EVIDENTIAL PROBABILITY AND OBJECTIVE BAYESIAN EPISTEMOLOGY

Gregory Wheeler and Jon Williamson

1 INTRODUCTION

Evidential probability (EP), developed by Henry Kyburg, offers an account of the impact of statistical evidence on single-case probability. According to this theory, observed frequencies of repeatable outcomes determine a probability interval that can be associated with a proposition. After giving a comprehensive introduction to EP in §2, in §3 we describe a recent variant of this approach, *second-order evidential probability* (2oEP). This variant, introduced in [Haenni *et al.*, 2010], interprets a probability interval of EP as bounds on the sharp probability of the corresponding proposition. In turn, this sharp probability can itself be interpreted as the degree to which one ought to believe the proposition in question.

At this stage we introduce objective Bayesian epistemology (OBE), a theory of how evidence helps determine appropriate degrees of belief (§4). OBE might be thought of as a rival to the evidential probability approaches. However, we show in §5 that they can be viewed as complimentary: one can use the rules of EP to narrow down the degree to which one should believe a proposition to an interval, and then use the rules of OBE to help determine an appropriate degree of belief from within this interval. Hence bridges can be built between evidential probability and objective Bayesian epistemology.

2 EVIDENTIAL PROBABILITY

2.1 Motivation

Rudolf Carnap [Carnap, 1962] drew a distinction between probability$_1$, which concerned rational degrees of belief, and probability$_2$, which concerned statistical regularities. Although he claimed that both notions of probability were crucial to scientific inference, Carnap practically ignored probability$_2$ in the development of his systems of inductive logic. Evidential probability (EP) [Kyburg, 1961; Kyburg and Teng, 2001], by contrast, is a theory that gives primacy to probability$_2$, and Kyburg's philosophical program was an uncompromising approach to see how far he could go with relative frequencies. Whereas Bayesianism springs from the view

Handbook of the Philosophy of Science. Volume 7: Philosophy of Statistics.
Volume editors: Prasanta S. Bandyopadhyay and Malcolm R. Forster. General Editors: Dov M. Gabbay, Paul Thagard and John Woods.

that probability$_1$ is all the probability needed for scientific inference, EP arose from the view that probability$_2$ is all that we really have.

The theory of evidential probability is motivated by two basic ideas: probability assessments should be based upon relative frequencies, to the extent that we know them, and the assignment of probability to specific individuals should be determined by everything that is known about that individual. Evidential probability is conditional probability in the sense that the probability of a sentence χ is evaluated given a set of sentences Γ_δ. But the evidential probability of χ given Γ_δ, written $\mathrm{Prob}(\chi, \Gamma_\delta)$, is a meta-linguistic operation similar in kind to the relation of provability within deductive systems.

The semantics governing the operator $\mathrm{Prob}(\cdot, \cdot)$ is markedly dissimilar to axiomatic theories of probability that take conditional probability as primitive, such as the system developed by Lester Dubbins [Dubbins, 1975; Arló-Costa and Parikh, 2005], and it also resists reduction to linear [de Finetti, 1974] as well as lower previsions [Walley, 1991]. One difference between EP and the first two theories is that EP is interval-valued rather than point-valued, because the relative frequencies that underpin assignment of evidential probability are typically incomplete and approximate. But more generally, EP assignments may violate coherence. For example, suppose that χ and φ are sentences in the object language of evidential probability. The evidential probability of $\chi \wedge \varphi$ given Γ_δ might fail to be less than or equal to the evidential probability that χ given Γ_δ.[1] A point to stress from the start is that evidential probability is a logic of statistical probability statements, and there is nothing in the activity of observing and recording statistical regularities that guarantees that a set of statistical probability statements will comport to the axioms of probability. So, EP is neither a species of Carnapian logical probability nor a kind of Bayesian probabilistic logic.[2,3] EP is instead a logic for approximate reasoning, thus it is more similar in kind to the theory of rough sets [Pawlak, 1991] and to systems of fuzzy logic [Dubois and Prade, 1980] than to probabilistic logic.

The operator $\mathrm{Prob}(\cdot, \cdot)$ takes as arguments a sentence χ in the first coordinate and a set of statements Γ_δ in the second. The statements in Γ_δ represent a knowledge base, which includes categorical statements as well as statistical generalities. Theorems of logic and mathematics are examples of categorical statements, but so too are contingent generalities. One example of a contingent categorical statement is the ideal gas law. EP views the propositions "$2 + 2 = 4$" and "$PV = nRT$"

[1] Specifically, the lower bound of $\mathrm{Prob}(\chi \wedge \varphi, \Gamma_\delta)$ may be strictly greater than the lower bound of $\mathrm{Prob}(\chi, \Gamma_\delta)$.

[2] See the essays by Levi and by Seidenfeld in [Harper and Wheeler, 2007] for a discussion of the sharp differences between EP and Bayesian approaches, particularly on the issue of conditionalization. A point sometimes overlooked by critics is that there are different *systems* of evidential probability corresponding to different conditions we assume to hold. Results pertaining to a qualitative representation of EP inference, for instance, assume that Γ_δ is consistent. A version of conditionalization holds in EP given that there is specific statistical statement pertaining to the relevant joint distribution. See [Kyburg, 2007] and [Teng, 2007].

[3] EP does inherit some notions from Keynes's [Keynes, 1921], however, including that probabilities are interval-valued and not necessarily comparable.

within a chemistry knowledge base as indistinguishable analytic truths that are built into a particular language adopted for handling statistical statements to do with gasses. In light of EP's expansive view of analyticity, the theory represents all categorical statements as universally quantified sentences within a guarded fragment of first-order logic [Andréka *et al.*, 1998].[4]

Statistical generalities within Γ_δ, by contrast, are viewed as direct inference statements and are represented by syntax that is unique to evidential probability. *Direct inference*, recall, is the probability assigned a target subclass given known frequency information about a reference population, and is often contrasted to *indirect inference*, which is the assignment of probability to a population given observed frequencies in a sample. Kyburg's ingenious idea was to solve the problem of indirect inference by viewing it as a form of direct inference. Since the philosophical problems concerning direct inference are much less contentious than those raised by indirect inference, the unusual properties and behavior of evidential probability should be weighed against this achievement [Levi, 2007].

Direct inference statements are statements that record the observed frequency of items satisfying a specified reference class that also satisfy a particular target class, and take the form of

$$\%\vec{x}(\tau(\vec{x}), \rho(\vec{x}), [l, u]).$$

This schematic statement says that given a sequence of propositional variables $\vec{x}$ that satisfies the reference class predicate ρ, the proportion of ρ that also satisfies the target class predicate τ is between l and u.

Syntactically, '$\tau(\vec{x}), \rho(\vec{x}), [l, u]$' is an open formula schema, where '$\tau(\cdot)$' and '$\rho(\cdot)$' are replaced by open first-order formulas, '$\vec{x}$' is replaced by a sequence of propositional variables, and '$[l, u]$' is replaced by a specific sub-interval of $[0, 1]$. The binding operator '%' is similar to the ordinary binding operators $(\forall, \exists)$ of first-order logic, except that '%' is a 3-place binding operator over the propositional variables appearing the *target formula* $\tau(\vec{x})$ and the *reference formula* $\rho(\vec{x})$, and binding those formulas to an interval.[5] The language $\mathcal{L}^{ep}$ of evidential probability then is a guarded first-order language augmented to include direct inference statements. There are additional formation rules for direct inference statements that are designed to block spurious inference, but we shall pass over these details of the theory.[6] An example of a direct inference statement that might appear in Γ_δ is

$$\%x(B(x), A(x), [.71, .83]),$$

which expresses that the proportion of As that are also Bs lies between 0.71 and 0.83.

As for semantics, a model M of $\mathcal{L}^{ep}$ is a pair, $\langle \mathcal{D}, \mathcal{I} \rangle$, where $\mathcal{D}$ is a two-sorted domain consisting of mathematical objects, $\mathcal{D}_m$, and a *finite* set of empirical objects, $\mathcal{D}_e$. EP assumes that there is a first giraffe and a last carbon molecule. $\mathcal{I}$ is

[4] A guarded fragment of first-order logic is a decidable fragment of first-order logic.

[5] Hereafter we relax notation and simply use an arbitrary variable 'x' for '$\vec{x}$'.

[6] See [Kyburg and Teng, 2001].

an interpretation function that is the union of two partial functions, one defined on $\mathcal{D}_m$ and the other on $\mathcal{D}_e$. Otherwise M behaves like a first-order model: the interpretation function $\mathcal{I}$ maps (empirical/mathematical) terms into the (empirical/mathematical) elements of $\mathcal{D}$, monadic predicates into subsets of $\mathcal{D}$, n-arity relation symbols into $\mathcal{D}^n$, and so forth. Variable assignments also behave as one would expect, with the only difference being the procedure for assigning truth to direct inference statements.

