Chapter 13
Epistemic Complexity from an Objective Bayesian Perspective

Jon Williamson

13.1 Introduction

This paper will focus on a particular kind of epistemic complexity, namely complexity of evidence. In particular we will look at the question of how complex evidence should impact on the strengths of an agent's beliefs.

It is a platitude to say that the strengths of our beliefs should depend on our available evidence, but it is notoriously hard to say exactly *how* evidence constrains appropriate degrees of belief. Bayesian epistemology begins to tackle this question, but typically considers only the simplest kinds of evidence, e.g., the case in which the evidence consists of a set of atomic propositions, or the case in which the evidence consists of a large database of good quality data. Reality, of course, is rarely if ever so simple. Evidence can be structured in a number of ways – causally, hierarchically, logically, for instance – and tends to be multifarious, a mixture of different kinds of structure from a mixture of different sources.

In this paper I will show how *objective Bayesianism* – one particular version of Bayesian epistemology – can help shed light on the precise relation between complex evidence and belief. Causal evidence will be considered in Section 13.4, hierarchically structured evidence in Section 13.5, logical structure in Section 13.6, and varied structure in Section 13.7. First, a crash-course on objective Bayesianism.

13.2 Objective Bayesian Epistemology

Some preliminaries An agent's *language* $\mathcal{L}$ is the means by which she expresses the propositions that concern her. Her *epistemic background* or *evidence* $\mathcal{E}$ is taken to consist of everything she *takes for granted* in her current operating context. This includes background knowledge, observations, theoretical assumptions and so on. (We will not assume that this evidence is in any way articulable, let alone articulable in $\mathcal{L}$.)

J. Williamson (✉)
Department of Philosophy and Centre for Reasoning, University of Kent, U.K.
e-mail: j.williamson@kent.ac.uk

A. Carsetti (ed.), *Causality, Meaningful Complexity and Embodied Cognition*, Theory and Decision Library A 46, DOI 10.1007/978-90-481-3529-5_13,

According to objective Bayesian epistemology, the agent's beliefs should satisfy certain norms, the first of which says:

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

Suppose, for example, that the agent's language $\mathcal{L}$ can express n different elementary (i.e., non logically complex) propositions $A_1, \ldots, A_n$. An *atomic state* ω on $\mathcal{L}$ is a sentence of the form $A_1^{j_1} \wedge \cdots \wedge A_n^{j_n}$ where $j_1, \ldots, j_n \in \{0, 1\}$, A_i^0 is $\neg A_i$ and A_i^1 is just A_i. Let Ω be the set of atomic states. Then the Probability norm says that the strengths of the agent's beliefs in the 2^n atomic states should be representable by non-zero real numbers that sum to 1; the degree to which she should believe an arbitrary proposition θ should be representable by the sum of her degrees of belief in those atomic states that logically imply θ. Thus the strengths of the agent's beliefs should be representable by a *probability function* over $\mathcal{L}$: a function P such that (i) $P(\omega) \geq 0$ for each ω, (ii) $\sum_{\omega \in \Omega} P(\omega) = 1$, and (iii) $P(\theta) = \sum_{\omega \in \Omega, \omega \models \theta} P(\omega)$.[1]

A second norm says that beliefs should fit with evidence:

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

Evidence $\mathcal{E}$ can constrain degrees of belief in a variety of ways. If $\mathcal{E}$ implies that proposition θ is true, then the agent should fully believe θ. More generally, if $\mathcal{E}$ implies that the empirical probability function P^* on $\mathcal{L}$ lies in a non-empty set $\mathbb{P}^*$ of probability functions, then the probability function $P_{\mathcal{E}}$ that represents the degrees of belief that the agent should adopt on the basis of $\mathcal{E}$ lies in the convex hull $[\mathbb{P}^*]$ of $\mathbb{P}^*$. Other kinds of constraint imposed by $\mathcal{E}$ will be discussed in subsequent sections of this paper. Let $\mathbb{E}$ denote the set of probability functions that are compatible with the agent's evidence (e.g., $\mathbb{E} = [\mathbb{P}^*]$). Then the calibration norm says that $P_{\mathcal{E}} \in \mathbb{E}$.[2]

The third norm says that beliefs should only be as bold as evidence warrants:

Equivocation Degrees of belief should otherwise be as equivocal as possible.

