Journal of Applied Logic 11 (2013) 468–486

Contents lists available at SciVerse ScienceDirect

Journal of Applied Logic

www.elsevier.com/locate/jal

From Bayesian epistemology to inductive logic

Jon Williamson

University of Kent, Canterbury, United Kingdom

ARTICLE INFO

Article history:
Available online 18 March 2013

Keywords:
Bayesianism
Objective Bayesianism
Bayesian epistemology
Maximum entropy
Maxent
Inductive logic
Probabilistic logic
Probability logic

ABSTRACT

Inductive logic admits a variety of semantics (Haenni et al. (2011) [7, Part 1]). This paper develops semantics based on the norms of Bayesian epistemology (Williamson, 2010 [16, Chapter 7]). Section 1 introduces the semantics and then, in Section 2, the paper explores methods for drawing inferences in the resulting logic and compares the methods of this paper with the methods of Barnett and Paris (2008) [2]. Section 3 then evaluates this Bayesian inductive logic in the light of four traditional critiques of inductive logic, arguing (i) that it is language independent in a key sense, (ii) that it admits connections with the Principle of Indifference but these connections do not lead to paradox, (iii) that it can capture the phenomenon of learning from experience, and (iv) that while the logic advocates scepticism with regard to some universal hypotheses, such scepticism is not problematic from the point of view of scientific theorising.

© 2013 Elsevier B.V. All rights reserved.

1. Bayesian epistemology as semantics for inductive logic

This section introduces the use of Bayesian epistemology as semantics for inductive logic. The material presented here is based on Williamson [16], to which the reader is referred for more details.

1.1. Bayesian epistemology: A primer

At root, Bayesian epistemology concerns the question of how strongly one should believe the various propositions that one can express. The Bayesian theory that answers this question can be developed in a number of ways, but it is usual to base the theory on the betting interpretation of degrees of belief. According to the betting interpretation, one believes proposition θ to degree x iff, were one to offer a *betting quotient* for θ—a number q such that one would pay qS to receive S in return should θ turn out to be true, where unknown stake $S \in \mathbb{R}$ may depend on q—then $q = x$.

This interpretation of degrees of belief naturally goes hand in hand with the claim that, were one to bet according to one's degrees of belief via the above betting set-up, then one shouldn't expose oneself to avoidable losses. In particular, arguably one's degrees of belief should minimise worst-case expected loss.

This starting point—the betting interpretation together with the loss-avoidance claim—can then be used to motivate various rational norms that answer the main question facing Bayesian epistemology, i.e., that specify how strongly one should believe the various propositions that one can express. Since the context of this paper is inductive logic, we will be particularly concerned with propositions expressed in a logical language. Suppose then that $\mathcal{L}_n$ is a propositional language on elementary propositions $A_1, \ldots, A_n$, with $S\mathcal{L}_n$ the set of propositions formed by recursively applying the usual logical connectives to the elementary propositions. Let Ω_n be the set of *atomic states* of $\mathcal{L}_n$, i.e., propositions ω_n of the form $\pm A_1 \wedge \cdots \wedge \pm A_n$, where $\pm A_i$ is either A_i or its negation.

E-mail address: j.williamson@kent.ac.uk.
URL: http://www.kent.ac.uk/secl/philosophy/jw/.

1570-8683/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jal.2013.03.006

One norm of rational belief, which we shall call the *Probability Norm*, says that one's degrees of belief should satisfy the axioms of probability, for otherwise, in the worst case, stakes are chosen that ensure positive expected loss—equivalently, stakes are chosen that ensure positive loss whichever atomic state turns out to be true (a so-called *Dutch book*). Thus degrees of belief should be probabilities in order to minimise worst-case expected loss:

Theorem 1. *Define function $P : S\mathcal{L}_n \to \mathbb{R}$ by $P(\theta) =$ a given agent's betting quotient for θ. The agent's bets on propositions expressible in $\mathcal{L}_n$ avoid the possibility of a Dutch book if and only if they satisfy the axioms of probability*:

P1. $P(\omega_n) \geqslant 0$ *for each* $\omega_n \in \Omega_n$,
P2. $P(\tau) = 1$ *for some tautology* $\tau \in S\mathcal{L}$,
P3. $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ *for each* $\theta \in S\mathcal{L}$.

See Williamson [16, Theorem 3.2] for a proof. This is known as the Dutch Book Theorem, or the Ramsey–de-Finetti Theorem. P1–3 offer one way of expressing the axioms of probability over the propositional language $\mathcal{L}_n$: this axiomatisation makes it clear that a probability function is determined by its values on the atomic states of $\mathcal{L}_n$.

A second norm, the *Calibration Norm* or *Principal Principle*, says that one's degrees of belief should be calibrated to known physical probabilities. In particular, if one knows just the physical probability $P^*(\theta)$ of θ then one should set one's betting quotient for θ to this physical probability, $P(\theta) = P^*(\theta)$, for otherwise in the worst case stakes will be chosen that render the expected loss positive.[1] While in the case of the Probability Norm, minimising worst-case expected loss is equivalent to avoiding sure loss (a Dutch book), in this case the link appears in a long run of bets rather than in the single bet on θ itself: if one were repeatedly to bet on θ-like events with the same betting quotient for each bet, then one would be susceptible to sure loss in the long run [16, pp. 39–42].

Of course in general one might know only of certain constraints on physical probability, rather than individual physical probabilities such as $P^*(\theta)$. In such a case the Calibration Norm would say that if one knows that the physical probability function P^* lies within some set $\mathbb{P}^*$ of probability functions, then one's belief function P should lie within the convex hull $\langle\mathbb{P}^*\rangle$ of this set of probability functions. The reason being that if one remains in the convex hull one's bets do not have demonstrably greater than the minimum worst-case expected loss (equivalently, one can't be forced to lose money in the long run), but outside the convex hull one can be sure of sub-optimal expected loss (equivalently, one can be forced to lose money in the long run).

A third norm, the *Equivocation Norm*, says that one should not adopt extreme degrees of belief unless forced to by the Probability Norm or the Calibration Norm: one's degrees of belief should be equivocate sufficiently between the basic possibilities that one can express (i.e., between the atomic states of $\mathcal{L}_n$). Equivalently, one's belief function P should satisfy the constraints imposed by the other norms and otherwise should be sufficiently close to the *equivocator* function that gives the same probability to each of the 2^n atomic states, $P_=(\omega_n) = 1/2^n$ for each $\omega_n \in \Omega_n$. (Distance between probability functions is measured by Kullback–Leibler divergence, $d(P, Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n)/Q(\omega_n)$.) Again this norm can be justified by minimising worst-case expected loss; the argument goes as follows [16, pp. 63–65]. In the absence of knowledge about one's losses, one should take the loss function to be logarithmic by default, i.e., one should assume that one will lose $-\log P(\omega_n)$ where ω_n is the atomic state that turns out true, for such a loss function is the only one that satisfies various desiderata that are natural to impose on a default loss function. But then, under some rather general conditions, the P that satisfies constraints imposed by other norms but minimises worst-case expected loss (the *robust Bayes* choice of P) is the P that is closest to the equivocator.

Note that we need the qualification that degrees of belief should be 'sufficiently' close to the equivocator in order to handle the case where there is no function closest to the equivocator. For example, one might know that a coin is biased in favour of heads, so that $P^*(H) > 1/2$ where H signifies heads at the next toss. Arguably then by the Calibration Norm one ought to believe H to degree greater than 1/2, $P(H) > 1/2$. But there is no degree of belief greater than 1/2 that is closest to $P_=(H) = 1/2$. Therefore the most one can expect of an agent is that $P(H)$ is sufficiently close to 1/2, where what counts as sufficiently close depends on pragmatic considerations such as the required accuracy of calculations.

In sum, the betting interpretation together with the loss-avoidance claim lead to three norms: Probability, Calibration and Equivocation. Bayesian epistemologists disagree as to whether to endorse all these norms. All accept the Probability Norm, most accept some version of Calibration Norm, but few accept the Equivocation Norm. The Probability Norm on its own leads to what is sometimes called *strict subjectivism*, Probability together with Calibration leads to *empirically-based subjectivism* and all three norms taken together lead to *objectivism*. Note that strict subjectivism doesn't imply that determining appropriate degrees of belief is entirely a question of subjective choice, inasmuch as the Probability Norm

[1] For simplicity of exposition we take physical probability P^* to be single-case here, defined over the propositions of the agent's language $\mathcal{L}$. But we could instead take physical probability to be generic, defined over repeatedly instantiatable outcomes, and then consider in place of P^* the single-case consequences of physical probability, i.e., the constraints that known generic physical probabilities impose on the agent's degree of belief. See Williamson [19] on this point.

imposes substantial constraints on which degrees of belief count as rational. Neither does objectivism imply that rational degrees of belief are totally objective in the sense of being uniquely determined—there may remain an element of subjective choice, as in the biased coin example introduced above, where the agent is free to choose between any sufficiently equivocal probability function.

In the light of the fact that the three norms have essentially the same motivation, it is hard to argue that this motivation warrants one norm but not another. Accordingly we will accept all three norms in what follows, exploring the consequences of *objective* Bayesian epistemology for inductive logic.

1.2. Predicate languages

Thus far we have introduced Bayesian epistemology in the context of a propositional language. But the framework extends quite naturally to a predicate language, as we shall now see. Suppose then that $\mathcal{L}$ is a first-order predicate language without equality, with finitely many predicate symbols and with countably many constant symbols $t_1, t_2, \ldots$, one for each element of the domain. For $n \geqslant 1$, let $\mathcal{L}_n$ be the finite predicate language involving only constants $t_1, \ldots, t_n$. Let $A_1, A_2, \ldots$ run through the atomic propositions of $\mathcal{L}$, i.e., propositions of the form Ut where U is a predicate symbol and t is a tuple of constant symbols of corresponding arity. Order the $A_1, A_2, \ldots$ as follows: any atomic proposition expressible in $\mathcal{L}_n$ but not expressible in $\mathcal{L}_m$ for $m < n$ should occur later in the ordering than those atomic propositions expressible in $\mathcal{L}_m$. Let $A_1, \ldots, A_{r_n}$ be the atomic propositions expressible in $\mathcal{L}_n$. An *atomic n-state* ω_n is an atomic state $\pm A_1 \wedge \cdots \wedge \pm A_{r_n}$ of $\mathcal{L}_n$. Let Ω_n be the set of atomic n-states.

While the Calibration Norm remains the same as in the propositional case, the Probability Norm needs modification because the predicate language requires an extra axiom of probability:

Theorem 2. *Define function $P : S\mathcal{L} \to \mathbb{R}$ by $P(\theta) =$ the agent's betting quotient for θ. The agent's bets on propositions expressible in $\mathcal{L}$ avoid the possibility of a Dutch book if and only if P is a probability function*:

PP1. $P(\omega_n) \geqslant 0$ *for each $\omega_n \in \Omega_n$ and each n,*
PP2. $P(\tau) = 1$ *for some tautology $\tau \in S\mathcal{L}$,*
PP3. $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ *for each quantifier-free proposition θ, for any n large enough that $\mathcal{L}_n$ contains all the atomic propositions occurring in θ, and*
PP4. $P(\exists x\theta(x)) = \sup_m P(\bigvee_{i=1}^m \theta(t_i))$.

See Williamson [16, Theorem 5.1] for a proof. PP4, the extra probability axiom, is known as *Gaifman's condition*. PP1–4 imply that $P(\exists x\theta(x)) = \lim_{m\to\infty} P(\bigvee_{i=1}^m \theta(t_i))$ and $P(\forall x\theta(x)) = \lim_{m\to\infty} P(\bigwedge_{i=1}^m \theta(t_i))$. As in the propositional case, a probability function is determined by its values on the atomic n-states, but in this case n varies over the natural numbers.

