J. Algorithms 63 (2008) 167–183

Contents lists available at ScienceDirect

Journal of Algorithms
Cognition, Informatics and Logic

www.elsevier.com/locate/jalgor

Objective Bayesian probabilistic logic

Jon Williamson

Philosophy, SECL, University of Kent, Canterbury, CT2 7NF, UK

A R T I C L E   I N F O

Article history:
Received 17 April 2008
Available online 6 September 2008

Keywords:
Objective Bayesianism
Maximum entropy
Maxent
Probabilistic logic
Probability logic
Bayesian network
Bayesian net
Credal network
Credal net

A B S T R A C T

This paper develops connections between objective Bayesian epistemology—which holds that the strengths of an agent's beliefs should be representable by probabilities, should be calibrated with evidence of empirical probability, and should otherwise be equivocal—and probabilistic logic. After introducing objective Bayesian epistemology over propositional languages, the formalism is extended to handle predicate languages. A rather general probabilistic logic is formulated and then given a natural semantics in terms of objective Bayesian epistemology. The machinery of objective Bayesian nets and objective credal nets is introduced and this machinery is applied to provide a calculus for probabilistic logic that meshes with the objective Bayesian semantics.

Contents

1. Introduction

Objective Bayesianism as developed in [9,10,20,21] is based around the following thesis:

Maximum Entropy Principle. An agent's degrees of belief should be representable by a probability function, from all those that satisfy constraints imposed by her evidence, that has maximum entropy.

E-mail address: j.williamson@kent.ac.uk.

0196-6774/$ – see front matter
doi:10.1016/j.jalgor.2008.07.001

Here the entropy H of a probability function P is defined by

$$H(P) = -\sum_{\omega\in\Omega} P(\omega)\log P(\omega)$$

in the case where Ω is a finite set of elementary outcomes.

Nilsson [15, §4] suggested that a maximum entropy probability function could be used to facilitate inference in a probabilistic logic. One aim of this paper is to develop that idea in some detail: to show how objective Bayesian probability provides a semantics for probabilistic logic. To this end the maximum entropy principle is motivated in Section 2, and extended from finite to denumerable outcome spaces in Section 3. Probabilistic logic is introduced in Section 4 and the objective Bayesian semantics is developed in Section 5.

A second aim of the paper is to show how one can answer questions posed in probabilistic logic. Section 6 introduces the machinery of objective Bayesian nets and objective credal nets: these are used to represent and reason with objective Bayesian probabilities. Section 7 shows how these nets can be used to provide a calculus for probabilistic logic.

2. Objective Bayesian epistemology

Objective Bayesian epistemology provides an answer to the following question:

- How strongly should an agent with evidence $\mathcal{E}$ believe the various propositions expressible in her language $\mathcal{L}$?

Initially we shall suppose that $\mathcal{L}$ can express finitely many elementary (i.e., non-logically-complex) propositions $A_1,\dots,A_n$, though we shall relax the finiteness condition when we consider predicate languages in Section 3. We shall also suppose that $\mathcal{L}$ can express logically complex propositions formed by applying the usual connectives of propositional logic to $A_1,\dots,A_n$. An *elementary outcome* or *atomic state* ω is a proposition of the form $\pm A_1\wedge\cdots\wedge\pm A_n$, where $+A_i$ is just A_i and $-A_i$ is $\neg A_i$. Ω is the set of all 2^n atomic states. The agent's *total evidence* or *epistemic background* $\mathcal{E}$ is understood as containing everything she takes for granted in her current operating context—including background knowledge, observations, theoretical assumptions, and so on. We need not assume that the evidence $\mathcal{E}$ is expressible in an agent's language $\mathcal{L}$.

Objective Bayesian epistemology imposes three norms on the strengths of an agent's beliefs: Probability, Calibration and Equivocation.

Probability. The strengths of the agent's beliefs should be representable by probabilities.

This norm posits that there is a probability function P on $\mathcal{L}$ such that $P(\theta)$ represents the degree to which the agent should believe θ, for each proposition θ expressible in $\mathcal{L}$. Here a *probability function* on $\mathcal{L}$ is a function P, from propositions expressible in $\mathcal{L}$ to real numbers, that satisfies the properties (i) $P(\omega)\geqslant 0$ for each $\omega\in\Omega$, (ii) $\sum_{\omega\in\Omega}P(\omega)=1$, and (iii) $P(\theta)=\sum_{\omega\models\theta}P(\omega)$ for each proposition θ expressible in $\mathcal{L}$. We use $\mathbb{P}_{\mathcal{L}}$ to denote the set of probability functions on $\mathcal{L}$. The Probability norm is typically justified by betting considerations: degrees of belief are indicative of betting intentions, and if the agent is to avoid bets that lose money whatever happens, her degrees of belief had better behave like probabilities.[1]

Calibration. The agent's degrees of belief should satisfy constraints imposed by her evidence.

It is thus supposed that there is some set $\mathbb{E}\subseteq\mathbb{P}_{\mathcal{L}}$ of probability functions on $\mathcal{L}$ that are compatible with evidence $\mathcal{E}$ and that the probability function $P_{\mathcal{E}}$ representing the agent's degrees of belief should lie in that set. $\mathbb{E}$ is determined as follows. First, if the evidence implies that the empirical probability function P^* on $\mathcal{L}$ lies in some set $\mathbb{P}^*$ of probability functions on $\mathcal{L}$, then the agent's belief function $P_{\mathcal{E}}$ should lie in the convex hull $[\mathbb{P}^*]$ of $\mathbb{P}^*$. Second, qualitative evidence of, e.g., causal, logical, hierarchical or ontological structure imposes certain structural constraints which force $P_{\mathcal{E}}$ to lie in a set $\mathbb{S}$ of probability functions on $\mathcal{L}$ that satisfy those constraints. Thus $\mathbb{E}=[\mathbb{P}^*]\cap\mathbb{S}$. The rationale behind taking convex hulls and the precise formulation of the structural constraints can be found in [21] and [26]. Note three things. First, $\mathbb{E}$ is always non-empty: an agent is never prohibited from holding any beliefs at all.[2] Second, $\mathbb{E}$ is always closed: it is hardly justifiable to deem irrational the limit point of rational belief functions. Third, in the context of probabilistic logic $\mathbb{E}$ is always convex: as will become apparent in Section 5, $\mathbb{P}^*$ is determined by the semantics of the logic and there are no structural constraints, so $\mathbb{E}=[\mathbb{P}^*]$ and we need not dwell on $\mathbb{S}$ in this paper. The motivation behind Calibration hinges on the use of degrees of belief for inference: well-calibrated degrees of belief lead to more reliable inferences in the long run (see, e.g., [8, §13.e]).

[1] See [16, Chapter 3] for this and other justifications of the Probability norm.

[2] If the evidence is inconsistent in the sense that $\mathbb{P}^*=\emptyset$ then some consistency maintenance procedure needs to be invoked. For example one might take $\mathbb{E}$ to be the space $\mathbb{P}_{\mathcal{L}}$ of all probability functions on $\mathcal{L}$; or the set of probability functions that satisfy maximally consistent subsets of evidence; or the set of probability functions that satisfy the more entrenched evidence. For ease of exposition we shall take $\mathbb{E}=\mathbb{P}_{\mathcal{L}}$ if $\mathbb{P}^*=\emptyset$; letting $[\emptyset]\stackrel{\mathrm{df}}{=}\mathbb{P}_{\mathcal{L}}$, the characterisation $\mathbb{E}=[\mathbb{P}^*]\cap\mathbb{S}$ still applies. However the approach developed below can be adjusted to handle other consistency maintenance procedures, if required.

Equivocation. The agent's degrees of belief should otherwise be as equivocal as possible.

Here 'as equivocal as possible' means as close as possible to the equivocator $P_=$ on $\mathcal{L}$, which gives each atomic state the same probability,

$$P_=(\omega) = \frac{1}{2^n}.$$

Distance between probability functions is measured by what is sometimes called *cross entropy* or *Kullback–Leibler divergence*,

$$d(P, Q) = \sum_{\omega \in \Omega} P(\omega) \log \frac{P(\omega)}{Q(\omega)},$$

where $0 \log 0$ is taken to be 0. Note that this is not a distance measure in the usual mathematical sense because it is not symmetric and does not satisfy the triangle inequality. The motivation behind Equivocation exploits the fact that belief is a basis for action: more extreme degrees of belief tend to trigger high-risk actions (where there is a lot to lose if the agent misjudges) while equivocal degrees of belief are associated with lower risks; and the agent should only take on risk to the minimum extent warranted by evidence [23].

In sum, the agent's degrees of belief should be representable by a probability function $P_{\mathcal{E}}$ on $\mathcal{L}$ that is in the set $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E}\colon\ d(P, P_=) \text{ is minimised}\}$. Since distance to the equivocator is minimised just when entropy is maximised, this gives:

Maximum Entropy Principle. The agent's degrees of belief should be representable by a probability function $P_{\mathcal{E}} \in \{P \in \mathbb{E}\colon\ H(P) \text{ is maximised}\}$.

Note that in the absence of structural constraints, $P_{\mathcal{E}}$ is uniquely determined. This is because entropy is a strictly concave function and it is being maximised over a closed and convex set of probability functions $\mathbb{E} = [\mathbb{P}^*]$, so it has a unique maximum. As noted above, there are no structural constraints in the application of objective Bayesianism to probabilistic logic.[3]

3. Predicate languages

While objective Bayesian epistemology has hitherto been developed for use on finite domains, and, in Bayesian statistics, on uncountable domains, it is also natural in the context of probabilistic logic to consider countably infinite domains, in particular first-order predicate languages. We shall see in this section that the analysis of the last section, which dealt with a propositional language, extends naturally to the case in which $\mathcal{L}$ is a predicate language.[4]

Accordingly we suppose now that $\mathcal{L}$ is a first-order predicate language (without equality). It is convenient to assume that each individual is picked out by a unique constant symbol t_i; we shall suppose that there is a countable infinity $t_1, t_2, \ldots$ of such constants, but only finitely many predicate symbols. For $n \geqslant 1$ let $\mathcal{L}_n$ be the finite predicate language involving the symbols of $\mathcal{L}$ but with only finitely many 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. We shall insist that the $A_1, A_2, \ldots$ are ordered 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.

As before we base objective Bayesian epistemology on the norms of Probability, Calibration and Equivocation.

Probability. The strengths of the agent's beliefs should be representable by probabilities.

This norm now requires that there be a probability function P on predicate language $\mathcal{L}$ such that $P(\theta)$ represents the degree to which the agent should believe θ, for each proposition θ expressible in $\mathcal{L}$. Here a *probability function* on a predicate language $\mathcal{L}$ is a function P, from propositions expressible in $\mathcal{L}$ to real numbers, that satisfies the properties (i) $P(\omega_n) \geqslant 0$ for each ω_n, (ii) for each n, $\sum_{\omega_n \in \Omega_n} P(\omega_n) = 1$, (iii) for each quantifier-free proposition θ, $P(\theta) = \sum_{\omega_n \models \theta} P(\omega_n)$ where n is large enough that $\mathcal{L}_n$ contains all the atomic propositions occurring in θ, and (iv) quantified statements are assigned probabilities via

[3] Objective Bayesianism has been criticised on account of the fact that the Maximum Entropy Principle may disagree with the principle of Bayesian Conditionalisation, which is a norm governing the updating of degrees of belief that is often advocated by subjective Bayesians; this objection is rebutted in [27]. It has also been criticised on account of the fact that the results of the Maximum Entropy Principle depend on the agent's language; this objection is rebutted in [25, §16] and is further discussed in the next section.