The basic idea behind the semantics for direct inference statements is that the statistical quantifier '%' ranges over the finite empirical domain $\mathcal{D}_e$, not the field terms l, u that denote real numbers in $\mathcal{D}_m$. This means that the only free variables in a direct inference statement range over a finite domain, which will allow us to look at proportions of models in which a sentence is true. A *satisfaction set* of an open formula φ whose only free n variables are empirical in the subset of $\mathcal{D}^n$ that satisfies φ.

A direct inference statement $\%x(\tau(x), \rho(x), [l, u])$ is true in M under variable assignment v iff the cardinality of the satisfaction sets for the open formula ρ under v is greater than 0 and the ratio of the cardinality of satisfaction sets for $\tau(x^*) \wedge \rho(x^*)$ over the cardinality of the satisfaction sets for $\rho(x)$ (under v) is in the closed interval $[l, u]$, where all variables of x occur in ρ, all variables of τ occur in ρ, and x^* is the sequence of variables free in ρ but not bound by $\%x$ [Kyburg and Teng, 2001].

The operator $\mathrm{Prob}(\cdot, \cdot)$ then provides a semantics for a nonmonotonic consequence operator [Wheeler, 2004; Kyburg *et al.*, 2007]. The structural properties enjoyed by this consequence operator are as follows:[7]

Properties of EP Entailment: Let $\models$ denote classical consequence and let $\equiv$ denote classical logical equivalence. Whenever $\mu \wedge \xi, \nu \wedge \xi$ are sentences of $\mathcal{L}^{ep}$,

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 $\xi \mathrel{|\!\approx} \nu$.

(KTW) Cautious Monotony: if $\mu \models \nu$ and $\mu \mathrel{|\!\approx} \xi$ then $\mu \wedge \xi \mathrel{|\!\approx} \nu$.

(KTW) Premise Disjunction: if $\mu \models \nu$ and $\xi \mathrel{|\!\approx} \nu$ then $\mu \vee \xi \mathrel{|\!\approx} \nu$.

(KTW) Conclusion Conjunction: if $\mu \models \nu$ and $\mu \mathrel{|\!\approx} \xi$ then $\mu \mathrel{|\!\approx} \nu \wedge \xi$.

As an aside, this qualitative EP-entailment relation presents challenges in handling disjunction in the premises since the KTW disjunction property admits a novel reversal effect similar to, but distinct from, Simpson's paradox [Kyburg *et al.*, 2007; Wheeler, 2007]. This raises a question over how best to axiomatize EP. One approach, which is followed by [Hawthorne and Makinson, 2007] and considered in

[7] Note that these properties are similar to, but strictly weaker than, the properties of the class of cumulative consequence relations specified by System P [Kraus *et al.*, 1990]. To yield the axioms of System P, replace the nonmonotonic consequence operator $\mathrel{|\!\sim}$ for $\models$ in the premise position of **[And*]**, **[Or*]**, and **[Cautious Monotonicity*]**.

[Kyburg *et al.*, 2007], is to replace Boolean disjunction by 'exclusive-or'. While this route ensures nice properties for $\not\approx$, it does so at the expense of introducing a dubious connective into the object language that is neither associative nor compositional.[8] Another approach explored in [Kyburg *et al.*, 2007] is a weakened disjunction axiom (KTW Or) that yields a sub-System P nonmonotonic logic and preserves the standard properties of the positive Boolean connectives.

Now that we have a picture of what EP is, we turn to consider the inferential behavior of the theory. We propose to do this with a simple ball-draw experiment before considering the specifics of the theory in more detail in the next section.

EXAMPLE 1. Suppose the proportion of white balls (W) in an urn (U) is known to be within $[.33, 4]$, and that ball t is drawn from U. These facts are represented in Γ_δ by the sentences, $\%x(W(x), U(x), [.33, .4])$ and $U(t)$.

(*i*) If these two statements are all that we know about t, i.e., they are the only statements in Γ_δ pertaining to t, then $\text{Prob}(W(t), \Gamma_\delta) = [.33, .4]$.

(*ii*) Suppose additionally that the proportion of plastic balls (P) that are white is observed to be between $[.31, .36]$, t is plastic, and that every plastic ball is a white ball. That means that $\%x(P(x), U(x), [.31, .36])$, $P(t)$, and $\forall x.P(x) \rightarrow W(x)$ are added to Γ_δ as well. Then there is conflicting statistical knowledge about t, since either:

1. the probability that ball t is white is between $[.33, .4]$, by reason of $\%x(W(x), U(x), [.33, .4])$, or
2. the probability that ball t is white is between $[.31, .36]$, by reason of $\%x(W(x), P(x), [.31, .36])$,

may apply. There are several ways that statistical statements may conflict and there are rules for handling each type, which we will discuss in the next section. But in this particular case, because it is known that the class of plastic balls is more *specific* than the class of balls in U and we have statistics for the proportion of plastic balls that are also white balls, the statistical statement in (2) dominates the statement in (1). So, the probability that t is white is in $[.31, .36]$.

(*iii*) Adapting an example from [Kyburg and Teng, 2001, 216], suppose U is partitioned into three cells, u_1, u_2, and u_3, and that the following compound experiment is performed. First, a cell of U is selected at random. Then a ball is drawn at random from that cell. To simplify matters, suppose that there are 25 balls in U and 9 are white such that 3 of 5 balls from u_1 are white, but only 3 of 10 balls in u_2 and 3 of 10 in u_3 are white. The following table summarizes this information.

[8] Example: 'A xor B xor C' is true if A, B, C are; and '(A xor B) xor C' is not equivalent to 'A xor (B xor C)' when A is false but B and C both true.

Table 1. Compound Experiment

	u_1	u_2	u_3	
W	3	3	3	9
$\overline{W}$	2	7	7	16
	5	10	10	25

We are interested in the probability that t is white, but we have a conflict. Given these over all precise values, we would have $\text{Prob}(W(t), \Gamma_\delta) = \frac{9}{25}$. However, since we know that t was selected by performing this compound experiment, then we also have the conflicting direct inference statement $\%x, y(W^*(x, y), U^*(x, y), [.4, .4])$, where U^* is the set of compound two stage experiments, and W^* is the set of outcomes in which the ball selected is white.[9] We should prefer the statistics from the compound experiment because they are *richer* in information. So, the probability that t is white is .4.

(*iv*) Finally, if there happens to be *no* statistical knowledge in Γ_δ pertaining to t, then we would be completely ignorant of the probability that t is white. So in the case of total ignorance, $\text{Prob}(W(t), \Gamma_\delta) = [0, 1]$.

We now turn to a more detailed account of how EP calculates probabilities.

2.2 Calculating Evidential Probability

In practice an individual may belong to several reference classes with known statistics. Selecting the appropriate statistical distribution among the class of potential probability statements is the *problem of the reference class*. The task of assigning evidential probability to a statement χ relative to a set of evidential certainties relies upon a procedure for eliminating excess candidates from the set of potential candidates. This procedure is described in terms of the following definitions.

Potential Probability Statement: A *potential probability statement* for χ with respect to Γ_δ is a tuple $\langle t, \tau(t), \rho(t), [l, u]\rangle$, such that instances of $\chi \leftrightarrow \tau(t)$, $\rho(t)$, and $\%x(\tau(x), \rho(x), [l, u])$ are each in Γ_δ.

Given χ, there are possibly many target statements of form $\tau(t)$ in Γ_δ that have the same truth value as χ. If it is known that individual t satisfies ρ, and known that between .7 and .8 of ρ's are also τ's, then $\langle t, \tau(t), \rho(t), [.7, .8]\rangle$ represents a potential probability statement for χ based on the knowledge base Γ_δ. Our focus

[9] Γ_δ should also include the categorical statements $\forall x, y(U^*\langle x, y\rangle \rightarrow W(y))$, which says that the second stage of U concerns the proportion of balls that are white, and three statements of the form $\ulcorner W^*(\mu, t) \leftrightarrow W(t)\urcorner$, where μ is replaced by u_1, u_2, u_3, respectively. This statement tells us that everything that's true of W^* is true of W, which is what ensures that this conflict is detected.

will be on the statistical statements $\%x(\tau(x), \rho(x), [l, u])$ in Γ_δ that are the basis for each potential probability statement.

Selecting the appropriate probability interval for χ from the set of potential probability statements reduces to identifying and resolving conflicts among the statistical statements that are the basis for each potential probability statement.

Conflict: Two intervals $[l, u]$ and $[l', u']$ *conflict* iff neither $[l, u] \subset [l', u']$ nor $[l, u] \supset [l', u']$. Two statistical statements conflict iff their intervals conflict.
Note that conflicting intervals may be disjoint or intersect. For technical reasons an interval is said to conflict with itself.

