves in either direction from the equivocator. Hence, $P_c$ has strictly the greatest $N$-entropy among all other $P_x$ for $x \in X \setminus \{c\}$.

Applying the above equivalence (10) we find that $P_c$ (since $c \in X$ is the closest to $P_=(\varphi)$) also has the greatest $n$-entropy among all $P_x$ for $x \in X$ for large enough $n$. $P_c$ has hence greater entropy than every other probability function $P \in \mathbb{E}_{\varphi^X} \setminus \{P_c\}$. □

One might hypothesise that one can generalise further still, simply by replacing premiss sentences $\varphi_1, \ldots, \varphi_k$ by their inductive equivalents $\varphi_1^n, \ldots, \varphi_k^n$, for sufficiently large $n$, at least in the case in which $\varphi_1, \ldots, \varphi_k$ are inductive non-contradictions. That is, one might hypothesise that $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx \psi^Y$ iff $\varphi_1^{n\,X_1}, \ldots, \varphi_k^{n\,X_k} \mid\!\approx \psi^Y$ for any $n \geq \max\{N_{\varphi_1}, \ldots, N_{\varphi_1}\}$ and $\varphi_1, \ldots, \varphi_k$ with positive measure.

Unfortunately, it is not possible to generalise in such a straightforward way. This is because premisses that are satisfiable may transform into premisses that are unsatisfiable.

*Example 43* If we have satisfiable premisses $(Ut_1 \vee \forall x Vx)^{.9}$, $(\neg Ut_1 \vee \forall x Vx)^{.9}$ and substitute the corresponding 1-supports, we obtain $(Ut_1)^{.9}$, $\neg(Ut_1)^{.9}$, which are not jointly satisfiable[7].

In order to determine the maximal entropy function for the premisses $(Ut_1 \vee \forall x Vx)^{.9}$, $(\neg Ut_1 \vee \forall x Vx)^{.9}$ we first observe that

$$Q(\omega) := P_= \left( \bigwedge_{i=1}^{n} U^{\epsilon_i} t_i \right) \cdot \begin{cases} 0, & \text{if } \omega \not\models \bigwedge_{i=1}^{n} Vt_i \\ 1, & \text{if } \omega \models \bigwedge_{i=1}^{n} Vt_i \end{cases}$$

is such that $Q(\forall x Vx) = 1$. Next note that the following function satisfies both premisses:

$$R := 0.8 \cdot Q + 0.2 \cdot P_=,$$

since $R(Ut_1 \vee \forall x Vx) = 0.8 \cdot 1 + 0.2 \cdot 0.5 = 0.9 = R(\neg Ut_1 \vee \forall x Vx)$.

Furthermore, every probability function $P$ such that $P(Ut_1 \vee \forall x Vx) = 0.9 = P(\neg Ut_1 \vee \forall x Vx)$ is such that $0.8 \leq P([Ut_1 \vee \forall x Vx] \wedge [\neg Ut_1 \vee \forall x Vx]) = P(\forall x Vx)$.

Finally, note that $R$ is the maximal entropy function by Theorem 16, since $R$ satisfies the premisses and is a limit in entropy because it is the $n$-entropy maximiser for all $n$. We leave this last claim as an exercise for the reader.

Similarly, one might have wondered whether in the categorical case, $\varphi_1, \ldots, \varphi_k \mathrel{|\!\!\!\approx} \psi^Y$ iff $\varphi_1^n, \ldots, \varphi_k^n \mathrel{|\!\!\!\approx} \psi^Y$ for any $n \geq \max\{N_{\varphi_1}, \ldots, N_{\varphi_k}\}$ and $\varphi_1, \ldots, \varphi_k$ with positive measure. We now see that this characterisation does not hold:

*Example 44* If we have satisfiable premisses $\varphi_1 := (Ut_1 \vee \forall x Vx)$, $\varphi_2 := (\neg Ut_1 \vee \forall x Vx)$ and substitute the corresponding 1-supports, we obtain $Ut_1$, $\neg Ut_1$, which are not jointly satisfiable. The maximal entropy function for the two given premisses is the function $Q$ from the above example.

Theorem 34 does not apply here since, although each of the premisses has positive measure, the measure of the conjunction of the premisses is zero:

$$\begin{aligned} P_=((Ut_1 \vee \forall x Vx) \wedge (\neg Ut_1 \vee \forall x Vx)) &= P_=((Ut_1 \wedge \neg Ut_1) \vee \forall x Vx) \\ &= P_=(\forall x Vx) = 0. \end{aligned}$$

## 7 Preservation of Inductive Tautologies

Having developed the limit in entropy method for determining maximal entropy functions, and having demonstrated concordance with Bayesian conditionalisation and Jeffrey conditionalisation, we will now discuss some of the general properties of the

[7] We see three plausible ways to define an inductive logic for inconsistent premisses: i) use the equivocator to draw inferences, ii) use the disjunction of all maximal consistent subsets of premisses as inductive premisses or iii) infer $\psi^Y$ for all $\psi \in S\mathcal{L}$ and all intervals $Y$. This example is not the place to debate the pros and cons of these three definitions, and we do not commit to one of them here. Instead, we content ourselves with pointing out that all three definitions applied to $(Ut_1)^{.9}$, $\neg(Ut_1)^{.9}$ give different inductive probabilities than the satisfiable premisses $(Ut_1 \vee \forall x Vx)^{.9}$, $(\neg Ut_1 \vee \forall x Vx)^{.9}$ yield in our approach.

maximal entropy approach. In this section, we outline some logical features of objective Bayesian inductive logic, while in Section 8 we will explore the extent to which inferences are invariant under permutations of the constants, and in Section 9 we investigate some cases involving categorical premisses with zero measure.

First we show that, in objective Bayesian inductive logic, inductive tautologies (i.e., probability 1 inferences in the absence of any premisses) are preserved after learning the probability of any proposition that is inductively consistent:

**Theorem 45** (Preservation of Inductive Tautologies, PIT) *If* $\overset{\circ}{|\!\!\approx} \theta$ *and* $\not\overset{\circ}{|\!\!\approx} \neg\varphi$, *then* $\varphi^c \overset{\circ}{|\!\!\approx} \theta$ *for any* $c \in (0, 1]$.

*Proof* To simplify notation, we use $P^\dagger$ denote the unique probability function with maximal entropy if there exists such a function.

First, note that applying the assumption $P_=(\theta) = 1$ to $P_=(\theta \wedge \varphi) + P_=(\neg\theta \wedge \varphi) = P_=(\varphi)$ entails $P_=(\theta \wedge \varphi) = P_=(\varphi)$ for all sentences $\varphi \in S\mathcal{L}$.

If $c = 1$, then by Theorem 34,

$$P^\dagger(\theta) = P_=(\theta|\varphi) = \frac{P_=(\theta \wedge \varphi)}{P_=(\varphi)} = \frac{P_=(\varphi)}{P_=(\varphi)} = 1.$$

So, $\varphi^1 \overset{\circ}{|\!\!\approx} \theta$.

If, on the other hand, $c \in (0, 1)$, then by Theorem 41,

$$\begin{aligned} P^\dagger(\theta) &= c \cdot P_=(\theta|\varphi) + (1-c) \cdot P_=(\theta|\neg\varphi) = c + (1-c) \\ &= 1. \end{aligned}$$

So, $\varphi^c \overset{\circ}{|\!\!\approx} \theta$. □

PIT implies that inductive contradictions are also preserved after learning the probability of any proposition that is not an inductive contradiction: if $\overset{\circ}{|\!\!\approx} \neg\theta$ and $\not\overset{\circ}{|\!\!\approx} \neg\varphi$, then $\varphi^c \overset{\circ}{|\!\!\approx} \neg\theta$ for any $c \in (0, 1]$.

PIT is loosely related to the Obstinacy principle of [27, p. 99], which provides a condition under which inferences from $\varphi_1^{X_1}, ..., \varphi_k^{X_k}$ are preserved upon learning $\pi_1^{W_1}, ..., \pi_j^{W_j}$. In the present setting, Obstinacy can be formulated as follows. Consider $\mathbb{E} := \{P : P \text{ satisfies } \varphi_1^{X_1}, ..., \varphi_k^{X_k}\}$ and $\mathbb{F} := \{P : P \text{ satisfies } \pi_1^{W_1}, ..., \pi_j^{W_j}\}$. Then:

**Theorem 46 (Obstinacy)** *If* $\operatorname{maxent}\mathbb{E} \subseteq \mathbb{F}$, *then* $\operatorname{maxent}\mathbb{E} \subseteq \operatorname{maxent}(\mathbb{E} \cap \mathbb{F})$.

*Proof* If $P \in \operatorname{maxent}\mathbb{E}$ then no function in $\mathbb{E}$ dominates $P$ in $n$-entropy for sufficiently large $n$ and $P \in \mathbb{E} \cap \mathbb{F} \neq \emptyset$. In particular, no function in $\mathbb{E} \cap \mathbb{F} \neq \emptyset$ dominates $P$ in $n$-entropy for sufficiently large $n$. Thus, $P \in \operatorname{maxent}(\mathbb{E} \cap \mathbb{F})$ and $\operatorname{maxent}\mathbb{E} \subseteq \operatorname{maxent}(\mathbb{E} \cap \mathbb{F})$. □

PIT can also be thought of as a variant of the Rational Monotonicity rule of inference in non-monotonic logic [25, §3.4]:

**Rational Monotonicity** If $\psi \approx\!\!\!\!\mid\; \theta$ and $\psi \not\approx\!\!\!\!\mid\; \neg\varphi$, then $\psi \wedge \varphi \approx\!\!\!\!\mid\; \theta$.

PIT specialises Rational Monotonicity to the case in which $\psi$ is an inductive tautology and then generalises it to the case in which $\varphi$ is uncertain.

PIT can also be interpreted as an absolute continuity condition [3, p. 422]: if $\theta$ has zero measure, i.e., $P_=(\theta) = 0$, then any $P^\dagger \in \text{maxent}\,\mathbb{E}_{\varphi^c}$ also gives zero probability to $\theta$, where $\varphi$ has positive measure and $c > 0$. Note that the concept of absolute continuity is usually developed in the framework of measure theory. The equivocator function $P_=$ corresponds to Lebesgue measure when probability functions on $\mathcal{L}$ are mapped to probability measures on the unit interval [44, §2.6.3]. Thus, 'zero measure' in the present sense (Definition 5) corresponds to zero Lebesgue measure.