The Equivocation Norm carries over from the propositional case as follows. Define the equivocator by $P_=(\omega_n) = 1/2^{r_n}$ for all n and ω_n. Define the *n-divergence* between P and Q to be $d_n(P, Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n)/Q(\omega_n)$. Say that P is *closer* to R than Q if there is some N such that for all $n \geqslant N$, $d_n(P, R) < d_n(Q, R)$. The Equivocation Norm says that P should be, from all the functions that satisfy the constraints imposed by the Probability and Calibration Norms, one that is sufficiently close to the equivocator. As in the propositional case it is a pragmatic question as to what counts as 'sufficiently' close to the equivocator.

1.3. Inductive logic

Following Haenni et al. [7], we will concern ourselves with inductive logics that yield entailment relationships of the form:

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

Here $\varphi_1, \ldots, \varphi_k, \psi$ are propositions of some given logical language, and the superscripts $X_1, \ldots, X_k, Y$ denote inductive qualities that attach to these propositions—characterising, e.g., their certainty, plausibility, reliability, weight of evidence, or probability. Such an entailment relationship can be read: if X_1 attaches to φ_1, ..., and X_k attaches to φ_k then Y attaches to ψ.

As to which entailment relationships hold depends very much on the semantics for the entailment relation $\mathrel{|\!\!\approx}$. Normally some given semantics will say something like, 'the entailment relationship holds if all interpretations that satisfy the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ also satisfy the conclusion ψ^Y', and will say what an interpretation is and what it is to satisfy the premisses and satisfy the conclusion. In a *probabilistic logic*, or *progic* for short, $X_1, \ldots, X_k, Y$ are sets of probabilities and interpretations are probability functions. Still, there are a wide range of semantics one might give for a progic: the standard semantics says that P satisfies θ^Z if and only if $P(\theta) \in Z$, but other semantics are provided by the theories of probabilistic argumentation (which is closely related to Dempster–Shafer theory), Kyburg's evidential probability, classical statistical inference, Bayesian statistical inference and Bayesian epistemology [7, Part I].

Bayesian epistemology provides semantics in the following way. First, one interprets the logical language in which $\varphi_1, \ldots, \varphi_k, \psi$ are expressed as the language of an agent. Then one can construe the premisses $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}$ as constituting the agent's evidence of physical probabilities: they say that $P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_k) \in X_k$. Finally, interpretations are belief functions, which, by the Probability Norm, are probability functions. By the Calibration Norm a belief function P satisfies the premisses iff it lies in the convex hull of all probability functions that satisfy the relevant evidential constraints, i.e., iff $P \in \langle \mathbb{P}^* \rangle$. Now if the premisses are consistent we can simply take $\mathbb{P}^* = \{P\colon\ P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k\}$. But if the premisses are inconsistent we cannot identify $\mathbb{P}^* = \emptyset$: inconsistent premisses tell us not that there is no chance function but rather that there is something wrong with the premisses. In this case, then, some consistency maintenance procedure needs to be employed. The simplest such procedure is to take $\mathbb{P}^*$ to include any function that lies in a maximal consistent subset of $\{P\colon\ P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k\}$. Of course, if a set of constraints is consistent then the maximal consistent subset of that set is that set itself, so whether or not the premisses are consistent, we can take $\mathbb{P}^* = \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\} \stackrel{\text{df}}{=} \{P\colon\ P$ satisfies some maximal consistent subset of $\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\}$. The conclusion ψ^Y is interpreted as an assertion about rational degree of belief rather than physical probability: $P(\psi) \in Y$. The norms dictate that a belief function that satisfies the premisses also satisfies the conclusion iff $P(\varphi) \in Y$ for all those $P \in \langle \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\rangle$ that are sufficiently equivocal.

It remains to say what counts as 'sufficiently' equivocal. From an epistemological point of view this depends on the context—if probabilities are only needed to 2 decimal places then this leeway can help determine what counts as sufficiently equivocal. But from the logical perspective there may simply be no contextual information available: an entailment relationship such as $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx \psi^Y$ says nothing on its own about considerations such as accuracy. Nevertheless, there are certain things we can say about what can count as sufficiently equivocal, as we shall now see.

Let $\mathbb{E}$ denote the set of probability functions that satisfy constraints imposed by the agent's evidence. In the above exposition of the Calibration Norm, $\mathbb{E} = \langle \mathbb{P}^* \rangle$.[2] Let $\downarrow\mathbb{E}$ denote the set of functions in $\mathbb{E}$ that are maximally equivocal (this set may be empty, as in the biased coin example above) and let $\Downarrow\mathbb{E}$ denote the set of functions in $\mathbb{E}$ that are sufficiently equivocal. Arguably,

E1: $\Downarrow\mathbb{E} \neq \emptyset$. An agent is always entitled to hold some beliefs.
E2: $\Downarrow\mathbb{E} \subseteq \mathbb{E}$. Probability functions calibrated with evidence that are sufficiently equivocal are calibrated with evidence.
E3: If $Q \in \Downarrow\mathbb{E}$ and $R \in \mathbb{E}$ is no less equivocal than Q then $R \in \Downarrow\mathbb{E}$. I.e., if Q is sufficiently equivocal then so is any function in $\mathbb{E}$ that is no less equivocal.
E4: If $\downarrow\mathbb{E} \neq \emptyset$ then $\Downarrow\mathbb{E} = \downarrow\mathbb{E}$. If it is possible to be maximally equivocal then one's belief function should be maximally equivocal.
E5: $\Downarrow\Downarrow\mathbb{E} = \Downarrow\mathbb{E}$. Any function, from those that are calibrated with evidence, that is sufficiently equivocal, is a function, from those that are calibrated with evidence and are sufficiently equivocal, that is sufficiently equivocal.

In the case of a propositional language, for example, these conditions imply that:

$$\Downarrow\mathbb{E} = \begin{cases} \downarrow\mathbb{E}\colon & \downarrow\mathbb{E} \neq \emptyset, \\ \{P \in \mathbb{E}\colon\ |d(P, P_=) - d(\bar{P}, P_=)| \leqslant \epsilon\}\colon & \text{otherwise}, \end{cases}$$

for some ϵ, where $\{\bar{P}\} = \downarrow\overline{\langle \mathbb{E} \rangle}$, i.e., where $\bar{P}$ is the unique function that is maximally equivocal from all those in the convex closure $\overline{\langle \mathbb{E} \rangle}$ of $\mathbb{E}$. In general, if there are no contextual factors available to determine parameters like ϵ that demarcate between sufficiently and insufficiently equivocal, the only option is to set:

$$\Downarrow\mathbb{E} = \begin{cases} \downarrow\mathbb{E}\colon & \downarrow\mathbb{E} \neq \emptyset, \\ \mathbb{E}\colon & \text{otherwise}. \end{cases}$$

We shall take this as the default option in the case of the Bayesian semantics for inductive logic.

We shall use the symbol $\mid\!\approx^{\circledS}$ to signify this Bayesian entailment relation, which can thus be characterised as follows: $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx^{\circledS} \psi^Y$ if and only if $P(\psi) \in Y$ for each maximally equivocal member P of $\langle \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\rangle$, if there are any maximally equivocal members, or for every member P of $\langle \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\rangle$ otherwise.

2. Inferences in Bayesian inductive logic

Perhaps the best way to get an intuitive feel for the above inductive logic is via a series of examples. In this section we will use examples to demonstrate how inferences can be performed in Bayesian inductive logic, and to compare this logic with a closely related logic.

[2] This version of the Calibration Norm presumes that all constraints on degrees of belief are mediated by evidence of physical probabilities. While this is an appropriate assumption in the context of our semantics for inductive logic, where premisses are interpreted exclusively in terms of physical probabilities, this assumption may be violated in other contexts [16, §3.3].

2.1. Categorical premisses in propositional inductive logic

Consider the following invalid argument in propositional logic:

$$\begin{array}{r} a \rightarrow b \\ b \\ \hline a \end{array}$$

We know the argument is invalid by considering its truth table:

a	b	$a \rightarrow b$	b	a
T	T	T	T	T
T	F	F	F	T
F	T	$\boxed{\text{T}}$	$\boxed{\text{T}}$	$\boxed{\text{F}}$
F	F	T	F	F

The two left-most columns go through all the truth assignments to the atomic propositions, while the next two columns give the corresponding truth values of the premisses and the right-most column gives the corresponding truth value of the conclusion. The argument is invalid because the third truth assignment makes the premisses true but the conclusion false.

Now suppose we are faced with the following question: to what extent do the premisses of the above argument entail the conclusion? This can be written:

$$a \rightarrow b, b \mid\!\approx a^{?}.$$

Note that the premisses are categorical in the sense that they are not qualified in any way by some inductive quality. In terms of the Bayesian semantics described above, an unqualified premiss proposition θ implies that $P^*(\theta) = 1$, where, as before P^* denotes chance or physical probability. Normally, this is all that θ implies.[3] So, from the point of view of Bayesian inductive logic, the question might be written,

$$a \rightarrow b,^{\{1\}} b^{\{1\}} \mid\!\approx^{\circledcirc} a^{?}.$$

This asks, given that $a \rightarrow b$ and b have physical probability 1, what degrees of belief would it be rational to attach to the conclusion a?

The Probability Norm says that degrees of belief over the propositional language $\{a, b\}$ had better be probabilities. Recall that the probability of any proposition is the sum of the probabilities of the atomic states that logically imply that proposition. In particular,

$$P(a) = P(a \wedge b) + P(a \wedge \neg b),$$
$$P(b) = P(a \wedge b) + P(\neg a \wedge b),$$
$$P(a \rightarrow b) = P(a \wedge b) + P(\neg a \wedge b) + P(\neg a \wedge \neg b).$$

This means that we can augment a truth table with the probabilities x_i of the atomic states ω_n to calculate the probability of any proposition:

$\pm a \wedge \pm b$	$P(\pm a \wedge \pm b)$	a	b	$a \rightarrow b$	b	a
$a \wedge b$	x_1	T	T	T	T	T
$a \wedge \neg b$	x_2	T	F	F	F	T
$\neg a \wedge b$	x_3	F	T	T	T	F
$\neg a \wedge \neg b$	x_4	F	F	T	F	F

$$P(a) = x_1 + x_2,$$
$$P(b) = x_1 + x_3,$$
$$P(a \rightarrow b) = x_1 + x_3 + x_4.$$

[3] There are exceptions though, as we shall see in Section 3. For example, θ might say that the first hundred randomly sampled ravens have been found to be black, in which case not only does θ imply that θ has probability 1 but, arguably, that the physical probability of the next sampled raven being black is close to 1. θ is called *simple* with respect to P^* if it only imposes the constraint $P^*(\theta) = 1$.