Cover: Let X be a set of intervals. An interval $[l, u]$ *covers* X iff for every $[l', u'] \in X$, $l \leq l'$ and $u' \leq u$. A cover $[l, u]$ of X is the *smallest cover*, $Cov(X)$, iff for all covers $[l^*, u^*]$ of X, $l^* \leq l$ and $u \leq u^*$.

Difference Set: (i) Let X be a non-empty set of intervals and $\mathcal{P}(X)$ be the powerset of X. A non-empty $Y \in \mathcal{P}(X)$ is *a difference set of* X iff Y includes every $x \in X$ that conflicts with some $y \in Y$. (ii) Let X be the set of intervals associated with a set Γ of statistical statements, and Y be the set of intervals associated with a set Λ of statistical statements. Λ is a difference set to Γ iff Y is closed under difference with respect to X.

EXAMPLE 2. An example might help. Let X be the set of intervals $[.30, .40]$, $[.35, .45]$, $[.325, .475]$, $[.50, .55]$, $[.30, .70]$, $[.20, .60]$, $[.10, .90]$. There are three sets closed under difference with respect to X:

(i) $\{[.30, .40], [.35, .45], [.325, .475], [.50, .55]\}$,

(ii) $\{[.30, .70], [.20, .60]\}$,

(iii) $\{[.10, .90]\}$.

The intuitive idea behind a difference set is to eliminate intervals from a set that are broad enough to include all other intervals in that set. The interval $[.10, .90]$ is the broadest interval in X. So, it only appears as a singleton difference set and is not included in any other difference set of X. It is not necessary that all intervals in a difference set X be pairwise conflicting intervals. Difference sets identify the set of all possible conflicts for each potential probability statement in order to find that conflicting set with the shortest cover.

Minimal Cover Under Difference: (i) Let X be a non-empty set of intervals and $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ the set of all difference sets of X. The *minimal cover under difference* of X is the smallest cover of the elements of $\mathcal{Y}$, i.e., the shortest cover in $\{Cov(Y_1), \ldots, Cov(Y_n)\}$.

(ii) Let X be the set of intervals associated with a set Γ of statistical statements, and $\mathcal{Y}$ be the set of all difference sets of X associated with a set Λ of statistical statements. Then the minimal cover under difference of Γ is the minimal cover under difference of X.

EP resolves conflicting statistical data concerning χ by applying two principles to the set of potential probability assignments, *Richness* and *Specificity*, to yield

a class of *relevant statements*. The (controversial) principle of *Strength* is then applied to this set of relevant statistical statements, yielding a unique probability interval for χ. For discussion of these principles, see [Teng, 2007].

We illustrate these principles in terms of a pair (φ, ϑ) of conflicting statistical statements for χ, and represent their respective reference formulas by ρ_φ and ρ_ϑ. The probability interval assigned to χ is the shortest cover of the relevant statistics remaining after applying these principles.

1. **[Richness]** If φ and ϑ conflict and ϑ is based on a marginal distribution while φ is based on the full joint distribution, eliminate ϑ.

2. **[Specificity]** If φ and ϑ both survive the principle of richness, and if $\rho_\varphi \subset \rho_\vartheta$, then eliminate $\langle \tau, \rho_\vartheta, [l, u] \rangle$ from *all* difference sets.

The principle of specificity says that if it is known that the reference class ρ_φ is included in the reference class ρ_ϑ, then eliminate the statement ϑ. The statistical statements that survive the sequential application of the principle of richness followed by the principle of specificity are called *relevant statistics*.

3. **[Strength]** Let Γ^{RS} be the set of relevant statistical statements for χ with respect to Γ_δ, and let the set $\{\Lambda_1, \ldots, \Lambda_n\}$ be the set of difference sets of Γ^{RS}. The principle of strength is the choosing of the minimal cover under difference of Γ^{RS}, i.e., the selection of the shortest cover in $\{Cov(\Lambda_1), \ldots, Cov(\Lambda_n)\}$.

The evidential probability of χ is the minimal cover under difference of Γ^{RS}.

We may define Γ_ϵ, the set of **practical certainties**, in terms of a body of evidence Γ_δ:

$$\Gamma_\epsilon = \{\chi : \exists\, l, u\ (\mathrm{Prob}(\neg\chi, \Gamma_\delta) = [l, u] \wedge u \leq \epsilon)\},$$

or alternatively,

$$\Gamma_\epsilon = \{\chi : \exists\, l, u\ (\mathrm{Prob}(\chi, \Gamma_\delta) = [l, u] \wedge l \geq 1 - \epsilon)\}.$$

The set Γ_ϵ is the set of statements that the evidence Γ_δ warrants accepting; we say a sentence χ is ϵ-*accepted* if $\chi \in \Gamma_\epsilon$. Thus we may add to our knowledge base statements that are nonmonotonic consequences of Γ_δ with respect to a threshold point of acceptance.

Finally, we may view the evidence Γ_δ to provide real-valued bounds on 'degrees of belief' owing to the logical structure of sentences accepted into Γ_δ. However, the probability interval $[l, u]$ associated with χ does not specify a range of equally rational degrees of belief between l and u: the interval $[l, u]$ itself is not a quantity, only l and u are quantities, which are used to specify bounds. On this view, no degree of belief within $[l, u]$ is defensible, which is in marked contrast to the view offered by Objective Bayesianism.

3 SECOND-ORDER EVIDENTIAL PROBABILITY

3.1 Motivation

Second-order evidential probability—developed in [Haenni *et al.*, 2010]—goes beyond Kyburg's evidential probability in two ways. First, it treats an EP interval as bounds on sharp probability. Second, it disentangles reasoning under uncertainty from questions of acceptance and rejection. Here we explain both moves in more detail.

3.1.0.1 Bounds on Degrees of Belief. Kyburg maintained that one can interpret an evidential probability interval for proposition χ as providing bounds on the degree to which an agent should believe χ, but he had reservations about this move:

> Should we speak of partial beliefs as 'degrees of belief'? Although probabilities are intervals, we could still do so. Or we could say that any 'degree of belief' satisfying the probability bounds was 'rational'. But what would be the point of doing so? We agree with Ramsey that logic cannot determine a real-valued a priori degree of belief in pulling a black ball from an urn. This seems a case where degrees of belief are not appropriate. No particular degree of belief is defensible. We deny that there are any appropriate a priori degrees of belief, though there is a fine a priori probability: $[0, 1]$. There are real valued *bounds* on degrees of belief, determined by the logical structure of our evidence. [Kyburg, 2003, p. 147]

Kyburg is making the following points here. Evidence rarely determines a unique value for an agent's degree of belief—rather, it narrows down rational belief to an interval. One can view this interval as providing bounds on rational degree of belief, but since evidence can not be used to justify the choice of one point over another in this interval, there seems to be little reason to talk of the individual points and one can instead simply treat the interval itself as a partial belief.

This view fits very well with the interpretation of evidential probability as some kind of measure of weight of evidence. (And evidential probability provides a natural measure of weight of evidence: the narrower the interval, the weightier the evidence.) Hence if evidence only narrows down probability to an interval, then there does indeed seem to be little need to talk of anything but the interval when measuring features of the evidence. But the view does not address how to fix a sharp degree of belief—intentionally so, since Kyburg's program was designed in part to show us how far one can go with relative frequency information alone. Even so, we may ask whether there is a way to use the resources of evidential probability to fix sharp degrees of belief. In other words, we might return to Carnap's original distinction between probability$_1$ and probability$_2$ and ask how a theory of the latter can be used to constrain the former. If we want to talk

Step 1	Evidence	$\{P(\varphi) \in [l_\varphi, u_\varphi]\}$
Step 2	Acceptance	$\Gamma_\delta = \{\varphi : l_\varphi \geq 1 - \delta\}$
Step 3	Uncertain reasoning	$\{P(\chi) \in [l_\chi, u_\chi]\}$
Step 4	Acceptance	$\Gamma_\varepsilon = \{\chi : l_\chi \geq 1 - \varepsilon\}$

Figure 1. The structure of (first-order) EP inferences.

not only of the quality of our evidence but also of our disposition to act on that evidence, then it would appear that we need a richer language than that provided by EP alone: while evidence—and hence EP—cannot provide grounds to prefer one point in the interval over another as one's degree of belief, there may be other, non-evidential grounds for some such preference, and formalising this move would require going beyond EP.