## 8 Invariance Under Permutations

Williamson [43, Proposition 5.10] shows that the maximal entropy approach is invariant under those finite and infinite permutations of the atomic sentences that list atomic sentences involving only $t_1, ..., t_n$ before those involving $t_{n+1}$ for each $n$. In this section, we explore invariance under permutations of the constants themselves.

**Definition 47** Let $f$ be a reordering of constants, i.e, $f$ is bijective. For $\varphi \in S\mathcal{L}$ we write $f(\varphi)$ for the result of reordering the constants in $\varphi$ according to $f$. We use $f(P)$ to denote the probability function obtained from $P$ by permuting the constants of $\varphi \in S\mathcal{L}$ according to $f$: $f(P)(\varphi(\vec{t})) := P(\varphi(f(\vec{t})))$ for all $\varphi \in S\mathcal{L}$.

**Lemma 48** *If $P \in \mathbb{P}$ and $f$ is a permutation, then $f(P) \in \mathbb{P}$.*

*Proof* It is clear that $f(P)$ satisfies P1 and P2.

Concerning P3, we need to show the last equality below holds. The first three equalities follow from the definition of $f(P)$. We use $\theta\left(f(t_i)/t_i, f(\vec{t})/\vec{t}\right)$ to denote the sentence one obtains from $\theta(t_i, \vec{t})$ by simultaneously replacing $t_i$ by $f(t_i)$ and $\vec{t}$ by $f(\vec{t})$.

$$\begin{aligned}
&\sup_m f(P)\left(\bigvee_{i=1}^m \theta(t_i, \vec{t})\right)\\
&= \sup_m P\left(f\left(\bigvee_{i=1}^m \theta(t_i, \vec{t})\right)\right)\\
&= \sup_m P\left(\bigvee_{i=1}^m f(\theta(t_i, \vec{t}))\right)\\
&= \sup_m P\left(\bigvee_{i=1}^m \theta\left(f(t_i)/t_i, f(\vec{t})/\vec{t}\right)\right)\\
&= f(P)(\exists x \theta(x, \vec{t})).
\end{aligned}$$

As usual, put $N := \max\{i : t_i \in \theta(\vec{t})\}$ and also let $N_f := \max\{j : t_j \in f(\theta(\vec{t}))\}$.

Let us now fix $m$ and consider $M_m \geq \max\{f(1), \ldots, f(m), N_f\}$, then $\bigvee_{i=1}^{m} \theta\left(f(t_i)/t_i, f(\vec{t})/\vec{t}\right) \models \bigvee_{i=1}^{M_m} \theta(t_i, f(\vec{t}))$ and thus

$$P\left(\bigvee_{i=1}^{m} \theta\left(f(t_i)/t_i, f(\vec{t})/\vec{t}\right)\right) \leq P\left(\bigvee_{i=1}^{M_m} \theta(t_i, f(\vec{t}))\right).$$

Similarly, let $J_m \geq \max\{f^{-1}(1), \ldots, f^{-1}(m), N_f\}$, then $\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t})) \models \bigvee_{i=1}^{J_m} f(\theta(t_i, \vec{t}))$ and so

$$P\left(\bigvee_{i=1}^{J_m} \theta\left(f(t_i)/t_i, f(\vec{t})/\vec{t}\right)\right) \geq P\left(\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t}))\right).$$

We next note that $(P(\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t}))))_{m\in\mathbb{N}}$ is an increasing non-negative sequence which converges by P3 to $P(\exists x\theta(x, f(\vec{t})))$. This entails that $\sup_m f(P)(\bigvee_{i=1}^{m} \theta(t_i, \vec{t}))$ also converges to $P(\exists x\theta(x, f(\vec{t})))$.

So,

$$\begin{aligned}\sup_m f(P)\left(\bigvee_{i=1}^{m} \theta(t_i, \vec{t})\right) &= \sup_m P\left(\bigvee_{i=1}^{m} \theta(t_i, f(\vec{t}))\right)\\ &= P(\exists x\theta(x, f(\vec{t})))\\ &= f(P)(\exists x\theta(x, \vec{t})),\end{aligned}$$

where the last equality is just definition of $f(P)$. Hence, $f(P)$ satisfies P3. □

The concept of 'greater entropy' is well defined in the sense that it is preserved under any permutation that preserves the probability functions that it permutes:

**Proposition 49** (Independence of ordering of constant symbols) *For any reordering of constants $f$ and probability functions $P$, $Q$ such that $f(P) = P$ and $f(Q) = Q$, $P$ has greater entropy than $Q$ if and only if $f(P)$ has greater entropy than $f(Q)$.*

*Proof* If $f(P) = P$ and $f(Q) = Q$ then $H_n(P) = H_n(f(P))$ and $H_n(Q) = H_n(f(Q))$. So, $H_n(P) > H_n(Q)$ if and only if $H_n(f(P)) > H_n(f(Q))$. Hence, $P$ has greater $n$-entropy than $Q$ for sufficiently large $n$ if and only if $f(P)$ has greater $n$-entropy than $f(Q)$ for sufficiently large $n$. □

On the other hand, if a permutation $f$ changes the two probability functions of interest, then the permuted functions can compare differently with respect to which has greater entropy:

**Proposition 50** (Dependence on ordering of constant symbols) *There exists an infinite reordering of constants $f$ and probability functions $P$, $Q$ such that $P$ has greater entropy than $Q$ but $f(Q)$ has greater entropy than $f(P)$.*

*Proof* To simplify matters we consider a language only containing a single relation symbol, $U$, which is unary. It is apparent from the proof that the proof strategy applies to all languages in our sense.

Let $f$ be the following bijection on $\mathbb{N}$. $f(2n+1) := 2n-1$ for all $n \geq 1$, $f(1) = 2$ and $f(2n) = 2n + 2$. Intuitively, the even numbers and 1 are postponed to the future and the odd numbers, with the exception of 1, are brought forward.

It is important in the following that for all $n$ it holds that $f$ is not a bijection on $\{1, \ldots, n\}$. For all even $n$ and $n = 1$ it holds that $f(n) > n$. For all other odd $n$ it holds that $f^{-1}(n) = n + 2 > n$. This fact will be used without further mention.

Next define a probability function $P \in \mathbb{P}$ by having all atomic sentences be independent of one another. This entails that $n$-entropies can be written as a sum of $n$ 1-entropies,

$$H_n(P) = -\sum_{i=1}^{n} P(Ut_i)\log(P(Ut_i)) + P(\neg Ut_i)\log(\neg P(Ut_i)).$$

This follows from, for example, Landes and Williamson [21, Equation 1].

For all $n \geq 1$ we now let

$$P(Ut_1) := \frac{1}{2} \qquad P(Ut_{2n}|\omega_{2n-1}) := \frac{1}{2} \qquad P(Ut_{2n+1}) := 1,$$

whenever $P(\omega_{2n-1}) > 0$.

We can then compute the $n$-entropies as follows for all $n \geq 1$

$$\begin{aligned} H_1(P) &= \log 2 \\ H_{2n}(P) &= (n+1)\log 2 \\ H_{2n+1}(P) &= H_{2n}(P) = (n+1)\log 2, \end{aligned}$$

since i) for all even $n$ and $n = 1$ it holds that $P$ is maximally equivocal on $\mathcal{L}_n$ conditional on $\mathcal{L}_{n-1}$ and ii) $P$ extends from $\mathcal{L}_n$ to $\mathcal{L}_{n+1}$ deterministically.

Ignoring the constant factor $\log 2$ , the $n$-entropies of $P$ can then be represented by the sequence $\langle 1, 2, 2, 3, 3, \ldots\rangle$. Figuratively speaking, the individual levels of $n$-entropy increase, $H_{n+1}(P) - H_n(P)$, are represented by $\langle 1, 1, 0, 1, 0, 1, 0, 1, 0, \ldots\rangle$. 0 here represents a deterministic behaviour and 1 represents a fully equivocal behaviour.

We now compute for all $n \geq 1$ the $n$-entropies of $f(P)$, $H_n(f(P))$, as follows

$$\begin{aligned} H_1(f(P)) &= 0 \\ H_{2n}(f(P)) &= n\log 2 \\ H_{2n+1}(f(P)) &= H_{2n}(f(P)) = n\log 2. \end{aligned}$$

Clearly, for all $n \geq 1$ it holds that $H_n(P) > H_n(f(P))$.

Figuratively speaking, the individual levels have the following entropies for $f(P)$ ignoring the constant factor $\log 2$: $\langle 0, 1, 0, 1, 0, 1, 0, 1, 0, \ldots\rangle$ and $n$-entropies

$\langle 0, 1, 1, 2, 2, 3, 3, 4, 4, \ldots \rangle$. Clearly, this last sequence is pointwise smaller than the corresponding sequence for $P$.

Now define a probability function $Q$ which also makes all constant symbols independent of each other. Implicitly define $Q$ by

$$\begin{aligned} H_1(Q) &= 0.6 \cdot \log 2 \\ H_2(Q) &= 1.2 \cdot \log 2 \\ H_3(Q) &= 1.8 \cdot \log 2 \\ H_{n+1}(Q) &= H_n(Q) + 0.5 \cdot \log 2 \end{aligned}$$

for all $n \geq 4$. That is, we need to find a value $Q(Ut_i)$ such that

$$-Q(Ut_i) \log Q(Ut_i) - (1 - Q(Ut_i)) \log(1 - Q(Ut_i)) = \alpha \log 2,$$

where $\alpha \in \{0.5, 0.6\}$.

We note that $Q$ is well defined under the assumption that $Q(Ut_i) \leq 0.5$ since i) 1-entropy is strictly concave and strictly increasing for $Q(Ut_i) \in [0, 0.5]$, ii) $H_1(P) \in [0, \log 2]$ for all $P \in \mathbb{P}$, iii) $H_1$ is continuous, iv) $H_1$ is a bijective map from $[0, 0.5]$ onto $[0, \log 2]$ and finally v) the intermediate value theorem holds.

Apparently, for $i \in \{1, 2, 3\}$ it holds that $H_i(P) > H_i(Q)$. That $H_i(P) > H_i(Q)$ holds for all greater $i$, too, follows from the definition of $Q$.