Here 'be as equivocal as possible' is just 'be as close as possible to maximally equivocal'.[3] The probability function that is maximally equivocal – the *equivocator* $P_=$ on $\mathcal{L}$ – is the function that gives each atomic state the same probability, $P_=(\omega) = 1/2^n$ for all $\omega \in \Omega$. The distance from one probability function to another is measured by *cross entropy*, $d(P, Q) = \sum_\omega P(\omega) \log P(\omega)/Q(\omega)$.

[1] This norm is typically justified by an appeal to a Dutch book argument or Cox's theorem – see, e.g., Paris (1994, Chapter 3).

[2] This norm is typically justified on the grounds that degrees of belief are used to make predictions, and calibrated degrees of belief lead to optimal predictions in the long run (Howson and Urbach, 1989, §13.e). Strictly speaking $P_{\mathcal{E}}$ depends on $\mathcal{L}$ as well as $\mathcal{E}$; we will write $P_{\mathcal{E}}^{\mathcal{L}}$ where we need to emphasise this dependence, but drop reference to $\mathcal{L}$ and $\mathcal{E}$ where the context permits. Williamson (2005, Chapter 12) discusses language change in the context of objective Bayesianism.

[3] This norm may be justified on the grounds that degrees of belief are used as a basis for action, extreme degrees of belief lead to riskier actions, and one should only take on risk to the extent that evidence demands – see Williamson (2007).

Distance to the equivocator, $d(P, P_=) = \sum_\omega P(\omega)\log(2^n P(\omega))$, is minimised just when *entropy* $H(P) = -\sum_\omega P(\omega)\log P(\omega)$ is maximised. Hence our three norms give us:

Maximum Entropy Principle An agent's degrees of belief should be representable by a probability function $P_\mathcal{E}$, from all those that satisfy constraints imposed by evidence $\mathcal{E}$, that has maximum entropy: $P_\mathcal{E} \in \{P \in \mathbb{E} : P = argmax H\}$.

Note that once we have these three norms we don't need any further principle to guide the updating of degrees of belief in the light of new evidence. As the evidence $\mathcal{E}$ changes to $\mathcal{E}'$, the agent's belief function will correspondingly change from $P_\mathcal{E}$, a maximally equivocal probability function from those compatible with $\mathcal{E}$, to $P_{\mathcal{E}'}$, a maximally equivocal probability function from those compatible with $\mathcal{E}'$. On a language $\mathcal{L}$ expressing n elementary propositions, $P_\mathcal{E}$ and $P_{\mathcal{E}'}$ are selected by successive applications of the maximum entropy principle (Williamson 2008).

13.3 Objective Bayesian Nets

Objective Bayesianism tells us how we should set our degrees of belief. Of course we can only be expected to follow the norms of objective Bayesianism to the extent that we *can* follow these norms. But following these norms is non-trivial: abiding by the maximum entropy principle is at first sight computationally demanding, since the number 2^n of atomic states grows exponentially with the number n of expressible elementary propositions. Fortunately, there are computational tools that mitigate this computational challenge. The machinery of *objective Bayesian nets* allows one to compute objective Bayesian probabilities more efficiently. In this section we shall introduce the concepts of Bayesian net and objective Bayesian net.

First some notation. For $i = 1, \ldots, n$ the propositional variable A_i takes one of two possible values, *true* or *false*; let a_i or a_i^1 signify the assignment $A_i = true$ and $\bar{a}_i$ or a_i^0 signify the assignment $A_i = false$. It is taken for granted that an agent's degree of belief that a proposition is true (respectively false) is just her degree of belief in the proposition itself (respectively in its negation): $P(a_i) = P(A_i)$ and $P(\bar{a}_i) = P(\neg A_i)$.

