Synthese (2008) 163:341–356
DOI 10.1007/s11229-007-9298-y

# Objective Bayesianism with predicate languages

**Jon Williamson**

Received: 9 November 2007 / Accepted: 9 November 2007 / Published online: 20 March 2008

**Abstract** Objective Bayesian probability is often defined over rather simple domains, e.g., finite event spaces or propositional languages. This paper investigates the extension of objective Bayesianism to first-order logical languages. It is argued that the objective Bayesian should choose a probability function, from all those that satisfy constraints imposed by background knowledge, that is closest to a particular frequency-induced probability function which generalises the $\lambda = 0$ function of Carnap's continuum of inductive methods.

**Keywords** Bayesianism · Objective Bayesianism · Bayesian epistemology · Formal epistemology · Inductive logic · Probabilistic logic

## 1 Objective Bayesianism

Objective Bayesianism is an epistemological theory that encapsulates the following three tenets. (i) Your degrees of belief should be representable by probabilities. (ii) Your degrees of belief should be compatible with your empirical evidence, in the sense of being calibrated with known frequencies and chances. (iii) Where such evidence does not fully determine your degrees of belief, they should be as equivocal as possible.

In the simplest case there is a well worked-out formalism underpinning these tenets (see, e.g., Williamson 2005a, Chap. 5, 2007b). If you represent your domain by a finite partition $\Omega$ of elementary outcomes, apply the *Maximum Entropy Principle* (*maxent* for short): (i) your degrees of belief should be representable by a probability function $p$ over the subsets of $\Omega$; (ii) if you know that the chance function $p^*$ lies in a set $\mathbb{P}^*$ of probability functions on this domain then $p$ should lie in the closed convex

J. Williamson (✉)
Department of Philosophy, SECL, University of Kent, Canterbury CT2 7NF, UK
e-mail: j.williamson@kent.ac.uk

hull $[\mathbb{P}^*]$ of $\mathbb{P}^*$; (iii) you should then choose a probability function $p$ in $[\mathbb{P}^*]$ that is closest to the maximally equivocal function $p^=$ which assigns equal probability to each $\omega \in \Omega$. Here distance between probability functions is measured using *cross entropy*, $d(p, p^=) = \sum_{\omega \in \Omega} p(\omega) \log p(\omega)/p^=(\omega)$, where $0 \log 0$ is taken to be 0. (It is not hard to see that in this simple case the function that is closest to $p^=$ is the function $p$ that has maximum *entropy*, $H = -\sum_{\omega \in \Omega} p(\omega) \log p(\omega)$, and that there is a unique maximum entropy probability function $p$. Consequently the entropy of a probability function is often taken as a measure of its uncertainty or lack of commitment. The Maximum Entropy Principle requires that your degrees of belief be representable by this maximally non-committal probability function).

Alas this simple domain structure is not rich enough to cover many interesting problems to which we might like to apply objective Bayesianism. This simple case would apply to an agent whose language is a finite propositional language: here $\Omega$ is the set of atomic states $\pm r_1 \wedge \cdots \wedge \pm r_n$ involving propositional variables $r_1, \ldots, r_n$, and the probability of a sentence $\theta$ is $p(\theta) = \sum_{\omega \in \Omega, \omega \models \theta} p(\omega)$. However, there are interesting questions—such as how to learn from experience, how to assess scientific theories—that need to be formulated using more expressive languages. This motivates extending objective Bayesianism to uncountable domains and to countable logical languages that go beyond propositional languages.

Consider first uncountable domains. Cross entropy can be straightforwardly extended to measure the distance between probability functions defined over a continuous partition: one just integrates instead of sums when calculating the distance. Thus the general recipe goes through as before: choose $p \in [\mathbb{P}^*]$ that is closest to the maximally equivocal probability function $p^=$. However there may be no unique maximally equivocal probability function. There may be more than one way to equivocate, as witnessed by Bertrand's paradox (Keynes 1921, §4.7; Gillies 2000, pp. 37–49). Or there may be a unique way to equivocate but one which is not representable using a probability function—this is case with the problem of improper priors encountered by objective Bayesian statisticians (Kass and Wasserman 1996, §4.2). In this latter situation there are typically many probability functions that are closest to the improper equivocator. This non-uniqueness is not a show-stopper for objective Bayesianism; it just means that on uncountable domains objective Bayesianism admits a certain amount of subjective choice as to which degrees of belief to adopt.[1]

The problems facing predicate languages are, at first sight, more formidable. One might try to treat predicating statements, such as $Ra$, $Sb$, as if they were propositional variables, $r_a, s_b$.[2] This determines a countable propositional language; one can then apply maxent to finite subsets of this language and take limits to extend the

[1] This is a point of difference between objective Bayesianism and the logical interpretation of probability of Keynes (1921) and Carnap (1950) which does require full-blown uniqueness of probability.

[2] A *first-order predicate language* contains predicate and relation symbols $R, S, T, \ldots$, constant symbols $a, b, c, \ldots$, and variable symbols $x, y, z, \ldots$, as well as the logical connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$ and existential and universal quantifiers $\exists, \forall$. Often function symbols $f, g, h, \ldots$ and an equality symbol $=$ are included too, and brackets are used to enable parsing. Sentences are constructed using strings of these symbols; thus $\forall x \exists y (Px \rightarrow (Q(y, a) \vee \neg fb = c))$ can be read 'for all $x$ there is some $y$ such that if $P$ holds of $x$ then either $y$ stands in relation $Q$ to $a$ or it is not the case that $c$ is the result of applying function $f$ to $b$.'

resulting probability function to the language as a whole (Paris and Vencovská 2003). Unfortunately when there is no background knowledge the resulting probability function (which corresponds to $\lambda = \infty$ in Carnap's continuum of inductive methods) suffers from an inability to capture learning from experience. An agent who is observing ravens will initially believe that the next raven is black to degree 1/2. But even if she goes on to observe 100 ravens, all black, she will only believe the next raven will be black to degree 1/2. Thus she is unswayed by evidence.