Figuratively speaking, the individual levels have the following entropies for $Q$ ignoring the constant factor $\log 2$: $\langle 0.6, 0.6, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, \ldots \rangle$ and $n$-entropies $\langle 0.6, 1.2, 1.8, 2.3, 2.8, 3.3, \ldots \rangle$. Clearly, this last sequence is pointwise smaller than the corresponding sequence for $P$.

We compute the $n$-entropies for $f(Q)$ as follows for all $1 \leq n \leq 4$

$$\begin{aligned} H_1(f(Q)) &= 0.6 \cdot \log 2 \\ H_2(f(Q)) &= 1.2 \cdot \log 2 \\ H_3(f(Q)) &= 1.7 \cdot \log 2 \\ H_4(f(Q)) &= 2.3 \cdot \log 2. \end{aligned}$$

For all larger $n \geq 5$ we observe

$$H_n(f(Q)) = H_4(f(Q)) + \frac{(n-4)}{2} \cdot \log 2 > 0.$$

Figuratively speaking, the individual levels have the following entropies for $Q$ ignoring the constant factor $\log 2$: $\langle 0.6, 0.6, 0.5, 0.6, 0.5, 0.5, 0.5, 0.5, 0.5, \ldots \rangle$ and $n$-entropies $\langle 0.6, 1.2, 1.7, 2.3, 2.8, 3.3, \ldots \rangle$. Clearly, this last sequence is pointwise greater than the corresponding sequence for $f(P)$. □

The proof shows in fact that for any language there exist probability functions $P, Q \in \mathbb{P}$ such that $P$ has greater entropy than $f(P)$ and $P$ has greater entropy than $Q$, yet $f(Q)$ has greater entropy than $f(P)$.

Interestingly, despite the possibility exposed by Proposition 50, our results show that in many natural cases, the function that has *maximal* entropy is invariant under reordering the constants:[8]

**Theorem 51** (Invariance under Permutations of Constant Symbols) *If* $1 > P_=(\varphi) > 0$ *and* $0 < c \leq 1$*, then for* $\{P^\dagger\} = \text{maxent}\,\mathbb{E}_{\varphi^c}$ *and* $\{P_f^\dagger\} = \text{maxent}\,\mathbb{E}_{f(\varphi)^c}$ *it holds that for all* $\psi \in S\mathcal{L}$ *that*

$$P^\dagger(\psi) = P_f^\dagger(f(\psi)).$$

*Proof* Let us first recall that by Lemma 48 we have $f(P) \in \mathbb{P}$. Furthermore, from the definition of $f(P)$ we immediately obtain that $P \in \mathbb{E}_{\varphi^c}$, if and only if $f(P) \in \mathbb{E}_{f(\varphi)^c}$.

After observing that $\models f(\varphi^m) \leftrightarrow f(\varphi)^m$ and that $\models f(\neg\varphi^m) \leftrightarrow f(\neg\varphi)^m$ for all large enough $m$, we apply Theorem 41 and find

$$\begin{aligned} P^\dagger &= c \cdot P_=(\cdot|\varphi^m) + (1-c) \cdot P_=(\cdot|\neg\varphi^m) \\ P_f^\dagger &= c \cdot P_=(\cdot|f(\varphi^m)) + (1-c) \cdot P_=(\cdot|f(\neg\varphi^m)). \end{aligned}$$

It now suffices to note that the equivocator function is as symmetrical as can be: for all $\chi, \rho \in QFS\mathcal{L}$ it holds that

$$P_=(\chi|\rho) = P_=(f(\chi)|f(\rho)).$$

Hence $P^\dagger(\chi) = P_f^\dagger(f(\chi))$ for all quantifier-free sentences $\chi \in QFS\mathcal{L}$. Gaifman's Theorem [9] then delivers the result that $P^\dagger(\cdot) = P_f^\dagger(f(\cdot))$. □

As might be expected, this result generalises easily to a single premiss with an attached uncertainty interval.

**Corollary 52** *If* $1 > P_=(\varphi) > 0$ *and interval* $\emptyset \neq X \subset [0, 1]$*, then for* $\{P^\dagger\} = \text{maxent}\,\mathbb{E}_{\varphi^X}$ *and* $\{P^\dagger{}_f\} = \text{maxent}\,\mathbb{E}_{f(\varphi)^X}$ *it holds that for all* $\psi \in S\mathcal{L}$ *that*

$$P^\dagger(\psi) = P_f^\dagger(f(\psi)).$$

*Proof* For both premisses a unique maximum entropy function exists which is equal to a Jeffrey conditionalisation of the equivocator. These Jeffrey (or simply Bayesian) conditionalisations are with respect to $\varphi^{N_\varphi}$, respectively, the logically equivalent $f(\varphi^{N_\varphi})$ and $(f(\varphi))^{N_\varphi}$. Furthermore, both Jeffrey conditionalisations are with respect to the same $x \in X$ (Corollary 42).

[8]Landes et al. [24, Footnote 2] show that Paris' approach to maximising entropy, which appeals to limits of entropy maximisers on finite languages, is invariant under finite and infinite permutations of constant symbols where it is well defined. They demonstrate that Paris' approach agrees with the maximal entropy approach in many cases, and conjecture that this agreement extends to all cases in which Paris' limiting function is well defined. In all such cases, invariance of this limit function implies that the maximal entropy approach is invariant under permutations of constants.

Finally, let us apply the proof of Theorem 51 to note that for all $\psi \in S\mathcal{L}$ it holds that

$$P^{\dagger}(\psi) = P^{\dagger}_{f}(f(\psi)).$$

□

## 9 Zero Measure Premisses

As Example 17 illustrates, there are cases of zero-measure premisses that are entirely unproblematic and that can be handled using the limit in entropy techniques introduced in Section 3.[9] However, some zero-measure premisses are more problematic, in that they generate sets $\mathbb{E}$ of probability functions in which there is no function with maximal entropy. We will focus on these pathological cases in this section. We first provide some examples of such cases and then we discuss how best to proceed when they arise. We argue that these cases suggest a refinement to the definition of maximal entropy and that they motivate drawing inferences from any function with sufficiently great entropy.

To simplify the exposition we assume in this section that the underlying language $\mathcal{L}$ contains only the single relation symbol employed in the respective propositions. The general case follows from the fact entropy maximisation is language invariant [27, Chapter 6], because maximal entropy functions equivocate over all sentences mentioning only relation symbols that are not mentioned by any premiss.

**Proposition 53** *For $\varphi = \exists x \forall y Uxy$ and any $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy than $P$. Hence,* $\text{maxent}\,\mathbb{E}_\varphi = \emptyset$.

*Proof* Suppose for contradiction that $\text{maxent}\,\mathbb{E}_\varphi \neq \emptyset$ and let $P \in \text{maxent}\,\mathbb{E}_\varphi$. A contradiction is achieved by first defining a probability function $P' \in \mathbb{E}_\varphi \setminus \{P\}$ such that $H_n(P') \geq H_n(P)$ for all large enough $n$. It is not necessarily the case that $P'$ has greater entropy than $P$. However, all probability functions that are a convex combination of $P$ and $P'$ are in $\mathbb{E}_\varphi$ ($\mathbb{E}_\varphi$ is convex) and have strictly greater $n$-entropy than $P$ for all large enough $n$ (because $H_n(\cdot)$ is strictly concave). Hence, all the convex combinations are in $\mathbb{E}_\varphi$ and have greater entropy than $P$. This yields a contradiction.

Note that $P_=(\varphi) = 0 < 1 = P(\varphi)$. Hence, $P \neq P_=$. Let us now define a probability function $P' \in \mathbb{E}$ by shifting all witnessing of $\exists x \forall y Uxy$ by one and then adding a constant $t_1$ such that $Ut_1t^*$ is independent from all other literals for all $t^* \neq t_1$. Intuitively, the literals $\pm Ut_it_k$ are replaced by $\pm Ut_{i+1}t_k$.

[9] More generally, if $\varphi$ is a universally quantified claim about a conjunction of literals then it has zero measure but can be handled straightforwardly [24].

Formally, let $\omega_n \in \Omega_n = \bigwedge_{i,k=1}^{n} U^{\epsilon_{i,k}} t_i t_k$ be an arbitrary $n$-state. Then define $P'$ by

$$P'(\omega_n) := P\left(\bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i-1,k}} t_{i-1} t_k\right) \cdot P_{=}\left(\bigwedge_{k=1}^{n} U^{\epsilon_{1,k}} t_1 t_k\right)$$
$$= \frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)}{2^n}.$$

Firstly, we note that $P'(\forall y U t_1 y) = \lim_{n\to\infty} P'(\bigwedge_{k=1}^{n} U t_1 t_k) = \lim_{n\to\infty} 2^{-n} = 0$. So, according to $P'$ the constant $t_1$ is not a witness of the existential premiss sentence $\varphi$.

We next show that $P \neq P'$. Firstly, note that since $\lim_{n\to\infty} P(\bigvee_{i=1}^{n} \forall y U t_i y) = P(\exists x \forall y U x y) = 1$, $\min\{i \in \mathbb{N} : P(\forall y U t_i y) > 0\}$ is a finite number. Armed with this observation, we note next that

$$\begin{aligned}
\min\{i \in \mathbb{N} : P'(\forall y U t_i y) > 0\} &= \min\left\{i \in \mathbb{N} : \lim_{n\to\infty} P'\left(\bigwedge_{k=1}^{n} U t_i t_k\right) > 0\right\}\\
&= \min\left\{i \in \mathbb{N} : \lim_{n\to\infty} P\left(\bigwedge_{k=1}^{n} U t_{i+1} t_k\right) > 0\right\}\\
&= 1 + \min\left\{i \in \mathbb{N} : \lim_{n\to\infty} P\left(\bigwedge_{k=1}^{n} U t_i t_k\right) > 0\right\}\\
&= 1 + \min\{i \in \mathbb{N} : P(\forall y U t_i y) > 0\}.
\end{aligned}$$

So, $P \neq P'$.

We also observe that for all $i \in \mathbb{N}$, $P'(\forall y U t_i y) = P(\forall y U t_{i+1} y)$ and furthermore $P'(\bigvee_{i\in I} \forall y U t_i y) = P(\bigvee_{i\in I} \forall y U t_{i+1} y)$ for all finite index sets $I$. So,

$$\begin{aligned}
P'(\exists x \forall y U x y) &= \lim_{n\to\infty} P'\left(\bigvee_{i=1}^{n} \forall y U t_i y\right)\\
&\geq \lim_{n\to\infty} P'\left(\bigvee_{i=2}^{n} \forall y U t_i y\right)\\
&= \lim_{n\to\infty} P\left(\bigvee_{i=1}^{n-1} \forall y U t_i y\right)\\
&= \lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} \forall y U t_i y\right)\\
&= 1.
\end{aligned}$$

This means that $P'(\exists x \forall y U x y) = 1$ and thus, as advertised, $P' \in \mathbb{E}_{\varphi}$.