A *Bayesian net* offers an efficient way of representing and manipulating a probability function. A Bayesian net on $A_1, \ldots, A_n$ consists of a directed acyclic graph whose nodes are $A_1, \ldots, A_n$, as in Fig. 13.1 for instance, together with the probability distribution $P(A_i|Par_i)$ of each variable conditional on its parents in the graph. An assumption called the *Markov Condition* is made; this says that each

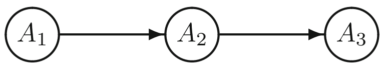

Fig. 13.1 A directed acyclic graph

variable is probabilistically independent of its non-descendants in the graph, conditional on its parents, written $A_i \perp\!\!\!\perp ND_i \mid Par_i$. Under this assumption, the Bayesian net determines a probability function over $\mathcal{L}$ via the identity $P(\omega) = P(A_1^{j_1} \wedge \cdots \wedge A_n^{j_n}) = \prod_{i=1}^n P(a_i^{j_i}|par^\omega)$, where par^ω is the assignment to Par_i that is determined by ω, and where $j_1, \ldots, j_n \in \{0, 1\}$. Conversely, any probability function P over a finite language can be represented by a Bayesian net: simply (i) determine the independencies that are satisfied by P, (ii) represent as many of these as possible by a directed acyclic graph satisfying the Markov condition with respect to P, and (iii) add the conditional probability functions $P(A_i|Par_i)$. A wide variety of algorithms have been developed for calculating probabilities from a Bayesian net. If the graph in the Bayesian net is relatively sparse, the size of the net can increase sub-exponentially with n, meaning that it may be computationally feasible to represent and reason with a probability function even where n is very large.

An *objective Bayesian net* is just a Bayesian net that represents an objective Bayesian probability function $P_\mathcal{E}$, which in turn represents degrees of belief that are appropriate on the basis of evidence $\mathcal{E}$. An objective Bayesian net can be constructed by (i) determining conditional independencies that $P_\mathcal{E}$ must satisfy; (ii) representing these independencies by a directed acyclic graph, and (iii) maximising entropy to find the conditional probability distributions $P_\mathcal{E}(A_i|Par_i)$. Fortunately a maximum entropy function $P_\mathcal{E}$ will normally satisfy a large number of probabilistic independencies. Construct an undirected graph by linking two variables if they both occur in the same constraint imposed by $\mathcal{E}$: then $X \perp\!\!\!\perp Y \mid Z$ for $P_\mathcal{E}$ if in this undirected graph the variables in Z separate those in X from those in Y. Hence the graph in an objective Bayesian net will typically be sparse and it will typically be feasible to handle objective Bayesian probabilities.

For example, suppose $\mathcal{E}$ imposes the following constraints: $P(A_1|\neg A_2) \geq 0.7$, $P(A_2 \vee A_4) = P(A_3)$, $P(\neg A_5 \wedge \neg A_3) = 0$, $P(A_4) \in [0.4, 0.5]$. Then Fig. 13.2 represents the independencies of $P_\mathcal{E}$. This can be transformed into a directed acyclic graph Fig. 13.3 that represents the same independencies via the Markov Condition. All that remains is to determine the conditional probability distributions. See Williamson (2005, Chapter 5) for a full algorithm for constructing an objective Bayesian net.

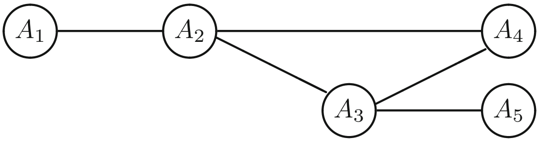

Fig. 13.2 An undirected graph representing the independencies of $P_\mathcal{E}$

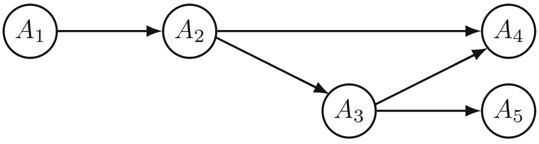

Fig. 13.3 A directed acyclic graph representing the independencies of $P_\mathcal{E}$ via the Markov Condition