The Calibration Norm says that degrees of belief should satisfy constraints imposed by evidence of physical probability. In the Bayesian semantics for inductive logic, the left-hand side of the entailment relation is interpreted as capturing the available information about physical probability. Since the premiss propositions have physical probability 1, $P((a \to b) \wedge b) = x_1 + x_3 = 1$, so our augmented truth table becomes:

$\pm a \wedge \pm b$	$P(\pm a \wedge \pm b)$	a	b	$a \to b$	b	a
$a \wedge b$	x_1	T	T	$\boxed{\text{T}}$	$\boxed{\text{T}}$	T
$a \wedge \neg b$	0	T	F	F	F	T
$\neg a \wedge b$	x_3	F	T	$\boxed{\text{T}}$	$\boxed{\text{T}}$	F
$\neg a \wedge \neg b$	0	F	F	T	F	F

Finally the Equivocation Norm, as applied to inductive logic, says that degrees of belief should otherwise equivocate, if possible, between the atomic states. Hence $x_1 = x_3 = 1/2$. The probability that attaches to the conclusion a is thus $x_1 + x_2 = 1/2$:

$\pm a \wedge \pm b$	$P(\pm a \wedge \pm b)$	a	b	$a \to b$	b	a
$a \wedge b$	1/2	T	T	T	T	$\boxed{\text{T}}$
$a \wedge \neg b$	0	T	F	F	F	$\boxed{\text{T}}$
$\neg a \wedge b$	1/2	F	T	T	T	F
$\neg a \wedge \neg b$	0	F	F	T	F	F

So,

$$a \to b, b \mid\!\!\!\approx a^{1/2}.$$

Note that this tells us what probability or probabilities attach to the conclusion, given the premisses; it can be written $P_{\{a \to b, b\}}(a) = 1/2$. It does not tell us the extent to which the premisses *increase* the probability of the conclusion. This latter concept, degree of *support* rather than degree of *partial entailment*, can be explicated as $P_{\{a \to b, b\}}(a) - P_{\emptyset}(a)$. Now the probability table for $P_{\emptyset}(a)$ is:

$\pm a \wedge \pm b$	$P(\pm a \wedge \pm b)$	a	b	$a \to b$	b	a
$a \wedge b$	1/4	T	T	T	T	$\boxed{\text{T}}$
$a \wedge \neg b$	1/4	T	F	F	F	$\boxed{\text{T}}$
$\neg a \wedge b$	1/4	F	T	T	T	F
$\neg a \wedge \neg b$	1/4	F	F	T	F	F

Clearly then $P_{\emptyset}(a) = 1/4 + 1/4 = 1/2$ and the degree of support is $1/2 - 1/2 = 0$. Thus the premisses are irrelevant to the conclusion in this example.[4]

Consider another invalid argument: either Miliband will be the next Prime Minister or Cameron will[5]; therefore, Miliband will be the next Prime Minister. To what extent does the premiss entail the conclusion? $m \vee c \mid\!\!\!\approx m^{?}$. Here $P(m \vee c) = 1$ so $P(\neg m \wedge \neg c) = 0$ and we have the following probability table:

$\pm m \wedge \pm c$	$P(\pm m \wedge \pm c)$	m	c	$m \vee c$	m
$m \wedge c$	1/3	T	T	T	$\boxed{\text{T}}$
$m \wedge \neg c$	1/3	T	F	T	$\boxed{\text{T}}$
$\neg m \wedge c$	1/3	F	T	T	F
$\neg m \wedge \neg c$	0	F	F	F	F

Hence $P_{\{m \vee c\}}(m) = 1/3 + 1/3 = 2/3$. Note that $P_{\emptyset}(m) = 1/2$ so the premiss does support the conclusion, to degree $2/3 - 1/2 = 1/6$.

[4] Note that, as $a \to b$ and b are simple with respect to P^*, $P_{\{a \to b, b\}}(a) = P_{\emptyset}(a | (a \to b) \wedge b)$. See Williamson [18] on this point.

[5] We shall treat this 'or' as inclusive disjunction, under the assumption that it is possible to have two Prime Ministers in a coalition government.

2.2. Qualified premisses in propositional inductive logic

We shall move to some examples of the situation in which the premisses are not categorical but have some inductive quality attached to them. Consider first the question,

$$a \wedge \neg b^{0.1} \mathrel{|\!\!\!\approx} b^{?}.$$

Here $a \wedge \neg b^{0.1}$ abbreviates $a \wedge \neg b^{\{0.1\}}$. Suppose, for instance that a is the proposition that *it rains today* and b is the proposition that *it will rain tomorrow*. Then $a \wedge \neg b^{0.1}$ is interpreted as saying that the chance of it raining today but not tomorrow is 0.1. The question is, how strongly should one believe that it will rain tomorrow, given this information?

The probability table is as follows:

$\pm a \wedge \pm b$	$P(\pm a \wedge \pm b)$	a	b	a	b
$a \wedge b$	0.3	T	T	T	T
$a \wedge \neg b$	0.1	T	F	T	F
$\neg a \wedge b$	0.3	F	T	F	T
$\neg a \wedge \neg b$	0.3	F	F	F	F

The probability $P(a \wedge \neg b) = 0.1$ follows by the Calibration Norm, while the others are due to the Equivocation Norm. Then $P_{\{a \wedge \neg b^{0.1}\}}(b) = 0.3 + 0.1 = 0.4$.

But one might also ask whether, given the chance information, getting rain today supports a forecast of rain tomorrow. One has to be a bit careful here: one cannot simply add a to the premisses because if a were the case then the chance of $a \wedge \neg b$ would be different. (This is more obvious in the case in which the chance information is $P^*(a) = 0.1$: then adding a as a premiss induces the constraint $P^*(a) = 1$ which clearly supersedes the original information.) But one can ask:

$$a \wedge \neg b^{0.1} \mathrel{|\!\!\!\approx} b|a^{?}.$$

Here a conclusion of the form $b|a^{Y}$ is interpreted as an assertion about conditional probability, namely $P(b|a) \in Y$. Now $P_{\{a \wedge \neg b^{0.1}\}}(b|a) = 0.3/0.4 = 0.75 > 0.4$, so getting rain today does indeed support the forecast of rain tomorrow.

Note, in contrast, that adding a to the premisses would yield $P_{\{a \wedge \neg b, {}^{0.1}a\}}(b) = 0.9$, as can be deduced from the probability table:

$\pm a \wedge \pm b$	$P(\pm a \wedge \pm b)$	a	b	a	b
$a \wedge b$	0.9	T	T	T	T
$a \wedge \neg b$	0.1	T	F	T	F
$\neg a \wedge b$	0	F	T	F	T
$\neg a \wedge \neg b$	0	F	F	F	F

So the use of conditional probability can yield different results to those derived by augmenting the premisses [17].

Consider another example. A herder randomly sampled a hundred goats in his herd and found that 80 of them were tetchy. He noticed that 5 of the non-tetchy goats were Angora and that 2 of those had horns. To what extent should he believe that the next horned Angora goat will be tetchy?

If we grant the facts of the story and that the given proportions of the sampled goats accurately reflect the proportions in the herd as a whole, then we are faced with the following question:

$$t,^{4/5} a|\neg t,^{1/4} h|a \wedge \neg t^{2/5} \mathrel{|\!\!\!\approx} t|h \wedge a^{?}.$$

Here we introduce conditional premisses: a premiss of the form $\theta|\varphi^{X}$ is interpreted as saying that $P^*(\theta|\varphi) \in X$, while, as before, a conclusion of the form $\theta|\varphi^{Y}$ is interpreted as saying that $P(\theta|\varphi) \in Y$.

We shall fill in the following probability table:

$\pm h \wedge \pm a \wedge \pm t$	$P(\pm h \wedge \pm a \wedge \pm t)$	h	a	t	$a \wedge \neg t$
$h \wedge a \wedge t$	x_1	T	T	T	F
$h \wedge a \wedge \neg t$	x_2	T	T	F	T
$h \wedge \neg a \wedge t$	x_3	T	F	T	F
$h \wedge \neg a \wedge \neg t$	x_4	T	F	F	F
$\neg h \wedge a \wedge t$	x_5	F	T	T	F
$\neg h \wedge a \wedge \neg t$	x_6	F	T	F	T
$\neg h \wedge \neg a \wedge t$	x_7	F	F	T	F
$\neg h \wedge \neg a \wedge \neg t$	x_8	F	F	F	F

To work out the probabilities of the atomic states, first we articulate the premisses as constraints on atomic states.

$$P(t) = x_1 + x_3 + x_5 + x_7 = 4/5,$$
$$P(a \mid\!\sim t) = \frac{P(a \wedge \neg t)}{P(\neg t)} = \frac{x_2 + x_6}{x_2 + x_4 + x_6 + x_8} = \frac{1}{4},$$
$$P(h|a \wedge \neg t) = \frac{P(h \wedge a \wedge \neg t)}{P(a \wedge \neg t)} = \frac{x_2}{x_2 + x_6} = \frac{2}{5}.$$

The first constraint can be satisfied by setting

$$x_1 = x_3 = x_5 = x_7 = 1/5,$$

since this constraint treats the atomic states that occur in it symmetrically, and those atomic states are not constrained by the other constraints (cf., Theorem 5).

The second constraint requires that

$$4x_2 + 4x_6 = x_2 + x_4 + x_6 + x_8.$$

I.e., $3x_2 + 3x_6 = x_4 + x_8$.

The third constraint requires that

$$5x_2 = 2x_2 + 2x_6.$$

I.e., $3x_2 = 2x_6$.

Substituting the third constraint into the second, $5x_6 = x_4 + x_8$. We can set $x_4 = x_8 = x$ say. Then $x_6 = (2/5)x$ and $x_2 = (2/3 \times 2/5)x = (4/15)x$. We have probability 1/5 to share among $h \wedge a \wedge \neg t, h \wedge \neg a \wedge \neg t, \neg h \wedge a \wedge \neg t, \neg h \wedge \neg a \wedge \neg t$. So $(4/15)x + (2/5)x + x + x = 1/5$, i.e., $(40/15)x = 1/5$ and $x = 3/40$. So we have:

$\pm h \wedge \pm a \wedge \pm t$	$P(\pm h \wedge \pm a \wedge \pm t)$	h	a	t
$h \wedge a \wedge t$	1/5	T	T	T
$h \wedge a \wedge \neg t$	1/50	T	T	F
$h \wedge \neg a \wedge t$	1/5	T	F	T
$h \wedge \neg a \wedge \neg t$	3/40	T	F	F
$\neg h \wedge a \wedge t$	1/5	F	T	T
$\neg h \wedge a \wedge \neg t$	3/100	F	T	F
$\neg h \wedge \neg a \wedge t$	1/5	F	F	T
$\neg h \wedge \neg a \wedge \neg t$	3/40	F	F	F

Now we can work out the degree to which the premisses partially entail the conclusion: $P(t|h \wedge a) = \frac{1/5}{1/5+1/50} = \frac{10}{11}$. Hence,

$$t,^{4/5} a|\neg t,^{1/4} h|a \wedge \neg t^{2/5} \mathrel{|\!\!\!\approx} t|h \wedge a^{10/11}.$$

Note that, in all these examples, the constraints imposed by the premisses are linear constraints on the atomic states. This implies that the set of probability functions satisfying these constraints is convex, and so there is no further need to take convex hulls in order to satisfy the Calibration Norm. In general, satisfying the Calibration Norm is a linear programming problem, and then satisfying the Equivocation Norm is an optimisation problem (minimisation of Kullback–Leibler divergence from the equivocator $P_=$, or, equivalently, maximisation of entropy). As suggested in Haenni et al. [7], there is scope to use probabilistic networks (in particular, credal networks) as the inference machinery for probabilistic logic: a credal network can be an efficient way of representing a convex set of probability functions—in our case the set of rational belief functions that are appropriate given the premisses—and there are efficient methods for calculating the set Y of probabilities that these functions give to a conclusion proposition. Note that in the above examples, calibration with the premisses yields a set of probability functions that is closed as well as convex. Consequently there is only one function in this set that is closest to the equivocator and a single probability that attaches to the conclusion proposition (see, e.g., [12]).