We now calculate $n$-entropies of $P$ and $P'$ and find for $n \geq 1$ that:

$$H_n(P) = -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 2\le r\le n\\ 1\le s\le n}} \sum_{\substack{\epsilon_u\in\{0,1\}\\ 1\le u\le n}} P\left(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k\right) \cdot \log\left(P\left(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k\right)\right)$$

$$\begin{aligned} H_n(P') &= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 1\le r,s\le n}} P'\left(\bigwedge_{i,k=1}^{n} U^{\epsilon_{i,k}} t_i t_k\right) \cdot \log\left(P'\left(\bigwedge_{i,k=1}^{n} U^{\epsilon_{i,k}} t_i t_k\right)\right) \\ &= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 2\le r\le n\\ 1\le s\le n}} \sum_{\substack{\epsilon_u\in\{0,1\}\\ 1\le u\le n}} P'\left(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k\right) \\ &\qquad \cdot \log\left(P'\left(\bigwedge_{k=1}^{n} U^{\epsilon_k} t_1 t_k \wedge \bigwedge_{i=2}^{n}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k\right)\right) \\ &= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 2\le r\le n\\ 1\le s\le n}} \sum_{\substack{\epsilon_u\in\{0,1\}\\ 1\le u\le n}} \frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)}{2^n} \\ &\qquad \cdot \log\left(\frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{k=1}^{n} U^{\epsilon_{i,k}} t_i t_k)}{2^n}\right). \end{aligned}$$

Holding the first summation fixed, we note that, since $n$-entropy is maximised by maximally equivocating, $H_n(P) \leq H_n(P')$. For example, if $P$ is flat on $\bigwedge_{k=1}^{n} U^{\epsilon_{n,k}} t_n t_k$, $P(\bigwedge_{k=1}^{n} U^{\epsilon_{n,k}} t_n t_k) = 2^{-n}$ for all $\epsilon_{n,k}$ with $1 \leq k \leq n$, and all these conjunctions are independent of $\bigwedge_{i=1}^{n-1}\bigwedge_{k=1}^{n} U^{\epsilon_{n,k}} t_i t_k$ for all $\epsilon$, then $H_n(P) = H_n(P')$.

Now define $Q := \frac{P+P'}{2}$. Since $\mathbb{E}_\varphi$ is convex and $P$, $P' \in \mathbb{E}_\varphi$, we observe that $Q \in \mathbb{E}_\varphi$.

Since $n$-entropy is a strictly concave function we conclude that $H_n(Q) > H_n(P)$ whenever $P$ and $P'$ disagree on $\mathcal{L}_n$. Since $P \neq P'$ there has to exist some finite $M$ and quantifier-free sentence $\psi \in QFS\mathcal{L}_M$ such that $P(\psi) \neq P'(\psi)$ (Gaifman's Theorem). Since $\mathcal{L}_m \subset \mathcal{L}_{m+1}$ for all $m$ we have that $P$ disagrees with $P'$ on $\mathcal{L}_m$ for all $m \geq M$. We have hence found a $Q \in \mathbb{E}$ such that $H_n(Q) > H_n(P)$ for all large enough $n$. Hence, $P \notin \text{maxent}\,\mathbb{E}_\varphi$. Contradiction. □

We generalise this result to higher quantifier complexity in Appendix 3. These results are summarised in the following theorem.

**Theorem 54** (Zero Measure Premisses) *For all $n \geq 1$ and*

$$\varphi = \exists v_{2n} \forall v_{2n-1} \ldots \exists v_2 \forall v_1 U v_1 v_2 \ldots v_{2n} \in \Sigma_{2n} \text{ or}$$
$$\varphi = \forall v_{2n+1} \ldots \exists v_2 \forall v_1 U v_1 v_2 \ldots v_{2n+1} \in \Pi_{2n+1},$$

*it holds that for all $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy. Hence,* $\text{maxent}\, \mathbb{E}_\varphi = \emptyset$.

Having introduced some pathological cases in which there is no maximal entropy function, we now turn to the question as to what to do in such cases.

For simplicity of exposition, we focus on the case in which we have a single premiss, $\varphi = \exists x \forall y U x y$, considered in Proposition 53. We call a proposition of the form $\forall y U t_i y$ a *witness proposition*. A probability function $P$ that satisfies $\varphi$ distributes probability 1 to the witness propositions, $\lim_{k\to\infty} P(\bigvee_{i=1}^k \forall y U t_i y) = P(\exists x \forall y U x) = 1$. We call a constant $t_i$ a *witness* if $P$ gives positive probability to the corresponding witness proposition $\forall y U t_i y$. Now, the equivocator function, which is the probability function with maximal entropy, gives $\varphi$ measure zero, $P_=(\exists x \forall y U x) = 0$, and thus it has no witnesses. Given $P$ that satisfies $\varphi$, one can construct a function $Q$ that has greater entropy than $P$ by making $Q$ 'closer to' the equivocator in one or both of two ways:

1. Delaying the witnesses. If there are infinitely many witnesses, then one can create $Q$ by increasing the index of each witness in an appropriate way in order to make $Q$ more like the equivocator than $P$ for each fixed $n$. For example, if $t_{i_1}, t_{i_2}, \ldots$ are the witnesses for $P$, one can construct $Q$ with witnesses $t_{i_2}, t_{i_3}, \ldots$, ensuring that $Q(\forall y U t_{i_1} y) = 0$ and $Q(\forall y U t_{i_j} y) = P(\forall y U t_{i_{j-1}} y)$ for each $j > 1$.
2. Flattening the distribution over witness propositions. Entropy can be increased by increasing the number of witnesses, if there are finitely many, and distributing probability more equally to the witnesses, decreasing the rate at which the probability of $\bigvee_{i=1}^k \forall y U t_i y$ converges to 1.

The approach taken in the proof of Proposition 53 involved a mixture of these strategies: delaying witnesses to give $P'$, and then flattening the distribution by taking a convex combination of $P$ and $P'$, to yield $Q$.

One might argue that although the first of these two strategies increases $n$-entropy for sufficiently large $n$, it does not on its own lead to a function that is more equivocal in an intuitive sense. Hence, this seems to be a case in which the formal concept of maximal entropy fails to adequately explicate the concept of being maximally equivocal. (In contrast, the second strategy is unproblematic: flattening the distribution over witness propositions does seem to be a genuine way of generating a more equivocal probability function.)

The explication of maximal entropy can however be refined to avoid this problem: we can deem $P$ to have greater entropy than $Q$ just when, for every reordering $f$ of the constants that do not appear in the premisses, $f(P)$ dominates $f(Q)$ in $n$-entropy for sufficiently large $n$. Note that this refinement relativises the greater-entropy relation to the premisses.

This refinement eradicates the first of the two strategies: delaying witnesses no longer increases entropy, because there are reorderings with respect to which the witnesses are not delayed. The refinement leaves intact the second kind of strategy.

If we accept this refinement, the question then becomes: what policy should be adopted when there is no maximal entropy function because of increases in entropy of the second kind?

Williamson [43, pp. 29–30] suggests a pragmatic policy: to take inferences to be determined by probability functions with *sufficiently* great entropy. Here, the cut-off between functions that have sufficiently great entropy and those that do not may depend on features of the problem or on the users of the logic, and may not be precise. Choosing a probability function with sufficiently great entropy amounts to a choice of $P$ such that $P(\bigvee_{i=1}^{k} \forall y U t_i y)$ converges to 1 sufficiently slowly as $k \longrightarrow \infty$.

Further desiderata may be imposed. For example, one might suggest equivocating between the constants by treating them equally. The thought here is that each constant should be a witness, because there needs to be at least one witness and the premiss gives no grounds for discriminating between constants that are witnesses and those that are not. This line of reasoning motivates giving each witness proposition the same probability $s > 0$ and making witness propositions probabilistically independent.[10] In which case, $P(\bigvee_{i=1}^{k} \forall y U t_i y) = 1 - (1-s)^k$, which converges to 1 as $k \longrightarrow \infty$, as required. Now, decreasing $s$ (and distributing the corresponding probability equally amongst $n$-states) will lead to a probability function with greater entropy—this is an application of the second of the two strategies outlined above. The pragmatic policy then amounts to drawing inferences from probability functions that correspond to values of $s > 0$ that are sufficiently small. One approach here is to take $s$ to be sufficiently small just when taking $s$ any smaller would not make a significant difference with respect to practical purposes.

In sum, we see that although these pathological examples require refinements to the overall approach, there is scope to devise policies that allow one to extend objective Bayesian inductive logic even to these difficult measure-zero cases.

## 10 Conclusion

Objective Bayesian inductive logic defines inductive entailment from a set of (possibly probabilistic) premisses in terms of maximal entropy probability functions that satisfy the given premisses. To be more precise, a set of premisses inductively entails a conclusion if every probability function with maximal entropy that satisfies the premisses also satisfies the conclusion. This is a very natural approach to inductive logic that has been studied extensively in the literature in the context of reasoning with propositional languages. An immediate task that arises with this approach is then to

[10] Such a distribution fits well with the maximal entropy approach, since it encapsulates symmetry and independence properties that have been used to motivate entropy maximisation [28, 31].

find these maximal entropy probability functions in order to perform inference. This is a straightforward, although possibly computationally expensive, problem when working with propositional languages. For more expressive languages, however, it is not clear how one should proceed to determine these maximal entropy probability functions. In this paper, we have studied this problem for premisses and conclusion that are given in terms of constraints on the probabilities of sentences of a first order language.

To do so we first introduced the notion of a *limit in entropy* and discussed its use for determining maximal entropy probability functions. We distinguished what we call the measure-zero sentences from those that have positive measure. Measure-zero sentences are sentences that are assigned probability zero by the equivocator function $P_=$. Intuitively, measure-zero sentences are those that have very few models. To be more precise, these are sentences for which the proportion of term structures with a countably infinite domain that satisfy them is negligible. We showed that for categorical premisses with positive measure, the maximal entropy approach agrees with Bayesian conditionalisation. This then generalises to Jeffrey conditionalisation when dealing with a non-categorical premiss that is given in terms of a constraint on the probability of some sentence. With these results in place we then showed that inductive tautologies are preserved on learning a premiss that involves an inductively consistent proposition. Moreover, although there is a sense in which comparative entropy may depend on the ordering of constants in the language, the probability functions with maximal entropy remain invariant under such permutations in the cases in which the maximal entropy approach agrees with Bayesian or Jeffrey conditionalisation.