Reconciling EP with a Bayesian approach has been considered to be highly problematic [Levi, 1977; Levi, 1980; Seidenfeld, 2007], and was vigorously resisted by Kyburg throughout his life. On the other hand, Kyburg's own search for an EP-compatible decision theory was rather limited [Kyburg, 1990]. It is natural then to explore how to modify evidential probability in order that it might handle point-valued degrees of belief and thereby fit with Bayesian decision theory. Accordingly second-order EP departs from Kyburg's EP by viewing evidential probability intervals as bounds on rational degree of belief, $P(\chi) \in \text{Prob}(\chi, \Gamma_\delta)$. In §5 we will go further still by viewing the results of EP as feeding into objective Bayesian epistemology.

3.1.0.2 Acceptance and Rejection. If we allow ourselves the language of point-valued degrees of belief, (first-order) EP can be seen to work like this. An agent has evidence which consists of some propositions $\varphi_1, \ldots, \varphi_n$ and information about their risk levels. He then accepts those propositions whose risk levels are below the agent's threshold δ. This leaves him with the evidential certainties, $\Gamma_\delta = \{\varphi_i : P(\varphi_i) \geq 1 - \delta\}$. From Γ_δ the agent infers propositions ψ of the form $P(\chi) \in [l, u]$. In turn, from these propositions the agent infers the practical certainties $\Gamma_\varepsilon = \{\chi : l \geq 1 - \varepsilon\}$. This sequence of steps is depicted in Figure 1.

There are two modes of reasoning that are intermeshed here: on the one hand the agent is using evidence to reason under uncertainty about the conclusion proposition ψ, and on the other he is deciding which propositions to accept and reject. The acceptance mode appears in two places: deciding which evidential propositions to accept and deciding whether to accept the proposition χ to which the conclusion ψ refers.

With second-order EP, on the other hand, *acceptance is delayed until all reasoning under uncertainty is completed.* Then we treat acceptance as a decision problem requiring a decision-theoretic solution—e.g., accept those propositions

Step 1	Evidence	$\Phi = \{P(\varphi) \in [l_\varphi, u_\varphi]\}$
Step 2	Uncertain reasoning	$\Psi = \{P(\chi) \in [l_\chi, u_\chi]\}$
Step 3	Acceptance	$\{\chi : \text{decision-theoretically optimal}\}$

Figure 2. The structure of 2oEP inferences.

whose acceptance maximises expected utility.[10] Coupling this solution with the use of point-valued probabilities we have second-order evidential probability (2oEP), whose inferential steps are represented in Figure 2.

There are two considerations that motivate this more strict separation of uncertain reasoning and acceptance.

First, such a separation allows one to chain inferences—something which is not possible in 1oEP. By 'chaining inferences' we mean that the results of step 2 of 2oEP can be treated as an input for a new round of uncertain reasoning, to be recombined with evidence and to yield further inferences. Only once the chain of uncertain reasoning is complete will the acceptance phase kick in. Chaining of inferences is explained in further detail in §3.2.

Second, such a separation allows one to keep track of the uncertainties that attach to the evidence. To each item of evidence φ attaches an interval $[l_\varphi, u_\varphi]$ representing the risk or reliability of that evidence. In 1oEP, step 2 ensures that one works just with those propositions φ whose risk levels meet the threshold of acceptance. But in 2oEP there is no acceptance phase before uncertain reasoning is initiated, so one works with the entirety of the evidence, including the risk intervals themselves. While the use of this extra information makes inference rather more complicated, it also makes inference more accurate since the extra information can matter—the results of 2oEP can differ from the results of 1oEP.

We adopt a decision-theoretic account of acceptance for the following reason. In 1oEP, each act of acceptance uniformly accepts those propositions whose associated risk is less than some fixed threshold: δ in step 2 and ε in step 4. (This allows statements to detach from their risk levels and play a role as logical constraints in inference.) But in practice thresholds of acceptance depend not so much on the step in the chain of reasoning as on the proposition concerned, and, indeed, the whole inferential set-up. To take a favourite example of Kyburg's, consider a lottery. The threshold of acceptance of the proposition *the lottery ticket that the seller is offering me will lose* may be higher than that of *the coin with a bias in favour of heads that I am about to toss will land heads* and lower than that of *the moon is made of blue cheese*. This is because nothing may hang on the

[10]Note that maximising expected utility is not straightforward in this case since bounds on probabilities, rather than the probabilities themselves, are input into the decision problem. EP-calibrated objective Bayesianism (§5) goes a step further by determining point-valued probabilities from these bounds, thereby making maximisation of expected utility straightforward. See [Williamson, 2009] for more on the combining objective Bayesianism with a decision-theoretic account of acceptance.

coin toss (in which case a 60% bias in favour of heads may be quite adequate for acceptance), while rather a lot hangs on accepting that the moon is made of blue cheese—many other propositions that I have hitherto granted will have to be revisited if I were to accept this proposition. Moreover, if I am going to use the judgement to decide whether to buy a ticket then the threshold of acceptance of the lottery proposition should plausibly depend on the extent to which I value the prize. Given these considerations, acceptance of a proposition can fruitfully be viewed as a decision problem, depending on the decision set-up including associated utilities [Williamson, 2009]. Again, while this is more complicated than the 1oEP solution of modelling acceptance using a fixed threshold, the subtleties of a full-blown decision-theoretic account can matter to the resulting inferences.

3.2 Calculating Second-order EP

In this section we will be concerned with developing some machinery to perform uncertain reasoning in second-order evidential probability (step 2 in Figure 2). See [Haenni *et al.*, 2010] for further details of this approach.

3.2.1 Entailment

Let $\mathcal{L}^\sharp$ be a propositional language whose propositional variables are of the form $\varphi^{[a,b]}$ for atomic propositions $\varphi \in \mathcal{L}$.[11] Here $\mathcal{L}$ is the language of (first-order) EP extended to include statements of the form $P(\chi) \in [l,u]$, and, for proposition φ of $\mathcal{L}$, $\varphi^{[a,b]}$ is short for $P(\varphi) \in [a,b]$. Hence in $\mathcal{L}^\sharp$ we can express propositions about higher-order probabilities, e.g., $P(\chi) \in [l,u]^{[a,b]}$ which is short for $P(P(\chi) \in [l,u]) \in [a,b]$. We write φ^a as an abbreviation of $\varphi^{[a,a]}$.

For $\mu, \nu \in \mathcal{L}^\sharp$ write $\mu \mathrel{|\!\approx_{2\mathrm{o}}} \nu$ if ν deductively follows from μ by appealing to the axioms of probability and the following EP-motivated axioms:

A1: Given $\varphi_1^1, \ldots, \varphi_n^1$, if $\mathrm{Prob}(\chi, \{\varphi_1, \ldots, \varphi_n\}) = [l,u]$ is derivable by (first-order) EP then infer ψ^1, where $\psi \in \mathcal{L}$ is the statement $P(\chi) \in [l,u]$.

A2: Given ψ^1 then infer $\chi^{[l,u]}$, where $\psi \in \mathcal{L}$ is the statement $P(\chi) \in [l,u]$.

Axiom A1 ensures that EP inferences carry over to 2oEP, while axiom A2 ensures that probabilities at the first-order level can constrain those at the second-order level.

The entailment relation $\mathrel{|\!\approx_{2\mathrm{o}}}$ will be taken to constitute *core second-order EP*. The idea is that when input evidence Φ consisting of a set of sentences of $\mathcal{L}^\sharp$, one infers a set Ψ of further such sentences using the above consequence relation. Note that although $\mathrel{|\!\approx_{2\mathrm{o}}}$ is essentially classical consequence with extra axioms, it is a nonmonotonic consequence relation since 1oEP is nonmonotonic. But 2oEP yields a strong logic inasmuch as it combines the axioms of probability with the rules of

[11] As it stands $\mathcal{L}^\sharp$ contains uncountably many propositional variables, but restrictions can be placed on a, b to circumscribe the language if need be.

EP, and so questions of consistency arise. Will there always be some probability function that satisfies the constraints imposed by 1oEP consequences of evidence? Not always: see [Seidenfeld, 2007] for some counterexamples. Consequently, some consistency-maintenance procedure needs to be invoked to cope with such cases. (Of course, some consistency maintenance procedure will in any case be required to handle certain inconsistent sets of evidential propositions, so there may be no extra burden here.) One option is to consider probability functions satisfying (EP consequences of) maximal satisfiable subsets of evidential statements, for example. In this paper we will not commit to a particular consistency-maintenance procedure; we leave this interesting question as a topic for further research.

3.2.2 Credal Networks

This entailment relation can be implemented using probabilistic networks, as we shall now explain. For efficiency reasons, we make the following further assumptions. First we assume that P is distributed uniformly over the EP interval unless there is evidence otherwise:

A3: If $\Phi \mathrel{|\!\!\approx_{2o}} \chi^{[l,u]}$ then $P(\chi^{[l',u']}|\Phi) = \frac{|[l,u]\cap[l',u']|}{|[l,u]|}$, as long as this is consistent with other consequences of Φ.