[4] This paper is not the first to extend objective Bayesianism to predicate languages. For example, [17] put forward an approach based on taking limits of maximum entropy probability functions. That line of work, however, deals with restricted classes of predicate languages—e.g., languages in which there are only unary predicates—and cases in which there is a restriction on the amount of evidence. For our application to probabilistic logic we need a more general framework, and so base our approach around the three norms of Section 2 rather than limits of maximum entropy functions.

$$P\big(\forall x\theta(x)\big) = \lim_{m\to\infty} P\left(\bigwedge_{i=1}^{m} \theta(t_i)\right),$$
$$P\big(\exists x\theta(x)\big) = \lim_{m\to\infty} P\left(\bigvee_{i=1}^{m} \theta(t_i)\right).$$

Note in particular that a probability function on predicate language $\mathcal{L}$ is determined by its values on the quantifier-free propositions of $\mathcal{L}$.[5]

Calibration. The agent's degrees of belief should satisfy constraints imposed by her evidence.

Again it is supposed that there is some set $\mathbb{E}$ of probability functions on $\mathcal{L}$ that are compatible with evidence $\mathcal{E}$ and that the probability function $P_{\mathcal{E}}$ representing the agent's degrees of belief should lie in that set. As before we take $\mathbb{E} = [\mathbb{P}^*] \cap \mathbb{S}$. In the context of probabilistic logic—and hence this paper—there are no structural constraints so $\mathbb{E} = [\mathbb{P}^*]$; this set is closed, convex and non-empty.

Equivocation. The agent's degrees of belief should otherwise be as equivocal as possible.

In the case of a predicate language we can define the equivocator $P_=$ on $\mathcal{L}$ by

$$P_=(\omega_n) = \frac{1}{2^{r_n}}$$

for all ω_n. We consider the n-distance between probability functions,

$$d_n(P, Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log \frac{P(\omega_n)}{Q(\omega_n)},$$

where as before $0 \log 0$ is taken to be 0. We say that P is *closer* to R than Q if there is some N such that for all $n \geqslant N$, $d_n(P, R) < d_n(Q, R)$. We write $P \prec Q$ if P is closer to the equivocator $P_=$ than Q. Now we define $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E}$: P is minimal with respect to $\prec\}$ as long as this set is non-empty, setting $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \mathbb{E}$ otherwise. Objective Bayesianism then requires that the agent's degrees of belief be representable by a probability function $P_{\mathcal{E}} \in \downarrow\mathbb{E}$.

Define the n-entropy $H_n(P)$ by

$$H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n).$$

We say that P has *greater entropy* than Q, written $P \gg Q$, if there is some N such that for all $n \geqslant N$, $H_n(P) > H_n(Q)$. We then have:

Maximum Entropy Principle. The agent's degrees of belief should be representable by a probability function $P_{\mathcal{E}} \in \{P \in \mathbb{E}$: P is maximal with respect to $\gg\}$ in cases where there is such a maximiser.

Discussion

Having presented the principal definitions, we shall in the remainder of this section discuss the key properties of the resulting framework. The reader primarily interested in the application of objective Bayesianism to probabilistic logics may wish to skip directly to the next section.

[5] Note too that (iv) ascribes probabilities to universally and existentially quantified propositions while, as mentioned in Section 2, these probabilities are interpreted via betting considerations. One might think that there is a tension here, because bets on universally quantified propositions are unlikely to be settled. If there are infinitely many elements of the domain then normally one cannot tell in a finite time whether a universally quantified proposition is true (there are exceptions, though—for instance if the proposition is a tautology). In which case a bet on the truth of such a proposition won't be settled and so betting considerations can hardly motivate a particular value for the agent's degree of belief in the proposition. Hence one can assign the proposition any degree of belief at all, contrary to the constraint invoked in (iv) above.

There are two possible responses. First, it suffices to point out that (iv) can be motivated by semantics rather than betting. The identities invoked by (iv) should hold in virtue of the meaning of 'for all' and 'there exists', given our assumption that each member of the domain is picked out by some constant symbol. These identities need not be justified by a Dutch book argument and so their justification does not require that bets on universally quantified propositions be settled in principle.

A second possible response involves denying the tension. The betting scenario is in any case a considerable idealisation of the following form: 'assuming an agent for whom the utility of money increases linearly with its size were compelled to bet for unknown positive or negative stakes, and assuming there were a stake-maker with exactly the same evidence as the agent trying to force the agent to lose money, and assuming no mistakes were made over the settling of the bets, and assuming that the agent cares equally about losses of the same magnitude whenever they are incurred in the future, ...' Since the last condition requires that the agent care about losses incurred after the end of universe, it doesn't take much to change this condition to 'assuming that the agent cares equally about losses of the same magnitude whenever they are incurred in the finite or infinite future.' In which case one could, if one wished, justify (iv) by Dutch book considerations after all.

Properties of the Closer Relation. First we shall investigate the *closer* relation defined above, showing that this notion does what one would expect of a closeness relation.

Proposition 3.1. *For fixed R the binary relation* $\cdot$ is closer than $\cdot$ to *R is irreflexive, asymmetric and transitive.*

Proof. Irreflexivity is immediate from the definition: $d_n(P,R) \not< d_n(P,R)$.

Asymmetry is also immediate: if $d_n(P,R) < d_n(Q,R)$ then $d_n(Q,R) \not< d_n(P,R)$. For transitivity, suppose P is closer than Q to R and Q is closer than S to R. Then there is some L such that for $n \geqslant L$, $d_n(P,R) < d_n(Q,R)$, and there is some M such that $n \geqslant M$ implies $d_n(Q,R) < d_n(S,R)$. Take N to be the maximum of L and M. Then for $n \geqslant N$, $d_n(P,R) < d_n(S,R)$, so P is closer than S to R. □

Proposition 3.2. *If P is closer than Q to 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 $P(\omega_n) \neq Q(\omega_n)$ for some ω_n. Let M be such that for $n \geqslant M$, $d_n(P,R) < d_n(Q,R)$. Take N to be the maximum of L and M. Now for $n \geqslant N$

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

The first inequality is a consequence of the information-theoretic log-sum inequality:

$$\sum_{i=1}^k x_i \log x_i/y_i \geqslant \left(\sum_{i=1}^k x_i\right) \log \left(\sum_{i=1}^k x_i\right) \Big/ \left(\sum_{i=1}^k y_i\right)$$

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. The second inequality is on account of P being closer than Q to R. □

Definition of the ↓ Operator. Note that we defined $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \{P \in \mathbb{E}\colon\ P$ is minimal with respect to $\prec\}$ if this set is non-empty, and $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \mathbb{E}$ otherwise. The question therefore arises as to whether there are in fact any cases in which $\mathbb{E}$ has no minimal elements, i.e., elements $P \in \mathbb{E}$ such that for all $Q \in \mathbb{E}$, $Q \not\prec P$. Suppose $\mathcal{L}$ is a language with a single unary predicate. Define P_i by

$$P_i(A_j|\omega_{j-1}) = \begin{cases} \frac{1}{2}, & j < i, \\ 1, & \text{otherwise} \end{cases}$$

for all j and all $\omega_{j-1} \in \Omega_{j-1}$. Then we have an infinite descending chain: for all i, $P_{i+1} \prec P_i$. So if $\mathbb{E} = \{P_1, P_2, \ldots\}$ then $\mathbb{E}$ has no minimal elements.

But is it possible that $\mathbb{E} = \{P_1, P_2, \ldots\}$? Recall that it is a normative constraint that $\mathbb{E}$ be closed, so the question is whether $\{P_1, P_2, \ldots\}$ contains its limit points—if not, $\mathbb{E}$ cannot be $\{P_1, P_2, \ldots\}$. In fact, this question depends on the notion of limit point that we invoke:

Definition 3.3 *(Strong limit point).* Probability function P is a *strong limit point* of $\mathbb{E}$ if for all $\varepsilon > 0$ there is some $Q \neq P$ in $\mathbb{E}$ such that $d_n(Q,P) < \varepsilon$ for all n.

Definition 3.4 *(Weak limit point).* Probability function P is a *weak limit point* of $\mathbb{E}$ if for all $\varepsilon > 0$ and for all n there is some $Q \neq P$ in $\mathbb{E}$ such that $d_n(Q,P) < \varepsilon$.[6]

[6] Strictly speaking Q must be zero wherever P is zero for $d_n(Q,P)$ to be well-defined. We might use Euclidean distance for d_n where this condition does not hold. In fact since in the context of limit points we are considering small distances, it doesn't much matter which notion of distance we use in these definitions.

Now if we consider the putative characterisation of $\mathbb{E}$ as $\{P_1, P_2, \ldots\}$, we see that this set has no strong limit points, but has the equivocator $P_=$ as a weak limit point. Plausibly, if $P_1, P_2, \ldots$ are appropriate candidates for a belief function then $P_=$ is also an appropriate candidate. In which case the kind of closure required for objective Bayesianism on predicate languages is closure under weak limit points. Then in fact $\mathbb{E} = \{P_=, P_1, P_2, \ldots\}$ and $\downarrow\mathbb{E} = \{P_=\}$.

Does every infinite descending chain have a weak limit point that is closer to the equivocator than any member of the chain? No. Again let $\mathcal{L}$ be a language with a single unary predicate. Define Q_i by

$$Q_i(A_j|\omega_{j-1}) = \begin{cases} \frac{1}{2}, & i < j \leqslant 2i, \\ 1, & \text{otherwise} \end{cases}$$

for all j, ω_{j-1}. This defines an infinite descending chain: for all i, $Q_{i+1} \prec Q_i$. The only weak limit point is Q defined by $Q(A_j|\omega_{j-1}) = 1$ for all j, ω_{j-1}. Hence in principle there might be some $\mathbb{E} = \{Q, Q_1, Q_2, \ldots\}$. But Q is a maximal element of $\mathbb{E}$: $Q_i \prec Q$ for all i. So $\mathbb{E}$ has no minimal elements.

If such a pathological case arises, the question of how to determine an appropriate belief function $P_{\mathcal{E}}$ becomes a concern. Arguably the most natural recommendation is to take $\downarrow\mathbb{E} = \mathbb{E}$, for if $\downarrow\mathbb{E}$ were a strict subset $\mathbb{F}$ of $\mathbb{E}$ in cases with infinite descending chains then idempotence could fail, $\downarrow\downarrow\mathbb{E} \neq \downarrow\mathbb{E}$. (If $\mathbb{F}$ is an infinite strict subset of an infinite descending chain then $\mathbb{F}$ itself would contain an infinite descending chain. On the other hand if $\mathbb{F}$ were a finite subset of an infinite descending chain, then if $\mathbb{E}$ contained infinitely many infinite descending chains, the union $\bigcup \mathbb{F}_i$ of the finite subsets could itself contain an infinite descending chain. Applying the $\downarrow$ operator again to an infinite descending chain would yield yet another strict subset, so $\downarrow\downarrow\mathbb{E} \neq \downarrow\mathbb{E}$.) This motivates defining $\downarrow\mathbb{E} \stackrel{\text{df}}{=} \mathbb{E}$ in cases where $\mathbb{E}$ lacks minimal elements.

Note that where $\downarrow\mathbb{E} = \mathbb{E}$ the Equivocation norm has no bite and we are essentially left with the Probability and Calibration norms. *Subjective Bayesian epistemology* appeals to Probability and Calibration but not Equivocation: constraints are weaker and the determination of an appropriate belief function is largely a question of subjective choice [8,19]. So objective and subjective Bayesian epistemology agree in those cases where there are infinite descending chains in $\mathbb{E}$ with respect to $\prec$. Since subjective Bayesian epistemology is already well studied, and since explicit mention of these pathological cases would obscure the exposition of the main points of this paper, we shall, in subsequent sections of this paper, mainly restrict our attention to evidence that does not admit infinite descending chains.