These results not only clarify which probabilities the maximal entropy probability functions assign for inductive inference but also give a constructive method for calculating the maximal entropy probabilities. On the one hand, this shows that the maximal entropy approach agrees with standard conceptions of baseline rationality, which appeal to conditionalisation. On the other, it witnesses the stability and generality of Bayesian conditionalisation as a process of probabilistic learning.

Finally, we turned our attention to inference from zero-measure premisses and identified a certain class of zero-measure sentences for which there is no maximal entropy probability function. This leaves the question of inductive inference from these pathological zero-measure premisses open. The issue is then to understand which inferences from zero measure premisses are rational and how to systematically characterize such inferences in terms of a unified inference process, and we developed a strategy for doing this.

Another interesting open question concerns what more can be said about inductive inference from multiple non-categorical premisses. Moreover, our results on objective Bayesian inductive logic have concerned languages containing only relation symbols. It is natural to extend these considerations to languages also containing a symbol for equality and function symbols, which have already been studied in Pure Inductive Logic [13, 17, 23, 32, 33]. Finally, our hope here is that these results can also suggest new avenues for investigating the open cases of the *entropy limit conjecture* that concerns the equivalence of the two main approaches to inductive inference introduced in Section 1.

## Appendix 1: Proofs of Proposition 15 and Theorem 16

First let us recount some basic information-theoretic facts.

The *n-divergence* of two probability functions $P$ and $Q$ is defined as the Kullback-Leibler divergence of $P$ from $Q$ on $\mathcal{L}_n$:

$$d_n(P, Q) \stackrel{\text{df}}{=} \sum_{\omega \in \Omega_n} P(\omega) \log \frac{P(\omega)}{Q(\omega)}.$$

A Pythagorean theorem holds for the $n$-divergence $d_n$ [7, Theorem 11.6.1]:

$$d_n(P, Q) \geq d_n(P, R_n) + d_n(R_n, Q),$$

for any convex $\mathbb{F} \subseteq \mathbb{P}$, if $P \in \mathbb{F}$ and $Q \notin \mathbb{F}$, where $R_n \in \arg\inf_{S \in \mathbb{F}} d_n(S, Q)$.

Consequently, for any $P \in \mathbb{E}$ and $Q_n \in \mathbb{H}_n$ [24, corollary 32]:

$$H_n(Q_n) - H_n(P) \geq d_n(P, Q_n).$$

Pinsker's inequality connects the $L_1$ distance to $n$-divergence (see, e.g., [7, Lemma 11.6.1]):

$$d_n(P, Q) \geq \frac{1}{2} \|P - Q\|_n^2 .$$

**Proposition 15** *If $P$ is a limit in entropy of $\mathbb{E}$ then there are $Q_n \in \mathbb{H}_n$ such that $\|Q_n - P\|_n \longrightarrow 0$ as $n \longrightarrow \infty$.*

*Proof* Putting our last two information-theoretic facts together we have that

$$\begin{aligned} H_n(Q_n) - H_n(P) \geq &\quad d_n(P, Q_n) \\ \geq &\quad \frac{1}{2} \|P - Q_n\|_n^2 , \end{aligned}$$

for $Q_n \in \mathbb{H}_n$ and $P \in \mathbb{E}$.

Now, if $P$ is a limit in entropy of $\mathbb{E}$ then there are $Q_n \in \mathbb{H}_n$ such that $|H_n(Q_n) - H_n(P)| \longrightarrow 0$ as $n \longrightarrow \infty$. Hence $\|P - Q_n\|_n^2$ also converge to zero, as required. □

**Theorem 16** *If $\mathbb{E}$ contains a limit in entropy $P$ then*

$$\text{maxent}\,\mathbb{E} = \{P\}.$$

*Proof* First we shall show that $P \in \text{maxent}\,\mathbb{E}$; later we shall see that there is no other member of $\text{maxent}\,\mathbb{E}$.

First, then, assume for contradiction that $P \notin \text{maxent}\,\mathbb{E}$. Then there is some $Q \in \mathbb{E}$ such that $Q$ has greater entropy than $P$. That is, for sufficiently large $n$, $H_n(Q_n) \geq H_n(Q) > H_n(P)$, where the $Q_n \in \mathbb{H}_n$ converge in entropy (and, by Proposition 15, in $L_1$) to $P$. N.b., $Q \neq P$. Hence, for sufficiently large $n$,

$$\begin{aligned} H_n(Q_n) - H_n(P) &> H_n(Q_n) - H_n(Q) \\ &\geq d_n(Q, Q_n) \\ &\geq \frac{1}{2} \|Q - Q_n\|_n^2 . \end{aligned}$$

Since the $Q_n$ converge in entropy to $P$, they converge in $L_1$ to $Q$. By the uniqueness of $L_1$ limit points, $Q = P$: a contradiction. Hence $P \in \text{maxent}\,\mathbb{E}$, as required.

Next we shall see that $P$ is the unique member of maxent $\mathbb{E}$. Suppose for contradiction that there is some $P^\dagger \in \text{maxent}\,\mathbb{E}$ such that $P^\dagger \neq P$. Then $P$ cannot eventually dominate $P^\dagger$ in $n$-entropy—i.e., there is some infinite set $J \subseteq \mathbb{N}$ such that for $n \in J$,

$$H_n(P^\dagger) \geq H_n(P).$$

Let $R \stackrel{\text{df}}{=} \lambda P^\dagger + (1 - \lambda)P$ for some $\lambda \in (0, 1)$. Now by the log-sum inequality [7, Theorem 2.7.1], for all $n \in J$ large enough that $P^\dagger(\omega_n) \neq P(\omega_n)$ for some $\omega_n \in \Omega_n$,

$$\begin{aligned} H_n(R) &> \lambda H_n(P^\dagger) + (1 - \lambda)H_n(P) \\ &\geq \lambda H_n(P) + (1 - \lambda)H_n(P) \\ &= H_n(P). \end{aligned}$$

Hence,

$$\begin{aligned} H_n(Q_n) - H_n(P) &> H_n(Q_n) - H_n(R) \\ &\geq d_n(R, Q_n), \end{aligned}$$

for large enough $n \in J$.

Now by Pinsker's inequality and the definition of $R$,

$$\begin{aligned} d_n(R, Q_n) &\geq \frac{1}{2} \|R - Q_n\|_n^2 \\ &= \frac{1}{2} \left\| P - Q_n + \lambda(P^\dagger - P) \right\|_n^2 \\ &= \frac{1}{2} \left( \sum_{\omega_n \in \Omega_n} \left| P(\omega_n) - Q_n(\omega_n) + \lambda(P^\dagger(\omega_n) - P(\omega_n)) \right| \right)^2 . \end{aligned}$$

Let $f_n(\varphi) \stackrel{\text{df}}{=} P(\varphi) - Q_n(\varphi) + \lambda(P^\dagger(\varphi) - P(\varphi))$ and $\rho_n \stackrel{\text{df}}{=} \bigvee_{f_n(\omega_n)>0} \omega_n$. Then,

$$\begin{aligned} \sum_{\omega_n \in \Omega_n} |f_n(\omega_n)| &= \sum_{\omega_n : f_n(\omega_n)>0} f_n(\omega_n) - \sum_{\omega_n : f_n(\omega_n) \leq 0} f_n(\omega_n) \\ &= \sum_{\omega_n : f_n(\omega_n)>0} f_n(\omega_n) - \sum_{\omega_n : f_n(\omega_n) \not> 0} f_n(\omega_n) \\ &= f_n(\rho_n) - f_n(\neg\rho_n) \\ &= 2 f_n(\rho_n) \end{aligned}$$

after substituting $P(\neg\rho_n) = 1 - P(\rho_n)$ etc.

Let us consider the behaviour of

$$f_n(\rho_n) = P(\rho_n) - Q_n(\rho_n) + \lambda(P^\dagger(\rho_n) - P(\rho_n))$$

as $n \longrightarrow \infty$. Now, $P(\rho_n) - Q_n(\rho_n) \longrightarrow 0$ as $n \longrightarrow \infty$, because $Q_n$ converges in $L_1$ to $P$. However, $\lambda(P^\dagger(\rho_n) - P(\rho_n)) \not\longrightarrow 0$ as $n \longrightarrow \infty$, as we shall now see. $P^\dagger \neq P$ by assumption, so they must differ on some quantifier-free sentence $\psi$, a sentence of $\mathcal{L}_m$, say. Suppose without loss of generality that $P^\dagger(\psi) > P(\psi)$ (otherwise take $\neg\psi$ instead) and let $\delta = P^\dagger(\psi) - P(\psi) > 0$. Now for $n \geq m$,

$$f_n(\rho_n) = \sum_{\omega_n : f_n(\omega_n) > 0} f_n(\omega_n) \geq \sum_{\omega_n \models \psi} f_n(\omega_n) = f_n(\psi).$$

Since $Q_n$ converges in $L_1$ to $P$ we can consider $n > m$ large enough that [7, Equation 11.137]:

$$\| Q_n - P \|_n = 2 \max_{\varphi \in S\mathcal{L}_n} (Q_n(\varphi) - P(\varphi)) < \lambda\delta.$$

In particular, since $\psi$ is quantifier-free, $Q_n(\psi) - P(\psi) \leq \max_{\varphi \in S\mathcal{L}_n}(Q_n(\varphi) - P(\varphi)) < \lambda\delta/2$. For any such $n$,

$$\begin{aligned} f_n(\rho_n) &\geq f_n(\psi) \\ &= P(\psi) - Q_n(\psi) + \lambda(P^\dagger(\psi) - P(\psi)) \\ &> -\frac{\lambda\delta}{2} + \lambda\delta \\ &= \frac{\lambda\delta}{2}. \end{aligned}$$

Putting the above parts together, we have that for sufficiently large $n \in J$,

$$H_n(Q_n) - H_n(P) > d_n(R, Q_n) \geq \frac{(2 f_n(\rho_n))^2}{2} > \frac{\lambda^2\delta^2}{2} > 0.$$

However, that these $H_n(Q_n) - H_n(P)$ are bounded away from zero contradicts the assumption that the $Q_n$ converge in entropy to $P$. Hence, $P$ is the unique member of maxent $\mathbb{E}$, as required. □

## Appendix 2. Alternative Proof of Corollary 20

This appendix provides a more direct proof of Corollary 20, which identifies an important scenario in which the equivocator function conditioned on a categorical constraint is the maximal entropy function.