In sum, if objective Bayesianism is to be applied to the question of learning from experience, a more sophisticated analysis is required. The goal of this paper is to put forward the principal elements of such an analysis. The plan is to consider successively richer predicate languages. The case of a finite language with a single unary predicate symbol is discussed in Sect. 2. The lessons learned from this case will then be applied to more complicated languages: languages that contain several unary predicate symbols in Sect. 3, languages with relation symbols in Sect. 4, infinite languages in Sect. 5, and languages with quantification in Sect. 6.

It is worth pointing out at an early stage that objective Bayesianism is not subjective Bayesianism. There are important differences between the two positions and one should draw analogies with caution. Subjective Bayesianism embodies tenet (i), that degrees of belief be probabilistic, as a core, though many 'empirically-based' subjective Bayesians go further by advocating tenet (ii), that beliefs be calibrated with evidence.[3] In a sense then, objective Bayesianism differs just in degree, adopting the further constraint (iii), that degrees of belief be otherwise equivocal. But this difference of degree veils important qualitative methodological differences. In particular (i) and (ii) are rather weak constraints: subject only to (i) and (ii), one may choose a belief function $p$ and then, on a minor change in one's evidence, formulate a new belief function $p'$ that is radically different to $p$. On account of this weakness, subjective Bayesians tend to advocate a further constraint, *Bayesian conditionalisation*: when your evidence changes by the addition of $e$ your new belief function $p'$ should be set to $p(\cdot|e)$, your old function conditional on the new evidence. This principle is notoriously problematic. It requires, for instance, that one must be able to formulate all future evidence as propositions in one's initial language; that one must totally specify one's initial belief function by asking oneself to what extent one believes each proposition conditional on each possible subsequent sequence of evidence; and that one never changes one's degrees of belief in the propositions that one believes to degree 0 or 1. Objective Bayesians need not be (and, in the version espoused here, are not) bound by Bayesian conditionalisation. By adopting (iii) in conjunction with (i) and (ii), objective Bayesian degrees of belief are highly if not fully constrained. On minor changes in evidence one's new degrees of belief will not admit changes as radical as under (i) and (ii) alone. For example, if the new evidence is in the domain of the probability function, this probability function gives non-zero probability to the

[3] Of course the word 'Bayesian' is used in different senses according to whether one is describing a theory of epistemology (concerning the strengths of our beliefs), an interpretation of probability (interpreting probabilities as degrees of belief), a branch of statistics (which advocates the use of prior probabilities), or a theory of confirmation (which explicates confirmation in terms of the probability of the hypothesis conditional on the evidence). Epistemology is our concern here.

new evidence, and learning the new evidence imposes the constraint that the evidence should be fully believed and no further constraints, then the maximum entropy update will agree with the results of conditionalisation (Williams 1980). Thus there is no need for a separate diachronic principle like conditionalisation. Without such a principle, objective Bayesians have certain freedoms not enjoyed by subjectivists: one's language can change in line with one's evidence (Williamson 2005a, Chap. 12); a belief function need not be fully specified (which has important computational advantages and makes it easier to maintain consistency); extreme probabilities can be revoked in the light of new evidence. Hence objectivism and subjectivism differ in important methodological respects.[4]

## 2 One predicate

In this section we shall consider the extension of objective Bayesianism to a very simple predicate language: a language which has a single unary predicate symbol, a large but finite number of constants and the usual logical connectives. This case is considered in detail in Williamson (2007a); the key strategies of that paper will be presented here, and defended in the face of some criticisms.

Recall the problem of learning from experience. Suppose one is examining ravens to see whether they are black. There is a single predicate symbol $B$, signifying *is black*, and constants $a_1, \ldots, a_k$ that refer to ravens in the order in which they are examined, where $k > 100$. Taking $\Omega = \{\pm Ba_1 \wedge \cdots \wedge \pm Ba_k\}$ and applying the Maximum Entropy Principle under no constraints yields $p(Ba_{101}) = 1/2$. While this seems reasonable enough, consider next the case in which the first hundred ravens have been observed to be black, $Ba_1 \wedge \cdots \wedge Ba_{100}$. This imposes the constraint $p'(Ba_1 \wedge \cdots \wedge Ba_{100}) = 1$, but maximising entropy under this constraint alone gives $p'(Ba_{101}) = 1/2$—no change, despite the overwhelming positive evidence.

While the problem of learning from experience might seem at first sight to reveal a flaw in the Maximum Entropy Principle, this is not in fact the case. Rather, the principle has been misapplied in this instance. To derive the problem it is assumed that initially there are no constraints, and that, once the ravens have been observed, there is a single constraint induced by the evidence. This overlooks important knowledge that is implicit in the language, namely that $Ba_1, \ldots, Ba_k$ are all related inasmuch as they are all applications of the same predicate. If this information is not taken into account then no connection between the observations can be made. If this information is to be ignored then maxent rightly renders the observations probabilistically independent—not a flaw after all.

[4] Note that some argue against objective Bayesianism on the grounds that there are cases in which objective Bayesian degrees of belief differ from those that would be obtained by Bayesian conditionalisation if one were to enlarge the probability space (see, e.g., Friedman and Shimony 1971; Dias and Shimony 1981; Seidenfeld 1979). Such arguments clearly beg the question: any difference does not in itself provide grounds to reject either prescription. A full discussion of this criticism of objective Bayesianism will be found in Williamson (2008).

Of course this implicit information should not simply be ignored. The question is, how should it be taken into account? What constraints does it impose on the agent's belief function $p$?

Arguably there are two types of constraint. First, learning that there will be a further observation $a_{k+1}$ should have no effect on one's degrees of belief concerning the first $k$ observations. More formally

$$p^{k+1}_{\downharpoonright k} = p^k,$$

i.e., one's probability function on the new language involving $a_1, \ldots, a_{k+1}$ should, when restricted to the language involving only $a_1, \ldots, a_k$, be equal to the function one should adopt if one just had this latter language. (In the terminology of Williamson (2005a), 'observed before' is an *influence relation*).