Let us now turn to two possible alternative characterisations of the Equivocation norm.

Limiting Distance. Define

$$d(P, Q) \stackrel{\text{df}}{=} \lim_{n\to\infty} d_n(P, Q).$$

As long as P is zero whenever Q is zero, $d(P, Q) \in [0, \infty]$ since for $m < n$,

$$d_n(P, Q) = d_m(P, Q) + \sum_{\omega_n \in \Omega_n} P(\omega_n) \log \frac{P(\omega'_n|\omega_m)}{Q(\omega'_n|\omega_m)} \geqslant d_m(P, Q) \geqslant 0$$

where ω_m is the m-state determined by ω_n and ω'_n is the 'remainder', i.e., ω_n is $\omega_m \wedge \omega'_n$.

One might suggest that the agent's degrees of belief should be representable by $P_{\mathcal{E}} \in \{P \in \mathbb{E}\colon\ d(P, P_=) \text{ is minimised}\}$. But as it stands this does not adequately explicate the concept of closeness to the equivocator, because in the case of a predicate language there are probability functions P, Q such that although one is intuitively closer to the equivocator than the other, $d(P, P_=) = d(Q, P_=)$. Suppose for example that $\mathcal{E}$ imposes the constraints $P(A_i|\pm A_1 \wedge \cdots \wedge \pm A_{i-1}) = 1$ for all $i \geqslant 2$ ($\mathbb{E}$ is then non-empty, closed and convex). Thus only $P(A_1)$ is unconstrained. Now $d(P, P_=) = \infty$ for all $P \in \mathbb{E}$. Yet intuitively there is a unique function in $\mathbb{E}$ that is closest to the equivocator, namely the function that sets $P(A_1) = 1/2$ and $P(A_i|\pm A_1 \wedge \cdots \wedge \pm A_{i-1}) = 1$ for all $i \geqslant 2$. Indeed this function is minimal with respect to $\prec$. This motivates taking $P_{\mathcal{E}} \in \{P \in \mathbb{E}\colon\ P \text{ is minimal with respect to } \prec\}$, as we did above, rather than $P_{\mathcal{E}} \in \{P \in \mathbb{E}\colon\ d(P, P_=) \text{ is minimised}\}$.

Example 3.5. Suppose $\mathbb{E} = \{P\colon\ P(\forall x Ux) = c\}$ for some fixed $c \in [0, 1]$. We have that $d(P, P_=) = \infty$ for all $P \in \mathbb{E}$. Define P by

$$P(Ut_1) = \frac{c+1}{2},$$

$$P(Ut_{i+1}|Ut_1 \wedge \cdots \wedge Ut_i) = \frac{(2^{i+1}-1)c+1}{(2^{i+1}-2)c+2},$$

$$P(Ut_{i+1}|\pm Ut_1 \wedge \cdots \wedge \pm Ut_i) = \frac{1}{2}$$

otherwise. Then P is the member of $\mathbb{E}$ that is closest to the equivocator.

The *closer* relation agrees with comparative distance in the sense of the following proposition:

Proposition 3.6. *$d(P, R) < d(Q, R)$ implies that P is closer than Q to R.*

Proof. Choose N such that $d_N(Q,R) > d(P,R)$. Then for $n \geqslant N$, $d_n(P,R) \leqslant d(P,R) < d_n(Q,R)$ so P is closer to R than Q. □

Pointwise Limit of Distance Minimisers. There is a second possible alternative characterisation of the equivocation norm, which proceeds as follows. Let $\mathbb{E}^n \stackrel{\mathrm{df}}{=} \{P_{\restriction \mathcal{L}_n}\colon\ P \in \mathbb{E}\}$, the set of probability functions on $\mathcal{L}_n$ that are restrictions of probability functions (on $\mathcal{L}$) in $\mathbb{E}$. Let P^n be a probability function in $\mathbb{E}^n$ that has minimum n-distance from the equivocator. (Equivalently, P^n is a probability function in $\mathbb{E}$ that has maximum n-entropy H_n.) Recall that $\mathbb{E}$ is always closed; if $\mathbb{E}$ is convex then P^n is uniquely determined: there is a unique minimiser of distance on a finite domain, from a non-empty, closed convex set of probability functions. Define

$$P^{\infty}(\omega_n) = \lim_{n\to\infty} P^n(\omega_n).$$

If this leads to a well-defined probability function, one might suggest that one deem P^{∞} to be an appropriate choice for a rational belief function $P_{\mathcal{E}}$.

In fact, where P^{∞} is well-defined it *will* be deemed appropriate by the approach taken here:

Proposition 3.7. *If P^{∞} exists then $P^{\infty} \in {\downarrow}\mathbb{E}$.*

Proof. First note that P^{∞} defines a probability function:

(i) $P^{\infty}(\omega_n) = \lim_{n\to\infty} P^n(\omega_n)$ and $P^n(\omega_n) \geqslant 0$ so $P^{\infty}(\omega_n) \geqslant 0$,
(ii) for each n,

$$\sum_{\omega_n\in\Omega_n} P^{\infty}(\omega_n) = \sum_{\omega_n\in\Omega_n} \lim_{m\to\infty} P^m(\omega_n) = \lim_{m\to\infty} \sum_{\omega_n\in\Omega_n} P^m(\omega_n) = \lim_{m\to\infty} 1 = 1.$$

Then principles (iii) and (iv) can be used to assign probabilities to arbitrary sentences.

Next note that $P^{\infty} \in \mathbb{E}$ since it is a weak limit point of members of $\mathbb{E}$ and $\mathbb{E}$ is closed under weak limits. One can see this as follows. For each P^n defined as above on $\mathcal{L}_n$ let P_n be some function on $\mathcal{L}$ that extends it and is in $\mathbb{E}$. By definition of P^{∞}, given ω_n and $\varepsilon > 0$, $|P^m(\omega_n) - P^{\infty}(\omega_n)| < \varepsilon$ for sufficiently large m—say for $m \geqslant M_{\omega_n}$. Letting $M = \max_{\omega_n\in\Omega_n} M_{\omega_n}$ we see that given $\varepsilon > 0$ and $m \geqslant M$, $|P^m(\omega_n) - P^{\infty}(\omega_n)| < \varepsilon$ for all $\omega_n \in \Omega_n$. Equivalently, given $\varepsilon > 0$, $d_n(P_m, P^{\infty}) < \varepsilon$ for sufficiently large m.[7] But this is just to say that P^{∞} is a weak limit point of the P_m.

If there are no minimal elements in $\mathbb{E}$ then ${\downarrow}\mathbb{E} = \mathbb{E}$ and $P^{\infty} \in {\downarrow}\mathbb{E}$ as required.

Otherwise, suppose for contradiction that $P^{\infty} \notin {\downarrow}\mathbb{E}$. ${\downarrow}\mathbb{E}$ is non-empty so there is a $Q \in {\downarrow}\mathbb{E}$ such that $Q \prec P^{\infty}$. So for sufficiently large n,

$$d_n\big(P^m, P_=\big) \leqslant d_n(Q, P_=) < d_n\big(P^{\infty}, P_=\big)$$

so

$$d_n\big(P^{\infty}, P_=\big) - d_n\big(P^m, P_=\big) \geqslant d_n\big(P^{\infty}, P_=\big) - d_n(Q, P_=) > 0.$$

But $d_n(P^{\infty}, P_=) - d_n(P^m, P_=) \to 0$ as $m \to \infty$ so $d_n(P^{\infty}, P_=) - d_n(Q, P_=) \to 0$ as $m \to \infty$. But P^{∞} and Q are independent of m, so $d_n(P^{\infty}, P_=) = d_n(Q, P_=)$. This contradicts $Q \prec P^{\infty}$, as required. □

Note that P^{∞} is not the limiting function of [17] and [2]. There the procedure is to take pointwise limits of functions that maximise n-entropy from all those satisfying evidence *where the evidence is re-expressed using $\mathcal{L}_n$*. Here the procedure is to take pointwise limits of functions that maximise n-entropy from all those satisfying evidence *expressed using $\mathcal{L}$*. The former case faces the finite model problem: while there may be probability functions that satisfy the evidence on an infinite domain, there may be no probability function that satisfies that evidence when reinterpreted as saying something about a finite domain. This problem arises when considering total orderings, for instance: if the evidence says that $\forall x \exists y Rxy$ where R is a strict total order then only an infinite language will yield probability functions that satisfy that evidence; hence one cannot satisfy such a proposition by taking limits of probability functions on finite languages that satisfy the proposition (but one can by taking limits of probability functions that are restrictions of functions on an infinite language that satisfy the proposition). In the approach outlined here, the finite model problem does not arise.

Order Invariance. Next we turn to the question of whether the *closer* relation is well-defined.

There is some flexibility in the ordering of the $A_1, A_2, \ldots$. Although atomic propositions expressible in $\mathcal{L}_m$ must occur before those expressible in $\mathcal{L}_n$ but not in $\mathcal{L}_m$, where $m < n$, there will typically be several orderings that satisfy this requirement. However, the *closer* relation is well-defined:

[7] As noted before, if for some ω_n, $P_m(\omega_n) \neq 0$ but $P^{\infty}(\omega_n) = 0$, then d_n is ill-defined but we can just appeal to Euclidean distance rather than cross entropy here.

Proposition 3.8. *The* closer *relation is independent of the precise ordering $A_1, A_2, \ldots$ of the atomic propositions.*

Proof. Suppose for contradiction that there are two orderings such that P is closer to R than Q under ordering 1 but Q is closer to R than P under ordering 2. Let N_1 and N_2 be the N for orderings 1 and 2 respectively.

Let n be the maximum of N_1 and N_2. Now $\mathcal{L}_n$ expresses the A_{N_1} of ordering 1 and the A_{N_2} of ordering 2 and all the predecessors of these propositions under both orderings.

Ordering $\mathcal{L}_n$ according to order 1 we have that $d_n(P,R) < d_n(Q,R)$ since $n \geqslant N_1$. Similarly if we order $\mathcal{L}_n$ according to order 2 we have that $d_n(Q,R) < d_n(P,R)$ since $n \geqslant N_2$. But this is a contradiction because cross-entropy distance is independent of ordering on a finite language. □

Note that we have assumed a fixed ordering of the constant symbols in the language. Although *closer* is well-defined, the question arises as to how this relation behaves on languages that differ only with respect to the ordering of the t_i. In fact in certain cases the *closer* relation does depend on the ordering of the t_i. Suppose for instance that $\mathcal{L}$ and $\mathcal{L}'$ have a single predicate U which is unary, but that in $\mathcal{L}$ the constants are ordered $t_1, t_3, t_2, t_5, t_4, t_7, t_6, t_9, \ldots$ and in $\mathcal{L}'$ they are ordered $t_2, t_4, t_1, t_6, t_3, t_8, t_5, t_{10}, \ldots$. Suppose that P and Q both render the Ut_i all probabilistically independent; let P be defined by $P(Ut_i) = 1$ if i is odd and $1/2$ otherwise, and let Q be defined by $Q(Ut_i) = 1$ if i is even and $1/2$ otherwise. Now on $\mathcal{L}$ we have that $Q \prec P$ but on $\mathcal{L}'$ we have that $P \prec Q$.