Second, we assume that items of evidence are independent unless there is evidence of dependence:

A4: If $\varphi_1^{[a_1,b_1]}, \ldots, \varphi_k^{[a_k,b_k]} \in \Phi$ then $P(\varphi_1^{[a_1,b_1]}, \ldots, \varphi_k^{[a_k,b_k]}) = P(\varphi_1^{[a_1,b_1]}) \cdots P(\varphi_k^{[a_k,b_k]})$, as long as this is consistent with other consequences of Φ.

These assumptions are not essential to second-order EP, but they make the probabilistic network implementation particularly straightforward.[12] Note that these assumptions are default rules; when determined by A1-4, the consequence relation $\mathrel{|\!\!\approx_{2o}}$ is nonmonotonic.

A *credal network* can be used to represent and reason with a set of probability functions [Cozman, 2000]. A credal network consists of (i) a directed acyclic graph whose nodes are variables $A_1, \ldots, A_n$ and (ii) constraints on conditional probabilities of the form $P(a_i \mid par_i) \in [l, u]$ where a_i is an assignment of a value to a variable and par_i is an assignment of values to its parents in the graph. It is assumed that each variable is probabilistically independent of its non-descendants conditional on its parents in the graph, written $A_i \perp\!\!\!\perp ND_i \mid Par_i$; this assumption is known as the *Markov Condition*.

[12] If items of evidence are known to be dependent then the corresponding nodes will be connected by arrows in the credal network representation outlined below. Any information that helps to quantify the dependence will help determine the conditional probability distributions associated with these arrows. If P is known to be distributed non-uniformly over the EP intervals then information about its distribution will need to be used to determine conditional probability distributions in the credal net.

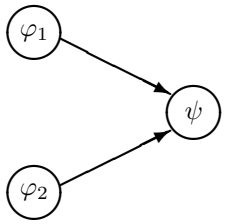
Credal networks are of fundamental importance for inference in probabilistic logic [Haenni *et al.*, 2010]. A logic is a *probabilistic logic* if its semantic interpretations are probability functions; the entailment relation of first-order EP does not constitute a probabilistic logic in this sense, but the entailment relation $\approx_{2\circ}$ of second-order EP does. In a probabilistic logic we are typically faced with the following sort of question: given premiss propositions $\varphi_1, \ldots, \varphi_n$ and their respective probabilities $X_1, \ldots, X_n$, what probability should we attach to a conclusion proposition ψ? This question can be written in the form

$$\varphi_1^{X_1}, \ldots, \varphi_n^{X_n} \mid\!\approx \psi^?$$

where $\mid\!\approx$ is the entailment relation of the probabilistic logic. For example, in second-order evidential probability we might be faced with the following question

$$\%x(Fx, Rx, [.2, .4])^{[.9,1]}, Rt \mid\!\approx_{2\circ} P(Ft) \in [.2, .4]^?$$

This asks, given evidence that (i) the proposition that the frequency of attribute F in reference class R is between .2 and .4 has probability at least .9, and (ii) t falls in reference class R, what probability interval should attach to the proposition that the probability that t has attribute F is between .2 and .4? In first-order EP, if $1 - \delta \geq .9$ then $\mathrm{Prob}(Ft, \Gamma_\delta) = [.2, .4]$ would be conclusively inferred (and hence treated as if it had probability 1). Clearly this disregards the uncertainty that attaches to the statistical evidence; the question is, what uncertainty should attach to the conclusion as a consequence? (This is a second-order uncertainty; hence the name *second-order* evidential probability.) One can construct a credal network to answer this question as follows. Let φ_1 be the proposition $\%x(Fx, Rx, [.2, .4])$, φ_2 be Rt and ψ be $P(Ft) \in [.2, .4]$. These can all be thought of as variables that take possible values True and False. The structure of 1oEP calculations determines the structure of the directed acyclic graph in the credal net:



The conditional probability constraints involving the premiss propositions are simply their given risk levels:

$$P(\varphi_1) \in [.9, 1],$$
$$P(\varphi_2) = 1.$$

Turning to the conditional probability constraints involving the conclusion proposition, these are determined by 1oEP inferences via axioms A1-3:

$$P(\psi | \varphi_1 \wedge \varphi_2) = 1,$$

$$P(\psi|\neg\varphi_1 \wedge \varphi_2) = P(\psi|\varphi_1 \wedge \neg\varphi_2) = P(\psi|\neg\varphi_1 \wedge \neg\varphi_2) = .2.$$

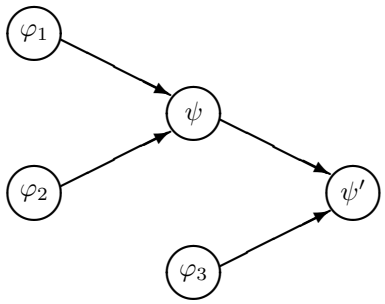
Finally, the Markov condition holds in virtue of A4, which implies that $\varphi_1 \perp\!\!\!\perp \varphi_2$. Inference algorithms for credal networks can then be used to infer the uncertainty that should attach to the conclusion, $P(\psi) \in [.92, 1]$. Hence we have:

$$\%x(Fx, Rx, [.2, .4])^{[.9,1]}, Rt \mathrel{|\!\approx_{2\circ}} P(Ft) \in [.2, .4]^{[.92,1]}$$

3.2.2.1 Chaining Inferences. While it is not possible to chain inferences in 1oEP, this is possible in 2oEP, and the credal network representation can just as readily be applied to this more complex case. Consider the following question:

$$\%x(Fx, Rx, [.2, .4])^{[.9,1]}, Rt, \%x(Gx, Fx, [.2, .4])^{[.6,.7]} \mathrel{|\!\approx_{2\circ}} P(Gt) \in [0, .25]^{?}$$

As we have just seen, the first two premisses can be used to infer something about Ft, namely $P(Ft) \in [.2, .4]^{[.92,1]}$. But now this inference can then be used in conjunction with the third premiss to infer something about Gt. To work out the probability bounds that should attach to an inference to $P(Gt) \in [0, .25]$, we can apply the credal network procedure. Again, the structure of the graph in the network is given by the structure of EP inferences:



Here φ_3 is $\%x(Gx, Fx, [.2, .4])$ and ψ' is $P(Gt) \in [0, .25]$; other variables are as before. The conditional probability bounds of the previous example simply carry over

$$P(\varphi_1) \in [.9, 1], P(\varphi_2) = 1,$$

$$P(\psi|\varphi_1 \wedge \varphi_2) = 1, P(\psi|\neg\varphi_1 \wedge \varphi_2) = .2 = P(\psi|\varphi_1 \wedge \neg\varphi_2) = P(\psi|\neg\varphi_1 \wedge \neg\varphi_2).$$

But we need to provide further bounds. As before, the risk level associated with the third premiss φ_3 provides one of these:

$$P(\varphi_3) \in [.6, .7],$$

and the constraints involving the new conclusion ψ' are generated by A3:

$$P(\psi'|\psi \wedge \varphi_3) = \frac{|[.2 \times .6 + .8 \times .1, .4 \times .7 + .6 \times .1] \cap [0, .25]|}{|[.2 \times .6 + .8 \times .1, .4 \times .7 + .6 \times .1]|} = .31,$$

$$P(\psi'|\neg\psi \wedge \varphi_3) = .27, P(\psi'|\psi \wedge \neg\varphi_3) = P(\psi'|\neg\psi \wedge \neg\varphi_3) = .25.$$

The Markov Condition holds in virtue of A4 and the structure of EP inferences. Performing inference in the credal network yields $P(\psi') \in [.28, .29]$. Hence

$$\%x(Fx, Rx, [.2, .4])^{[.9,1]}, Rt, \%x(Gx, Fx, [.2, .4])^{[.6,.7]} \approx_{2\circ} P(Gt) \in [0, .25]^{[.28,.29]}.$$

This example shows how general inference in 2oEP can be: we are not asking which probability bounds attach to a 1oEP inference in this example, but rather which probability bounds attach to an inference that cannot be drawn by 1oEP. The example also shows that the probability interval attaching to the conclusion can be narrower than intervals attaching to the premisses.

4 OBJECTIVE BAYESIAN EPISTEMOLOGY

4.1 Motivation

We saw above that evidential probability concerns the impact of evidence upon a conclusion. It does not on its own say how strongly one should *believe* the conclusion. Kyburg was explicit about this, arguing that evidential probabilities can at most be thought of as 'real-valued *bounds* on degrees of belief, determined by the logical structure of our evidence' [Kyburg, 2003, p. 147]. To determine rational degrees of belief themselves, one needs to go beyond EP, to a normative theory of partial belief.