Second, past observations of the same predicate should be a guide to future observations: the greater the number of positive observations in the past, the stronger one should believe the next observation will be positive. Consider sequences of evidence of the form $B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n$, where $\varepsilon_i \in \{0, 1\}$, $B^1a_i$ is just $Ba_i$, and $B^0a_i$ is $\neg Ba_i$. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$, and let $x^{\varepsilon}_{n+1} \overset{\mathrm{df}}{=} p\,(Ba_{n+1}|B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n)$. Note that $x^{\varepsilon}_{n+1} = p_{\varepsilon}(Ba_{n+1})$, the degree to which one should believe $Ba_{n+1}$ having observed only $B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n$—in such cases maxent agrees with Bayesian conditionalisation.[5] Define $\varepsilon^+ \overset{\mathrm{df}}{=} \sum_{j=1}^{n} \varepsilon_j$, the number of observed positive instances. Then the second kind of constraint can be explicated thus:

$$x^{\varepsilon}_{n+1} \geq x^{\varepsilon'}_{n+1} + \tau_n, \quad \text{if } \varepsilon^+ > \varepsilon'^+,$$

where $\tau_n \geq 0$ is called the $n$th *inductive influence threshold*.

Assuming for the moment that appropriate inductive influence thresholds are known, maximising entropy subject only to these two constraints yields

$$x^{\varepsilon}_{n+1} = \frac{1 + \tau_n(\varepsilon^+ - \varepsilon^-)}{2},$$

where $\varepsilon^- \overset{\mathrm{df}}{=} n - \varepsilon^+$ is the number of observed negative instances and where $\tau_0 = 0$. Note that this determines the joint probability distribution $p$ via the identity

$$p\left(B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_k}a_k\right) = \prod_{n=0}^{k-1} p\left(B^{\varepsilon_{n+1}}a_{n+1}|B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n\right).$$

[5] This identity is derivable where the conditioned statement has non-zero probability. When $p\left(B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n\right) = 0$ one can simply define $p\left(Ba_{n+1}|B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n\right) \overset{\mathrm{df}}{=} p_{\varepsilon}\left(Ba_{n+1}\right)$.

In the absence of any further background knowledge, we can deduce plausible values for the inductive influence thresholds. If we assume (I) that $p(Ba_1) = p(Ba_2) = \cdots = p(Ba_k) = x$ say, and (II) that the inductive influence thresholds vary continuously with $x$, then $\tau_n = 1/n$ for $n \geq 1$ (Williamson 2007a, §9). Consequently,

$$x_1^{\varepsilon} = p(Ba_1) = 1/2, \tag{1}$$

$$x_{n+1}^{\varepsilon} = \mathrm{freq}_n(B) \overset{\mathrm{df}}{=} \frac{\varepsilon^+}{n}, \quad \mathrm{if} n \geq 1, \tag{2}$$

where $\mathrm{freq}_n(B)$ is just the frequency of $B$ in the first $n$ observations. (Note that these two equations imply that

$$p(B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_k}a_k) = \begin{cases} 1/2 & : \ \varepsilon_1 = \cdots = \varepsilon_k \\ 0 & : \ \text{otherwise} \end{cases}$$

Interestingly, the induced probability function corresponds to $\lambda = 0$ in Carnap's continuum of inductive methods $x_{n+1}^{\varepsilon} = (\varepsilon^+ + \lambda/2)/(n + \lambda)$, not to $\lambda = \infty$ as suggested by the original problem of learning from experience.

By means of this analysis we see that objective Bayesians can, after all, learn from experience: given evidence just of a 100 ravens, all black, the agent should be certain that the 100 and first raven will be black.

One might object to this resolution, arguing that we have just replaced an inability to learn from experience by the other extreme, namely a tendency to be led astray by experience. Thus if only a single raven is observed, and that raven is observed to be black, the recommendation seems to be that one should be certain that the next raven will be black, $p(Ba_2|Ba_1) = 1$. This appears to be rather over-hasty.

In response one should note that the above analysis depends on there being no further background knowledge. Of course, this is almost never the case in practice. In the example we know that ravens are involved and that they are tested to see whether they are black. Even if we know nothing about the colour of ravens, if we know a bit of biology then we know that colour may vary widely over a species, and even if it does not, if ravens were normally black there would be the distinct possibility of albino ravens. Even if it is not known to what the predicates refer, we know that for natural languages the chance of picking a population and a predicate $R$ such that $\mathrm{freq}(R) = 1$ is rather slender. Such items of generic knowledge should lead to more cautious assignments of belief than $p(Ba_2|Ba_1) = 1$. (When background knowledge takes this rather vague form it is not clear what precise constraints it imposes on an agent's degrees of belief. This challenge of knowledge imprecision affects objective Bayesianism and empirically-based subjective Bayesianism alike, and of course all the sciences. The most one can ask of objective Bayesianism is that, when input precisely formulated background knowledge, it can derive precise degrees of belief).

The above response sounds rather hollow without a well-defined procedure for determining degrees of belief when one does have (precisely-formulated) knowledge that goes beyond the two constraints imposed by the structure of the language. In that case, it is not clear how to determine the inductive influence thresholds. But progress can nevertheless be made by re-framing the problem as follows. Thus far it has been argued that the structure of the language imposes two constraints on degrees of belief; in the absence of further constraints an agent's belief function is determined by frequencies, $x_{n+1}^{\varepsilon} = \varepsilon^{+}/n$. To put it another way, on a logical language with a single unary predicate and finitely many constants, the most equivocal probability function $p^{=}$ is that determined by these frequencies via Eqs. 1 and 2 ($p^{=}$ is equivocal in the sense that condition I holds above). This reformulation suggests a more general solution that applies to the case in which there is further knowledge: in the presence of further constraints, choose the probability function, from all those that satisfy these constraints, that is closest to $p^{=}$, the maximally equivocal function with respect to this language. Since the further constraints narrow down a closed convex set of probability functions and cross entropy distance is a strictly convex function, this process will determine a unique probability function. (There is a technicality here: under the usual definition of cross entropy distance, uniqueness requires the assumption that there is some probability function satisfying the constraints that is zero wherever $p^{=}$ is zero; at the end of this section we shall see how the definition of cross entropy distance can be modified to avoid this assumption). Thus the procedure for determining a probability function under no further constraints induces a general procedure for determining a probability function under arbitrary constraints. In sum, both objections—the objection that the objective Bayesian cannot learn from experience and the objection that the objective Bayesian learns too quickly from experience—can be countered.