On the other hand there is much agreement across orderings, as can be seen from the following consideration. Call two orderings *commensurable* if given any constant symbol there is a finite set of constants that contains that constant and is closed with respect to taking ancestors under each ordering; an analogue of the proof of Proposition 3.8 shows that *closer* is independent of ordering where the orderings under consideration are commensurable.

In our context it is important to point out that this dependence of the *closer* relation on the ordering of the constants does not imply a dependence of the recommendations of objective Bayesian epistemology on the ordering of the constants. The above example yields dependence on ordering because the probability functions are defined in terms of features of the indices of the constants which are not expressible within the language. If the evidence $\mathcal{E}$ imposes finitely many constraints on the probabilities of propositions in $\mathcal{L}$ (as is the case in the probabilistic logic considered in Sections 4, 5) then $\downarrow\mathbb{E}$ will not be sensitive to the order of the constants. (This is because finitely many constraints can only mention finitely many constant symbols, and hence can only distinguish finitely many atomic propositions. So sensitivity to order can only occur for sufficiently small n. However the *closer* relation depends on sufficiently large n.)

While this is all well and good for the context of this paper, the question still remains as to what to do should $\downarrow\mathbb{E}$ be sensitive to order of the constants in other situations. That different agents with different languages adopt different degrees of belief is of course no problem in itself. What is more problematic is that in natural language there may be no natural order of the constant symbols (i.e., names). If the agent's natural language can be explicated by any one of a number of predicate languages $\mathcal{L}^i$ and rational degree of belief is sensitive to this choice of predicate language then some protocol is required to handle this sensitivity. The natural protocol is to take $\downarrow\mathbb{E}$ to be $\bigcup_i \downarrow^i\mathbb{E}$ where $\downarrow^i\mathbb{E}$ is the set of probability functions satisfying evidence and closest to the equivocator with respect to language $\mathcal{L}^i$. If there is nothing in the agent's context that determines a most appropriate $\downarrow^i\mathbb{E}$, then there is clearly no rational requirement that the agent's beliefs be representable by a function from one $\downarrow^i\mathbb{E}$ rather than another, and the agent is free to choose among the full range of the $\downarrow^i\mathbb{E}$.

Equidistance. Having discussed the question of indeterminacy with respect to the ordering of the constants, we now turn to a second kind of indeterminacy: given a fixed language (and hence a fixed ordering of the constant symbols) is $\downarrow\mathbb{E}$ always a singleton? Or are there cases in which evidence fails to uniquely determine the agent's degrees of belief? We have seen that $\downarrow\mathbb{E}$ can be a non-singleton if there are infinite descending chains (in which case $\downarrow\mathbb{E}$ is taken to be $\mathbb{E}$); but there are other kinds of non-uniqueness.

Definition 3.9 *(Equidistant).* P and Q are *equidistant* from R if neither is closer than the other to R.

Proposition 3.10. *For fixed R the binary relation* equidistant *is reflexive and symmetric but not transitive in general and so not an equivalence relation.*

Proof. Reflexivity and symmetry follow directly from the definitions. To construct a counterexample to transitivity, choose P, Q and S such that $d_n(P,R)$ and $d_n(Q,R)$ oscillate in magnitude, as do $d_n(Q,R)$ and $d_n(S,R)$, but where $d_n(P,R)$ and $d_n(S,R)$ do not. □

Definition 3.11 *(Stably equidistant).* P and Q are *stably equidistant* from R iff there is some N such that for all $n \geqslant N$, $d_n(P,R) = d_n(Q,R)$.

Proposition 3.12. *For fixed R,* stably equidistant *defines an equivalence relation. If distinct P and Q are stably equidistant from R then any proper convex combination S of P and Q is closer than either P or Q to R.*

Proof. That *stably equidistant* is an equivalence relation follows directly from the definitions.

We need to show that there is some N such that for $n \geqslant N$, $d_n(S,R) < d_n(P,R), d_n(Q,R)$. Let L be the smallest n such that $P(\omega_n) \neq Q(\omega_n)$ for some ω_n. Let M be the smallest j such that for $n \geqslant j$, $d_n(P,R) = d_n(Q,R)$. Take N to be the maximum of L and M. For $n \geqslant N$,

$$\begin{aligned} d_n(S,R) &< \lambda d_n(P,R) + (1-\lambda) d_n(Q,R) \\ &= \lambda d_n(P,R) + (1-\lambda) d_n(P,R) = d_n(P,R). \end{aligned}$$

The inequality follows as in the proof of Proposition 3.2. The following equality is due to P and Q being stably equidistant from R.

Similarly, $d_n(S,R) < d_n(Q,R)$. □

Definition 3.13 *(Unstably equidistant). P and Q are unstably equidistant from R if they are equidistant from R but not stably equidistant from R.*

Proposition 3.14. *For fixed R,* unstably equidistant *is irreflexive, symmetric but not transitive in general. P and Q being unstably equidistant from R does* not *imply that any proper convex combination S of P and Q is closer than either P or Q to R. But a proper convex combination S of P and Q can be no further from R than P or Q. Nor can S and P (respectively S and Q) be stably equidistant from R.*

Proof. Irreflexivity: P and P are *stably* equidistant from R. Symmetry: P and Q are unstably equidistant from R if neither of $d_n(P,R)$ and $d_n(Q,R)$ dominate the other for sufficiently large n, nor are they equal for sufficiently large n; this is clearly symmetric. Failure of transitivity follows from Proposition 3.10.

We shall show that a convex combination of unstably equidistant probability functions need not be closer to R by providing a counterexample. Take $\mathcal{L}$ to have a single unary predicate, take R to be $P_=$, the equivocator, and define two families P_ζ and Q_ζ of probability functions parameterised by $\zeta \in \mathbb{N}$ as follows. Take the atomic propositions A_i to be probabilistically independent with respect to both P_ζ and Q_ζ. Given n let $l_\zeta(n)$ be such that $\zeta^{l_\zeta(n)} \leqslant n < \zeta^{l_\zeta(n)+1}$. Let

$$P_\zeta(A_n) = \begin{cases} \frac{1}{2} & l_\zeta(n) \text{ is odd}, \\ 1 & l_\zeta(n) \text{ is even}, \end{cases}$$
$$Q_\zeta(A_n) = \begin{cases} 1 & l_\zeta(n) \text{ is odd}, \\ \frac{1}{2} & l_\zeta(n) \text{ is even}. \end{cases}$$

We shall show first that for sufficiently large ζ, P_ζ and Q_ζ are unstably equidistant. We need to show that $d_i(P,P_=) > d_i(Q,P_=)$ for i in some infinite $I \subseteq \mathbb{N}$, and also that $d_j(P,P_=) < d_j(Q,P_=)$ for $j \in J \subseteq \mathbb{N}$ where J is infinite. Let $I = \{\zeta^{2k+1} - 1\colon\ k \in \mathbb{N}\}$ and $J = \{\zeta^{2k} - 1\colon\ k \in \mathbb{N}\}$.

Recall that, for any probability function T and for $m < n$, if we let

$$d'_n(T,P_=) = \sum_{\omega_n \in \Omega_n} T(\omega_n) \log\bigl(2^{n-m} T(\omega'_n | \omega_m)\bigr)$$

then

$$d_n(T,P_=) = d_m(T,P_=) + d'_n(T,P_=) \geqslant d_m(T,P_=) \geqslant 0$$

where ω_m is the m-state determined by ω_n and ω'_n is the remainder, i.e., ω_n is $\omega_m \wedge \omega'_n$. Also, $d_m(T,P_=) \leqslant m \log 2$.

Let $n \in I$ so that $n = \zeta^{2k+1} - 1$ for some k. Let $m = \zeta^{2k} - 1$, so $n - m = (\zeta - 1)\zeta^{2k}$. Consider the atomic states ω_m that P_ζ awards positive probability. P_ζ gives all these r states the same probability, $1/r$. Moreover, for each such state ω_m there is only one state with positive probability that extends it, namely that in which ω'_n is $A_{m+1} \wedge \cdots \wedge A_n$, i.e., ω_n is $\omega_m \wedge A_{m+1} \wedge \cdots \wedge A_n$. Hence

$$d'_n(P_\zeta,P_=) = \sum_{\omega_m} P_\zeta(\omega_m) \log 2^{n-m} = \frac{r}{r}(n-m)\log 2 = (\zeta - 1)\zeta^{2k} \log 2.$$

So $d_n(P_\zeta,P_=) \geqslant (\zeta - 1)\zeta^{2k} \log 2$ since $d_m(P_\zeta,P_=) \geqslant 0$. But $Q_\zeta(\omega'_n|\omega_m) = 2^{-(n-m)}$ so $d'_n(Q_\zeta,P_=) = 0$. Moreover, $d_m(Q_\zeta,P_=) \leqslant m \log 2 = (\zeta^{2k} - 1)\log 2$. Consequently $d_n(Q_\zeta,P_=) < \zeta^{2k} \log 2$. Hence if $\zeta \geqslant 2$, $d_n(P_\zeta,P_=) > d_n(Q_\zeta,P_=)$ for all $n \in I$. Similarly, $d_n(P_\zeta,P_=) < d_n(Q_\zeta,P_=)$ for $n \in J$. Thus P_ζ and Q_ζ are indeed unstably equidistant for large enough ζ.

Letting $S_\zeta = \frac{1}{2}P_\zeta + \frac{1}{2}Q_\zeta$, we will see that S_ζ and Q_ζ are unstably equidistant for sufficiently large ζ (similarly S_ζ and P_ζ). Again take $n \in I$ so that $n = \zeta^{2k+1} - 1$ for some k, and let $m = \zeta^{2k} - 1$, so $n - m = (\zeta - 1)\zeta^{2k}$. Define $\Omega_n^P = \{\omega_n \in \Omega_n\colon\ P_\zeta(\omega_n) > 0\}$ and $\Omega_n^Q = \{\omega_n \in \Omega_n\colon\ Q_\zeta(\omega_n) > 0\}$. It will be convenient to take Ω_n^P and Ω_n^Q to be disjoint for all n (this will be used in Eq. (1) below): we can achieve this by creating a dummy variable A_0 and stipulating that $P_\zeta(A_0) = 1$ and $Q_\zeta(A_0) = 0$, and that A_0 is independent of all other A_i with respect to both P_ζ and Q_ζ. Note that P_ζ assigns equal probability to each member of Ω_n^P and similarly for Q_ζ ((3) below). Also, if $\omega_n \in \Omega_n^P$ then $S_\zeta(\omega_n) = P_\zeta(\omega_n)/2 + Q_\zeta(\omega_n)/2 = P_\zeta(\omega_n)/2$; similarly, if $\omega_n \in \Omega_n^Q$ then $S_\zeta(\omega_n) = Q_\zeta(\omega_n)/2$ ((2) below). Now

$$\begin{aligned}
d'_n(S_\zeta, P_=) &= \sum_{\omega_n \in \Omega_n^P} S_\zeta(\omega_n) \log\bigl(2^{n-m} S_\zeta(\omega'_n|\omega_m)\bigr) + \sum_{\omega_n \in \Omega_n^Q} S_\zeta(\omega_n) \log\bigl(2^{n-m} S_\zeta(\omega'_n|\omega_m)\bigr) &(1)\\
&= \sum_{\omega_n \in \Omega_n^P} \frac{P_\zeta(\omega_n)}{2} \log\left(2^{n-m} \frac{P_\zeta(\omega_n)/2}{P_\zeta(\omega_m)/2}\right) + \sum_{\omega_n \in \Omega_n^Q} \frac{Q_\zeta(\omega_n)}{2} \log\left(2^{n-m} \frac{Q_\zeta(\omega_n)/2}{Q_\zeta(\omega_m)/2}\right) &(2)\\
&= \sum_{\omega_m \in \Omega_m^P} \frac{P_\zeta(\omega_m)}{2} \log\left(2^{n-m} \frac{P_\zeta(\omega_m)}{P_\zeta(\omega_m)}\right) + \sum_{\omega_n \in \Omega_m^Q} \frac{2^{-(n-m)} Q_\zeta(\omega_m)}{2} \log\left(2^{n-m} \frac{2^{-(n-m)} Q_\zeta(\omega_m)}{Q_\zeta(\omega_m)}\right)\\
&= \frac{1}{2} \log 2^{n-m} + \frac{2^{-(n-m)}}{2} \log\bigl(2^{n-m} 2^{-(n-m)}\bigr) &(3)\\
&= \frac{n-m}{2} \log 2\\
&= \frac{(\zeta-1)\zeta^{2k}}{2} \log 2.
\end{aligned}$$

Hence $d_n(S_\zeta, P_=) \geqslant \frac{1}{2}(\zeta-1)\zeta^{2k} \log 2$. But we saw above that $d_n(Q_\zeta, P_=) < \zeta^{2k} \log 2$. So for $n \in I$ and $\zeta \geqslant 3$, $d_n(S_\zeta, P_=) > d_n(Q_\zeta, P_=)$. Similarly when $n \in J$, $d_n(S_\zeta, P_=) < d_n(Q_\zeta, P_=)$. Therefore S_ζ, Q_ζ are unstably equidistant.