2.3. Quantified premisses in predicate inductive logic

Consider a predicate language. Let us ask, $\forall x\theta(x)^c \mathrel{|\!\!\!\approx} \theta(t_1)^?$ for some quantifier-free proposition θ and some number c in the unit interval. For instance, if the chance of all ravens being black is 0.7, then how strongly should you believe that raven t_1 is black?

Let us use the notation θ_n for $\theta(t_1) \wedge \cdots \wedge \theta(t_n)$. Then by the Probability Norm, the premiss implies the constraints $P(\theta_n) \geqslant c$ for all n, and $P(\theta_n) \to c$ as $n \to \infty$. So let us write $P(\theta_n) = c + x_n$, where $x_n \geqslant 0$ and $x_n \to 0$ as $n \to \infty$ (so $P(\neg\theta_n) = 1 - c - x_n$). Let $[\varphi]_n \stackrel{\text{df}}{=} \{\omega_n \in \Omega_n\colon\ \omega_n \models \varphi\}$ and $|\varphi|_n \stackrel{\text{df}}{=} |[\varphi]_n| = |\{\omega_n \in \Omega_n\colon\ \omega_n \models \varphi\}|$. Now,

$$P(\theta_n) = \sum_{\omega_n \in [\theta_n]_n} P(\omega_n) = c + x_n$$

and

$$P(\neg\theta_n) = \sum_{\omega_n \in [\neg\theta_n]_n} P(\omega_n) = 1 - c - x_n.$$

Since the premiss is symmetric with respect to the $\omega_n \in [\theta_n]_n$, the Equivocation Norm will give them the same probability:

$$P(\omega_n) = \frac{c + x_n}{|\theta_n|_n}.$$

Similarly, for $\omega_n \in [\neg\theta_n]_n$,

$$P(\omega_n) = \frac{1 - c - x_n}{|\neg\theta_n|_n}.$$

One might think that the most equivocal probability function satisfying these constraints will set $x_n = 0$ for all n sufficiently large (i.e., for all n such that $c/|\theta_n|_n \geqslant P_=(\omega_n) = 1/2^{r_n}$), since this function will be closest to the equivocator. However, if $c < 1$ then setting $x_n = 0$ would yield a function that is not a probability function, in violation of the Probability Norm. Consider, for example, a language with a single unary predicate U, and let $\theta(x)$ be Ux: then $|\theta_n|_n = 1$ so if $x_n = 0$ for sufficiently large n we would have

$$\begin{aligned} c &= P(\theta_n) \\ &= P(Ut_1 \wedge \cdots \wedge Ut_n) \\ &\neq P(Ut_1 \wedge \cdots \wedge Ut_n \wedge Ut_{n+1}) + P(Ut_1 \wedge \cdots \wedge Ut_n \wedge \neg Ut_{n+1}) \\ &= c + \frac{1-c}{2^{n+1}-1}, \end{aligned}$$

for such n, in violation of axiom PP3 of Section 1. Similar reasoning can be used to show that if $c < 1$ then $x_n > 0$ for all n. Note that if θ is tautologous then for the premiss to be consistent, c must be 1 and so $x_n = 0$ for all n.

In order to calculate x_k for some fixed k, consider $n \geqslant k$ and:

$$\begin{aligned} P(\theta_k) &= \sum_{\omega_n \models \theta_k \wedge \theta_n} P(\omega_n) + \sum_{\omega_n \models \theta_k \wedge \neg\theta_n} P(\omega_n) \\ &= \sum_{\omega_n \in [\theta_n]_n} P(\omega_n) + \sum_{\omega_n \models \theta_k \wedge \neg\theta_n} P(\omega_n) \\ &= c + x_n + \big(|\theta_k|_n - |\theta_n|_n\big)\frac{1-c-x_n}{2^{r_n} - |\theta_n|_n} \\ &= c + x_n + |\theta_k|_n \frac{1-c-x_n}{2^{r_n} - |\theta_n|_n} - \frac{1-c-x_n}{2^{r_n}/|\theta_n|_n - 1} \\ &\to c + (1-c)\lim_{n\to\infty} \frac{|\theta_k|_n}{2^{r_n} - |\theta_n|_n} \quad \text{as } n \to \infty, \end{aligned}$$

as long as θ is such that $2^{r_n}/|\theta_n|_n \to \infty$. But $x_k = P(\theta_k) - c$ is a constant for fixed k—it does not vary with n. So $x_k = (1-c)\lim_{n\to\infty} \frac{|\theta_k|_n}{2^{r_n} - |\theta_n|_n}$.

For example, if $\theta(x)$ is Ux for unary predicate U then $2^{r_n}/|\theta_n|_n \to \infty$ and $\frac{|\theta_k|_n}{2^{r_n} - |\theta_n|_n} \to 2^{-k}$ so

$$\forall x U x^c \mathrel{|\!\!\approx} Ut_1^{c+(1-c)/2}.$$

To consider another toy example, suppose that the chance of all men being mortal is $\frac{2}{3}$; to what extent should one believe that Socrates is mortal given that he is human? I.e.,

$$\forall x(Hx \to Mx)^{2/3} \mathrel{|\!\!\approx} Ms|Hs^{?}$$

where H stands for Human, M for Mortal and s for Socrates.

Here $\theta(x)$ is $Hx \to Mx$ and $|\theta_k|_n = 3^k 4^{n-k}$ so $x_k = 1/3 \times (3/4)^k$. We have then that $P(Hs \to Ms) = 11/12$, so $P(Hs \wedge Ms) = P(\neg Hs \wedge Ms) = P(\neg Hs \wedge \neg Ms) = 11/36$, and $P(\neg(Hs \to Ms)) = P(Hs \wedge \neg Ms) = 1/12$. Therefore, $P(Ms|Hs) = P(Hs \wedge Ms)/(P(Hs \wedge Ms) + P(Hs \wedge \neg Ms)) = 11/14$:

$$\forall x(Hx \to Mx)^{2/3} \mathrel{|\!\!\approx} Ms|Hs^{11/14}.$$

Let us consider one related example. Suppose that 80% of sampled individuals are men (i.e., human). To what extent does the proposition that all men are mortal entail the conclusion that Socrates, a randomly sampled individual, is mortal? Assuming that the frequency information warrants the estimate $P^*(Ht_i) = 4/5$ for each sampled individual—in particular, $P^*(Hs) = 4/5$—one might phrase this as the question:

$$Hs^{4/5} \mathrel{|\!\!\approx} Ms|\forall x(Hx \to Mx)^?.$$

But this question cannot be answered directly since $P_{\{Hs^{4/5}\}}(\forall x(Hx \to Mx)) = 0$ and the conditional probability of the conclusion is undefined. In this case, though, $\forall x(Hx \to Mx)$ being the case would not in itself change the chance $P^*(Hs) = 4/5$, so one can use the alternative formulation:

$$Hs^{4/5}, \forall x(Hx \to Mx) \mathrel{|\!\!\approx} Ms^?.$$

The first premiss implies that $P(Hs \wedge Ms) + P(Hs \wedge \neg Ms) = 4/5$ while the second premiss implies that $P(Hs \wedge Ms) + P(\neg Hs \wedge Ms) + P(\neg Hs \wedge \neg Ms) = 1$. $P(Hs \wedge \neg Ms) = 0$ so $P(Hs \wedge Ms) = 4/5$ and, equivocating, $P(\neg Hs \wedge Ms) = P(\neg Hs \wedge \neg Ms) = 1/10$. Hence $P_{\{Hs^{4/5}, \forall x(Hx \to Mx)\}}(Ms) = 4/5 + 1/10 = 9/10$, in answer to our question.

2.4. A comparison with the Barnett–Paris method

We close this section by mentioning some examples provided in Rad [13], a work that offers a nice comparison of the method presented here with the leading alternative method for predicate languages, the Barnett–Paris method of Barnett and Paris [2]. Given a predicate language $\mathcal{L}$, this alternative method (i) interprets the premisses of an inductive argument as constraints on rational belief over a *finite* predicate language $\mathcal{L}_n$, then (ii) calculates the rational belief function P^n on this finite language by choosing from the functions that satisfy the constraints, the P^n that maximises entropy (equivalently, that is closest to the equivocator in terms of n-divergence), and finally (iii) deems P^∞, defined by $P^\infty(\omega_m) = \lim_{n\to\infty} P^n(\omega_m)$, to be the rational belief function over $\mathcal{L}$ itself. Note that the Barnett–Paris method presumes premiss constraints that delimit a closed, convex set of functions, in order that P^n always be defined; hence their method is not applicable to the biased coin example mentioned in Section 1, where the set of constrained functions is not closed and where there is no P^n with maximum entropy. But if the constraint set is closed and convex, and if the language in question contains only *unary* predicates, the two approaches will yield the same answers to questions in inductive logic [13, Theorem 29].

Rad [13, §4.1] considers the following non-unary question, phrased in a predicate language augmented with an equality symbol that satisfies the axioms of equality[6]:

$$\forall x \neg Rxx, \forall xy\big(\neg(x = y) \to (Rxy \vee Ryx)\big),$$
$$\forall xyz\big((Rxy \wedge Ryz) \to Rxz\big), \forall x\exists y Rxy \mathrel{|\!\!\approx} \bigwedge_{i,j=1}^{r} \pm Rt_it_j{}^?.$$

Here the premisses specify that the binary relation R is a strict linear order that extends indefinitely in at least one direction. As Rad shows, the approach advocated in this paper will give an answer of $1/r!$ or 0 according to whether $\bigwedge_{i,j=1}^{r} \pm Rt_it_j$ is consistent with some such linear order or not (since it equivocates between all strict linear orders). On the other hand, the Barnett–Paris method will not be able to answer this question at all, since any attempt to treat the premisses as constraints on a *finite* sublanguage $\mathcal{L}_n$ leads to the difficulty that the premisses have no finite model, i.e., the premisses are inconsistent if the domain is finite.[7] (This *finite model problem* also besets the *random worlds* method of Grove et al. [6], Bacchus et al. [1], which provides semantics for inductive logic by considering the limit as $n \to \infty$ of the proportion of all $\mathcal{L}_n$-models of the premisses that satisfy the conclusion.)

The finite model problem is just one obstacle for the Barnett–Paris method—there are others. Rad [13, §3.2] shows that even where the premisses are restricted to categorical Π_2 sentences—premisses of the form $\forall x\exists y\theta(x, y)$ where θ is quantifier-free—, the Barnett–Paris limit may not exist. On the other hand, the Barnett–Paris method can be used to answer all questions with unqualified premisses from the class of quantifier-free sentences, or the class of Σ_1 sentences (propositions of the form $\exists x\theta(x)$), or the class of Π_1 sentences ($\forall x\theta(x)$), and it is known that at least in the first two cases, the Barnett–Paris method will agree with the method presented here.

In more complex cases, there is no guarantee that the method of this paper will lead to a unique maximally equivocal function. For example, Rad [13, §4.3] shows that for the question:

$$\exists x\forall y Rxy \mathrel{|\!\!\approx} \psi^?,$$

---

6 As Rad notes, this kind of example can also be phrased in a language without equality by treating some relation in that language as a surrogate for equality.

7 Recall that it is presumed that each element of the domain is picked out by a constant of the language; consequently if the language is finite, so is the domain.

for any function satisfying the premiss there is always a more equivocal function satisfying that premiss. In that sense, such a question is like the biased coin question of Section 1 and our policy is to consider any function satisfying the premiss as sufficiently equivocal (in the absence of contextual information that can offer more guidance). Hence while the approach presented here can answer such questions, it will not always attach a precise probability value to a conclusion of interest. For instance,

$$\exists x \forall y Rxy \mathrel{|\!\!\approx} Rt_1t_2^{[0,1]}$$

since the premiss can be satisfied whether or not t_1 and t_2 stand in relation R. Hence the conclusions in such cases can be rather weak—they agree with the conclusions that would be generated by an empirically-based subjective Bayesian semantics (Section 1), which does not appeal to the Equivocation Norm at all. But, as any subjective Bayesian will agree, weak does not imply wrong: in any particular situation there simply may be no considerations (neither evidential nor pragmatic) that further narrow down how strongly a given conclusion should be believed.