There is a further objection that is closely related to this latter charge of over-hasty learning. Carnap's $\lambda = 0$ function, which coincides with the maximally equivocal function advocated here, has often been criticised on the grounds that it gives probability 1 to all future observations being the same as the first. This seems to clash with intuition. Three responses are particularly pertinent. First, the previous reply applies equally here: typically there is knowledge that rules out this maximally equivocal function, such as knowledge that in the past the first observation has been a rather poor guide to the outcome of subsequent observations. Hence in the framework advocated here the $\lambda = 0$ function is a reference point rather than a representation of the degrees of belief that a realistic agent should adopt. Second, since the objective Bayesian updates degrees of belief by maxent rather than Bayesian conditionalisation, probability 1 statements are defeasible. So when it turns out that observations are not all the same as the first, one can simply revise one's degrees of belief accordingly. This is not an option for the advocate of Bayesian conditionalisation. Third, probability theory often clashes with intuition: if darts are thrown uniformly at random at a dartboard then then there is probability 0 of a dart hitting a particular point of a continuous dartboard, whatever the point, even though there is probability 1 of the dart hitting some point of the dartboard; a subjective Bayesian must give probability 1 to the proposition that in the long run her degrees of belief will tend to become perfectly calibrated with frequency, however ridiculous her prior (Dawid 1982); probabilistic considerations imply that a solid ball can be taken apart into finitely many pieces which can then be rearranged to construct a ball twice as large as the original (see, e.g., Wagon 1985). Further examples of probabilistic violations of intuition abound in the psychology literature (see, e.g., Tversky and Kahneman 1974). So if intuition is to be a deciding factor then it decides against all flavours of Bayesianism, and indeed against the use of probability at all.

Before proceeding to languages with more than one predicate, we turn to the technical point mentioned above. Cross entropy distance is normally defined as $d(p, p^{=}) = \sum_{\omega \in \Omega} p(\omega) \log p(\omega)/p^{=}(\omega)$. This is infinite if there is some $\omega \in \Omega$ such that $p(\omega) > 0$ but $p^{=}(\omega) = 0$. In the case of a propositional language the equivocator, which sets $p^{=}(\omega) = 1/|\Omega|$, is never zero so this case never arises. However the frequency-induced equivocator advocated above is zero on several $\omega$. Now if the constraints force $p(\omega) > 0$ for some $\omega$ such that $p^{=}(\omega) = 0$, then every probability function that satisfies the constraints is infinitely far from the equivocator, so distance from the equivocator cannot be used to further narrow down the choice of $p$. This is a limitation of using cross entropy as a measure of distance between probability functions: it provides a poor measure in such cases.[6] However, this limitation can be overcome by modifying the definition of distance, as follows. First observe that

$$d(p, p^{=}) = \sum_{n=0}^{k-1} \sum_{\varepsilon=(\varepsilon_1,\ldots,\varepsilon_n)} d^{\varepsilon}_{n+1}(p, p^{=}),$$

where

$$d^{\varepsilon}_{n+1}(p, p^{=}) \overset{\text{df}}{=} \sum_{\varepsilon_{n+1}=0}^{1} p\left(B_1^{\varepsilon_1} a_1 \cdots B_{n+1}^{\varepsilon_{n+1}} a_{n+1}\right) \times \log \frac{p\left(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\right)}{p^{=}\left(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\right)}.$$

$d^{\varepsilon}_{n+1}(p, p^{=})$ can be thought of as a local contribution to the total distance. It is a good measure of this local contribution, except where $p^{=}\big(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\big) = 0$ for some $\varepsilon_{n+1} \in \{0, 1\}$. In that case the measure is obviously too coarse: $d^{\varepsilon}_{n+1}(p, p^{=}) = \infty$ unless $p\left(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\right) = p^{=}\big(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\big)$. On the other hand it is intuitively clear what should happen in this situation: the closer $p\big(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\big)$ is to $p^{=}\big(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\big)$, the smaller the local contribution to the total distance. This motivates the use of a better distance measure in such cases. For example, we might define

$$d'^{\varepsilon}_{n+1}(p, p^{=}) = \Big[p\left(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\right) - p^{=}\left(B_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid B_1^{\varepsilon_1} a_1 \wedge \cdots \wedge B_n^{\varepsilon_n} a_n\right)\Big]^2.$$

[6] There are other technical limitations: cross entropy is not a distance function in the usual mathematical sense because it is not symmetric and does not satisfy the triangle inequality. Note though that one can take $d(p, q) + d(q, p)$ as a symmetric measure of distance; this is the Kullback–Leibler divergence of Kullback and Leibler (1951, p. 81).

This would generate a new distance measure:

$$\begin{aligned} D^{\varepsilon}_{n+1}(p, p^{=}) &= \begin{cases} d^{\varepsilon}_{n+1}(p, p^{=}) & : \quad d^{\varepsilon}_{n+1}(p, p^{=}) < \infty \\ d'^{\varepsilon}_{n+1}(p, p^{=}) & : \quad \text{otherwise} \end{cases} \\ D(p, p^{=}) &= \sum_{n=0}^{k-1} \sum_{\varepsilon \in \{0,1\}^n} D^{\varepsilon}_{n+1}(p, p^{=}). \end{aligned}$$

Clearly $D(p, p^{=}) \geq 0$ with equality iff $p = p^{=}$. Moreover, there is no $p$ for which $D(p, p^{=})$ is infinite, and the new distance function agrees with the old where the old is finite. $D$ is by no means an attractive *general* measure of distance because of the discontinuity in $D(p, q)$ as $q\left(B^{\varepsilon_{n+1}}_{n+1} a_{n+1} \mid B^{\varepsilon_1}_1 a_1 \wedge \cdots \wedge B^{\varepsilon_n}_n a_n\right)$ reaches 0. But because the second place of the distance relation is fixed to $p^{=}$ and because we are only interested in comparative distance relations (minimising distance rather than using exact magnitudes of distance), it is quite adequate here. Nothing much hinges on the exact choice of $d'^{\varepsilon}_{n+1}$ since different measures agree as to comparative distance relations in the local binary situation to which it applies.