13.4 Causal Structure

In Section 13.2 we saw that evidence of empirical probability constrains degrees of belief in a rather straightforward way: the set of probability functions compatible with evidence is just the convex hull of the set of functions in which (according to the evidence) the empirical probability function lies – written $\mathbb{E} = [\mathbb{P}^*]$. But evidence can contain information other than information about empirical probability, and the question arises as to what constraints $\mathcal{E}$ imposes on degrees of belief in such cases. In this section we shall look at the case in which evidence of causal relations is available to the agent.

Suppose for example that the agent grants the following evidence $\mathcal{E}$: smoking causes bronchitis and lung cancer, each of which cause chest pains; 30% of the population are smokers, 4% of the population but 10% of smokers get bronchitis, 2% of the population but 5% of smokers get lung cancer, 5% of the population but 99% of those with bronchitis or lung cancer have chest pains, Bob is a non-smoker with chest pains. Suppose further that $\mathcal{L}$ can express the elementary propositions S: *Bob is a smoker*, B: *Bob has bronchitis*, L: *Bob has lung cancer*, C: *Bob has chest pains*. The agent's causal evidence can be represented as in Fig. 13.4.

Causal evidence imposes constraints on degrees of belief in the following way. Causality is an *influence relation* in the sense that learning just of new non-influences provides no grounds for changing degrees of belief (Williamson 2005). More precisely, if the language $\mathcal{L}$ is extended to $\mathcal{L}'$, which expresses a new proposition, and it is known that the corresponding variable is not a cause of any of the former variables, and other information in $\mathcal{E}$ does not indicate otherwise, then the agent's degrees of belief over the former language should not change: $P_{\mathcal{E}}^{\mathcal{L}'}(\theta) = P_{\mathcal{E}_{\mathcal{L}}}^{\mathcal{L}}(\theta)$ for each sentence θ of $\mathcal{L}$, where $\mathcal{E}_{\mathcal{L}}$ is the evidence in $\mathcal{E}$ that concerns $\mathcal{L}$. Hence causal evidence imposes equality constraints on degrees of belief.

In our example $P_{\mathcal{E}}^{\mathcal{L}}(S) = P_{\mathcal{E}_{\{S\}}}^{\{S\}}(S)$ is but one such constraint. In fact these equality constraints ensure that the objective Bayesian net can be constructed by taking the causal graph Fig. 13.4 as the directed acyclic graph, and by iteratively maximising entropy to find the conditional probability distributions. See Williamson (2005, §5.8) for a detailed description of the procedure for constructing an objective Bayesian net in the presence of causal constraints. Ignoring for the moment the information that Bob is a non-smoker with chest pains, the objective Bayesian net has conditional probability distributions specified by:

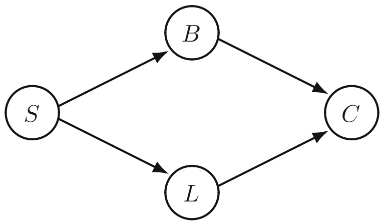

Fig. 13.4 Smoking causes Bronchitis and Lung Cancer, each of which cause Chest Pains

$$P(s) = \frac{3}{10};$$

$$P(b|s) = \frac{1}{10}, P(b|\bar{s}) = \frac{1}{70};$$

$$P(l|s) = \frac{1}{20}, P(l|\bar{s}) = \frac{1}{140};$$

$$P(c|bl) = \frac{99}{100}, P(c|\bar{b}l) = \frac{99}{100}, P(c|b\bar{l}) = \frac{99}{100}, P(c|\bar{b}\bar{l}) = \frac{40491}{659100}.$$

Now if we take the information specific to Bob into account by instantiating S to $\bar{s}$ and C to c in the network we get $P_{\mathcal{E}}(b) = P(b|c\bar{s}) \simeq 0.65$ and $P_{\mathcal{E}}(l) = P(l|c\bar{s}) \simeq 0.33$.