Objective Bayesian epistemology is just such a normative theory [Rosenkrantz, 1977; Jaynes, 2003; Williamson, 2005]. According to the version of objective Bayesianism presented in [Williamson, 2005], one's beliefs should adhere to three norms:

Probability: The strengths of one's beliefs should be representable by probabilities. Thus they should be measurable on a scale between 0 and 1, and should be additive.

Calibration: These degrees of belief should fit one's evidence. For example, degrees of belief should be calibrated with frequency: if all one knows about the truth of a proposition is an appropriate frequency, one should believe the proposition to the extent of that frequency.

Equivocation: One should not believe a proposition more strongly than the evidence demands. One should equivocate between the basic possibilities as far as the evidence permits.

These norms are imprecisely stated: some formalism is needed to flesh them out.

4.1.0.2 Probability. In the case of the Probability norm, the mathematical calculus of probability provides the required formalism. Of course mathematical probabilities attach to abstract events while degrees of belief attach to propositions, so the mathematical calculus needs to be tailored to apply to propositions. It is usual to proceed as follows — see, e.g., [Paris, 1994]. Given a predicate language $\mathcal{L}$ with constants t_i that pick out all the members of the domain, and sentences θ, φ of $\mathcal{L}$, a function P is a *probability function* if it satisfies the following axioms:

P1: If $\models \theta$ then $P(\theta) = 1$;

P2: If $\models \neg(\theta \wedge \varphi)$ then $P(\theta \vee \varphi) = P(\theta) + P(\varphi)$;

P3: $P(\exists x \theta(x)) = \lim_{n\to\infty} P(\bigvee_{i=1}^{n} \theta(t_i))$.

P1 sets the scale, P2 ensures that probability is additive, and P3, called *Gaifman's condition*, sets the probability of 'θ holds of something' to be the limit of the probability of 'θ holds of one or more of $t_1, ..., t_n$', as n tends to infinity. The Probability norm then requires that the strengths of one's beliefs be representable by a probability function P over (a suitable formalisation of) one's language. Writing $\mathbb{P}$ for the set of probability functions over $\mathcal{L}$, the Probability norm requires that one's beliefs be representable by some $P \in \mathbb{P}$.

4.1.0.3 Calibration. The Calibration norm says that the strengths of one's beliefs should be appropriately constrained by one's evidence $\mathcal{E}$. (By evidence we just mean everything taken for granted in the current operating context—observations, theory, background knowledge etc.) This norm can be explicated by supposing that there is some set $\mathbb{E} \subseteq \mathbb{P}$ of probability functions that satisfy constraints imposed by evidence and that one's degrees of belief should be representable by some $P_{\mathcal{E}} \in \mathbb{E}$. Now typically one has two kinds of evidence: quantitative evidence that tells one something about physical probability (frequency, chance etc.), and qualitative evidence that tells one something about how one's beliefs should be structured. In [Williamson, 2005] it is argued that these kinds of evidence should be taken into account in the following way. First, quantitative evidence (e.g., evidence of frequencies) tells us that the physical probability function P^* must lie in some set $\mathbb{P}^*$ of probability functions. One's degrees of belief ought to be similarly constrained by evidence of physical probabilities, subject to a few provisos:

C1: $\mathbb{E} \neq \emptyset$.

If evidence is inconsistent this tells us something about our evidence rather than about physical probability, so one cannot conclude that $\mathbb{P}^* = \emptyset$ and one can hardly insist that $P_{\mathcal{E}} \in \emptyset$. Instead $\mathbb{P}^*$ must be determined by some consistency maintenance procedure—one might, for example, take $\mathbb{P}^*$ to be determined by maximal consistent subsets of one's evidence—and neither $\mathbb{P}^*$ nor $\mathbb{E}$ can ever be empty.

C2: If $\mathcal{E}$ is consistent and implies proposition θ that does not mention physical probability P^*, then $P(\theta) = 1$ for all $P \in \mathbb{E}$.

This condition merely asserts that categorical evidence be respected—it prevents $\mathbb{E}$ from being too inclusive. The qualification that θ must not mention physical probability is required because in some cases evidence of physical probability should be treated more pliably:

C3: If $P, Q \in \mathbb{P}^*$ and $R = \lambda P + (1 - \lambda)Q$ for $\lambda \in [0, 1]$ then, other things being equal, one should be permitted to take R as one's belief function $P_{\mathcal{E}}$.

Note in particular that C3 implies that, other things being equal, if $P \in \mathbb{P}^*$ then $P \in \mathbb{E}$; it also implies C1 (under the understanding that $\mathbb{P}^* \neq \emptyset$). C3 is required to handle the following kind of scenario. Suppose for example that you have evidence just that an experiment with two possible outcomes, a and $\neg a$, has taken place. As far as you are aware, the physical probability of a is now 1 or 0 and no value in between. But this does not imply that your degree of belief in a should be 1 or 0 and no value in between—a value of $\frac{1}{2}$, for instance, is quite reasonable in this case. C3 says that, in the absence of other overriding evidence, $\langle \mathbb{P}^* \rangle \subseteq \mathbb{E}$ where $\langle \mathbb{P}^* \rangle$ is the convex hull of $\mathbb{P}^*$. The following condition imposes the converse relation:

C4: $\mathbb{E} \subseteq \langle \mathbb{P}^* \rangle$.

Suppose for example that evidence implies that either $P^*(a) = 0.91$ or $P^*(a) = 0.92$. While C3 permits any element of the interval $[0.91, 0.92]$ as a value for one's degree of belief $P_{\mathcal{E}}(a)$, C4 confines $P_{\mathcal{E}}(a)$ to this interval—indeed a value outside this interval is unwarranted by this particular evidence. Note that C4 implies C2: θ being true implies that its physical probability is 1, so $P(\theta) = 1$ for all $P \in \mathbb{P}^*$, hence for all $P \in \langle \mathbb{P}^* \rangle$, hence for all $P \in \mathbb{E}$.

In the absence of overriding evidence the conditions C1–4 set $\mathbb{E} = \langle \mathbb{P}^* \rangle$. This sheds light on how *quantitative* evidence constrains degrees of belief, but one may also have qualitative evidence which constrains degrees of belief in ways that are not mediated by physical probability. For example, one may know about causal influence relationships involving variables in one's language: this may tell one something about physical probability, but it also tells one other things—e.g., that if one extends one's language to include a new variable that is not a cause of the current variables, then that does not on its own provide any reason to change one's beliefs about the current variables. These constraints imposed by evidence of influence relationships, discussed in detail in [Williamson, 2005], motivate a further principle:

C5: $\mathbb{E} \subseteq \mathbb{S}$ where $\mathbb{S}$ is the set of probability functions satisfying structural constraints.

We will not dwell on C5 here since structural constraints are peripheral to the theme of this paper, namely to connections between objective Bayesian epistemology and evidential probability. It turns out that the set $\mathbb{S}$ is always non-empty, hence C1–5 yield:

Calibration: One's degrees of belief should be representable by $P_{\mathcal{E}} \in \mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$.

4.1.0.4 Equivocation. The third norm, Equivocation, can be fleshed out by requiring that $P_{\mathcal{E}}$ be a probability function, from all those that are calibrated with evidence, that is as close as possible to a totally equivocal probability function $P_=$ called the *equivocator* on $\mathcal{L}$. But we need to specify the equivocator and also what we mean by 'as close as possible'. To specify the equivocator, first create an ordering $a_1, a_2, \ldots$ of the atomic sentences of $\mathcal{L}$—sentences of the form Ut where U is a predicate or relation and t is a tuple of constants of corresponding arity—such that those atomic sentences involving constants $t_1, \ldots t_{n-1}$ occur earlier in the ordering than those involving t_n. Then we can define the equivocator $P_=$ by $P_=(a_j^{e_j} \mid a_1^{e_1} \wedge \cdots \wedge a_{j-1}^{e_{j-1}}) = 1/2$ for all j and all $e_1, \ldots e_j \in \{0, 1\}$, where a_i^1 is just a_i and a_i^0 is $\neg a_i$. Clearly $P_=$ equivocates between each atomic sentence of $\mathcal{L}$ and its negation. In order to explicate 'as close as possible' to $P_=$ we shall appeal to the standard notion of distance between probability functions, the *n-divergence* of P from Q:

$$d_n(P, Q) \stackrel{\text{df}}{=} \sum_{e_1, \ldots, e_{r_n}=0}^{1} P(a_1^{e_1} \wedge \cdots \wedge a_{r_n}^{e_{r_n}}) \log \frac{P(a_1^{e_1} \wedge \cdots \wedge a_{r_n}^{e_{r_n}})}{Q(a_1^{e_1} \wedge \cdots \wedge a_{r_n}^{e_{r_n}})}.$$

Here $a_1, ..., a_{r_n}$ are the atomic sentences involving constants $t_1, ..., t_n$; we follow the usual convention of taking $0 \log 0$ to be 0, and note that the n-divergence is not a distance function in the usual mathematical sense because it is not symmetric and does not satisfy the triangle inequality—rather, it is a measure of the amount of information that is encapsulated in P but not in Q. We then say that P is closer to the equivocator than Q if there is some N such that for $n \geq N$, $d_n(P, P_=) < d_n(Q, P_=)$. Now we can state the Equivocation norm as follows. For a set $\mathbb{Q}$ of probability functions, denote by $\downarrow\mathbb{Q}$ the members of $\mathbb{Q}$ that are closest to the equivocator $P_=$.[13] Then,

E1: $P_{\mathcal{E}} \in \downarrow\mathbb{E}$.