## 3 Multiple predicates

Having dealt with the case of a language with a single unary predicate symbol, in this section we shall see how the approach might be extended to cover more predicates.

This situation is more complicated in the following respect. In the single-predicate case there is a connection between all the observations—they are all applications of the same predicate. In Sect. 2 it was suggested that this connection imposes two constraints: an invariance in degrees of belief under knowledge that there will be further observations, and a dependence imposed by the need for the past to be a guide to the future. With multiple predicates, though, there is another type of connection to be taken into account. Not only is it the case that $Ba_m$ and $Ba_n$ are connected (being instantiations of the same predicate), but also $Ba_n$ and $Ra_n$ are connected because they both concern the same individual. Unlike the intra-predicate constraints—in which earlier observations influence later observations but not vice versa—this inter-predicate connection is symmetric.

Despite this added complexity, the strategy for dealing with the single-predicate case carries over to the multiple-predicate case. Thus far we have seen that not only should objective Bayesian degrees of belief be calibrated with long-run frequency via tenet (ii) of Sect. 1, but in the absence of further constraints degrees of belief should also be calibrated with frequency in the short run (Sect. 2). For a single predicate $B$ there is an equivocator $p^{=}$ induced by frequency considerations, and one should set one's belief function to be as close as possible to this equivocator, subject to the constraints imposed by one's other background knowledge. Now if there are two predicates, $R$ for 'raven' and $B$ for 'black', consistency with the single-predicate case requires that the equivocator on the extended language give probabilities that coincide with those of Sect. 2. Perhaps the most natural candidate for such an equivocator is the following frequency-induced function (further justification for this choice will be provided in Sect. 4):

$$\begin{aligned} x^{\delta,\varepsilon}_{n+1} &\overset{\text{df}}{=} p\left(Ra_{n+1}|R^{\delta_1}a_1 \wedge \cdots \wedge R^{\delta_n}a_n \wedge B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n\right) \\ &= p_{\delta\varepsilon}\left(Ra_{n+1}\right) = \text{freq}_n(R) = \frac{\delta^+}{n}, \end{aligned}$$

$$\begin{aligned} y^{\delta:i,\varepsilon}_{n+1} &\overset{\text{df}}{=} p\left(Ba_{n+1}|R^{\delta_1}a_1 \wedge \cdots \wedge R^{\delta_n}a_n \wedge R^{i}a_{n+1} \wedge B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_n}a_n\right) \\ &= p_{\delta:i,\varepsilon}\left(Ba_{n+1}\right) = \begin{cases} \text{freq}_n(B|R) = \left(\sum_{i=1}^{n}\delta_i\varepsilon_i\right)\Big/\delta^+ & : \ i=1 \\ \text{freq}_n(B|\neg R) = \left(\sum_{i=1}^{n}(1-\delta_i)\varepsilon_i\right)\Big/\delta^- & : \ i=0 \end{cases} \end{aligned}$$

Here $\delta : i = (\delta_1, \ldots, \delta_n) : i \overset{\text{df}}{=} (\delta_1, \ldots, \delta_n, i)$ for $i \in \{0, 1\}$, and $\text{freq}_0(X)$ is taken to be 1/2, so that $p(Ra_1) = 1/2 = p(Ba_1|R^i a_1)$. Then

$$p\left(R^{\delta_1}a_1 \wedge \cdots \wedge R^{\delta_k}a_k \wedge B^{\varepsilon_1}a_1 \wedge \cdots \wedge B^{\varepsilon_k}a_k\right) = \prod_{i=1}^{k} x_i^{\delta,\varepsilon} y_i^{\delta,\varepsilon}.$$

More generally, if there are $r$ (unary) predicates $R_1, \ldots, R_r$ then the frequency-induced equivocator is defined by

$$\begin{aligned} x^{\varepsilon,j}_{n+1} &\overset{\text{df}}{=} p\Bigg(R_{j+1}a_{n+1}|R_1^{\varepsilon_1^1}a_1 \wedge \cdots \wedge R_1^{\varepsilon_n^1}a_n \wedge \cdots \wedge R_r^{\varepsilon_1^r}a_1 \wedge \cdots \wedge R_r^{\varepsilon_n^r}a_n \\ &\qquad \wedge R_1^{\varepsilon_{n+1}^1}a_{n+1} \wedge \cdots \wedge R_j^{\varepsilon_{n+1}^j}a_{n+1}\Bigg) \\ &= \text{freq}_n\left(R_{j+1}|R_1^{\varepsilon_{n+1}^1} \cdots R_j^{\varepsilon_{n+1}^j}\right) \end{aligned}$$

where this is taken to equal 1/2 if $n = 0$. Note that this is undefined if the condition has probability zero, $\text{freq}_n\left(R_1^{\varepsilon_{n+1}^1} \cdots R_j^{\varepsilon_{n+1}^j}\right) = 0$; in that case it is natural to set

$$x^{\varepsilon,j}_{n+1} \overset{\text{df}}{=} \frac{1}{m}\sum_{M} \text{freq}_n(R_{j+1}|M),$$

where the $M$ range over the $m$ maximal subsets of $\left\{R_1^{\varepsilon_{n+1}^1} \cdots R_j^{\varepsilon_{n+1}^j}\right\}$ for which $\text{freq}_n(M) \neq 0$.

## 4 Relations

Relations can be treated analogously to predicates. In fact the above procedure may be taken to characterise the equivocator $p^{=}$ in the case in which $R_1, \ldots, R_r$ are predicate or relation symbols and the $a_i$ are tuples of constant symbols. The objective Bayesian recipe is then to adopt as a representation of one's degrees of belief the probability function that is closest to this equivocator, from all those that satisfy constraints imposed by background knowledge.

There are, however, some considerations that are peculiar to relations.