13.5 Hierarchical Structure

Causal structure provides one kind of evidential complexity, but there are others. In this section we shall look at evidence of hierarchical structure. Hierarchical structure occurs in descriptions of mechanisms. For instance, in describing mechanisms in the human body we often need to talk simultaneously about processes that occur at the level of the body as a whole (e.g., the circulation of the blood), those at the level of the cell (e.g., oxygenation of haemoglobin), and those at the level of the genome (e.g., mutation of a single nucleotide of the β-globin gene). Hierarchical structure also occurs in describing causal relationships, because causal relations can themselves act as causes and effects. For example, *smoking causing cancer* causes governments to restrict tobacco advertising, which prevents smoking and thereby prevents cancer (Fig. 13.5). This example shows that the same variable can occur at more than one level in the hierarchy.

Consider a simple example of hierarchically structured evidence. The National Farmer's Union needs to decide whether to lobby government for more subsidies. The evidence globally is that lobbying L is a cause of national agricultural policy A (Fig. 13.6). Here A is a hierarchical variable: one assignment a corresponds to the case in which farming F causes subsidy S (Fig. 13.7); a second assignment $\bar{a}$ is the case in which there is no link between farming and subsidy (Fig. 13.8). The evidence is that if farming causes subsidy, $P^*(s|f) = 1$, but 5% of the population receive

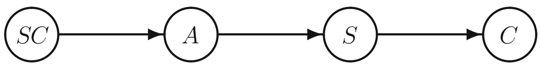

Fig. 13.5 SC: smoking causes cancer; A: tobacco advertising; S: smoking; C: cancer

Fig. 13.6 Lobbying L causes Agricultural Policy A

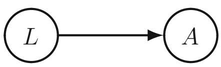

Fig. 13.7 a: Farming F causes Subsidy S

Fig. 13.8 $\bar{a}$: Farming F unrelated to Subsidy S

subsidies in any case, since fishing and other industries are subsidised. 10% of the population are farmers, and lobbying raises the probability of getting policy a by 20%.

In order to make sense of such an example we need to be clear about how evidence of hierarchical structure constrains degrees of belief. Call variable A *superior* to variable B if A occurs at a higher level in the hierarchy to B. Plausibly, *superiority* is an influence relation: learning of a new variable that is not superior to any of the current variables (and is not an influence in another respect – e.g., a causal influence) provides no grounds for changing degrees of belief concerning the current variables. So if the agent's language changes from $\mathcal{L}$ to $\mathcal{L}'$ and it is known that new propositions are not hierarchically superior to the old, then the agent's degrees of belief over the old language should not change: $P_{\mathcal{E}}^{\mathcal{L}'}(\theta) = P_{\mathcal{E}_{\mathcal{L}}}^{\mathcal{L}}(\theta)$ for each sentence θ of $\mathcal{L}$, where $\mathcal{E}_{\mathcal{L}}$ is the evidence in $\mathcal{E}$ that concerns $\mathcal{L}$. Hence hierarchical evidence imposes equality constraints on degrees of belief in the same way that causal evidence imposes such constraints.

Our example contains a mixture of causal and hierarchical evidence, but since both are evidence of influence relations, both can be treated alike. In this case the objective Bayesian net is a hierarchical or *recursive* Bayesian net (Williamson 2005). At the higher level is a network based on Fig. 13.6 – here the conditional probabilities are:

$$P(l) = \frac{1}{2};$$

$$P(a|l) = \frac{3}{5}, P(a|\bar{l}) = \frac{2}{5}.$$

At the lower level, the network for a, based on Fig. 13.7, has probabilities

$$P_a(f) = \frac{1}{10};$$

$$P_a(s|f) = 1, P_a(s|\bar{f}) = \frac{1}{20}.$$

The network for $\bar{a}$, based on Fig. 13.8, has probabilities

$$P_{\bar{a}}(f) = \frac{1}{10};$$

$$P_{\bar{a}}(s) = \frac{1}{20}.$$

Given this hierarchical net, the probability of a farmer receiving subsidy after lobbying, $P(s|lf)$, is 0.62, while with no lobbying $P(s|\bar{l}f) = 0.46$. These probabilities can be helpful for calculating the change lobbying will make to the expected subsidy, and thus helpful for the decision facing the National Farmer's Union.