3. Critiques of inductive logic

In this section we will examine some of the chief criticisms of inductive logic and investigate the extent to which the inductive logic of this paper—i.e., objective Bayesian inductive logic—survives these criticisms.

It was suggested in the 1940s and 50s that a viable inductive logic could further our understanding of the way in which scientific hypotheses are confirmed by evidence and that it could also provide a normative foundation for reasoning under uncertainty, with application to artificial intelligence. Rudolf Carnap was the main architect of the programme to develop inductive logic [3], and, by the 1980s, a substantial body of research had been built up in this area (see, e.g., Carnap and Jeffrey [4], Jeffrey [8]).

However, the view amongst philosophers at that time was that the programme had failed. This is for several reasons. Notably, Carnap did not isolate a single inductive logic, but a family of inductive logics, and an element of choice was required to pick out an appropriate logic, making inductive reasoning relative to subjective and pragmatic concerns. This was problematic from Carnap's perspective since he interpreted probability in inductive logic as a *logical relation between propositions*, along the lines of Keynes [9]—an interpretation that has no obvious role for an agent or subjective choice. But as we have discussed at the end of the last section, an element of subjective choice is not in itself problematic from the perspective of the Bayesian semantics, since Bayesian probability *is* defined relative to an agent, and Bayesian probability need not be fully constrained by the agent's evidence and language.

While the question of subjectivity is less problematic in our context, four other criticisms of the inductive logic programme deserve more discussion. First, it was thought that the inferences of an inductive logic are problematically dependent on the underlying logical language. Second, inductive logics are normally based on invariance principles, which can be thought of as applications of the principle of indifference—a principle that is known to lead to paradoxes. Third, it was recognised that Carnap's inductive logic could not simultaneously capture inductive relations between logically overlapping propositions (e.g., between *the bird is black or the bird is white*, and, *the bird is black*) and inductive relations between logically disjoint propositions (e.g., between *all ravens so far observed are black*, and, *the next observed raven is black*) [14]. Fourth, attempts to develop inductive logics in which universal hypotheses (e.g., *all ravens are black*) receive inductive support were viewed as complex and only partially satisfactory. We shall discuss these objections in turn. While we will focus on predicate languages, most of we what will have to say will apply equally to propositional languages.

3.1. Language dependence

Since inductive logics as presented here are defined with regard to propositions of a logical language $\mathcal{L}$, the question arises as to what extent inferences in an inductive logic depend on the underlying language. For the semantics presented here, there is a clear sense in which inferences *do not* depend on the underlying language: if one can formulate an inference in more than one language then the two formulations will agree as to whether the premisses entail the conclusion.

Theorem 3. *Given predicate languages $\mathcal{L}^1$ and $\mathcal{L}^2$, suppose that $\varphi_1, \ldots, \varphi_n, \psi$ are propositions of both $\mathcal{L}^1$ and $\mathcal{L}^2$. Let $\mathrel{|\!\!\approx}^1$ be the Bayesian entailment relation with respect to $\mathcal{L}^1$ and $\mathrel{|\!\!\approx}^2$ be the Bayesian entailment relation with respect to $\mathcal{L}^2$. Then $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx}^1 \psi^Y$ if and only if $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx}^2 \psi^Y$.*

This result, which is proved in Appendix A, is important as it shows that there is normally no need to spell out the underlying language for any particular inference, and it justifies the use of the notation $\mathrel{|\!\!\approx}$ rather than $\mathrel{|\!\!\approx}^1$, $\mathrel{|\!\!\approx}^2$, etc.

However, there is another sense in which inferences *do* depend on the underlying language. Suppose $\mathcal{L}^1$ has two unary predicates, *Green* and *Blue*, but that $\mathcal{L}^2$ just has one unary predicate, *Grue*, which is synonymous with *Green or Blue*. Then we have that

$$\mathrel{|\!\!\approx}^1 Green(t_1) \vee Blue(t_1)^{3/4}$$

but

$$\mathrel{|\!\!\approx}^2 Grue(t_1)^{1/2}.$$

Yet this appears to confer two different levels of plausibility on what is in effect the same conclusion proposition. Which is the correct inference?

The Bayesian answer to this question is that both are correct. Granting nothing at all, if your language is $\mathcal{L}_1$ then you ought to believe that t_1 is Green or Blue to degree 3/4, but if your language is $\mathcal{L}_2$ you ought to believe that t_1 is Grue to degree 1/2. The Bayesian can justify this stance by pointing out that, just as degrees of belief should depend on explicit evidence (because such evidence provides us with information about the world), so too should degrees of belief depend on language (because language also provides information about the world). Evidence tells us facts about the world, and one evidence base can be preferred to another if it tells us more, or if what it tells us is more accurate; language tells us about how the world can be carved up, and one language can be preferred to another if it carves up the world more effectively.[8]

But the Bayesian would also say that both inferences are lacking: there is an important piece of information, namely that Grue is synonymous with Green or Blue, that has not been taken into account in the two inferences given above. Taking this information into account requires adopting a language $\mathcal{L}^3$ in which one can express the proposition $\forall x, \mathit{Grue}(x) \leftrightarrow (\mathit{Green}(x) \vee \mathit{Blue}(x))$. Then one can formulate the inference

$$\forall x, \mathit{Grue}(x) \leftrightarrow \big(\mathit{Green}(x) \vee \mathit{Blue}(x)\big) \mathrel{|\!\!\!\approx^{3}} \mathit{Grue}(t_1)^{3/4}.$$

But we also find that

$$\forall x, \mathit{Grue}(x) \leftrightarrow \big(\mathit{Green}(x) \vee \mathit{Blue}(x)\big) \mathrel{|\!\!\!\approx^{3}} \mathit{Green}(t_1) \vee \mathit{Blue}(t_1)^{3/4},$$

so there is no inconsistency. Note that this inference validates that of $\mathcal{L}^1$, the richer of the two original languages.

Any dependence on the underlying language has typically been regarded as highly problematic in the literature on inductive logic and objective Bayesian methods (see, e.g., Seidenfeld [15]). But is such dependence as there is under the Bayesian semantics really problematic? Perhaps not, given the following triviality result. Define a *synonymy map* between predicate languages $\mathcal{L}$ and $\mathcal{L}'$ to be a consistent, countable set of propositions of the form $\theta_i \leftrightarrow \theta'_i$ where the θ_i are propositions of $\mathcal{L}$ and the θ'_i are propositions of $\mathcal{L}'$.[9] Then demanding invariance under all possible synonymy maps is simply too tall an order, because the only entailment relations which satisfy that demand are trivial relations:

Theorem 4. *If an entailment relation $\mathrel{|\!\!\!\approx}$ of a probabilistic logic with underlying predicate language $\mathcal{L}$ is invariant under all synonymy maps between $\mathcal{L}$ and $\mathcal{L}'$, for all $\mathcal{L}'$, then,*

1. $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\!\approx} \psi^Y$ *if and only if* $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\!\approx} \neg\psi^Y$,
2. $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\!\approx} \psi^Y$ *implies that* $\{0, 1\} \subseteq Y$.

The proof is in Appendix A.

In sum, Theorem 3 shows that inferences in the Bayesian inductive logic of this paper are independent of the underlying language. However, they are not invariant under arbitrary synonymy maps. Theorem 4 shows that one cannot demand this stronger invariance condition without trivialising inductive logic.

3.2. The principle of indifference

Inductive logics have also been criticised for succumbing to paradoxes associated with the Principle of Indifference. Since the Equivocation Norm appears to be a kind of indifference principle, it appears that the Bayesian inductive logic of this paper might be prone to these paradoxes too. Here we shall make the connection with the Principle of Indifference precise, and assess the extent to which this connection is problematic.

First, the Principle of Indifference itself:

> The principle of indifference asserts that if there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an *equal* probability. These *equal* probabilities must be assigned to each of several arguments, if there is an absence of positive ground for assigning *unequal* ones. [9, p. 45]

Keynes restricted the Principle to *indivisible* alternatives:

> In short, the principle of indifference is not applicable to a pair of alternatives, if we know that either of them is capable of being further split up into a pair of possible but incompatible alternatives of the same form as the original pair. [9, p. 66]

[8] Note that Bayesian epistemology provides norms that connect evidence and language on the one hand, and rational degrees of belief on the other. It says little or nothing about norms concerning evidence (e.g., one ought to gather evidence) and norms concerning language (e.g., one ought to be able to express what one wishes to reason about).

[9] Note that a synonymy map of the form $\forall x, \mathit{Grue}(x) \leftrightarrow (\mathit{Green}(x) \vee \mathit{Blue}(x))$ can be written as $\{\mathit{Grue}(t_i) \leftrightarrow (\mathit{Green}(t_i) \vee \mathit{Blue}(t_i))\colon\ i = 1, 2, \ldots\}$.

On a finite propositional or predicate language $\mathcal{L}_n$, the set Ω_n of atomic states represents the partition of indivisible alternatives. The situation is more complicated on an infinite predicate language $\mathcal{L}$ because there is no partition of indivisible alternatives: for every set Ω_n of atomic n-states, there is another set Ω_i for $i > n$ that splits up the original alternatives. Nevertheless, one can formulate the Principle of Indifference in a way that applies equally to the infinite and the finite case. In the setting of inductive logic, we can state the Principle of Indifference as follows:

POI. If atomic n-states ω_n^* and $\omega_n^\dagger$ are treated symmetrically by the premisses, then $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!|\; \omega_n^{*Y}$ if and only if $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \approx\!\!\!|\; \omega_n^{\dagger Y}$.

Here we can say that ω_n^* and $\omega_n^\dagger$ are *treated symmetrically by the premisses* just in case, for any probability function P satisfying the premisses, there is another function satisfying the premisses which swaps the probabilities of ω_n^* and $\omega_n^\dagger$ but which otherwise agrees with P as far as possible: if $P \in \langle \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\rangle$ then there is some $Q \in \langle \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\rangle$ such that

$$Q(\omega_i) = \begin{cases} P(f(\omega_i)): & \omega_i \models \omega_n^*, \\ P(f^{-1}(\omega_i)): & \omega_i \models \omega_n^\dagger, \\ P(\omega_i): & \text{otherwise}, \end{cases}$$

for some bijection $f : [\omega_n^*] \to [\omega_n^\dagger]$ from states that deductively entail ω_n^* (here $[\theta] \stackrel{\text{df}}{=} \bigcup_{i=1}^\infty [\theta]_i = \{\omega_i \in \Omega_i\colon\ \omega_i \models \theta,\ i = 1, 2, \ldots\}$) to states that deductively entail $\omega_n^\dagger$, such that $f(\omega_i) \in \Omega_i$ for each $\omega_i \in \Omega_i \cap [\omega_n^*]$ and $\omega_i \models \omega_j$ iff $f(\omega_i) \models f(\omega_j)$ for $\omega_i, \omega_j \in [\omega_n^*]$.

Interestingly, one can show (see Appendix A) that the inductive logic presented here does satisfy the Principle of Indifference:

Theorem 5. *The objective Bayesian entailment relation* $\approx\!\!\!|^{\circ}$, *defined in Section* 1, *satisfies* POI.

To what extent is satisfying the Principle of Indifference problematic, in the context of our Bayesian inductive logic?