First, it is rare to repeatedly sample $m$-tuples $a_i$ for $m > 1$. Normally when we learn from experience using relations we fix all but one place in the relation and perform monadic induction: if of five things sampled, all have been more-slithy-than the Jabberwock, I'll strongly believe that the next thing will also be more-slithy-than the Jabberwock. Here, although more-slithy-than $(x, y)$ is a binary relation, monadic learning takes place with more-slithy-than ($x$, Jabberwock). Thus we must admit the possibility that tuples of constants $a_i$ and $a_j$ have members in common when $i \neq j$. This has two consequences. For one thing, the same tuple may be examined more than once: if $a_{n+1} = a_i$ for $i < n + 1$ then $x_{n+1}^{\varepsilon,j} \in \{0, 1\}$ and so may differ from $\mathrm{freq}_n\left(R_{j+1}|R_1^{\varepsilon^1_{n+1}} \cdots R_j^{\varepsilon^j_{n+1}}\right)$. Having raised this possibility we shall not consider it further—for ease of exposition we shall assume that $a_i$ and $a_j$ differ in some respect for $i \neq j$. The other consequence is that for a language with constant symbols $a_1, \ldots, a_k$, the space of elementary outcomes is $\Omega = \left\{\bigwedge_{i=1}^r \bigwedge_a R_i^{\varepsilon_a^i} a : \varepsilon_a^i \in \{0, 1\}\right\}$, where the $a$ range over all sequences $(a_1, \ldots, a_m)$ where $m$ is the arity of the corresponding relation. Cross entropy distance between probability functions is summed over these elementary outcomes, which are usually known as *state descriptions*.

A second peculiarity concerns *order relations*. Some relations are used to order the elements of the domain, and one can normally tell this syntactically, from the use of prefixes such as 'more' or 'less'. Thus one can tell syntactically that 'more-slithy-than $(x, y)$' denotes an order relation, despite not knowing the meaning of 'slithy' and hence the precise order relation that is denoted. When the order yields a total order, one would expect the relation to be instantiated in around 50% of samples from a large domain: choosing $a$ and $b$ at random, it is as likely as not that more-slithy-than $(a, b)$. Hence if neither $x$ nor $y$ is fixed to a constant such as Jabberwock, one should not learn from experience on more-slithy-than $(x, y)$: if in five samples of $x$ and $y$, $x$ was more-slithy-than $y$, I'd attribute that to a freak sample and would not conclude that of the next two things I sample the first will be more-slithy-than the second. Properties and non-order relations, on the other hand, don't admit such obvious symmetry between positive and negative instantiations. Clearly, when it is known that a relation $R$ yields a total order, this knowledge should be factored into the calculation of objective Bayesian degrees of belief, perhaps via a default constraint of the form $p(Ra) = 1/2$. More generally, relations often have some structure—e.g., transitivity, irreflexivity—that is known from the outset (Bar-Hillel 1951; Carnap 1951). This information must be taken into account when determining objective Bayesian degrees of belief, by imposing the appropriate constraints and choosing the probability function, from all those that satisfy these constraints, that is closest to the frequency-induced equivocator.

The choice of the frequency-induced function as the equivocator $p^{=}$ is inevitable (modulo what to do when the conditioned event has probability zero) in the following respect. In Sect. 2 we required that the past be a guide to the future in the sense that $x^{\varepsilon}_{n+1} \geq x^{\varepsilon'}_{n+1} + \tau_n$ where $\tau_n$ is the $n$th inductive influence threshold. Suppose we require further of a potential equivocator $p^{=}$ that these thresholds are strictly positive (condition II of Sect. 2 achieves the same thing), and that the following principles hold:

**SX**: (Spectrum exchange ability) $p^{=}(\sigma)$ depends only on the *spectrum* of the state description $\sigma$, i.e., the multiset of the sizes of the equivalence classes of the indistinguishability relation (two constants are indistinguishable in $\sigma$ if they satisfy the same predicates and stand in the same relations to other constants).

**GJSP**: (Generalised Johnson's sufficientness postulate) $p^{=}(\sigma_{n+1}|\sigma_n)$ depends only on $n$ and the number of constants that are indistinguishable in $\sigma_{n+1}$ from $a_{n+1}$, where $\sigma_{n+1}$ is a state description involving $a_1, \ldots, a_{n+1}$ and $\sigma_n$ is its restriction to $a_1, \ldots, a_n$.

Then the frequency-induced function is the only possible choice for an equivocator. In the case of a language with relation symbols that are all binary and with no predicate symbols this follows directly from the proof of Vencovská (2006, Theorem 4); the same method can be used to demonstrate the general case.[7] Both SX and special cases of GJSP have been investigated in depth and found to be plausible—for the former principle see Nix (2005), Nix and Paris (2007), Paris and Vencovská (2007), Landes et al. (2007); for the latter see Johnson (1932), Carnap (1952), Hill et al. (2002), Vencovská (2006).

Finally note that $m$-ary functions can be construed as $(m + 1)$-ary relations. Thus functions with a finite domain and range fit into the framework developed above.

## 5 Infinite languages

Next we turn to the extension of this framework to the case in which there is a countable infinity of predicate symbols, relation symbols and constant symbols. At this point we depart from the simple situation in which the set of elementary outcomes is finite.

Consider an ordering $\alpha_1, \alpha_2, \ldots$ of the atomic sentences of the language—each $\alpha_i$ is of the form $Ra$ for some predicate or relation $R$ and for some tuple $a$ of constant symbols. Let $R_i$ be the predicate or relation symbol occurring in $\alpha_i$ and $a_i$ be the tuple of constants in $\alpha_i$, so $\alpha_i$ can be written $R_i a_i$.