In sum, one can show that P and Q being unstably equidistant from R does *not* imply that any proper convex combination S of P and Q is closer than either P or Q to R, by taking $R = P_=$ and P, Q, S to be $P_\zeta, Q_\zeta, S_\zeta$ defined as above for any fixed $\zeta \geqslant 3$.

It remains to show that convex combination $S = \lambda P + (1-\lambda)Q$ of unstably equidistant P and Q is no further from R than P or Q, and neither are S, P (respectively S, Q) stably equidistant from R. First note that since P and Q are equidistant there are infinite sets I and J of natural numbers such that $d_n(P, R) \leqslant d_n(Q, R)$ if $n \in I$, and $d_n(P, R) \geqslant d_n(Q, R)$ if $n \in J$. Since P and Q are unstably equidistant they are distinct, so, as explained in the proof of Proposition 3.12, for sufficiently large n

$$d_n(S, R) < \lambda d_n(P, R) + (1-\lambda) d_n(Q, R).$$

Now if such $n \in I$, $d_n(S, R) < d_n(Q, R)$, so S is no further from R than Q, nor are S and Q stably equidistant from R. Similarly, for large enough $n \in J$, $d_n(S, R) < d_n(P, R)$, so the same is true of S and P. □

Corollary 3.15. *If $\mathbb{E}$ is a convex set of probability functions then so is $\downarrow\mathbb{E}$.*

Proof. If $P, Q \in \downarrow\mathbb{E}$ then P, Q, S must be (unstably) equidistant from the equivocator for any convex combination S of P and Q (since S cannot be closer to the equivocator by definition of $\downarrow\mathbb{E}$, and neither can it be further from the equivocator by Proposition 3.14). So $S \in \downarrow\mathbb{E}$. □

While for a propositional language $\downarrow\mathbb{E}$ is a singleton (there any closed convex set of probability functions contains a single function that is closest to the equivocator), with a predicate language this need not be the case. We have seen that non-uniqueness can arise in the presence of unstably equidistant probability functions, as well as in the presence of infinite descending chains.[8] In this paper we make no presumption that $\downarrow\mathbb{E}$ is a singleton. Non-uniqueness of $P_\mathcal{E} \in \downarrow\mathbb{E}$ poses no conceptual difficulty, but in some applications of Bayesian epistemology—including the application to probabilistic logic considered in this paper—care needs to be taken to keep track of all the potential belief functions.[9]

4. Probabilistic logics

Broadly speaking there are three main kinds of probabilistic logic, or *progic* for short. In an *internal* progic, a logical language $\mathcal{L}$ includes function symbols that are interpreted as probability functions. One then performs standard logical inference in order to draw conclusions that say something about these probability functions. An internal progic—where probabilities are internal to the language—is thus useful for reasoning *about* uncertainty [6]. In contrast, in an *external* progic the logical language $\mathcal{L}$ does not involve probabilities but instead probabilities are ascribed to the propositions of $\mathcal{L}$—the probabilities are external to the language. One then performs probabilistic inference to draw conclusions that say something about the probabilities of propositions. An external progic is thus used for reasoning *under* uncertainty [16]. The third case

[8] Having said this, preliminary results indicate that $\downarrow\mathbb{E}$ is a singleton in the restricted environment of probabilistic logic on unary languages (Paris & Rad, personal communication).

[9] Of course uniqueness is much more rarely obtained with subjective Bayesian epistemology, the main rival to objective Bayesian epistemology.

is that of a *mixed* progic, which contains probabilities internal to the language and external to the language; a mixed progic can be used to reason about and under uncertainty.

In order to be fully general we shall consider a mixed progic in this paper. But for clarity of exposition we shall consider a mixed progic of a fixed structure; the discussion of this paper can be generalised to other forms of mixed progic, or indeed particularised to internal and external progics.

Let $\mathcal{L}$ be a propositional or predicate language of the kind considered in previous sections. (Recall that we considered only predicate languages with no function symbols, so there are no probabilities internal to $\mathcal{L}$.) We will be primarily interested in inferences of the following form:

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^Y. \tag{4}$$

Here $\varphi_1, \ldots, \varphi_n, \psi$ are propositions of $\mathcal{L}$, and $X_1, \ldots, X_n, Y$ are sets of probabilities that attach to these propositions. $\mathrel{|\!\!\approx}$ is an unspecified entailment relation—in Section 5 we will give $\mathrel{|\!\!\approx}$ an objective Bayesian interpretation.

Thus far the probabilities are external to $\mathcal{L}$. Note however that the entailment relation $\mathrel{|\!\!\approx}$ relates propositions together with attached probabilities. So the premisses and conclusion are expressions in some language richer than $\mathcal{L}$. Accordingly consider a propositional language $\mathcal{L}^\sharp$ whose propositional variables are expressions of the form φ^X where φ is a proposition of $\mathcal{L}$ and X is a set of probabilities. (Note that as it stands there are uncountably many propositional variables, though one could always circumscribe the language $\mathcal{L}^\sharp$ to reduce its cardinality.) By applying the usual connectives of propositional logic, we can generate propositions μ of $\mathcal{L}^\sharp$ of arbitrary logical complexity. The metalanguage $\mathcal{L}^\sharp$ can be thought of as the language of the entailment relation $\mathrel{|\!\!\approx}$, and this entailment relation can be extended to inferences of the more general form

$$\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \nu \tag{5}$$

where $\mu_1, \ldots, \mu_n, \nu$ are arbitrary propositions of $\mathcal{L}^\sharp$. Here the probabilities are internal to the language $\mathcal{L}^\sharp$, though they are external to the language $\mathcal{L}$. Thus the entailment relation $\mathrel{|\!\!\approx}$ yields a mixed progic.

Clearly Eqs. (4) and (5) might be interpreted in a variety of ways according to one's understanding of the uncertainties involved. In Section 5 we shall develop an interpretation that is motivated by objective Bayesian epistemology. Then we shall turn to the practicalities of inference in probabilistic logic.

Classical logic faces an inferential question of the following form: given premisses and conclusion, do the premisses entail the conclusion? A proof theory is normally invoked to answer this kind of question. But a progic of the sort outlined above more typically faces a question of a rather different form: given premisses $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$ and a conclusion proposition ψ of $\mathcal{L}$, what set of probabilities Y should attach to ψ? This question can be represented thus:

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^?. \tag{6}$$

More generally, such a question may take the form:

$$\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \psi^? \tag{7}$$

where $\mu_1, \ldots, \mu_n$ are arbitrary propositions of $\mathcal{L}^\sharp$ and as before ψ is a proposition of $\mathcal{L}$. Since these are primarily questions about probability rather than logic, answering such a question will tend to invoke techniques from probability theory rather than proof theory. In Section 6 we shall introduce the machinery of *objective Bayesian nets* that, as explained in Section 7, can be invoked to answer such a question.

5. Objective Bayesian semantics

It is possible to interpret a question of the form of Eq. (6) by appealing to objective Bayesian epistemology as follows.

First interpret $\mathcal{L}$ as the language of an agent, and the left-hand side of Eq. (6) as characterising her evidence. Thus the premisses $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$ are construed as facts about empirical probability determined by the agent's evidence $\mathcal{E}$: a premiss of the form φ^X is interpreted as $P^*(\varphi) \in X$, where P^* is empirical probability, and the evidence $\mathcal{E}$ is taken to imply that $P^*(\varphi_1) \in X_1, \ldots, P^*(\varphi_n) \in X_n$.[10] The Calibration principle then deems the set $\mathbb{E}$ of probability functions that are compatible with this evidence to be the convex closure of the set $\mathbb{P}^* = \{P \in \mathbb{P}_{\mathcal{L}}\colon\ P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n\}$, i.e., $\mathbb{E} = [\mathbb{P}^*]$.[11]

[10] One may of course be sceptical about the existence of a single-case empirical probability function P^* since any characterisation of P^* will need to overcome the metaphysical reference class problem. But, as argued in [21], one can cash out the Calibration principle in terms of general-case frequency instead. The formulation of the Calibration principle will be more complicated in this case—it will need to tackle a more benign epistemological version of the reference class problem for instance—but the details need not detain us here, since here we are simply given the constraints in the formulation of Eq. (6).

[11] Note that statistical theory may be required to draw out the consequences of $\mathcal{E}$ for P^*. For example, if $\mathcal{E}$ contains evidence that the first hundred observed ravens were black: $B(t_1), \ldots, B(t_{100})$, together with certain statistical hypotheses about P^* that are granted by the agent, then this evidence may imply that $P^*(B(t_i))$ is close to 1 for all $i > 100$. This consequence of $\mathcal{E}$ in turn imposes constraints on the agent's degrees of belief. According to this point of view, learning from experience is a result of evidential considerations rather than of probabilistic logic itself. See [27, §5] for the justification of this point of view, and [3,22,24] for the opposing view, i.e., for attempts to integrate learning from experience into the agent's belief function directly rather than indirectly via her evidence.

We can now go on to interpret Eq. (6) as asking how strongly an agent whose evidence is characterised by the premisses should believe the conclusion proposition ψ. The objective Bayesian answer is that the agent's degree of belief in ψ should be representable by the probability $P_{\mathcal{E}}(\psi)$ where the agent's belief function $P_{\mathcal{E}} \in \downarrow\mathbb{E}$, i.e., $P_{\mathcal{E}}$ is a maximally equivocal probability function from all those compatible with the premisses. Thus the optimal set Y of probabilities to attach to the conclusion proposition is $Y = \{P_{\mathcal{E}}(\psi)\colon\ P_{\mathcal{E}} \in \downarrow[\mathbb{P}^*]\}$.

It is straightforward to extend this approach to handle questions of the form of Eq. (7), $\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \psi^?$. As before the premisses are interpreted as statements about empirical probability and $\mathbb{P}^* = \{P \in \mathbb{P}_{\mathcal{L}}\colon\ \mu_1, \ldots, \mu_n \text{ hold}\}$. As before, the optimal answer to the question is $Y = \{P_{\mathcal{E}}(\psi)\colon\ P_{\mathcal{E}} \in \downarrow[\mathbb{P}^*]\}$.