This principle is discussed at more length in [Williamson, 2008]. It can be construed as a version of the maximum entropy principle championed by Edwin Jaynes. Note that while some versions of objective Bayesianism assume that an agent's degrees of belief are uniquely determined by her evidence and language, we make no such assumption here: $\downarrow\mathbb{E}$ may not be a singleton.

[13] If there are no closest members (i.e., if chains are all infinitely descending: for any member P of $\mathbb{Q}$ there is some P' in $\mathbb{Q}$ that is closer to the equivocator than P) the context may yet determine an appropriate subset $\downarrow\mathbb{Q} \subseteq \mathbb{Q}$ of probability functions that are *sufficiently close* to the equivocator; for simplicity of exposition we shall ignore this case in what follows.

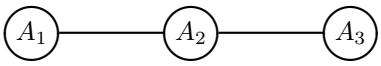


Figure 3. Constraint graph.

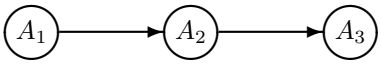


Figure 4. Graph satisfying the Markov Condition.

4.2 Calculating Objective Bayesian Degrees of Belief

Just as credal nets can be used for inference in 2oEP, so too can they be used for inference in OBE. The basic idea is to use a credal net to represent $\downarrow\mathbb{E}$, the set of rational belief functions, and then to perform inference to calculate the range of probability values these functions ascribe to some proposition of interest. These methods are explained in detail in [Williamson, 2008]; here we shall just give the gist.

For simplicity we shall describe the approach in the base case in which the evidence consists of interval bounds on the probabilities of sentences of the agent's language $\mathcal{L}$, $\mathcal{E} = \{P^*(\varphi_i) \in [l_i, u_i] : i = 1, \ldots, k\}$, $\mathcal{E}$ is consistent and does not admit infinite descending chains; but these assumptions can all be relaxed. In this case $\mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S} = \mathbb{P}^*$. Moreover, the evidence can be written in the language $\mathcal{L}^\sharp$ introduced earlier: $\mathcal{E} = \{\varphi_1^{[l_1,u_1]}, \ldots, \varphi_k^{[l_k,u_k]}\}$, and the question facing objective Bayesian epistemology takes the form

$$\varphi_1^{[l_1,u_1]}, \ldots, \varphi_k^{[l_k,u_k]} \mid\!\!\approx_{\text{OBE}} \psi^?$$

where $\mid\!\!\approx_{\text{OBE}}$ is the entailment relation defined by objective Bayesian epistemology as outlined above. As explained in [Williamson, 2008], this entailment relation is nonmonotonic but it is well-behaved in the sense that it satisfies all the System-P properties of nonmonotonic logic.

The method is essentially this. First construct an undirected graph, the *constraint graph*, by linking with an edge those atomic sentences that appear in the same item of evidence. One can read off this graph a list of probabilistic independencies that any function in $\downarrow\mathbb{E}$ must satisfy: if node A separates nodes B and C in this graph then $B \perp\!\!\!\perp C \mid A$ for each probability function in $\downarrow\mathbb{E}$. This constraint graph can then be transformed into a directed acyclic graph for which the Markov Condition captures many or all of these independencies. Finally one can calculate bounds on the probability of each node conditional on its parents in the graph by using entropy maximisation methods: each probability function in $\downarrow\mathbb{E}$ maximises entropy subject to the constraints imposed by $\mathcal{E}$, and one can identify the probability it gives to one variable conditional on its parents using numerical optimisation methods [Williamson, 2008].

To take a simple example, suppose we have the following question:

$$\forall x(Ux \to Vx)^{3/5}, \forall x(Vx \to Wx)^{3/4}, Ut_1{}^{[0.8,1]} \mid\!\!\approx_{\text{OBE}} Wt_1^?$$

A credal net can be constructed to answer this question. There is only one constant symbol t_1, and so the atomic sentences of interest are Ut_1, Vt_1, Wt_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 credal network. (In fact, since the conditional probabilities are precisely determined rather than bounded, we have a special case of a credal net called a *Bayesian net*.) The Markov Condition holds since separation in the constraint graph implies probabilistic independence. Standard inference methods then give us $P(A_3) = 11/15$ as an answer to our question.

5 EP-CALIBRATED OBJECTIVE BAYESIANISM

5.1 Motivation

At face value, evidential probability and objective Bayesian epistemology are very different theories. The former concerns the impact of evidence of physical probability, Carnap's probability$_2$, and concerns acceptance and rejection; it appeals to interval-valued probabilities. The latter theory concerns rational degree of belief, probability$_1$, and invokes the usual point-valued mathematical notion of probability. Nevertheless the core of these two theories can be reconciled, by appealing to second-order EP as developed above.

2oEP concerns the impact of evidence on rational degree of belief. Given statistical evidence, 2oEP will infer statements about rational degrees of belief. These statements can be viewed as constraints that should be satisfied by the degrees of belief of a rational agent with just that evidence. So 2oEP can be thought of as mapping statistical evidence $\mathcal{E}$ to a set $\mathbb{E}$ of rational belief functions that are compatible with that evidence. (This is a non-trivial mapping because frequencies attach to a sequence of outcomes or experimental conditions that admit repeated instantiations, while degrees of belief attach to propositions. Hence the epistemological reference-class problem arises: how can one determine appropriate single-case probabilities from information about generic probabilities? Evidential probability is a theory that tackles this reference-class problem head on: it determines a probability interval that attaches to a sentence from statistical evidence about repetitions.)

But this mapping from $\mathcal{E}$ to $\mathbb{E}$ is just what is required by the Calibration norm of OBE. We saw in §4 that OBE maps evidence $\mathcal{E}$ to $\mathbb{E} = \langle \mathbb{P}^* \rangle \cap \mathbb{S}$, a set of probability functions calibrated with that evidence. But no precise details were given as to how $\langle \mathbb{P}^* \rangle$, nor indeed $\mathbb{P}^*$, is to be determined. In special cases this is

straightforward. For example, if one's evidence is just that the chance of a is $\frac{1}{2}$, $P^*(a) = 1/2$, then $\langle \mathbb{P}^* \rangle = \mathbb{P}^* = \{P \in \mathbb{P} : P(a) = 1/2\}$. But in general, determining $\langle \mathbb{P}^* \rangle$ is not a trivial enterprise. In particular, statistical evidence takes the form of information about generic frequencies rather than single-case chances, and so the reference-class problem arises. It is here that 2oEP can be plugged in: if $\mathcal{E}$ consists of propositions of $\mathcal{L}^\sharp$—i.e., propositions, including statistical propositions, to which probabilities or closed intervals of probabilities attach—then $\langle \mathbb{P}^* \rangle$ is the set of probability functions that satisfy the $\approx_{2\text{o}}$ consequences of $\mathcal{E}$.

C6: If $\mathcal{E}$ is a consistent set of propositions of $\mathcal{L}^\sharp$ then $\langle \mathbb{P}^* \rangle = \{P : P(\chi) \in [l, u]$ for all χ, l, u such that $\mathcal{E} \approx_{2\text{o}} \chi^{[l,u]}\}$.

We shall call OBE that appeals to calibration principles C1–6 *epistemic-probability-calibrated objective Bayesian epistemology*, or EP-OBE for short. We shall denote the corresponding entailment relation by $\approx^{\text{EP}}_{\text{OBE}}$.

We see then that there is a sense in which EP and OBE can be viewed as complementary rather than in opposition. Of course, this isn't the end of the matter. Questions still arise as to whether EP-OBE is the right way to flesh out OBE. One can, for instance, debate the particular rules that EP uses to handle reference classes (§2.2). One can also ask whether EP tells us everything we need to know about calibration. As mentioned in §4.1, further rules are needed in order to handle structural evidence, fleshing out C5. Moreover, both 1oEP and 2oEP take statistical statements as input; these statements themselves need to be inferred from particular facts—indeed EP, OBE and EP-OBE each presume a certain amount of statistical inference. Consequently we take it as understood that Calibration requires more than just C1–6.