Paradoxes of the Principle of Indifference arise when the Principle is applied to different partitions in order to yield different probabilities for the same proposition. Keynes' restriction to indivisible alternatives was intended to avoid this phenomenon. In our framework, such paradoxes cannot arise: Theorem 5 shows that POI is a *consequence* of the Bayesian semantics, not a means of setting probabilities, so there is no question of choosing different partitions to obtain different probabilities for the same proposition. Probabilities over sentences of $\mathcal{L}$ are set by the norms of Bayesian epistemology, and these norms set them in a consistent way. Moreover, Theorem 3 shows that the probabilities required by inductive logic are independent of the precise language in which one can formulate the premiss and conclusion propositions.

It might be objected that paradoxes of the Principle of Indifference can still arise when one changes the conceptualisation of a particular problem to an equivalent but different conceptualisation in which the partition of indivisible alternatives is different. In the context of inductive logic, this means changing the language and at the same time asserting an equivalence between certain propositions of the new language and of the old—i.e., by introducing a synonymy map. The *Grue* example considered above is an instance of this sort of paradox: when one is indifferent to the states of the *Grue* language one gets different probabilities to when one is indifferent to the states of the *Green–Blue* language, if the synonymy map is taken into account.

As explained above, there is a response one can give to this sort of 'paradox'. First, the Bayesian can deny it is genuinely paradoxical by maintaining that degrees of belief really ought to display this sort of behaviour—they should depend on language as well as explicit evidence. Second, one can note that any demand that inductive logic should be immune to this sort of behaviour is untenable since there is no non-trivial inductive logic that satisfies such a demand (Theorem 4).

In sum, the Bayesian inductive logic presented here does satisfy the Principle of Indifference, but that is arguably no reason to dismiss the inductive logic.

3.3. Learning from experience

Any inductive logic that satisfies POI has the property that $\approx\!\!\!|\; \psi^Y$ if and only if $P_=(\psi) \in Y$, i.e., if there are no premisses, the logic is determined by the equivocator function. This property has been roundly criticised (see, e.g., Carnap [3, p. 81]; Salmon [14]), on the grounds that it apparently fails to capture the phenomenon of learning from experience. In particular, the equivocator yields probability $\frac{1}{2}$ that the 101st observed raven will be found to be black, $\approx\!\!\!|^{\circ}\, Br_{101}{}^{.5}$, which seems reasonable, but also yields probability $\frac{1}{2}$ that the 101st observed raven will be found to be black, conditional on the first 100 ravens having been found to be black, $\approx\!\!\!|^{\circ}\, Br_{101}|Br_1 \wedge \cdots \wedge Br_{100}{}^{.5}$, which seems unreasonable because it represents an apparent inability to increase a probability on the basis of good evidence that ravens tend to be black.

In response to this objection, it suffices to point out that in the context of the inductive logic considered here, this apparent inability to capture learning from experience is based on a simple misinterpretation. The problem is a confusion between conditions and evidence: the condition $Br_1 \wedge \cdots \wedge Br_{100}$ appearing in the conclusion should not be interpreted as evidence or experience—it is the premisses that are intended to reflect evidence. Thus the situation has been misrepresented. According to the Bayesian semantics presented here, if 100 ravens are observed and found to be black, this needs to be translated into constraints on physical probability, of the form $P^*(Br_i) \in X$. Clearly $Br_1 \wedge \cdots \wedge Br_{100}$ provides the constraints $P^*(Br_1) = 1, \ldots, P^*(Br_{100}) = 1$. But statistical theory can be used to derive constraints on $P^*(Br_i)$ for $i > 100$: if the physical probability of ravens being found to be black is granted to be independent and identically distributed, confidence-interval estimation methods motivate claims of the form $P^*(P^*(Br_i) \geqslant 1 - \delta) = 1 - \epsilon$ for $i > 100$. Given some threshold $1 - \epsilon_0$ of acceptance, an agent can then choose δ_0 such that $P^*(P^*(Br_i) \geqslant 1 - \delta_0) = 1 - \epsilon_0$, i.e., such that the claim that $P^*(Br_i) \geqslant 1 - \delta_0$ reaches the threshold of acceptance. Then the question under consideration is better represented as

$$Br_1{}^1, \ldots, Br_{100}{}^1, Br_{101}{}^{[1-\delta_0,1]}, Br_{102}{}^{[1-\delta_0,1]}, \ldots \mid\!\approx Br_{101}{}^?$$

(here we extend the framework in the obvious way to allow infinitely many premisses).

The inductive logic represented here will choose the maximally equivocal point in the interval $[1 - \delta_0, 1]$ to yield the inference:

$$Br_1{}^1, \ldots, Br_{100}{}^1, Br_{101}{}^{[1-\delta_0,1]}, Br_{102}{}^{[1-\delta_0,1]}, \ldots \mid\!\approx Br_{101}{}^{1-\delta_0},$$

which is clearly a much more reasonable value than $\frac{1}{2}$, inasmuch as it does represent an ability to learn from experience.

In contrast, a conclusion of the form $Br_{101}|Br_1 \wedge \cdots \wedge Br_{100}{}^Y$ is simply a formal abbreviation of the claim that $P(Br_1 \wedge \cdots \wedge Br_{100} \wedge Br_{101})/P(Br_1 \wedge \cdots \wedge Br_{100}) \in Y$ for every $P \in \Downarrow\mathbb{E}$. It is $\mathbb{E}$, which is obtained via the premisses, that captures evidence or experience. In general, under an objective Bayesian interpretation, conditional probabilities are not always interpretable as conditional beliefs [17].

While the problem of capturing the phenomenon of learning from experience is not a genuine problem for the Bayesian semantics, it does highlight a peculiarity of this semantics—its strong connection with statistical theory, via the Calibration Norm. This connection between statistical theory and Bayesianism is explained in detail in Williamson [19]. The resulting logic, then, represents a rather radical departure from the Carnapian programme, which aimed to develop inductive logic independently of statistical theory.

3.4. Universal hypotheses

Another oft-criticised consequence of POI is that many universal hypotheses are given zero probability, $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx \forall x\theta(x)^0$.

As to whether such a universal generalisation is given zero probability depends both on θ and on the premisses. There are θ which do not lead to this property—e.g., $\mid\!\approx \forall x(Ux \vee \neg Ux)^1$ and $\mid\!\approx \forall x((Ux \vee \neg Ux) \wedge Vt_1)^{.5}$. But $\mid\!\approx \forall x\theta(x)^0$ when $|\theta_n|_n/2^{r_n} \to 0$ as $n \to \infty$, where (see Section 2) $|\theta_n|_n = |\{\omega_n \in \Omega_n\colon\ \omega_n \models \theta(t_1) \wedge \cdots \wedge \theta(t_n)\}|$. For instance, $\mid\!\approx \forall xUx^0$ for unary predicate U. Moreover, there are premisses that can give a universal conclusion positive probability—e.g., $\forall xVx, \forall x(Vx \to Ux) \mid\!\approx \forall xUx^1$.

The problem is that this zero-probability phenomenon can be counterintuitive, since, for instance, finding the first 100 observed ravens to be black offers no support to the conclusion that all ravens are black:

$$Br_1{}^1, \ldots, Br_{100}{}^1, Br_{101}{}^{[1-\delta_0,1]}, Br_{102}{}^{[1-\delta_0,1]}, \ldots \mid\!\approx \forall xBx^0.$$

In general, we have a phenomenon which we might call *no generalisations in, no generalisations out*, i.e., if premisses are to raise the probability of a universally quantified proposition away from zero, then those premisses must themselves involve quantifiers:

Theorem 6. *Suppose that, for $\theta(x)$ quantifier-free, $\mid\!\approx \forall x\theta(x)^0$ but $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mid\!\approx \forall x\theta(x)^Y$ where $\inf Y > 0$. Then $\varphi_1, \ldots, \varphi_k$ are not all quantifier-free.*

The proof is in Appendix A.

Can the proponent of the Bayesian semantics for inductive logic bite the bullet here and simply accept that observations (which are arguably representable by quantifier-free sentences) should not raise one's belief in a universal generalisation? Such a move looks rather implausible when one considers scientific theories, which seem to routinely invoke universal generalisations.

But there are two considerations that suggest that this move bears closer scrutiny. First, philosophy of science takes a different stance towards universal generalisations these days, in comparison with the era in which Carnap put forward his inductive logics, which were roundly criticised for awarding universal generalisations probability zero. In Carnap's time, largely under the influence of the logical empiricists, scientific theories were widely understood in terms of collections of universal generalisations (perhaps augmented by statements specifying boundary conditions, bridge laws and so on). Consequently, any inductive logic that gave universal generalisations probability zero was taken to be refuted by scientific practice. But in the 1980s and 1990s the explication of scientific laws in terms of exceptionless generalisations was found to be untenable, as the ubiquity of ceteris paribus laws and pragmatic laws became widely recognised [5,11]. More recently still, the Hempelian deductive-nomological account of explanation, which saw scientific explanations as deductions from universal generalisations, has been replaced by a mechanistic view of explanation, where a phenomenon is held to be explained when the mechanism responsible for that phenomenon has been adequately pointed out (see, e.g., Machamer et al. [10]). Science is increasingly understood largely in terms of a body of mechanisms or nomological machines rather than a body of strict laws or universal generalisations. In sum, the relevance of Theorem 6 to science is less obvious now than it would have appeared a few decades ago.

Second, there are a variety of epistemic attitudes one can take towards universal generalisations. Bayesian epistemology, for instance, draws a sharp distinction between what is *believed* and what is *granted* [16, §1.4.1]. Given what is already granted, Bayesian epistemology provides rational norms for narrowing down appropriate degrees of belief: in terms of inductive logic, it tells us how strongly one should believe a conclusion proposition having granted some premisses. Moreover, the norms that govern what one should take for granted in the first place are divorced from the norms of rational belief. On the one hand, grounds for *granting* propositions may include their coherence, simplicity, strength, accuracy, technical convenience, unifying power and so on, while on the other hand, propositions should only be *believed* to the extent warranted by their Bayesian probability relative to what is already taken for granted. In view of this, while scientists who grant certain universal hypotheses should believe the consequences of those hypotheses, it does not follow that if a universal hypothesis has low Bayesian probability, one should not subsequently take it for granted for the purposes of scientific inquiry. Thus the Bayesian can argue that one should remain sceptical about universal hypotheses that are not supported by what is already granted, yet one can go on to grant those same universal hypotheses for reasons other than strength of rational belief.

So the proponent of the Bayesian semantics can respond to the objection concerning universal generalisations. Universal generalisations appear to play less of a role in science than previously thought, and, in any case, that observations fail to increase the degree to which one should believe a universal generalisation does not preclude a role for that generalisation in a scientific theory understood as a body of propositions that are rationally granted.

4. Summary

This paper has sought to introduce the Bayesian semantics for inductive logic, to explore inferences in this logic and to examine the prospects of the logic in the light of some criticisms that have been levelled against inductive logics in the past. We have seen that there is an important sense in which inferences in this Bayesian inductive logic are independent of language, and that to demand a stronger kind of language independence—invariance under synonymy maps—is to demand too much of any inductive logic. Although this logic satisfies the Principle of Indifference, the indifference partition is well-specified and there do not seem to be genuinely problematic repercussions, once the points about language independence are borne in mind. This logic can capture the phenomenon of learning from experience. Finally, although the logic endorses a certain amount of scepticism concerning universal generalisations, this scepticism is defensible and does not deem scientific practice to be irrational.

Acknowledgements

This research was supported by a British Academy Research Development Award, and by a UK Arts and Humanities Research Council research project grant.