13.6 Logical Structure

In the case of logical structure, we shall consider two kinds of evidential complexity. The first kind involves evidence of logical implications. The second kind involves evidence concerning the probabilities of logically complex propositions.

Logical Influence

The first kind – evidence of logical implications – proceeds analogously to the cases of causal structure (Section 13.4) and hierarchical structure (Section 13.5). To take a rather elementary example, suppose the agent's evidence includes the knowledge that $\theta \rightarrow \varphi$ and that θ logically implies φ. As well as this logical structure, the agent knows that if Socrates was a man then he was mortal ($A \rightarrow B$), and that it is as at least twice as likely as not that Socrates existed and was a man (A).

Now logical connection is an influence relation: a new proposition that does not, together with some current propositions, logically imply a current proposition, provides no grounds for changing degrees of belief over the current propositions. So if the agent's language changes from $\mathcal{L}$ to $\mathcal{L}'$ and it is known that new propositions are not influences (logical or otherwise) of the old, then the agent's degrees of belief over the old language should not change: $P_{\mathcal{E}}^{\mathcal{L}'}(\theta) = P_{\mathcal{E}_{\mathcal{L}}}^{\mathcal{L}}(\theta)$ for each sentence θ of $\mathcal{L}$, where $\mathcal{E}_{\mathcal{L}}$ is the evidence in $\mathcal{E}$ that concerns $\mathcal{L}$. Hence logical evidence imposes equality constraints on degrees of belief in the same way that causal or hierarchical evidence imposes such constraints.

In our example, the objective Bayesian net has the graph of Fig. 13.9. (Here C can be considered to be a hierarchical variable where assignment c corresponds to a net whose graph has nodes A and B and an arrow from A to B.) The probabilities are

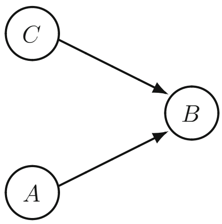

Fig. 13.9 C: if Socrates was a man then he was mortal; A: Socrates was a man; B: Socrates was mortal

$$P(c) = 1;$$

$$P(a) = \frac{2}{3};$$

$$P(b|ca) = 1, P(b|\bar{c}a) = \frac{1}{2}, P(b|c\bar{a}) = \frac{1}{2}, P(b|\bar{c}\bar{a}) = \frac{1}{2}.$$

In particular, the agent should believe that Socrates was mortal to degree $P(b) = 5/6$. See Williamson (2005, Chapter 11) for a full discussion of logical influence.

Predicate Languages

The second kind of complexity arises where the agent's evidence concerns logically complex propositions. The framework of Section 13.2 already handles arbitrary propositions in the propositional calculus. For instance, if the evidence says just that the physical probability of proposition θ is at least 0.8, $P^*(\theta) \geq 0.8$, then the agent's degrees of belief should be representable by probability function $P_{\mathcal{E}}$ which is closest to the equivocator, from all those in $\mathbb{E} = [\mathbb{P}^*] = \mathbb{P}^* = \{P : P(\theta) \geq 0.8\}$. But the question arises as to how handle evidence and beliefs concerning propositions with predicates, relations, constants, variables, quantifiers, etc. – i.e., propositions expressed in a predicate language.