**Corollary 20** *If $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large $n$ then*

$$\text{maxent}\, \mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}.$$

*Proof* There are two cases: either $P_=(\varphi) = 1$ or $P_=(\varphi) < 1$.

If $P_=(\varphi) = 1$ then $P_= \in \mathbb{E}_\varphi$ and $P_=(\cdot|\varphi) = P_=(\cdot)$. $P_=$ is the unique member of maxent $\mathbb{E}_\varphi$ because the equivocator function has greater entropy than any other probability function, so maxent $\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}$, as required.

If $P_=(\varphi) < 1$ then we can proceed as follows.

Since $P_=(\varphi) > 0$, $P_=(\cdot|\varphi)$ is well defined. $P_=(\varphi|\varphi) = 1$ so $P_=(\cdot|\varphi) \in \mathbb{E}$. Thus $\mathbb{E}_\varphi \neq \emptyset$.

Suppose for contradiction that maxent $\mathbb{E}_\varphi \neq \{P_=(\cdot|\varphi)\}$. Then in $\mathbb{E}_\varphi$ there must be some $P^\dagger \neq P_=(\cdot|\varphi)$ that is not eventually dominated in entropy by $P_=(\cdot|\varphi)$. That is, there is some infinite $J \subseteq \mathbb{N}$ such that $H_n(P^\dagger) \geq H_n(P_=(\cdot|\varphi))$ for all $n \in J$. (To see this consider that there are three cases: (i) if maxent $\mathbb{E}_\varphi = \emptyset$ then every member of $\mathbb{E}_\varphi$ is eventually dominated by some other in entropy, so $P_=(\cdot|\varphi)$ is dominated by some $P^\dagger$ and $P^\dagger$ is not dominated by $P_=(\cdot|\varphi)$; (ii) if $P_=(\cdot|\varphi) \notin$ maxent $\mathbb{E}_\varphi = \{P^\dagger, \ldots\}$ then $P^\dagger$ is not dominated by $P_=(\cdot|\varphi)$; (iii) if maxent $\mathbb{E}_\varphi = \{P_=(\cdot|\varphi), P^\dagger, \ldots\}$ then $P^\dagger$ is not dominated by $P_=(\cdot|\varphi)$.)

Define a probability function $Q \stackrel{\text{df}}{=} \lambda P^\dagger + (1-\lambda)P_=(\cdot|\varphi)$ for some $\lambda \in (0, 1)$. By the log-sum inequality [7, Theorem 2.7.1], for all $n \in J$ large enough that $P^\dagger(\omega) \neq P_=(\omega|\varphi)$ for some $\omega \in \Omega_n$,

$$\begin{aligned} H_n(Q) &> \lambda H_n(P^\dagger) + (1-\lambda)H_n(P_=(\cdot|\varphi)) \\ &\geq \lambda H_n(P_=(\cdot|\varphi)) + (1-\lambda)H_n(P_=(\cdot|\varphi)) \\ &= H_n(P_=(\cdot|\varphi)). \end{aligned}$$

However, that $H_n(Q) > H_n(P_=(\cdot|\varphi))$ for sufficiently large $n \in J$ contradicts the assumption that $\mathbb{H}_n$ contains $P_=(\cdot|\varphi)$ for sufficiently large $n$. Hence maxent $\mathbb{E}_\varphi = \{P_=(\cdot|\varphi)\}$, as required. □

## Appendix 3. Zero Measure Premisses of Higher Quantifier Complexity

**Proposition 55** ($\Sigma_{2m}$) *For* $\varphi = \exists x_{2m} \forall x_{2m-1} \ldots \forall x_1 U x_{2m} x_{2m-1} \ldots x_1 \in \Sigma_{2m}$ *it holds that for all* $P \in \mathbb{E}_\varphi$ *there exists a probability function* $Q \in \mathbb{E}_\varphi$ *which has greater entropy. Hence,* maxent $\mathbb{E}_\varphi = \emptyset$.

*Proof* For ease of notation we will write $Ut_i\vec{t}$ for $Ut_i t_{k_{2m-1}} \ldots t_{k_1}$ and $\bigwedge_{t=1}^n Ut_i\vec{t}$ for $\bigwedge_{k_{2m-1}=1}^n \cdots \bigwedge_{k_1=1}^n Ut_i t_{k_{2m-1}} \ldots t_{k_1}$.

Suppose for contradiction that maxent $\mathbb{E} \neq \emptyset$ and let $P \in$ maxent $\mathbb{E}$. Note that $P_=(\varphi) = 0 < 1 = P(\varphi)$. Hence, $P \neq P_=$.

Let us now define a probability function $P' \in \mathbb{E}$ by shifting all witnessing of $\exists x_{2m} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U\vec{x}$ by one and then adding a constant $t_1$ such that $Ut_1\vec{t}$ is independent from all other literals for all $\vec{t}$. Intuitively, the literals $\pm Ut_i\vec{t}$ are replaced by $\pm Ut_{i+1}\vec{t}$.

Formally, let $\omega_n \in \Omega_n = \bigwedge_{i,t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}$ be an arbitrary $n$-state. Then define $P'$ by

$$P'(\omega_n) := P\left(\bigwedge_{i=2}^{n}\bigwedge_{t=1}^{n} U^{\epsilon_{i-1,\vec{t}}} t_{i-1}\vec{t}\right) \cdot P_{=}\left(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t}\right)$$
$$= \frac{P(\bigwedge_{i=1}^{n-1}\bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})}{2^{n^{2m-1}}}.$$

Firstly, we note that

$$P'(\forall x_{2m-1}\exists x_{2m-2}\ldots\forall x_1 \; Ut_1\vec{x}) = \lim_{n\to\infty} P'\left(\bigwedge_{j=1}^{n} \exists x_{2m-2}\ldots\forall x_1 Ut_1 t_j \vec{x}\right)$$
$$= \lim_{n\to\infty} P_{=}\left(\bigwedge_{j=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_1 t_j \vec{x}\right) = 0. \quad (11)$$

So, according to $P'$ the constant $t_1$ is not a witness of the existential premiss sentence $\varphi$.

We next show that $P \neq P'$. Firstly, note that

$$\lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} \forall y \exists x_{2m-2}....\forall x_1 Ut_1 y\vec{x}\right) = P(\exists z\forall y\exists x_{2m-2}....\forall x_1 Uzy\vec{x})) = 1$$

and thus there is a smallest $i \in \mathbb{N}$ for which $P(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i\vec{x}) > 0$. With this and Eq. 11, we have

$$\min\{i \in \mathbb{N} \;:\; P'(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i x_{2m-1}\vec{x}) > 0\}$$
$$= \min\left\{i \in \mathbb{N} \;:\; \lim_{n\to\infty} P'\left(\bigwedge_{k=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_i t_k \vec{x}\right) > 0\right\}$$
$$= \min\left\{i \in \mathbb{N} \;:\; \lim_{n\to\infty} P\left(\bigwedge_{k=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_{i-1} t_k \vec{x}\right) > 0\right\}$$
$$= 1 + \min\left\{i \in \mathbb{N} \;:\; \lim_{n\to\infty} P\left(\bigwedge_{k=1}^{n} \exists x_{2m-2}....\forall x_1 Ut_i t_k \vec{x}\right) > 0\right\}$$
$$= 1 + \min\{i \in \mathbb{N} \;:\; P(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i x_{2m-1}\vec{x}) > 0\}.$$

So, $P \neq P'$.

We also observe that for all $i \geq 2$,

$$P'(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i\vec{x}) = P(\forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_{i-1}\vec{x})$$

and furthermore,

$$P'\left(\bigvee_{i\in I} \forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_i\vec{x}\right) = P\left(\bigvee_{i\in I} \forall x_{2m-1}\exists x_{2m-2}....\forall x_1 Ut_{i-1}\vec{x}\right)$$

for all finite index sets $I$. So,

$$\begin{aligned} P'(\exists y \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U y \vec{x}) &= \lim_{n\to\infty} P'\left(\bigvee_{i=1}^{n} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}\right) \\ &\geq \lim_{n\to\infty} P'\left(\bigvee_{i=2}^{n} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}\right) \\ &= \lim_{n\to\infty} P\left(\bigvee_{i=1}^{n-1} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}\right) \\ &= \lim_{n\to\infty} P\left(\bigvee_{i=1}^{n} \forall x_{2m-1} \exists x_{2m-2} .... \forall x_1 U t_i \vec{x}\right) \\ &= 1. \end{aligned}$$

This means that $P'(\exists x \forall y U x y) = 1$ and thus, as advertised, $P' \in \mathbb{E}$.

We now calculate $n$-entropies of $P$ and $P'$ and find for $n \geq 1$ that:

$$\begin{aligned} H_n(P) = &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 2 \leq i \leq n}} \sum_{\epsilon_{1,\vec{t}} \in \{0,1\}} P\left(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}\right) \\ &\cdot \log\left(P\left(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}\right)\right) \\ H_n(P') = &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 1 \leq i \leq n}} P'\left(\bigwedge_{i,t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}\right) \cdot \log\left(P'\left(\bigwedge_{i,t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}\right)\right) \\ = &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 2 \leq i \leq n}} \sum_{\epsilon_{1,\vec{t}} \in \{0,1\}} P'\left(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}\right) \\ &\cdot \log\left(P'\left(\bigwedge_{t=1}^{n} U^{\epsilon_{1,\vec{t}}} t_1 \vec{t} \wedge \bigwedge_{i=2}^{n} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t}\right)\right) \\ = &- \sum_{\substack{\epsilon_{i,\vec{t}} \in \{0,1\} \\ 2 \leq i \leq n}} \sum_{\epsilon_{1,\vec{t}} \in \{0,1\}} \frac{P(\bigwedge_{i=1}^{n-1} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})}{2^{n^{2m-1}}} \\ &\cdot \log\left(\frac{P(\bigwedge_{i=1}^{n-1} \bigwedge_{t=1}^{n} U^{\epsilon_{i,\vec{t}}} t_i \vec{t})}{2^{n^{2m-1}}}\right). \end{aligned}$$

Holding the first summation fixed, we note that, since $n$-entropy is maximised by maximally equivocating, $H_n(P) \leq H_n(P')$. Now define $Q := \frac{P+P'}{2}$. Since $\mathbb{E}$ is convex and $P, P' \in \mathbb{E}$, we observe that $Q \in \mathbb{E}$.