In turn we can interpret an entailment claim of the form of Eq. (4) as holding if and only if an agent with evidence characterised by the premisses $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n}$ should believe the conclusion proposition ψ to some degree within the set Y. Thus $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\!\approx} \psi^Y$ iff $\{P_{\mathcal{E}}(\psi)\colon\ P_{\mathcal{E}} \in \downarrow[\mathbb{P}^*]\} \subseteq Y$. More generally we can consider an entailment claim of the form of Eq. (5)—i.e., $\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \nu$ where $\mu_1, \ldots, \mu_n, \nu$ are propositions of $\mathcal{L}^\sharp$—to hold if and only if the conclusion ν holds of the degrees of belief of an agent with evidence characterised by the premisses $\mu_1, \ldots, \mu_n$.

There is an important special case. We will call the set $\{\mu_1, \ldots, \mu_n\}$ of premiss propositions of $\mathcal{L}^\sharp$ *regular* iff it gives rise to a set $\mathbb{P}^*$ that is closed, convex and non-empty. Premisses are regular, for example, if they are mutually consistent, they involve propositions of $\mathcal{L}$ that are all quantifier-free, and they involve sets of probabilities that are all closed and convex. If the premisses are regular then $\mathbb{E} = \mathbb{P}^*$ and the set of probabilities to attach to the conclusion proposition is $Y = \{P_{\mathcal{E}}(\psi)\colon\ P_{\mathcal{E}} \in \downarrow\mathbb{P}^*\}$. This case arises naturally when the premisses are observations of empirical probabilities (i.e., $P^*(\varphi_i) = x_i$; the X_i are singletons) or estimates of empirical probabilities (i.e., $P^*(\varphi_i) \in [x_i - \varepsilon, x_i + \varepsilon]$; the X_i are closed intervals).

Example 5.1. Consider the question $\forall x Ux^{3/5} \mathrel{|\!\!\approx} Ut_1^?$. Objective Bayesian epistemology interprets the expression on the left-hand side as $P^*(\forall x Ux) = 3/5$; this premiss is regular, so $\mathbb{E} = [\mathbb{P}^*] = \mathbb{P}^* = \{P\colon\ P(\forall x Ux) = 3/5\}$. There is one function P in $\mathbb{E}$ that is closest to the equivocator, as described in Example 3.5. This function gives $P(Ut_i) = 4/5$ for each constant t_i. Hence the answer to the question is 4/5.

In the remainder of the paper we shall ignore cases in which premisses permit infinite descending chains with respect to $\prec$. As mentioned in Section 3, such cases trivialise in the sense that $\downarrow\mathbb{P}^* = \mathbb{P}^*$; hence they are of less interest.

Discussion

Having presented the objective Bayesian semantics, we now investigate some of the properties of the resulting entailment relation. The reader more interested in techniques for answering questions of the form of Eqs. (6) and (7) may wish to skip to the next section.

First note that $\mathrel{|\!\!\approx}$ is a genuine semantic entailment relation in the sense that there are suitable notions of *model* and *satisfies* such that $\mu_1, \ldots, \mu_n$ entails ν if and only if every model of $\mu_1, \ldots, \mu_n$ satisfies ν. For proposition μ of $\mathcal{L}^\sharp$ let $\mathbb{P}(\mu) \stackrel{\text{df}}{=} \{P \in \mathbb{P}_{\mathcal{L}}\colon\ \mu \text{ holds}\}$. A probability function P is said to *satisfy* μ if $P \in \mathbb{P}(\mu)$. Let the set of *models* of μ be defined by $\mathbb{M}(\mu) \stackrel{\text{df}}{=} \downarrow[\mathbb{P}(\mu)]$.[12] Then indeed under the objective Bayesian semantics $\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \nu$ iff every model of the left-hand side satisfies the right-hand side, i.e., iff $\mathbb{M}(\mu_1, \ldots, \mu_n) \subseteq \mathbb{P}(\nu)$.

Definition 5.2 *(Decomposable).* Let $\mathrel{|\!\!\approx}$ be an arbitrary entailment relation and let $\mathbb{M}(\mu)$ be the corresponding notion of a set of models of μ. $\mathrel{|\!\!\approx}$ is called *decomposable* iff for all propositions $\mu_1, \ldots, \mu_n$ of the language of the entailment relation, $\mathbb{M}(\mu_1, \ldots, \mu_n) = \mathbb{M}(\mu_1) \cap \cdots \cap \mathbb{M}(\mu_n)$.

Definition 5.3 *(Monotonic).* An entailment relation $\mathrel{|\!\!\approx}$ is *monotonic* iff for all propositions $\nu, \mu_1, \ldots, \mu_m, \ldots, \mu_n$ of the language of $\mathrel{|\!\!\approx}$ (where $m < n$), we have that $\mu_1, \ldots, \mu_m \mathrel{|\!\!\approx} \nu$ implies $\mu_1, \ldots, \mu_m, \ldots, \mu_n \mathrel{|\!\!\approx} \nu$.

Proposition 5.4. *A decomposable entailment relation is monotonic.*

Proof. Suppose that $\mathrel{|\!\!\approx}$ is decomposable and that $\mu_1, \ldots, \mu_m \mathrel{|\!\!\approx} \nu$. Now

$$\begin{aligned}\mathbb{M}(\mu_1, \ldots, \mu_n) &= \mathbb{M}(\mu_1) \cap \cdots \cap \mathbb{M}(\mu_n) \\ &\subseteq \mathbb{M}(\mu_1) \cap \cdots \cap \mathbb{M}(\mu_m) = \mathbb{M}(\mu_1, \ldots, \mu_m) \subseteq \mathbb{P}(\nu).\end{aligned}$$

So $\mu_1, \ldots, \mu_n \mathrel{|\!\!\approx} \nu$, as required. $\square$

[12] As before we take $[\emptyset] = \mathbb{P}_{\mathcal{L}}$.

Now entailment under the objective Bayesian semantics is nonmonotonic: $A_1^{[0,1]} \mathrel{|\!\!\approx} A_1^{\{0.5\}}$ but it is not the case that $A_1^{[0,1]}, A_1^{\{1\}} \mathrel{|\!\!\approx} A_1^{\{0.5\}}$, where A_1 is a propositional variable (respectively atomic proposition) of propositional (respectively predicate) language $\mathcal{L}$. Hence objective Bayesian entailment is not decomposable.

Although nonmonotonic, this notion of entailment still satisfies a number of interesting and useful properties. In order to characterise these properties we shall need to appeal to some notions that are common in the literature on nonmonotonic logics—see, e.g., [12], [13, §3.2] and [7] for background. A probability function P on $\mathcal{L}$ can be construed as a valuation on $\mathcal{L}^\sharp$: P assigns the value True to proposition μ of $\mathcal{L}^\sharp$ if $P \in \mathbb{P}(\mu)$, i.e., if P satisfies μ. Define a decomposable entailment relation $\models$ by $\mathbb{M}_{\models}(\mu) = \mathbb{P}(\mu)$ for proposition μ of $\mathcal{L}^\sharp$. In particular, if μ is of the form φ^X, then $P \in \mathbb{M}_{\models}(\mu)$ iff $P(\varphi) \in X$. Recall that for probability functions P and Q, $P \prec Q$ iff P is closer to the equivocator than Q. Now $(\mathbb{P}_{\mathcal{L}}, \prec, \models)$ is a *preferential model*: $\mathbb{P}_{\mathcal{L}}$ is a set of valuations on $\mathcal{L}^\sharp$, $\prec$ is an irreflexive, transitive relation over $\mathbb{P}_{\mathcal{L}}$, and $\models$ is a decomposable entailment relation. Moreover, this preferential model is *smooth*: if $P \in \mathbb{M}_{\models}(\mu)$ then either P is minimal with respect to $\prec$ in $\mathbb{P}(\mu)$ or there is a $Q \prec P$ in $\mathbb{P}(\mu)$ that is minimal. Hence this model determines a *preferential consequence relation* $\mathrel{|\!\!\sim}$ as follows: $\mu \mathrel{|\!\!\sim} \nu$ iff P satisfies ν for every $P \in \mathbb{P}_{\mathcal{L}}$ that is minimal among those probability functions that satisfy μ. Wherever $\{\mu, \nu\}$ is regular, $\mathrel{|\!\!\sim}$ will agree with $\mathrel{|\!\!\approx}$. Consequently on regular propositions $\mathrel{|\!\!\approx}$ will satisfy the properties of preferential consequence relations, often called *system-P* properties—see, e.g., [12]:

Proposition 5.5 *(Properties of entailment). Let $\models$ denote entailment in classical logic and let $\equiv$ denote classical logical equivalence. Whenever $\{\mu, \nu, \xi\}$ is regular,*

Right Weakening: *if $\mu \mathrel{|\!\!\approx} \nu$ and $\nu \models \xi$ then $\mu \mathrel{|\!\!\approx} \xi$.*
Left Classical Equivalence: *if $\mu \mathrel{|\!\!\approx} \nu$ and $\mu \equiv \xi$ then $\xi \mathrel{|\!\!\approx} \nu$.*
Cautious Monotony: *if $\mu \mathrel{|\!\!\approx} \nu$ and $\mu \mathrel{|\!\!\approx} \xi$ then $\mu \wedge \xi \mathrel{|\!\!\approx} \nu$.*
Premiss Disjunction: *if $\mu \mathrel{|\!\!\approx} \nu$ and $\xi \mathrel{|\!\!\approx} \nu$ then $\mu \vee \xi \mathrel{|\!\!\approx} \nu$.*
Conclusion Conjunction: *if $\mu \mathrel{|\!\!\approx} \nu$ and $\mu \mathrel{|\!\!\approx} \xi$ then $\mu \mathrel{|\!\!\approx} \nu \wedge \xi$.*

6. Objective Bayesian nets

The calculus for probabilistic logic developed in Section 7 will appeal to the concepts of *objective Bayesian net* and *objective credal net*, which will be introduced in this section.

A *Bayesian net* is a representation of a probability function. Bayesian nets were developed to represent probability functions defined over sets of variables [14,18], but they can equally be applied to represent probability functions over logical languages, which are our concern here. Let $\mathcal{L}_n$ be a propositional language on the propositional variables $A_1, \ldots, A_{r_n}$ (in which case $r_n = n$), or a finite predicate language with atomic propositions $A_1, \ldots, A_{r_n}$. Suppose P is a probability function defined on $\mathcal{L}_n$. A Bayesian net representation of P consists of a directed acyclic graph whose nodes are $A_1, \ldots, A_{r_n}$, together with the conditional probability distribution $P(A_i|Par_i)$, induced by P, of each A_i conditional on its parents Par_i in the graph. The graph must be constructed in such a way that the following condition holds:

Markov Condition. Each A_i is probabilistically independent of its non-descendants in the graph, conditional on its parents, written $A_i \perp\!\!\!\perp ND_i \mid Par_i$.

The Bayesian net determines P via the identity

$$P(\omega_n) = \prod_{i=1}^{r_n} P\left(A_i^{\omega_n} | Par_i^{\omega_n}\right)$$

where $A_i^{\omega_n}$ and $Par_i^{\omega_n}$ are the states of A_i and its parents that are consistent with ω_n.