Appendix A. Proofs

Proof of Theorem 3. We suppose that $\varphi_1, \dots, \varphi_n, \psi$ are propositions of predicate languages $\mathcal{L}^1$ and $\mathcal{L}^2$, that $\mathrel{|\!\overset{\circ}{\approx}}^1$ is the entailment relation with respect to $\mathcal{L}^1$ and that $\mathrel{|\!\overset{\circ}{\approx}}^2$ is the entailment relation with respect to $\mathcal{L}^2$. We need to show that $\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \mathrel{|\!\overset{\circ}{\approx}}^1 \psi^Y$ if and only if $\varphi_1^{X_1}, \dots, \varphi_k^{X_k} \mathrel{|\!\overset{\circ}{\approx}}^2 \psi^Y$.

Without loss of generality we may assume that $\mathcal{L}^1$ and $\mathcal{L}^2$ have the same constant symbols $t_1, t_2, \dots$. This is because we already assume that for each language there is a bijection between constant symbols and elements of the domain; hence there is a bijection between the constant symbols of the two languages.

Let $\mathcal{L}$ be the language formed from the predicate symbols that occur in both $\mathcal{L}^1$ and $\mathcal{L}^2$, together with the constant symbols $t_1, t_2, \dots$. Note that $\varphi_1, \dots, \varphi_n, \psi$ are propositions of $\mathcal{L}$ since they are propositions of both $\mathcal{L}^1$ and $\mathcal{L}^2$.

Write $\omega_n^1 \in \Omega_n^1$, an atomic n-state of $\mathcal{L}^1$, as $\omega_n \wedge \omega_n^{1+}$, the conjunction of an atomic n-state of $\mathcal{L}$ and an atomic n-state of $\mathcal{L}^{1+}$, the language that has those predicates in $\mathcal{L}^1$ but not in $\mathcal{L}$. Similarly for language $\mathcal{L}^2$. Let $\mathbb{P}, \mathbb{P}^1, \mathbb{P}^2$ be the set of probability functions on $\mathcal{L}, \mathcal{L}^1, \mathcal{L}^2$ respectively. In general we shall use superscripts to distinguish differences in underlying language.

Definition 7 *(Equivocal projection).* For $i, j \in \{1, 2\}$ and probability function P^i on $\mathcal{L}^i$, define the *equivocal projection* of P^i onto $\mathcal{L}^j$ to be the probability function $P^{i \to j}$ such that

$$P^{i \to j}(\omega_n^j) = P^i(\omega_n) P_{=}^j(\omega_n^{j+} | \omega_n) = \frac{P^i(\omega_n)}{|\Omega_n^{j+}|},$$

for all $\omega_n^j \in \Omega_n^j$ and $n \geqslant 1$.

Lemma 8. *$P^{i\to j}(\theta) = P^i(\theta)$ for each proposition θ of $\mathcal{L}$.*

Proof.

$$\begin{aligned}
P^{i\to j}(\theta) &= \sum_{\omega_n^j \models \theta} P^i(\omega_n) P_=^j\big(\omega_n^{j+}\big|\omega_n\big) \\
&= \sum_{\omega_n \models \theta} P^i(\omega_n) \sum_{\omega_n^{j+}} P_=^j\big(\omega_n^{j+}\big|\omega_n\big) \\
&= \sum_{\omega_n \models \theta} P^i(\omega_n) \\
&= P^i(\theta). \qquad \square
\end{aligned}$$

Lemma 9. $d_n^j(P^{i\to j}, P_=^j) = d_n(P^i, P_=)$.

Proof.

$$\begin{aligned}
d_n^j\big(P^{i\to j}, P_=^j\big) &= \sum_{\omega_n^j \in \Omega_n^j} P^i(\omega_n) P_=^j\big(\omega_n^{j+}\big|\omega_n\big) \log \frac{P^i(\omega_n) P_=^j(\omega_n^{j+}|\omega_n)}{P_=^j(\omega_n) P_=^j(\omega_n^{j+}|\omega_n)} \\
&= \sum_{\omega_n \in \Omega_n} P^i(\omega_n) \log \frac{P^i(\omega_n)}{P_=^j(\omega_n)} \\
&\quad + \sum_{\omega_n^j \in \Omega_n^j} P^i(\omega_n) P_=^j\big(\omega_n^{j+}\big|\omega_n\big) \log \frac{P_=^j(\omega_n^{j+}|\omega_n)}{P_=^j(\omega_n^{j+}|\omega_n)} \\
&= \sum_{\omega_n \in \Omega_n} P^i(\omega_n) \log \frac{P^i(\omega_n)}{P_=(\omega_n)} \\
&= d_n\big(P^i, P_=\big). \qquad \square
\end{aligned}$$

Lemma 10. *If $P^i \in \downarrow\mathbb{E}^i$ then $P^{i\to j} \in \downarrow\mathbb{E}^j$.*

Proof. Note first that $P^{i\to j} \in \mathbb{E}^j$ by Lemma 8, since $\varphi_1, \ldots, \varphi_k$ are propositions of $\mathcal{L}$.

We shall prove that $P^{i\to j} \in \downarrow\mathbb{E}^j$ by contradiction. Suppose otherwise that $Q^j \in \mathbb{E}^j$ is closer than $P^{i\to j}$ to the equivocator on $\mathcal{L}^j$. Then for sufficiently large n,

$$\begin{aligned}
d_n^j\big(Q^j, P_=^j\big) &= d_n\big(Q^j, P_=\big) + \sum_{\omega_n^j \in \Omega_n^j} Q^j\big(\omega_n^j\big) \log \frac{Q^j(\omega_n^{j+}|\omega_n)}{P_=^j(\omega_n^{j+}|\omega_n)} \\
&< d_n^j\big(P^{i\to j}, P_=^j\big) \\
&= d_n\big(P^i, P_=\big)
\end{aligned}$$

by Lemma 9, so $d_n(Q^j, P_=) < d_n(P^i, P_=)$. But $d_n(Q^j, P_=) = d_n^i(Q^{j\to i}, P_=^i)$ by Lemma 9, and $d_n(P^i, P_=) \leqslant d_n^i(P^i, P_=^i)$, so $d_n^i(Q^{j\to i}, P_=^i) < d_n^i(P^i, P_=^i)$ for sufficiently large n, and $Q^{j\to i} \in \mathbb{E}^i$ is thus closer to the equivocator than P^i on $\mathcal{L}^i$, contradicting the assumption that $P^i \in \downarrow\mathbb{E}^i$. Hence $P^{i\to j} \in \downarrow\mathbb{E}^j$ after all. $\square$

To prove the theorem we shall consider two cases.

(i) Suppose first that $\downarrow\mathbb{E}^1 = \emptyset$, so that $\Downarrow\mathbb{E}^1 = \mathbb{E}^1$, where $\mathbb{E}^1 = \langle \biguplus\{P \in \mathbb{P}^1\colon\ P(\varphi_1) \in X_1, \ldots, P(\varphi_k) \in X_k\}\rangle$. In this case there is no function in $\mathbb{E}^1$ closest to the equivocator and so $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ 1}} \psi^Y$ holds iff *all* probability functions lying in the convex hull of the premisses satisfy the conclusion.

It is also the case that $\downarrow\mathbb{E}^2 = \emptyset$. For if we suppose otherwise that there is some $P^2 \in \downarrow\mathbb{E}^2$, then $P^{2\to 1} \in \downarrow\mathbb{E}^1$ by Lemma 10, which contradicts the assumption that $\downarrow\mathbb{E}^1 = \emptyset$. Consequently, $\Downarrow\mathbb{E}^2 = \mathbb{E}^2$ and $\mathrel{|\!\!\approx^{\circ 2}}$ is determined in the same way as $\mathrel{|\!\!\approx^{\circ 1}}$, by considering *all* probability functions in the convex hull of the premisses.

Now suppose $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 1} \psi^Y$ and that $P^2 \in \mathbb{E}^2$. Now $P^2(\psi) = P^{2\to 1}(\psi)$ by Lemma 8, and $P^{2\to 1}(\psi) \in Y$ because $P^{2\to 1} \in \mathbb{E}^1$ by Lemma 8 and $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 1} \psi^Y$. So $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 2} \psi^Y$. Similarly, if $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 2} \psi^Y$ then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 1} \psi^Y$, as required.

(ii) The second case is that in which $\downarrow\mathbb{E}^1 \neq \emptyset$, so $\Downarrow\mathbb{E}^1 = \downarrow\mathbb{E}^1$ (and, by the above reasoning, $\downarrow\mathbb{E}^2 \neq \emptyset$ so $\Downarrow\mathbb{E}^2 = \downarrow\mathbb{E}^2$).

Suppose that $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 1} \psi^Y$ and that $P^2 \in \downarrow\mathbb{E}^2$. Now $P^2(\psi) = P^{2\to 1}(\psi)$ by Lemma 8, and $P^{2\to 1}(\psi) \in Y$ because $P^{2\to 1} \in \downarrow\mathbb{E}^1$ by Lemma 10 and $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 1} \psi^Y$. So $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 2} \psi^Y$. Similarly, if $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 2} \psi^Y$ then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ 1} \psi^Y$, as required. □

Proof of Theorem 4. To demand that an entailment relation $\mid\!\approx$ of a probabilistic logic with underlying predicate language $\mathcal{L}$ is invariant under all synonymy maps between $\mathcal{L}$ and $\mathcal{L}'$, for all $\mathcal{L}'$, is to demand that, if, for some $\mathcal{L}'$ and some synonymy map σ between $\mathcal{L}$ and $\mathcal{L}'$, σ implies that $\varphi_1 \leftrightarrow \varphi_1', \ldots, \varphi_k \leftrightarrow \varphi_k', \psi \leftrightarrow \psi'$ where $\varphi_1,\ldots,\varphi_k,\psi$ are propositions of $\mathcal{L}$ and $\varphi_1',\ldots,\varphi_k',\psi'$ are propositions of $\mathcal{L}'$, then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx \psi^Y$ if and only if $\varphi_1'^{X_1},\ldots,\varphi_k'^{X_k} \mid\!\approx' \psi'^Y$.

1. Let $\mathcal{L}'$ be $\mathcal{L}$ and consider the synonymy map $\sigma = \{\psi \leftrightarrow \neg\psi\}$. Then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx \psi^Y$ if and only if $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx \neg\psi^Y$, as required.

2. Let $\mathcal{L}'$ be $\mathcal{L}$ with new unary predicate U and consider the synonymy map $\sigma_0 = \{\psi \leftrightarrow (Ut_1 \wedge \neg Ut_1)\}$. Then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx \psi^Y$ if and only if $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx Ut_1 \wedge \neg Ut_1{}^Y$. But any probability function P must set $P(Ut_1 \wedge \neg Ut_1) = 0$, so $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx Ut_1 \wedge \neg Ut_1{}^Y$ implies that $0 \in Y$.

Now consider the synonymy map $\sigma_1 = \{\psi \leftrightarrow (Ut_1 \vee \neg Ut_1)\}$. Then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx \psi^Y$ if and only if $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx Ut_1 \vee \neg Ut_1^Y$. But any probability function P must set $P(Ut_1 \vee \neg Ut_1) = 1$, so $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx Ut_1 \vee \neg Ut_1{}^Y$ implies that $1 \in Y$. Hence $\{0,1\} \in Y$, as required. □

Regarding part 2, one might think that this consequence can be avoided if only we forbid tautologies and contradictions from featuring in synonymy maps. But such a move would be over-restrictive because perfectly natural synonymy maps need to mention tautologies. For example, enumerate the propositions of $\mathcal{L}$ by $\theta_1, \theta_2, \ldots,$ and consider a language $\mathcal{L}'$ whose constants t_i are intended to pick out the propositions θ_i of $\mathcal{L}$, and which has a unary predicate T with intended meaning *tautologous*. Then it is natural to consider the synonymy map $\sigma = \{\theta_i \leftrightarrow Tt_i\colon\ \theta_i \text{ is a tautology}\}$. Clearly, disallowing tautologies from appearing in synonymy maps would be a step too far, because it would rule out a synonymy map such as this.