Since $n$-entropy is a strictly concave function we conclude that $H_n(Q) > H_n(P)$ whenever $P$ and $P'$ disagree on $\mathcal{L}_n$. Since $P \neq P'$ there has to exist some finite $M$ and quantifier-free sentence $\psi \in QFS\mathcal{L}_M$ such that $P(\psi) \neq P'(\psi)$ (Gaifman's Theorem). Since $\mathcal{L}_m \subset \mathcal{L}_{m+1}$ for all $m$ we have that $P$ disagrees with $P'$ on $\mathcal{L}_m$ for all $m \geq M$. We have hence found a $Q \in \mathbb{E}$ such that $H_n(Q) > H_n(P)$ for all large enough $n$. Hence, $P \notin \text{maxent}\,\mathbb{E}$. Contradiction. □

**Proposition 56** ($\Pi_3$) *For $\varphi = \forall x \exists y \forall z Sxyz \in \Pi_3$ it holds that for all $P \in \mathbb{E}_\varphi$ there exists a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy. Hence,* $\text{maxent}\,\mathbb{E}_\varphi = \emptyset$.

*Proof* Let us first note that

$$\begin{aligned}\mathbb{E}_\varphi &= \{P \in \mathbb{P} \ : \ P(\varphi) = 1\} \\ &= \{P \in \mathbb{P} \ : \ P(\exists y \forall z S t_1 y z) = 1, \ P(\exists y \forall z S t_2 y z) = 1, \dots, \}. \qquad (12)\end{aligned}$$

Assume for contradiction that $P \in \text{maxent}\,\mathbb{E}_\varphi$. Since $P_=(\varphi) = 0$, $P$ cannot be the equivocator. However, since $P \in \mathbb{E}_\varphi$, it must also hold that for all $t_i$ ($i \in \mathbb{N}$) there has to exist some minimal $t_{k_i^*}$ ($k_i^* \geq 1$) such that $P(\forall z S t_i t_{k_i^*} z) > 0$.

We now define a probability function $Q \in \mathbb{E}_\varphi$ which has greater entropy than $P$, which contradicts that $P \in \text{maxent}\,\mathbb{E}_\varphi$. First, we postpone for all $i$ the witnessing (see Proposition 53) to $k_i^* + 1$. This is again achieved by first defining a probability function $P' \in \mathbb{E}_\varphi \setminus \{P\}$ such that $H_n(P') \geq H_n(P)$ for all large enough $n$:

$$P'\left(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l\right) := \frac{P(\bigwedge_{k=1}^{n-1} \bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l)}{2^n}.$$

As we saw in Proposition 53, $P'(\exists y \forall z S t_i y z) = 1$ for all $i \in \mathbb{N}$. Furthermore, for all $i \in \mathbb{N}$ there exists an $n_i \in \mathbb{N}$ and $\epsilon_{k,l} \in \{0, 1\}^{n_i \times n_i}$ such that $P'(\bigwedge_{k=1}^{n_i} \bigwedge_{l=1}^{n_i} S^{\epsilon_{k,l}} t_i t_k t_l) \neq P(\bigwedge_{k=1}^{n_i} \bigwedge_{l=1}^{n_i} S^{\epsilon_{k,l}} t_i t_k t_l)$.

Given the way we wrote $\mathbb{E}_\varphi$ (see Eq. 12), we see that every extension of $P'$ to a probability function—which so far has not been defined on the entire language—will be in $\mathbb{E}_\varphi$ since membership in $\mathbb{E}_\varphi$ solely depends on sub-states where the first constant is fixed to some $t_i$.

We now define $P'$ on an arbitrary $n$-state $\omega_n$ of the language, and hence on the entire language by

$$P'(\omega_n) := \prod_{i=1}^{n} P'\left(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_i t_k t_l\right).$$

Because of the additivity of the entropy function [8, P. 63], we also find for all $n \in \mathbb{N}$ that

$$H_n(P') = -\sum_{i=1}^{n} \sum_{\substack{\epsilon_{i,r,s} \in \{0,1\} \\ 1 \leq r \leq n \\ 1 \leq s \leq n}} P'\left(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_1 t_k t_l\right) \cdot \log\left(P'\left(\bigwedge_{k=1}^{n} \bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_1 t_k t_l\right)\right).$$

Since the entropy function is maximised for independent variables we also find:

$$H_n(P) \geq -\sum_{i=1}^{n} \sum_{\substack{\epsilon_{i,r,s}\in\{0,1\}\\ 1\leq r\leq n\\ 1\leq s\leq n}} P'\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_i t_k t_l\right)\cdot \log\left(P'\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} S^{\epsilon_{i,k,l}} t_i t_k t_l\right)\right).$$

Now recall that we saw in Proposition 53 that the following inequality holds for all large enough fixed $i \in \mathbb{N}$:

$$\begin{aligned}
{}_iH_n(P') &:= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 1\leq r\leq n\\ 1\leq s\leq n}} P'\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l\right)\cdot \log\left(P'\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l\right)\right)\\
&= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 1\leq r\leq n\\ 1\leq s\leq n}} P'\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} U^{\epsilon_{k,l}} t_k t_l\right)\cdot \log\left(P'\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} U^{\epsilon_{k,l}} t_k t_l\right)\right)\\
&\geq -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 1\leq r\leq n\\ 1\leq s\leq n}} P\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} U^{\epsilon_{k,l}} t_k t_l\right)\cdot \log\left(P\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} U^{\epsilon_{k,l}} t_k t_l\right)\right)\\
&= -\sum_{\substack{\epsilon_{r,s}\in\{0,1\}\\ 1\leq r\leq n\\ 1\leq s\leq n}} P\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l\right)\cdot \log\left(P\left(\bigwedge_{k=1}^{n}\bigwedge_{l=1}^{n} S^{\epsilon_{k,l}} t_i t_k t_l\right)\right) := {}_iH_n(P).
\end{aligned}$$

So, we have for all large enough $n \in \mathbb{N}$ that

$$H_n(P') = \sum_{i=1}^{n} {}_iH_n(P') \geq \sum_{i=1}^{n} {}_iH_n(P) \geq H_n(P).$$

We again put $Q := \frac{P+P'}{2}$ and note that since $P \neq P'$, $Q \neq P$. Since $P' \in \mathbb{E}_\varphi$ we easily find by applying the convexity of $\mathbb{E}_\varphi$ that $Q \in \mathbb{E}_\varphi$. Furthermore, $H_n(Q) > H_n(P)$ for all large enough $n \in \mathbb{N}$ since $Q$ is a convex combination of $P$ and $P'$ and $H_n(P') \geq H_n(P)$ for all $n \in \mathbb{N}$. □

**Proposition 57** ($\Pi_{2m+3}$) *For* $\varphi = \forall v_1 \exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R v_1 w_1 \ldots v_m w_m x y z \in \Pi_{2m+1}$ *and for all* $P \in \mathbb{E}_\varphi$ *there exists a probability function* $Q \in \mathbb{E}_\varphi$ *which has greater entropy than* $P$. *Hence,* maxent $\mathbb{E}_\varphi = \emptyset$.

*Proof* The proof proceeds by induction on the quantifier complexity $m$.

The *base case* $m = 0$ is Proposition 56.

The *induction step* for $m \geq 1$ assumes the result for $m - 1 \geq 0$. The proof follows the blueprint laid out in the base case.

Let us first note that

$$\begin{aligned}\mathbb{E}_\varphi &= \{P \in \mathbb{P} \; : \; P(\varphi) = 1\} \\ &= \{P \in \mathbb{P} \; : \; P(\exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_1 w_1 \ldots v_m w_m xyz) = 1, \\ &\qquad P(\exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_2 w_1 \ldots v_m w_m xyz) = 1, \\ &\qquad \ldots, \}. \end{aligned} \tag{13}$$

Assume for contradiction that $P \in \operatorname{maxent} \mathbb{E}_\varphi$. Since $P_=(\varphi) = 0$, $P$ cannot be the equivocator. However, since $P \in \mathbb{E}_\varphi$, it must also hold that for all $t_i$ ($i \in \mathbb{N}$) there has to exist some minimal $t_{k_i^*}$ ($k_i^* \geq 1$) such that $P(\forall v_2 \exists w_2 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_i t_{k_i^*} v_2 w_2 \ldots v_m w_m xyz) > 0$. We now postpone this witnessing as usual.

We begin by assigning probabilities to substates fixing $t_i$

$$P'\left(\bigwedge_{b_1=1}^{n}\bigwedge_{a_2=1}^{n}\cdots\bigwedge_{a_{m+1}=1}^{n}\bigwedge_{b_{m+1}=1}^{n}\bigwedge_{a_{m+2}=1}^{n} R^{\epsilon_{b_1,a_2,\ldots,a_{m+2}}} t_i t_{b_1} t_{a_2} \ldots t_{a_{m+2}}\right) := \frac{P(\bigwedge_{b_1=1}^{n-1}\bigwedge_{a_2=1}^{n}\cdots\bigwedge_{a_{m+1}=1}^{n}\bigwedge_{b_{m+1}=1}^{n}\bigwedge_{a_{m+2}=1}^{n} R^{\epsilon_{b_1,a_2,\ldots,a_{m+2}}} t_i t_{b_1} t_{a_2} \ldots t_{a_{m+2}})}{2^n}.$$

Again, $P'(\exists w_1 \ldots \forall v_m \exists w_m \forall x \exists y \forall z R t_i w_1 \ldots v_m w_m xyz) = 1$ for all $i \in \mathbb{N}$. Furthermore, for all $i \in \mathbb{N}$ there exist an $n_i \in \mathbb{N}$ and $\vec{\epsilon} \in \{0,1\}^{n_i^{2m+2}}$ such that

$$P'\left(\bigwedge_{b_1=1}^{n_i}\bigwedge_{a_2=1}^{n_i}\cdots\bigwedge_{a_{m+1}=1}^{n_i}\bigwedge_{b_{m+1}=1}^{n_i}\bigwedge_{a_{m+2}=1}^{n_i} R^{\epsilon_{b_1,a_2,\ldots,a_{m+2}}} t_i t_{b_1} t_{a_2} \ldots t_{a_{m+2}}\right) :\neq P'\left(\bigwedge_{b_1=1}^{n_i}\bigwedge_{a_2=1}^{n_i}\cdots\bigwedge_{a_{m+1}=1}^{n_i}\bigwedge_{b_{m+1}=1}^{n_i}\bigwedge_{a_{m+2}=1}^{n_i} R^{\epsilon_{b_1,a_2,\ldots,a_{m+2}}} t_i t_{b_1} t_{a_2} \ldots t_{a_{m+2}}\right).$$

In particular, $P' \neq P$.