And questions arise as to whether the alterations to EP that are necessary to render it compatible with OBE are computationally practical. Second-order EP replaces the original theory of acceptance with a decision theoretic account which will incur a computational burden. Moreover, some thought must be given as to which consistency maintenance procedure should be employed in practice. Having said this, we conjecture that there will be real inference problems for which the benefits will be worth the necessary extra work.

5.2 Calculating EP-Calibrated Objective Bayesian Probabilities

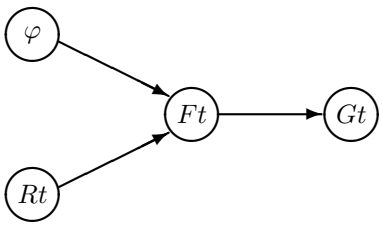
Calculating EP-OBE probabilities can be achieved by combining methods for calculating 2oEP probabilities with methods for calculating OBE probabilities. Since credal nets can be applied to both formalisms independently, they can also be applied to their unification. In fact in order to apply the credal net method to OBE, some means is needed of converting statistical statements, which can be construed as constraints involving generic, repeatably-instantiatable variables, to constraints involving the single-case variables which constitute the nodes of the objective Bayesian credal net; only then can the constraint graph of §4.2 be constructed. The 2oEP credal nets of §3.2 allow one to do this, since this kind

of net incorporates both statistical variables and single-case variables as nodes. Thus 2oEP credal nets are employed first to generate single-case constraints, at which stage the OBE credal net can be constructed to perform inference. This fits with the general view of 2oEP as a theory of how evidence constrains rational degrees of belief and OBE as a theory of how further considerations—especially equivocation—further constrain rational degrees of belief.

Consider the following very simple example:

$$\%x(Fx, Rx, [.2, .5]), Rt, \forall x(Fx \to Gx)^{3/4} \mid\!\sim^{\text{EP}}_{\text{OBE}} Gt^{?}$$

Now the first two premisses yield $Ft^{[.2,.5]}$ by EP. This constraint combines with the third premiss to yield an answer to the above question by appealing to OBE. This answer can be calculated by constructing the following credal net:



Here φ is the first premiss. The left-hand side of this net is the 2oEP net, with associated probability constraints

$$P(\varphi) = 1, P(Rt) = 1,$$

$$P(Ft|\varphi \wedge Rt) \in [.2, .5], P(Ft|\neg\varphi \wedge Rt) = 0 = P(Ft|\varphi \wedge \neg Rt) = P(Ft|\neg\varphi \wedge \neg Rt).$$

The right-hand side of this net is the OBE net with associated probabilities

$$P(Gt|Ft) = 7/10, P(Gt|\neg Ft) = 1/2.$$

Standard inference algorithms then yield an answer of 7/12 to our question:

$$\%x(Fx, Rx, [.2, .5]), Rt, \forall x(Fx \to Gx)^{3/4} \mid\!\sim^{\text{EP}}_{\text{OBE}} Gt^{7/12}$$

6 CONCLUSION

While evidential probability and objective Bayesian epistemology might at first sight appear to be chalk and cheese, on closer inspection we have seen that their relationship is more like horse and carriage—together they do a lot of work, covering the interface between statistical inference and normative epistemology.

Along the way we have taken in an interesting array of theories—first-order evidential probability, second-order evidential probability, objective Bayesian epistemology and EP-calibrated OBE—that can be thought of as nonmonotonic logics.

2oEP and OBE are probabilistic logics in the sense that they appeal to the usual mathematical notion of probability. More precisely, their entailment relations are probabilistic: premisses entail the conclusion if every model of the premisses satisfies the conclusion, where models are probability functions. This connection with probability means that credal networks can be applied as inference machinery. Credal nets yield a perspicuous representation and the prospect of more efficient inference [Haenni *et al.*, 2010].

ACKNOWLEDGEMENTS

We are grateful to the Leverhulme Trust for supporting this research, and to Prasanta S. Bandyopadhyay, Teddy Seidenfeld and an anonymous referee for helpful comments.

BIBLIOGRAPHY

[Andréka *et al.*, 1998] H. Andréka, J. van Benthem, and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27:217–274, 1998.

[Arló-Costa and Parikh, 2005] H. Arló-Costa and R. Parikh. Conditional probability and defeasible inference. *Journal of Philosophical Logic*, 34:97–119, 2005.

[Carnap, 1962] R. Carnap. *The Logical Foundations of Probability*. University of Chicago Press, 2nd edition, 1962.

[Cozman, 2000] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[de Finetti, 1974] B. de Finetti. *Theory of Probability: A critical introductory treatment*. Wiley, 1974.

[Dubbins, 1975] L. E. Dubbins. Finitely additive conditional probability, conglomerability, and disintegrations. *Annals of Probability*, 3:89–99, 1975.

[Dubois and Prade, 1980] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Kluwer, North Holland, 1980.

[Haenni *et al.*, 2010] R. Haenni, J.-W. Romeijn, G. Wheeler, and J. Williamson. *Probabilistic Logic and Probabilistic Networks*. The Synthese Library, Springer, 2010.

[Harper and Wheeler, 2007] W. Harper and G. Wheeler, editors. *Probability and Inference: Essays In Honor of Henry E. Kyburg, Jr.* King's College Publications, London, 2007.

[Hawthorne and Makinson, 2007] J. Hawthorne and D. Makinson. The quantitative/qualitative watershed for rules of uncertain inference. *Studia Logica*, 86(2):247–297, 2007.

[Jaynes, 2003] E. T. Jaynes. *Probability theory: the logic of science*. Cambridge University Press, Cambridge, 2003.

[Keynes, 1921] J. M. Keynes. *A Treatise on Probability*. Macmillan, London, 1921.

[Kraus *et al.*, 1990] S. Kraus, D. Lehman, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Kyburg and Teng, 2001] H. E. Kyburg, Jr. and C. M. Teng. *Uncertain Inference*. Cambridge University Press, Cambridge, 2001.

[Kyburg *et al.*, 2007] H. E. Kyburg, Jr., C. M. Teng, and G. Wheeler. Conditionals and consequences. *Journal of Applied Logic*, 5(4):638–650, 2007.

[Kyburg, 2003] H. E. Kyburg Jr. Are there degrees of belief? *Journal of Applied Logic*, 1:139–149, 2003.

[Kyburg, 1961] H. E. Kyburg, Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, Middletown, CT, 1961.

[Kyburg, 1990] H. E. Kyburg, Jr. *Science and Reason*. Oxford University Press, New York, 1990.

[Kyburg, 2007] H. E. Kyburg, Jr. Bayesian inference with evidential probability. In William Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, pages 281–296. King's College, London, 2007.

[Levi, 1977] I. Levi. Direct inference. *Journal of Philosophy*, 74:5–29, 1977.

[Levi, 1980] I. Levi. *The Enterprise of Knowledge*. MIT Press, Cambridge, MA, 1980.

[Levi, 2007] I. Levi. Probability logic and logical probability. In William Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.*, pages 255–266. College Publications, 2007.

[Paris, 1994] J. B. Paris. *The uncertain reasoner's companion*. Cambridge University Press, Cambridge, 1994.

[Pawlak, 1991] Z. Pawlak. *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer, Dordrecht, 1991.

[Rosenkrantz, 1977] R. D. Rosenkrantz. *Inference, method and decision: towards a Bayesian philosophy of science*. Reidel, Dordrecht, 1977.

[Seidenfeld, 2007] T. Seidenfeld. Forbidden fruit: When Epistemic Probability may not take a bite of the Bayesian apple. In William Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* King's College Publications, London, 2007.

[Teng, 2007] C. M. Teng. Conflict and consistency. In William L. Harper and Gregory Wheeler, editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg, Jr.* King's College Publications, London, 2007.

[Walley, 1991] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[Wheeler, 2004] G. Wheeler. A resource bounded default logic. In James Delgrande and Torsten Schaub, editors, *NMR 2004*, pages 416–422, 2004.

[Wheeler, 2007] G. Wheeler. Two puzzles concerning measures of uncertainty and the positive Boolean connectives. In José Neves, Manuel Santos, and José Machado, editors, *Progress in Artificial Intelligence, 13th Portuguese Conference on Artificial Intelligence*, LNAI 4874, pages 170–180, Berlin, 2007. Springer-Verlag.

[Williamson, 2005] J. Williamson. *Bayesian nets and causality: philosophical and computational foundations*. Oxford University Press, Oxford, 2005.

[Williamson, 2008] J. Williamson. Objective Bayesian probabilistic logic. *Journal of Algorithms in Cognition, Informatics and Logic*, 63:167–183, 2008.

[Williamson, 2009] J. Williamson. Aggregating judgements by merging evidence. *Journal of Logic and Computation*, 19(3): 461–473, 2009.