The set of elementary outcomes is $\Omega = \left\{\varepsilon \overset{\text{df}}{=} (\varepsilon_1, \varepsilon_2, \ldots) : \varepsilon_i \in \{0, 1\}, i = 1, 2, \ldots\right\} = \{0, 1\}^{\infty}$. Here $\varepsilon_i$ signifies a value attaching to $\alpha_i$. A *cylinder of rank n* is a subset of elementary outcomes of the form $A = \{\varepsilon \in \Omega : (\varepsilon_1, \ldots, \varepsilon_n) \in H\}$ where $H \subseteq \{0, 1\}^n$. The set $\mathcal{C}_0$ of cylinders of all ranks is a field of subsets of $\Omega$ and a probability measure can be defined over this field. Every finitely additive probability measure defined on $\mathcal{C}_0$ is in fact countably additive (Billingsley 1979, Theorem 2.3), determined by its values on the *thin cylinders* $\{\varepsilon \in \Omega : (\varepsilon_1, \ldots, \varepsilon_n) = H\}$ where $H \in \{0, 1\}^n$, and uniquely extendible to the sigma field $\mathcal{C}$ generated by $\mathcal{C}_0$ (Billingsley 1979, Theorem 3.1).

[7] I am grateful to Jürgen Landes for this last point.

Any sentence $\theta$ in our extended language corresponds to a cylinder $C_\theta$ in $\mathcal{C}$ (Williamson 2002, §2). If the elementary outcomes are thought of as truth valuations of the atomic sentences $\alpha_i$, then $C_\theta$ is the set of valuations under which $\theta$ is true. Given a probability measure $p$ on $\mathcal{C}$, one can define $p(\theta) \overset{\text{df}}{=} p(C_\theta)$. Conversely a function $p$ that assigns a value in [0, 1] to each sentence of the language is a probability function if $p(C_\theta) \overset{\text{df}}{=} p(\theta)$ defines a probability measure on $\mathcal{C}$.

The state descriptions in our extended language are expressions of the form $R_1^{\varepsilon_1} a_1 \wedge \cdots \wedge R_n^{\varepsilon_n} a_n$. Each such state description $\sigma_n$ corresponds to a thin cylinder $C_{\sigma_n}$ in $\mathcal{C}$. Since its values on the thin cylinders determine a probability measure on $\mathcal{C}$, its values on the state descriptions determine a probability function on the language and $p(\theta) = \sum_{\sigma_n \models \theta} p(\sigma_n)$, where $n$ is chosen large enough that all the $\alpha_i$ in $\theta$ occur in $\alpha_1, \ldots, \alpha_n$. A function $p$ that assigns a value in [0, 1] to each state description determines a probability function if (a) $\sum_{\sigma_1} p(\sigma_1) = 1$ and (b) if $m < n$ then $p(\sigma_m) = \sum_{\sigma_n \models \sigma_m} p(\sigma_n)$.

The frequency-induced equivocator for the language is determined exactly as before:

$$p^{=}\left(R_1^{\varepsilon_1} a_1 \wedge \cdots \wedge R_n^{\varepsilon_n} a_n\right) = \prod_{i=1}^{n} \text{freq}_{i-1}\left(R_i^{\varepsilon_i} \mid R_1^{\varepsilon_1} \cdots R_{i-1}^{\varepsilon_{i-1}}\right),$$

with qualifications as described at the end of Sect. 3.

The definition of distance from the equivocator needs extending to cope with an infinite language. Previously, $D(p, p^{=}) = \sum_{n=0}^{k-1} \sum_{\varepsilon} D_{n+1}^{\varepsilon}(p, p^{=})$; now we require that $D(p, p^{=}) = \sum_{n=0}^{\infty} \sum_{\varepsilon} D_{n+1}^{\varepsilon}(p, p^{=})$. Note that this extension changes some of the properties of $D$. For instance, there are now probability functions $p$ for which $D(p, p^{=}) = \infty$, e.g., $p$ defined by $p\left(R_{n+1}^{\varepsilon_{n+1}} a_{n+1} \mid R_1^{\varepsilon_1} a_1 \cdots R_n^{\varepsilon_n} a_n\right) = 1$ for the sequence $\varepsilon_1 = 0, \varepsilon_{i+1} = 1 - \varepsilon_i$ $(i = 1, 2, \ldots)$ and for all $n = 0, 1, \ldots$. If the constraints only admit probability functions that are infinitely far from the equivocator then there will be nothing to choose between these functions; objective Bayesian probability will not be uniquely determined. Having said that, such a situation seems rather far-fetched—it would be a surprise if realistic background knowledge induced this kind of pathological set of constraints.

## 6 Quantifiers

Finally we extend the language further by including variables and the quantifiers ∃, ∀. The probability space remains as defined in Sect. 5, and we have that[8]

$$p(\exists x \theta x) \overset{\text{df}}{=} p(C_{\exists x \theta x}) = \sup_n p\left(\bigvee_{i=1}^{n} \theta a_i\right),$$

$$p(\forall x \theta x) \overset{\text{df}}{=} p(C_{\forall x \theta x}) = p(\neg \exists x \neg \theta x).$$

[8] Note that these definitions are only plausible under the assumption that each element of the domain is picked out by some constant in the language.

Any probability function over the whole language is uniquely determined by its values on the quantifier-free sentences of the language, which are in turn determined by its values on the state descriptions (Gaifman 1964; Paris 1994, Chap. 11), because as mentioned above a probability measure on the sigma field $\mathcal{C}$ of cylinders is determined by the probabilities of the thin cylinders.

Since the probability space remains the same, the frequency-induced equivocator is just as defined above, as is the measure of distance between probability functions.

Note that an agent's degrees of belief should satisfy all the axioms of probability, including countable additivity (Williamson 1999). Now limiting frequency does not necessarily satisfy countable additivity: suppose that unary predicates $R_1, R_2, \ldots$ are mutually exclusive and exhaustive and $R_i$ holds only of $a_i$; then $\mathrm{freq}_\infty(R_i) = 0$ for all $i$, contradicting countable additivity. But the frequency-induced equivocator $p^=$ is defined on the field of cylinder sets, and, as described above, we get countable additivity for free here for all finitely-additive probability measures on this domain. Consequently, unlike in the case of uncountable domains (Sect. 1), we do not have non-uniqueness of potential belief functions arising from a non-countably-additive equivocator.

## 7 Conclusion

We have seen that there is a plausible objective Bayesian strategy for handling predicate languages: choose as a representation of your degrees of belief the probability function, from all those that satisfy constraints imposed by background knowledge, that is closest to the frequency-induced equivocator. This fits well with (i) objective Bayesianism on finite propositional languages, which advocates choosing a belief function closest to the equivocator on such languages (though in this case the equivocator is the function that assigns equal probability to each atomic state). It also fits well with tenet (ii) of objective Bayesianism which advocates calibrating degrees of belief with best available estimates of frequencies and chances; any other choice of equivocator on a predicate language would apparently conflict with this tenet.