An *objective Bayesian net* (or *obnet* for short) is a Bayesian net representation of a function $P_{\mathcal{E}}$ on $\mathcal{L}_n$ that, according to objective Bayesian epistemology, represents appropriate degrees of belief for an agent with language $\mathcal{L}_n$ and evidence $\mathcal{E}$. An obnet can be constructed by (i) determining independencies that must be satisfied by $P_{\mathcal{E}}$, (ii) representing these independencies by a directed acyclic graph that satisfies the Markov Condition, and then (iii) determining the conditional distributions $P_{\mathcal{E}}(A_i|Par_i)$.

Task (i) is straightforward as we shall now see:

Definition 6.1 *(Constraint graph).* The *constraint graph* for $\mathcal{E}$ on $\mathcal{L}_n$ is constructed by taking the $A_1, \ldots, A_{r_n}$ as nodes and linking two nodes with an edge if they occur in the same constraint in $\mathcal{E}$. (In the case in which $\mathcal{L}_n$ is a predicate language and constraints involve quantifiers, first substitute each occurrence of $\forall x\theta(x)$ by $\bigwedge_t \theta(t)$, where the conjunction ranges over k-tuples t of constants from $t_1, \ldots, t_n$, k being the arity of the tuple x of variables; similarly substitute each occurrence of $\exists x\theta(x)$ by $\bigvee_t \theta(t)$.)

Proposition 6.2. *For all $P_{\mathcal{E}} \in \downarrow\mathbb{E}$, separation in the constraint graph implies conditional independence in $P_{\mathcal{E}}$: for $X, Y, Z \subseteq \{A_1, \ldots, A_n\}$, if Z separates X from Y in the constraint graph then $X \perp\!\!\!\perp Y \mid Z$ for $P_{\mathcal{E}}$.*

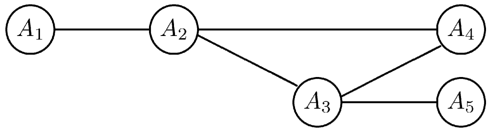

Fig. 1. Constraint graph.

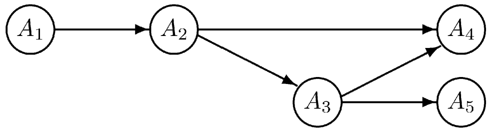

Fig. 2. Graph satisfying the Markov Condition.

Proof. See [21, Theorem 5.1]. □

For example, suppose the evidence $\mathcal{E}$ takes the form: $P^*(A_1 \wedge \neg A_2) \in [0.8, 0.9]$, $P^*((\neg A_4 \vee A_3) \rightarrow A_2) = 0.2$, $P^*(A_5 \vee A_3) \in [0.3, 0.6]$, $P^*(A_4) = 0.7$. Then the constraint graph of Fig. 1 represents conditional independencies that must be satisfied by any maximally equivocal function compatible with this evidence.

Step (ii), representing these independencies by a directed acyclic graph, can be performed by the following algorithm[13]:

Algorithm 6.3.

Input: An undirected graph $\mathcal{G}$.

- Triangulate $\mathcal{G}$ to give $\mathcal{G}^T$.
- Reorder the variables according to maximum cardinality search.
- Let $D_1, \ldots, D_l$ be the cliques of $\mathcal{G}^T$, ordered according to highest labelled vertex.
- Let $E_j = D_j \cap (\bigcup_{i=1}^{j-1} D_i)$ and $F_j = D_j \backslash E_j$, for $j = 1, \ldots, l$.
- Construct a directed acyclic graph $\mathcal{H}$ by taking propositional variables as nodes, and
  - Add an arrow from each vertex in E_j to each vertex in F_j, for $j = 1, \ldots, l$.
  - Add further arrows, from lower numbered variables to higher numbered variables, to ensure that there is an arrow between each pair of vertices in D_j, $j = 1, \ldots, l$.

Output: A directed acyclic graph $\mathcal{H}$.

When input a constraint graph this algorithm produces a directed acyclic graph that satisfies the Markov Condition with respect to any maximally equivocal function compatible with this evidence [21, Theorem 5.3]. A graph produced by this algorithm typically looks much like the constraint graph that is input: e.g., when input Fig. 1, Fig. 2 is one possible output of the algorithm.

Step (iii) in the construction of an obnet is the determination of the conditional distributions $P(A_i|Par_i)$. Here any of a variety of techniques (for instance, numerical methods or Lagrange multiplier methods) can be used to determine the values of these parameters that maximise entropy subject to constraints imposed by evidence. As pointed out above, on a *finite* language $\mathcal{L}_n$ there is a unique entropy maximiser. Hence these conditional distributions are uniquely determined. We will not explore the determination of these conditional distributions in any detail in this paper; the importance of a Bayesian net lies in the graph rather than the numerical parameters: by capturing probabilistic independencies graphically, they can be exploited to render representing and reasoning with probability more tractable.

Bayesian nets and objective Bayesian nets are thus far defined over a finite language $\mathcal{L}_n$. But in this paper we also consider the case in which the agent's language $\mathcal{L}$ is an infinite predicate language. Now as we saw in Section 3, an infinite language $\mathcal{L}$ can be handled naturally by considering the sequence $\mathcal{L}_1, \mathcal{L}_2, \ldots$ of finite languages generated by the ordering of the constant symbols: each $\mathcal{L}_n$ involves only constants $t_1, \ldots, t_n$. A probability function P over $\mathcal{L}$ is determined by its values over the $\mathcal{L}_n$ ($n = 1, 2, \ldots$) so it is determined by the sequence of Bayesian nets representing P over the $\mathcal{L}_n$ ($n = 1, 2, \ldots$). These Bayesian nets can be constructed in such a way that each net contains its predecessor in the sequence as a strict subnet; hence we can define the corresponding *infinitary* Bayesian net to be the Bayesian net on the infinite domain $A_1, A_2, \ldots$ that has each member of the sequence of (finite) Bayesian nets as a subnet. In turn, then, $P_{\mathcal{E}} \in \downarrow\mathbb{E}$ may be represented by a sequence of obnets, or an infinitary obnet. In practice, of course, one will at any stage only be interested in propositions involving the symbols of some fixed $\mathcal{L}_n$, so one can restrict attention to the finite subnet on that particular $\mathcal{L}_n$.

A second consideration arises with an infinite predicate language $\mathcal{L}$: when the agent's language is finite $\downarrow\mathbb{E}$ will be a singleton, but this does not hold in general in the infinite case. So while a single obnet suffices to represent $\downarrow\mathbb{E}$ in the finite case, a single (infinitary) obnet may not suffice to represent $\downarrow\mathbb{E}$ in the infinite case. Note, however, that thanks to Proposition 6.2 all members of $\downarrow\mathbb{E}$ may be represented by (infinitary) obnets on the same graph. Furthermore, $\downarrow\mathbb{E}$ is convex (Corollary 3.15). These two facts imply that $\downarrow\mathbb{E}$ can be represented by a *credal net*. A credal net, like a Bayesian net, contains a directed acyclic graph. But where a Bayesian net uniquely specifies the conditional probability distributions, a credal net only narrows down conditional probabilities to closed intervals. So the credal net can be thought of as the set of Bayesian

[13] See, e.g., [14] for the graph theoretic terminology.

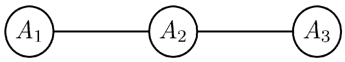

Fig. 3. Constraint graph.

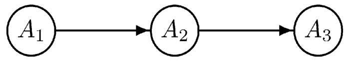

Fig. 4. Graph satisfying the Markov Condition.

nets that have the graph of the credal net and whose conditional probability distributions satisfy the constraints in the credal net. The credal net can thus be used to represent the set of probability functions that are determined by these associated Bayesian nets.[14] An *objective credal net* (or *ocnet* for short) is a credal net that represents the set $\downarrow\mathbb{E}$ that is of fundamental interest to objective Bayesianism. In the case in which $\mathcal{L}$ is an infinite predicate language, the ocnet will be infinitary and may be non-trivial in the sense that it represents more than one probability distribution. The graph of this infinitary ocnet can be determined by Definition 6.1 and Algorithm 6.3.

7. A calculus for probabilistic logic

We now have all the tools we need to say how questions of the form of Eq. (6), and more generally Eq. (7), can be answered.

The general strategy is as follows. We interpret a question of the form $\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mathrel{|\!\approx} \psi^?$ by means of the objective Bayesian semantics of Section 5. As we saw, this gives an answer in principle to the above question: the objective Bayesian semantics attaches the set $Y = \{P_{\mathcal{E}}(\psi)\colon\ P_{\mathcal{E}} \in \downarrow[\mathbb{P}^*]\}$ to the conclusion proposition ψ, where $\mathbb{P}^* = \{P\colon\ P(\varphi_1) \in X_1, \ldots, P(\varphi_n) \in X_n\}$. The question remains as to how to determine Y in practice. We answer this question by first representing $\mathbb{E} = \downarrow[\mathbb{P}^*]$ by an objective credal net, and then using this net to calculate the probability interval Y that attaches to ψ.

Example 7.1. Suppose we have a question of the form:

$$A_1 \wedge \neg A_2{}^{[0.8,0.9]}, (\neg A_4 \vee A_3) \rightarrow A_2{}^{0.2}, A_5 \vee A_3{}^{[0.3,0.6]}, A_4{}^{0.7} \mathrel{|\!\approx} A_5 \rightarrow A_1{}^?.$$

This is short for the following question: given that $A_1 \wedge \neg A_2$ has probability between 0.8 and 0.9 inclusive, $(\neg A_4 \vee A_3) \rightarrow A_2$ has probability 0.2, $A_5 \vee A_3$ has probability in $[0.3, 0.6]$ and A_4 has probability 0.7, what probability should $A_5 \rightarrow A_1$ have? As explained in Section 5, this question can be given an objective Bayesian interpretation: supposing the agent's evidence says $P^*(A_1 \wedge \neg A_2) \in [0.8, 0.9]$, $P^*((\neg A_4 \vee A_3) \rightarrow A_2) = 0.2$, $P^*(A_5 \vee A_3) \in [0.3, 0.6]$, $P^*(A_4) = 0.7$, how strongly should she believe $A_5 \rightarrow A_1$? Now an objective credal net can be constructed to answer this question (in fact the net is an objective Bayesian net: $\mathcal{L}$ is a finite propositional language so $\downarrow\mathbb{E}$ is a singleton). First construct undirected constraint graph Fig. 1 by linking variables that occur in the same constraint. Next, follow Algorithm 6.3 to transform the undirected graph into a directed acyclic graph satisfying the Markov Condition, such as Fig. 2. The third step is to maximise entropy to determine the probability distribution of each variable conditional on its parents in the directed graph. This yields the objective Bayesian net. Finally we use the net to calculate the probability of the conclusion

$$\begin{aligned} P(A_5 \rightarrow A_1) &= P(\neg A_5 \wedge A_1) + P(A_5 \wedge A_1) + P(\neg A_5 \wedge \neg A_1) \\ &= P(A_1) + P(\neg A_5|\neg A_1)\big(1 - P(A_1)\big). \end{aligned}$$

Thus we must calculate $P(A_1)$ and $P(\neg A_5|\neg A_1)$ from the net, which can be done using standard Bayesian or credal net inference algorithms.