We now define $P'$ on an arbitrary $n$-state $\omega_n$ of the language, and hence on the entire language, by fixing $\vec{\epsilon}_i \in \{0,1\}^{n^{2m+2}}$ for $1 \leq i \leq n$ and letting

$$P'(\omega_n) := \prod_{i=1}^{n} P'\left(\bigwedge_{\vec{\epsilon}_i \in \{0,1\}^{n^{2m+2}}} R^{\vec{\epsilon}_i} t_i \vec{t}\right).$$

Because of the additivity of the entropy function [8, p. 63], we also find for all $n \in \mathbb{N}$ that

$$\begin{aligned} H_n(P') &= -\sum_{i=1}^{n} \sum_{\vec{\epsilon}_i \in \{0,1\}^{n^{2m+2}}} P'\left(\bigwedge_{\vec{\epsilon}_i \in \{0,1\}^{n^{2m+2}}} R^{\vec{\epsilon}_i} t_i \vec{t}\right) \cdot \log\left(P'\left(\bigwedge_{\vec{\epsilon}_i \in \{0,1\}^{n^{2m+2}}} R^{\vec{\epsilon}_i} t_i \vec{t}\right)\right) \\ &: = \sum_{i=1}^{n} {}_i H_{n,2m+2}(P'). \end{aligned}$$

We now use the proof of Proposition 55 to obtain that for all $i$ and all large enough $n$ (depending on $i$),

$$ {}_iH_{n,2m+2}(P') \geq {}_iH_{n,2m+2}(P). $$

${}_iH_{n,2m}(P)$ is the $n$-entropy of a probability function $P$ on a language containing one $(2m + 2)$-ary relation symbol $U$, $\varphi = \exists w_1 \forall v_2 \exists w_2 \ldots \exists w_{m+1} \forall v_{m+2} U w_1 v_2 w_2 \ldots w_{m+1} v_{m+2} \in \Pi_{2m+2}$ and $P \in \mathbb{E}_\varphi$.

Since $n$-entropy is maximised by probability functions with as many probabilistic independences as possible, we again have:

$$ H_n(P) \geq \sum_{i=1}^{n} {}_iH_{n,2m+2}(P), $$

which overall gives the inequality:

$$ H_n(P') = \sum_{i=1}^{n} {}_iH_{n,2m+2}(P') \geq \sum_{i=1}^{n} {}_iH_{n,2m+2}(P) \geq H_n(P). $$

Taking $Q$ to be any convex combination of $P$ and $P'$, we see that $H_n(Q) > H_n(P)$ for all large enough $n$. This entails that $Q$ has greater entropy than $P$. □

**Acknowledgments** We are grateful to Jeff Paris and Alena Vencovská for very helpful advice.

**Author Contribution** All authors significantly contributed to this manuscript. All authors have read and approved the uploaded version of the manuscript.

**Funding** Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. Jürgen Landes is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) 405961989 and 432308570 and NextGenerationEU funding for the project "Practical Reasoning for Human-Centred Artificial Intelligence". Soroush Rafiee Rad's work is supported by the Dutch Institute for Emergent Phenomena (DIEP) cluster at the University of Amsterdam and partly by the Deutsche Forschungsge-meinschaft (DFG, German Research Foundation) 432308570. Jon Williamson is funded by the Leverhulme Trust (grant RPG-2019-059) and the Deutsche Forschungsgemeinschaft (DFG, grant 432308570).

**Data Availability** This manuscript does not use data.

## Declarations

**Conflict of Interest** No author reports a conflict of interest.

## References


1. Balestrino, A., Caiti, A., & Crisostomi, E. (2006). Efficient numerical approximation of maximum entropy estimates. *International Journal of Control*, *79*(9), 1145–1155.
2. Barnett, O., & Paris, J. B. (2008). Maximum entropy inference with quantified knowledge. *Logic Journal of IGPL*, *16*(1), 85–98.
3. Billingsley, P. (1979). *Probability and measure* (3rd (1995) edn). New York: Wiley.
4. Carnap, R. (1952). *The continuum of inductive methods*. Chicago: University of Chicago Press.
5. Caticha, A., & Giffin, A. (2006). Updating probabilities. In *Proceedings of MaxEnt* (Vol. 872 pp. 31–42).
6. Chen, B., Hu, J., & Zhu, Y. (2010). Computing maximum entropy densities: a hybrid approach. *Signal Processing: An International Journal*, *4*(2), 114–122.
7. Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory* (2nd (2006) edn). New York: Wiley.
8. Csiszár, I. (2008). Axiomatic characterizations of information measures. *Entropy*, *10*(3), 261–273.
9. Gaifman, H. (1964). Concerning measures in first order calculi. *Israel Journal of Mathematics*, *2*(1), 1–18.
10. Goldman, S. A. (1987). Efficient methods for calculating maximum entropy distributions Master's thesis. Electrical Engineering and Computer Science, Massachusetts Institute of Technology.
11. Goldman, S. A., & Rivest, R. (1988). A non-iterative maximum entropy algorithm. In L. Kanal, & J. Lemmer (Eds.) *Uncertainty in Artificial Intelligence 2* (pp. 133–148). North-Holland: Elsevier.
12. Haenni, R., Romeijn, J.-W., Wheeler, G., & Williamson, J. (2011). *Probabilistic logics and probabilistic networks. Synthese Library*. Dordrecht: Springer.
13. Howarth, E., & Paris, J. B. (2019). Pure inductive logic with functions. *Journal of Symbolic Logic*, *84*(4), 1382–1402.
14. Howson, C. (2014). Finite additivity, another lottery paradox and conditionalisation. *Synthese*, *191*(5), 989–1012.
15. Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, *106*(4), 620–630.
16. Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge: Cambridge University Press.
17. Landes, J. (2009). The principle of spectrum exchangeability within inductive logic. PhD thesis, Manchester Institute for Mathematical Sciences.
18. Landes, J. (2021a). A triple uniqueness of the maximum entropy approach. In J. Vejnarová, & N. Wilson (Eds.) *Proceedings of ECSQARU, volume 12897 of LNAI* (pp. 644–656). Cham: Springer.
19. Landes, J. (2021b). The entropy-limit (conjecture) for $\Sigma_2$-premisses. *Studia Logica*, *109*(2), 423–442.
20. Landes, J., & Williamson, J. (2015). Justifying objective Bayesianism on predicate languages. *Entropy*, *17*(4), 2459–2543.
21. Landes, J., & Williamson, J. (2016). Objective Bayesian nets from consistent datasets. In A. Giffin, & K.H. Knuth (Eds.) *Proceedings of MaxEnt*, (Vol. 1757 pp. 020007-1–020007-8). AIP.
22. Landes, J., & Williamson, J. (2022). Objective bayesian nets for integrating consistent datasets. *Journal of Artificial Intelligence Research*, *74*, 393–458.
23. Landes, J., Paris, J. B., & Vencovská, A. (2009). Representation theorems for probability functions satisfying spectrum exchangeability in inductive logic. *International Journal of Approximate Reasoning*, *51*(1), 35–55.
24. Landes, J., Rafiee Rad, S., & Williamson, J. (2021). Towards the entropy-limit conjecture. *Annals of Pure and Applied Logic*, *172*(2), 102870.
25. Lehmann, D., & Magidor, M. (1992). What does a conditional knowledge base entail? *Artificial Intelligence*, *55*(1), 1–60.
26. Ormoneit, D., & White, H. (1999). An efficient algorithm to compute maximum entropy densities. *Econometric Reviews*, *18*(2), 127–140.
27. Paris, J. B. (1994). *The uncertain reasoner's companion*. Cambridge: Cambridge University Press.
28. Paris, J. B. (1998). Common sense and maximum entropy. *Synthese*, *117*(1), 75–93.
29. Paris, J. B., & Rafiee Rad, S. (2010). A note on the least informative model of a theory. In F. Ferreira, B. Löwe, E. Mayordomo, L. Mendes Gomes, J. B. Paris, & S. Rafiee Rad (Eds.) *Proceedings of CiE* (pp. 342–351). Berlin: Springer.
30. Paris, J. B., & Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, *4*(3), 183–223.

31. Paris, J. B., & Vencovská, A. (1997). In defense of the maximum entropy inference process. *International Journal of Approximate Reasoning*, *17*(1), 77–103.
32. Paris, J., & Vencovská, A. (2015). *Pure inductive logic*. Cambridge: Cambridge University Press.
33. Paris, J. B., & Vencovská, A. (2019). Six problems in pure inductive logic. *Journal of Philosophical Logic*, *48*(4), 731–747.
34. Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo: Morgan Kaufmann.
35. Rafiee Rad, S. (2009). Inference processes for probabilistic first order languages. PhD thesis, Manchester Institute for Mathematical Sciences.
36. Rafiee Rad, S. (2017). Equivocation axiom on first order languages. *Studia Logica*, *105*(1), 121–152.
37. Rafiee Rad, S. (2018). Maximum entropy models for $\Sigma_1$ sentences. *Journal of Applied Logics - IfCoLoG Journal of Logics and their Applications*, *5*(1), 287–300.
38. Rafiee Rad, S. (2021). On probabilistic characterisation of models of first order theories. *Annals of Pure and Applied Logic*, *172*(1).
39. Seidenfeld, T. (1986). Entropy and uncertainty. *Philosophy of Science*, *53*(4), 467–491.
40. Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.
41. Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, *31*(2), 131–144.
42. Williamson, J. (2008). Objective Bayesian probabilistic logic. *Journal of Algorithms*, *63*(4), 167–183.
43. Williamson, J. (2010). *In defence of objective Bayesianism*. Oxford: Oxford University Press.
44. Williamson, J. (2017). *Lectures on inductive logic*. Oxford: Oxford University Press.

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