While this proposal overcomes the objection that objective Bayesianism cannot account for learning from experience, there is clearly more to do before a convincing case can be made that objective Bayesianism can handle the assessment of scientific theories. Of course, objective Bayesianism inherits challenges that face other flavours of Bayesianism: in particular, how to delimit background knowledge and render it precise, and how to determine the constraints imposed on degrees of belief by precisely-formulated background knowledge. It also faces its own challenges (Williamson 2007c)—e.g., how to compute the objective Bayesian probability function (*objective Bayesian nets* may be of some help here—see Williamson 2005b). But objective Bayesianism has certain advantages over other varieties of Bayesianism: it seems better set to account for the apparent objectivity of scientific confirmation, and, since Bayesian conditionalisation is avoided, it avoids objections that are attributable to this principle. Objective Bayesianism also has advantages over the logical interpretation of probability—uniqueness of objective Bayesian probability is not essential, and, unlike many of the classic approaches to inductive logic, it does not merely apply to the case in which an agent has no background knowledge at all (e.g., as described in Sect. 4, objective Bayesianism can handle known relational structure).

**Acknowledgements** This research was conducted as a part of the academic network *Probabilistic Logic and Probabilistic Networks* supported by The Leverhulme Trust. I am very grateful to Jürgen Landes, Jeff Paris and Alena Vencovská for helpful comments.

## References

Bar-Hillel, Y. (1951). A note on state-descriptions. *Philosophical Studies, 2*, 72–75.

Billingsley, P. (1979). *Probability and measure*, 3rd (1995) ed. New York: Wiley.

Carnap, R. (1950). *Logical foundations of probability*. London: Routledge and Kegan Paul.

Carnap, R. (1951). The problem of relations in inductive logic. *Philosophical Studies, 2*, 75–80.

Carnap, R. (1952). *The continuum of inductive methods*. Chicago, IL: University of Chicago Press.

Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association, 77*, 604–613. With discussion.

Dias, P. M. C., & Shimony, A. (1981). A critique of Jaynes' maximum entropy principle. *Advances in Applied Mathematics, 2*(2), 172–211.

Friedman, K., & Shimony, A. (1971). Jaynes's maximum entropy prescription and probability theory. *Journal of Statistical Physics, 3*(4), 381–384.

Gaifman, H. (1964). Concerning measures in first order calculi. *Israel Journal of Mathematics, 2*, 1–18.

Gillies, D. (2000). *Philosophical theories of probability*. London and New York: Routledge.

Hill, M., Paris, J., & Wilmers, G. (2002). Some observations on induction in predicate probabilistic reasoning. *Journal of Philosophical Logic, 31*, 43–75.

Johnson, W. E. (1932). Probability: The deductive and inductive problems. *Mind, 41*(164), 409–423.

Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association, 91*, 1343–1370.

Keynes, J. M. (1921). *A treatise on probability*. London: Macmillan (1948).

Kullback, S., & Leibler, R.(1951). On information and sufficiency. *The Annals of Mathematical Statistics, 22*, 79–86.

Landes, J., Paris, J., & Vencovská, A. (2007). The principle of conformity and spectrum exchangeability. In *Proceedings of the Sixth Conference on Foundations of the Formal Sciences: Reasoning about Probabilities and Probabilistic Reasoning*, Amsterdam.

Nix, C. (2005). Probabilistic induction in the predicate calculus. Ph.D. thesis, University of Manchester.

Nix, C., & Paris, J. (2007). A note on binary inductive logic. *Journal of Philosophical Logic*, *36*, 735–771.

Paris, J. B. (1994). *The uncertain reasoner's companion*. Cambridge: Cambridge University Press.

Paris, J. B., & Vencovská, A. (2003). The emergence of reasons conjecture. *Journal of Applied Logic, 1*(3–4), 167–195.

Paris, J. B., & Vencovská, A. (2007). From unary to binary inductive logic. In *Proceedings of the Second Indian Conference on Logic and its Relationship with Other Disciplines*, IIT Bombay.

Seidenfeld, T. (1979). Why I am not an objective Bayesian. *Theory and Decision, 11*, 413–440.

Tversky, A., Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Vencovská, A. (2006). Binary induction and Carnap's continuum. In *Proceedings of the 7th Workshop on Uncertainty Processing (WUPES)*, Mikulov. http://www.mtr.utia.cas.cz/wupes06/articles/.

Wagon, S. (1985). *The Banach-Tarski paradox*. Cambridge: Cambridge University Press.

Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science, 31*, 131–144.

Williamson, J. (1999). Countable additivity and subjective probability. *British Journal for the Philosophy of Science, 50*(3), 401–416.

Williamson, J. (2002). Probability logic. In D. Gabbay, R. Johnson, H. J. Ohlbach, & J. Woods (Eds.), *Handbook of the logic of argument and inference: The turn toward the practical* (pp. 397–424). Amsterdam: Elsevier.

Williamson, J. (2005a). *Bayesian nets and causality: Philosophical and computational foundations*. Oxford: Oxford University Press.

Williamson, J. (2005b). Objective Bayesian nets. In S. Artemov, H. Barringer, A. S. d'Avila Garcez, L. C. Lamb, & J. Woods (Eds.), *We will show them! Essays in honour of Dov Gabbay* (Vol. 2, pp. 713–730). London: College Publications.

Williamson, J. (2007a). Inductive influence. *British Journal for the Philosophy of Science*, *58*, 689–708.

Williamson, J. (2007b). 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.* (pp. 151–179) London: College Publications.

Williamson, J. (2007c). Philosophies of probability: Objective Bayesianism and its challenges. In A. Irvine (Ed.), *Handbook of the philosophy of mathematics* (Vol. 4). Elsevier.

Williamson, J. (2008). Objective Bayesianism, Bayesian conditionalisation and voluntarism. *Synthese* (in press).