Here is a simple example of a predicate language question:

Example 7.2. Suppose we have a question of the form:

$$\forall x(Ux \rightarrow Vx)^{3/5},\ \forall x(Vx \rightarrow Wx)^{3/4},\ Ut_1{}^{[0.8,1]} \mathrel{|\!\approx} Wt_1^?.$$

Again, an objective Bayesian net can be constructed to answer this question. There is only one constant symbol t_1, so we can focus on a finite predicate language $\mathcal{L}_1$. Let A_1 be Ut_1, A_2 be Vt_1 and A_3 be Wt_1. Then the constraint graph $\mathcal{G}$ is depicted in Fig. 3 and the corresponding directed acyclic graph $\mathcal{H}$ is depicted in Fig. 4. It is not hard to see that $P(A_1) = 4/5$, $P(A_2|A_1) = 3/4$, $P(A_2|\neg A_1) = 1/2$, $P(A_3|A_2) = 5/6$, $P(A_3|\neg A_2) = 1/2$; together with $\mathcal{H}$, these probabilities yield a Bayesian network. Standard inference methods then give us $P(A_3) = 11/15$ as an answer to our question.

The general procedure for determining Y can be represented by the following high-level algorithm:

[14] This set of probability functions is sometimes called the *complete extension* of the credal net. The *strong extension* of a credal net is the convex closure of the complete extension. In our case the two coincide: $\downarrow\mathbb{E}$ can be represented by the complete extension of a credal net; since $\downarrow\mathbb{E}$ is closed and convex, this is also the strong extension.

Algorithm 7.3.
Input: A question of the form $\mu_1, \dots, \mu_n \approx\!\!\!\mid \psi^?$, where $\mu_1, \dots, \mu_n$ are propositions of $\mathcal{L}^\sharp$ and ψ is a proposition of $\mathcal{L}$.[15]

1. If $\mathcal{L}$ is an infinite predicate language let n be the smallest j such that $\mu_1, \dots, \mu_n$ are propositions of $\mathcal{L}_j^\sharp$ and ψ is a proposition of $\mathcal{L}_j$. Otherwise let $\mathcal{L}_n = \mathcal{L}$.
2. Construct the constraint graph $\mathcal{G}$ for constraints $\mu_1, \dots, \mu_n$ on $\mathcal{L}_n$ (Definition 6.1).
3. Transform this graph into a directed acyclic graph $\mathcal{H}$ by means of Algorithm 6.3.
4. Determine the corresponding conditional probability intervals for an objective credal net representation of $\downarrow\mathbb{E}$:
   (a) Determine the credal net on $\mathcal{H}$ that represents the functions, from those that satisfy the Markov Condition with respect to $\mathcal{H}$, that are in $\mathbb{E}$ (i.e., that satisfy the constraints $\mu_1, \dots, \mu_n$ where these are consistent) [5, Algorithm 9.1].
   (b) Use Monte-Carlo methods to narrow the intervals in this credal net to represent the probability functions, from those in the above net, that are closest to the equivocator.
5. Use this credal net to determine the interval Y that attaches to ψ (Algorithm 7.4 below).

Output: Approximate Y such that $\mu_1, \dots, \mu_n \approx\!\!\!\mid \psi^Y$.

In step 4(b), entropy is determined by the network parameters via the identity

$$H(P) = -\sum_{i=1}^{r_n} \sum_{\omega_i} \left(\prod_{A_j \in Anc_i} P\left(A_j^{\omega_i} | Par_j^{\omega_i}\right) \right) \log P\left(A_i^{\omega_i} | Par_i^{\omega_i}\right),$$

where the ω_i range over the states of $Anc_i \stackrel{\text{df}}{=} \{A_i$ and its ancestors in $\mathcal{H}\}$.

We now turn to step 5 of this algorithm: using the net to perform inference. In principle this step can be implemented in a variety of ways, but here we advocate a particular approach based on the following considerations. Inference with credal nets can be computationally much more complex than inference with Bayesian nets: the credal net may represent a convex set of probability functions with a large number of extremal points, each of which needs to be kept track of to perform inference. Hence it is natural to use numerical approximation methods, rather than exact inference methods, when working with credal nets. Moreover, computational complexity can also be mitigated by a *compilation* methodology, i.e., by trying to do much of the computational work during a computationally expensive compilation phase that is performed once offline, leaving a cheaper query answering phase that is performed online (perhaps many times, if answering several queries with the same premisses but different conclusion propositions ψ). Accordingly, hill-climbing numerical methods for compiled credal nets, developed in [4] and applied to probabilistic logic in [5, §8.2], are well suited. There is only space to outline these methods here—the reader is urged to consult these references for the details. These methods work perfectly well in the extreme case in which the credal net is in fact a Bayesian net; since we may not know in advance whether this case will arise, this is a distinct advantage.

In brief, step 5 of Algorithm 7.3 can be implemented by:

Algorithm 7.4.
Input: A credal net on $\mathcal{L}_n$ and a proposition ψ of $\mathcal{L}_n$.

- Compilation phase:
  – Transform the credal net into a d-DNNF (deterministic Decomposable Negation Normal Form) net.
- Inference phase:
  – Transform ψ into $\psi_1 \vee \dots \vee \psi_l$ where the ψ_i are mutually exclusive conjunctions of literals—e.g., via Abraham's algorithm [1].
  – Use hill-climbing in the d-DNNF net to calculate approximate bounds on each $P(\psi_i)$.
  – Calculate bounds on $P(\psi)$ via the identity $P(\psi) = \sum_{i=1}^{l} P(\psi_i)$.

Output: $Y = \{P(\psi)\colon\ P$ is subsumed by the input credal net$\}$.

We see then that, given a question of the form of Eq. (6) or Eq. (7), an objective credal net can be constructed to provide an answer. Of course, in the worst case logical inference and probabilistic inference are each computationally intractable (in the sense that computing answers takes time exponential in n), so their combination in a probabilistic logic is no more tractable. But that is the worst case; if the constraint graph is sparse, the objective credal net calculus offers the potential for efficient inference.

[15] We make no assumptions about $\mu_1, \dots, \mu_n$ here, other than that they admit no infinite descending chains (if they do then $\downarrow\mathbb{E} = \mathbb{E} = [\mathbb{P}^*]$ and step 1 is no longer appropriate). Recall that if the premisses are inconsistent then $\mathbb{E} = \mathbb{P}_{\mathcal{L}}$ and $\downarrow\mathbb{E} = \{P_=\}$; if they are not regular then $\mathbb{E} = [\mathbb{P}^*] \neq \mathbb{P}^*$.

8. Summary

Objective Bayesian epistemology offers a systematic account of the normative constraints on an agent's degrees of belief. This account is based on three norms: Probability, Calibration and Equivocation (Section 2). Note that objective Bayesianism is objective in the sense that these norms impose very strong constraints on degrees of belief. So strong, in fact, that there is no need to impose any further norm (such as Bayesian conditionalisation) for updating degrees of belief [27]. In contrast, under subjective Bayesian epistemology—which advocates only Probability and Calibration—degrees of belief are largely a question of personal choice and further constraints are required to guide updating. As we saw, though, objective Bayesian degrees of belief are not always uniquely determined—in some cases there remains an element of subjective choice. This is a point in common with the subjective approach and this distinguishes objective Bayesianism from the logical interpretation of probability [11], which construes probability to be a unique relation between propositions.

This account can be extended from the usual case of a propositional language to the case in which the agent's language is an infinite first-order predicate language (Section 3). The three norms carry over with minor modifications, and one can even formulate a version of the maximum entropy principle. Subtleties arise to do with closeness of probability functions, but the resulting theory does all that one might expect of it.

Objective Bayesian epistemology also provides a natural semantics (Section 5) for probabilistic logic (Section 4): simply interpret the premisses as evidence and the conclusion as a claim about rational degree of belief; the conclusion follows from the premisses if an agent with evidence explicated by the premisses ought to have degrees of belief that satisfy the conclusion.

In a probabilistic logic it is natural to ask which set of probabilities attaches to a conclusion proposition, given a set of premisses. Objective Bayesian epistemology provides the answer in theory, and an objective credal net (Section 6) can be used to calculate the answer in practice (Section 7).

Acknowledgments

This research was supported by the Leverhulme Trust. I am grateful to Jeff Paris, Soroush Rad and two anonymous referees for very helpful comments.

References

[1] J.A. Abraham, An improved algorithm for network reliability, IEEE Trans. Reliability 28 (1979) 58–61.
[2] O. Barnett, J. Paris, Maximum entropy inference with quantified knowledge, Log. J. IGPL 16 (2008) 85–98.
[3] R. Carnap, The Continuum of Inductive Methods, University of Chicago Press, Chicago, IL, 1952.
[4] R. Haenni, Climbing the hills of compiled credal networks, in: G. de Cooman, J. Vejnarová, M. Zaffalon (Eds.), ISIPTA'07, 5th International Symposium on Imprecise Probabilities and Their Applications, Prague, Czech Republic, 2007, pp. 213–222.
[5] R. Haenni, J.-W. Romeijn, G. Wheeler, J. Williamson, Probabilistic logic and probabilistic networks, in press; www.kent.ac.uk/secl/philosophy/jw/2006/progicnet.htm, 2008.
[6] J.Y. Halpern, Reasoning about Uncertainty, MIT Press, Cambridge, MA, 2003.
[7] J. Hawthorne, D.C. Makinson, The quantitative/qualitative watershed for rules of uncertain inference, Studia Logica 86 (2007) 247–297.
[8] C. Howson, P. Urbach, Scientific Reasoning: The Bayesian Approach, Open Court, Chicago IL, 1989; second 1993 edition.
[9] E.T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (4) (1957) 620–630.
[10] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, 2003.
[11] J.M. Keynes, A Treatise on Probability, Macmillan (1948), London, 1921.
[12] S. Kraus, D. Lehmann, M. Magidor, Nonmonotonic reasoning, preferential models and cumulative logics, Artificial Intelligence 44 (1990) 167–207.
[13] D. Makinson, Bridges from Classical to Nonmonotonic Logic, King's College Publications, London, 2005.
[14] R.E. Neapolitan, Probabilistic Reasoning in Expert Systems: Theory and Algorithms, Wiley, New York, 1990.
[15] N.J. Nilsson, Probabilistic logic, Artificial Intelligence 28 (1986) 71–87.
[16] J.B. Paris, The Uncertain Reasoner's Companion, Cambridge University Press, Cambridge, 1994.
[17] J.B. Paris, A. Vencovská, The emergence of reasons conjecture, J. Appl. Log. 1 (3–4) (2003) 167–195.
[18] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988.
[19] F.P. Ramsey, Truth and probability, in: H.E. Kyburg, H.E. Smokler (Eds.), Studies in Subjective Probability, Robert E. Krieger Publishing Company, Huntington, New York, 1926, pp. 23–52; second 1980 edition.
[20] R.D. Rosenkrantz, Inference, Method and Decision: Towards a Bayesian Philosophy of Science, Reidel, Dordrecht, 1977.
[21] J. Williamson, Bayesian Nets and Causality: Philosophical and Computational Foundations, Oxford University Press, Oxford, 2005.
[22] J. Williamson, Inductive influence, British J. Philos. Sci. 58 (4) (2007) 689–708.
[23] J. Williamson, Motivating objective Bayesianism: from empirical constraints to objective probabilities, in: W.L. Harper, G.R. Wheeler (Eds.), Probability and Inference: Essays in Honour of Henry E. Kyburg Jr., College Publications, London, 2007, pp. 151–179.
[24] J. Williamson, Objective Bayesianism with predicate languages, Synthese 163 (3) (2008) 341–356.
[25] J. Williamson, Philosophies of probability, in: A. Irvine (Ed.), Handbook of the Philosophy of Mathematics, in: Handbook of the Philosophy of Science, vol. 4, Elsevier, Amsterdam, 2008.
[26] J. Williamson, Epistemic complexity from an objective Bayesian perspective, La Nuova Critica (2008), in press.
[27] J. Williamson, Objective Bayesianism, Bayesian conditionalisation and voluntarism, Synthese (2009), in press.