Proof of Theorem 5. We shall need to appeal to the following lemma of Williamson [16, §5.3]:

Lemma 11. *If $P \neq Q$ and if, for sufficiently large n, $d_n(P,R) \leqslant d_n(Q,R)$ then any proper convex combination of P and Q, i.e., $S = \lambda P + (1-\lambda)Q$ for $\lambda \in (0,1)$, is closer than Q to R.*

Proof. In order to show that S is closer than Q to R we need to show that there is some N such that for $n \geqslant N$, $d_n(S,R) < d_n(Q,R)$.

Let L be the smallest n such that $P(\omega_n) < Q(\omega_n)$ for some ω_n. Let M be such that for $n \geqslant M$, $d_n(P,R) \leqslant d_n(Q,R)$. Take N to be the maximum of L and M. Now for $n \geqslant N$

$$\begin{aligned}
d_n(S,R) &= \sum_{\omega_n} \bigl[\lambda P(\omega_n) + (1-\lambda)Q(\omega_n)\bigr] \log \frac{\lambda P(\omega_n) + (1-\lambda)Q(\omega_n)}{\lambda R(\omega_n) + (1-\lambda)R(\omega_n)} \\
&< \sum_{\omega_n} \lambda P(\omega_n) \log \frac{\lambda P(\omega_n)}{\lambda R(\omega_n)} + (1-\lambda)Q(\omega_n) \log \frac{(1-\lambda)Q(\omega_n)}{(1-\lambda)R(\omega_n)} \\
&= \lambda d_n(P,R) + (1-\lambda) d_n(Q,R) \\
&\leqslant \lambda d_n(Q,R) + (1-\lambda) d_n(Q,R) \\
&= d_n(Q,R).
\end{aligned}$$

The first inequality is a consequence of the information-theoretic log-sum inequality: $\sum_{i=1}^k x_i \log x_i/y_i \geqslant (\sum_{i=1}^k x_i) \times \log(\sum_{i=1}^k x_i)/(\sum_{i=1}^k y_i)$ with equality iff x_i/y_i is constant, where $x_1,\ldots,x_k, y_1,\ldots,y_k$ are non-negative real numbers. Here x_i/y_i is not constant because $n \geqslant L$. The second inequality follows since $n \geqslant M$. □

We need to show that, if ω_n^* and $\omega_n^\dagger$ are treated symmetrically by the premisses, then $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ} \omega_n^{*Y}$ if and only if $\varphi_1^{X_1},\ldots,\varphi_k^{X_k} \mid\!\approx^{\circ} \omega_n^{\dagger Y}$.

Suppose for contradiction that this is not the case: without loss of generality we may suppose that there are some $\omega_n^*, \omega_n^\dagger, Y$ such that ω_n^* and $\omega_n^\dagger$ are treated symmetrically by the premisses, yet $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ}} {\omega_n^*}^Y$ and $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \not\mathrel{|\!\!\approx^{\circ}} {\omega_n^\dagger}^Y$. (The situation in which $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \not\mathrel{|\!\!\approx^{\circ}} {\omega_n^*}^Y$ and $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ}} {\omega_n^\dagger}^Y$ can be treated in exactly the same way.)

There are two cases.

(i) $\downarrow\mathbb{E} = \emptyset$. Then $\Downarrow\mathbb{E} = \mathbb{E} = \langle \biguplus\{\varphi_1^{X_1}, \ldots, \varphi_k^{X_k}\}\rangle$. Let $P \in \mathbb{E}$ be such that $P(\omega_n^\dagger) \notin Y$. Note that $P(\omega_n^*) \in Y$. Since ω_n^* and $\omega_n^\dagger$ are treated symmetrically by the premisses, we can choose $Q \in \mathbb{E}$ that swaps the values that P gives to ω_n^* and $\omega_n^\dagger$. Now $Q(\omega_n^*) = P(\omega_n^\dagger) \notin Y$, which contradicts the claim that $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ}} {\omega_n^*}^Y$, as required.

(ii) $\downarrow\mathbb{E} \neq \emptyset$. Then $\Downarrow\mathbb{E} = \downarrow\mathbb{E}$. Let $P \in \downarrow\mathbb{E}$ be such that $P(\omega_n^\dagger) \notin Y$. Choose $Q \in \mathbb{E}$ that swaps the values that P gives to ω_n^* and $\omega_n^\dagger$. Note that $P(\omega_n^*) \in Y$ so $Q \neq P$. Choose some proper convex combination S of P and Q. Now $P, Q \in \mathbb{E}$ hence so is S. Moreover, by definition of Q, $d_i(P, P_=) = d_i(Q, P_=)$ for all $i = 1, 2, \ldots$. Hence by Lemma 11, S is closer to the equivocator than either P or Q. This contradicts the claim that $P \in \downarrow\mathbb{E}$. □

Proof of Theorem 6. Suppose that, for $\theta(x)$ quantifier-free, $\mathrel{|\!\!\approx^{\circ}} \forall x\theta(x)^0$ but $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ}} \forall x\theta(x)^Y$ where $\inf Y > 0$. We need to show that $\varphi_1, \ldots, \varphi_k$ are not all quantifier-free.

That $\mathrel{|\!\!\approx^{\circ}} \forall x\theta(x)^0$ means that

$$P_=\big(\forall x\theta(x)\big) = \lim_{n\to\infty} P_=(\theta_n) = \lim_{n\to\infty} \frac{|\theta_n|_n}{2^{r_n}} = 0,$$

where, as before, θ_n is $\theta(t_1) \wedge \cdots \wedge \theta(t_n)$ and $|\theta_n|_n = |\{\omega_n \in \Omega_n\colon\ \omega_n \models \theta_n\}|$.

Consider now the entailment relationship $\varphi_1^{X_1}, \ldots, \varphi_k^{X_k} \mathrel{|\!\!\approx^{\circ}} \forall x\theta(x)^Y$. We shall suppose that $\varphi_1, \ldots, \varphi_k$ are all quantifier-free and we will see that $0 \in Y$, as required.

Choose l large enough that the quantifier-free formulae $\varphi_1, \ldots, \varphi_k, \theta(x)$ are all expressible in $\mathcal{L}_l$ (i.e., all their constant symbols are among $t_1, \ldots, t_l$). Note that $\Downarrow\mathbb{E} \neq \emptyset$ (see principle E1 of Section 1). For some $P \in \Downarrow\mathbb{E}$, define Q by:

$$Q(\omega_n) \stackrel{\text{df}}{=} P(\omega_l) P_=\big(\omega_n^+ | \omega_l\big) = \frac{P(\omega_l)}{2^{r_n - r_l}},$$

where, as before, ω_n is $\omega_l \wedge \omega_n^+$. $Q \in \mathbb{E}$ because it agrees with P as to the probabilities of the premiss propositions $\varphi_1, \ldots, \varphi_k$. Also, $Q \in \Downarrow\mathbb{E}$: if $\downarrow\mathbb{E} = \emptyset$ then $\Downarrow\mathbb{E} = \mathbb{E}$ so $Q \in \Downarrow\mathbb{E}$; otherwise note that for $n \geqslant l$, $d_n(Q, P_=) = d_l(P, P_=) \leqslant d_n(P, P_=)$ so Q is at least as close to the equivocator as P is, in which case $Q \in \Downarrow\mathbb{E} = \downarrow\mathbb{E}$ since $P \in \Downarrow\mathbb{E}$.

Now,

$$\begin{aligned}
Q(\theta_n) &= \frac{1}{2^{r_n - r_l}} \sum_{\omega_n \models \theta_n} P(\omega_l) \\
&= \frac{1}{2^{r_n - r_l}} \sum_{\omega_l \in \Omega_l} P(\omega_l) |\omega_l \wedge \theta_n|_n \\
&\leqslant \frac{1}{2^{r_n - r_l}} \sum_{\omega_l \in \Omega_l} P(\omega_l) |\theta_n|_n \\
&= \frac{|\theta_n|_n}{2^{r_n - r_l}} \\
&= 2^{r_l} \frac{|\theta_n|_n}{2^{r_n}} \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

since $\lim_{n\to\infty} \frac{|\theta_n|_n}{2^{r_n}} = P_=(\forall x\theta(x)) = 0$.

Hence $Q(\forall x\theta(x)) = 0$ and $0 \in Y$. □

References

[1] F. Bacchus, A.J. Grove, J.Y. Halpern, D. Koller, From statistical knowledge bases to degrees of belief, Artificial Intelligence 87 (1996) 75–143.
[2] O. Barnett, J. Paris, Maximum entropy inference with quantified knowledge, Logic Journal of the IGPL 16 (1) (2008) 85–98.
[3] R. Carnap, On inductive logic, Philosophy of Science 12 (2) (1945) 72–97.
[4] R. Carnap, R.C. Jeffrey (Eds.), Studies in Inductive Logic and Probability, vol. 1, University of California Press, Berkeley CA, 1971.
[5] N. Cartwright, How the Laws of Physics Lie, Clarendon Press, Oxford, 1983.
[6] A.J. Grove, J.Y. Halpern, D. Koller, Random worlds and maximum entropy, Journal of Artificial Intelligence Research 2 (1994) 33–88.
[7] R. Haenni, J.-W. Romeijn, G. Wheeler, J. Williamson, Probabilistic Logics and Probabilistic Networks, Synthese Library, Springer, Dordrecht, 2011.
[8] R.C. Jeffrey (Ed.), Studies in Inductive Logic and Probability, vol. 2, University of California Press, Berkeley, CA, 1980.
[9] J.M. Keynes, A Treatise on Probability (1921), Macmillan, London, 1948.
[10] P. Machamer, L. Darden, C. Craver, Thinking about mechanisms, Philosophy of Science 67 (2000) 1–25.

[11] S.D. Mitchell, Pragmatic laws, Philosophy of Science Proceedings 64 (1997) S468–S479.
[12] J.B. Paris, The Uncertain Reasoner's Companion, Cambridge University Press, Cambridge, 1994.
[13] S.R. Rad, Inference processes for probabilistic first order languages, PhD thesis, Department of Mathematics, University of Manchester, 2009, available at http://www.maths.manchester.ac.uk/~jeff/theses/srthesis.pdf.
[14] W.C. Salmon, Carnap's inductive logic, The Journal of Philosophy 64 (21) (1967) 725–739.
[15] T. Seidenfeld, Entropy and uncertainty, Philosophy of Science 53 (4) (1986) 467–491.
[16] J. Williamson, In Defence of Objective Bayesianism, Oxford University Press, Oxford, 2010.
[17] J. Williamson, An objective Bayesian account of confirmation, in: D. Dieks, W.J. Gonzalez, S. Hartmann, T. Uebel, M. Weber (Eds.), Explanation, Prediction, and Confirmation. New Trends and Old Ones Reconsidered, Springer, Dordrecht, 2011, pp. 53–81.
[18] J. Williamson, Objective Bayesianism, Bayesian conditionalisation and voluntarism, Synthese 178 (2011) 67–85.
[19] J. Williamson, Why frequentists and Bayesians need each other, Erkenntnis (2011), http://dx.doi.org/10.1007/s10670-011-9317-8.