If $\mathcal{L}$ is a predicate language, then the objective Bayesian method can be developed by appealing to the same three norms introduced in Section 13.2. Let $A_1, A_2, \ldots$ enumerate the *atomic propositions* of $\mathcal{L}$, i.e., the statements of the form Ut where U is a predicate or relation symbol and $t = (t_1, \ldots, t_k)$ is a tuple of constants of corresponding arity. An atomic n-state ω_n is an atomic state involving the first n of these atomic propositions: ω_n has the form $A_1^{j_1} \wedge \cdots \wedge A_n^{j_n}$ where $j_1, \ldots, j_n \in \{0, 1\}$. Let Ω_n be the set of atomic n-states.

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

Here a probability function is a function P such that (i) $P(\omega_n) \geq 0$ for all ω_n, (ii) for each n, $\sum_{\omega_n \in \Omega_n} P(\omega_n) = 1$, (iii) for quantifier-free θ, $P(\theta) = \sum_{\omega_n \in \Omega_n, \omega_n \models \theta} P(\omega_n)$ where n is chosen large enough such that $A_1, \ldots, A_n$ includes all the atomic propositions in θ. Note that quantified sentences can be assigned probabilities as follows: $P(\exists x \theta(x)) = \lim_{m \to \infty} P(\bigvee_{i=1}^m \theta(t_i))$ and $P(\forall x \theta(x)) = \lim_{m \to \infty} P(\bigwedge_{i=1}^m \theta(t_i))$, where the $t_1, t_2, \ldots$ are the constant symbols, and where it is assumed that each element of the domain is named by precisely one constant symbol.

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

Here, as before, the set $\mathbb{E}$ of probability functions compatible with evidence $\mathcal{E}$ is determined as follows. First take the convex closure of the set $\mathbb{P}^*$ of probability functions in which the empirical probability function is presumed to lie. Then remove those functions which do not satisfy the equality constraints imposed by structural evidence – e.g., evidence of causal, hierarchical or logical influence considered above. Hence $\mathbb{E} = [\mathbb{P}^*] \cap \mathbb{S}$ where $\mathbb{S}$ is the set of probability functions that satisfy the structural constraints.

Equivocation Degrees of belief should otherwise be as equivocal as possible.

Again, 'be as equivocal as possible' is just 'be as close as possible to maximally equivocal'. The equivocator is defined by $P_=(\omega_n) = 1/2^n$ for all ω_n. Let $d_n(P, Q) = \sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n)/Q(\omega_n)$. Then take P to be closer to the equivocator than Q if there is some N such that for all $n \geq N, d_n(P, P_=) < d_n(Q, P_=)$. Thus the recipe is just as for the propositional case outlined in Section 13.2: the agent's degrees of belief should be representable by a probability function $P_\mathcal{E}$ from $\mathbb{E}$ that is closest to the equivocator.

Note that P is closer to the equivocator than Q if there is some N such that for all $n \geq N$, $H_n(P) > H_n(Q)$, where H_n is the n-entropy defined by $H_n(P) = -\sum_{\omega_n \in \Omega_n} P(\omega_n) \log P(\omega_n)$. If we deem P to have *greater entropy than* Q if this condition holds (i.e., $\exists N, \forall n \geq N, H_n(P) > H_n(Q)$), then we have a version of the maximum entropy principle for predicate languages:

Maximum Entropy Principle An agent's degrees of belief should be representable by a probability function $P_\mathcal{E}$, from all those that satisfy constraints imposed by evidence $\mathcal{E}$, that has maximum entropy in the sense outlined above.

Consider a simple example. Suppose that the agent's evidence says that *all men are mortal* has empirical probability at least $3/4$, *All those who are virtuous are men* has empirical probability at least $3/5$, and that Socrates is virtuous has probability $4/5$. The graph of the resulting objective Bayesian net is depicted in Fig. 13.10. The corresponding probabilities are

$$P(v) = \frac{4}{5};$$

$$P(h|v) = \frac{3}{4}, P(h|\bar{v}) = \frac{1}{2};$$

$$P(m|h) = \frac{5}{6}, P(m|\bar{h}) = \frac{1}{2}.$$

It turns out then that the agent should believe that Socrates is mortal to degree $P(m) = 11/15$. See Haenni et al. (2010) for more on objective Bayesianism with predicate languages, and on how to construct objective Bayesian nets in such cases.

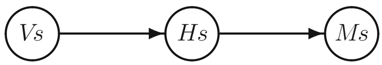

Fig. 13.10 V: virtuous; H: (hu)man; M: mortal; s: Socrates

13.7 Varied Evidence

Examples concerning the mortality of Socrates can seem remote from practical applications; in this section we shall look at a more realistic case study which exhibits a variety of kinds of evidence.

We will consider the application of objective Bayesian nets to breast cancer prognosis, described in detail in Nagl et al. (2008). The problem here is that a patient has breast cancer and an agent must make an appropriate treatment decision. Some treatments have harsh side-effects and it would not be justifiable to inflict these on low-risk patients. Broadly speaking, the higher the probability of recurrence of the cancer, the more aggressive the treatment that should be given. So it is important to determine the degree to which the agent should believe the patient's cancer will recur.

This is a genuine case of epistemic complexity in the sense that the evidence available is multifarious and exhibits various kinds of structure. Evidence includes the following. There are a variety of clinical datasets describing the clinical symptoms and disease progress of past patients. There are genomic datasets describing the presence or absence of molecular markers in past patients. There are scientific papers that provide evidence of causal relations, mechanisms, and statistical information that quantifies the strength of connection between the variables under study. Causal relationships and mechanisms can also be elicited from experts in the field, such as clinicians and researchers in cancer systems biology. And there are also a whole host of prior medical informatics systems which provide a variety of evidence: e.g., evidence of ontological relationships between variables in medical ontologies, evidence of logical relationships in medical argumentation systems.[4]

Traditional machine learning methodology would take one of two standard courses. One option is to choose the best piece of data – e.g., a clinical dataset – and to build a model – e.g., a Bayesian net – that represents the distribution of that data. The resulting model would then be used as a basis for decision. Clearly this approach ignores much of the available evidence, and will not yield useful results if the chosen data is not plentiful, accurate and relevant. A second option is to build a model from each piece of evidence and to combine the results – e.g., by each model taking a vote on the recommended decision and somehow aggregating these votes. There are several difficulties with this approach. One is that most machine learning methods only take a dataset as input; consequently the qualitative causal evidence and the evidence concerning hierarchical mechanisms is likely to be ignored. A second is that the resulting models may be based on mutually inconsistent assumptions, in which case it is not clear that they should be combined at all. A third difficulty is that the problem of aggregating the judgements of the various models is itself fraught (Williamson 2009). In contrast, the approach taken in Nagl et al. (2008)

[4] Ontological or semantic evidence may be understood in terms of influence relations, just as can causal, hierarchical and logical evidence – see Williamson (2005, §11.4).

is to construct a *single* model – an objective Bayesian net – that takes into account the full range of evidence. We considered four evidential sources, which will now be described.

The first source is the SEER study, a clinical dataset involving 3 million patients in the US from 1975–2003; of these 4731 were breast cancer patients. This dataset measures the following variables: Age, Tumour size (mm), Grade (1–3), HR Status (Oestrogen/Progesterone receptors), Lymph Node Tumours, Surgery, Radiotherapy, Survival (months), Status (alive/dead). A sample of the dataset appears below.

Age	T size	Grade	HR	LN	Surgery	Radiotherapy	Survival	Status
70–74	22	2	1	1	1	1	37	1
45–49	8	1	1	0	2	1	41	1
...	...	...	...	...	...	...	...	...

If standard machine learning methods for learning a Bayesian net that represents the empirical probability distribution of this dataset were invoked, they would generate a net with a graph similar to that of Fig. 13.11. In our case, however, we treat this empirical distribution as a constraint on appropriate degrees of belief. An agent's degree of belief in any sentence that involves only variables measured in this dataset should match the empirical probability of that sentence as determined by the dataset: $P_{\mathcal{E}}(\theta) = P^*(\theta)$ for all θ involving just variables in the dataset.

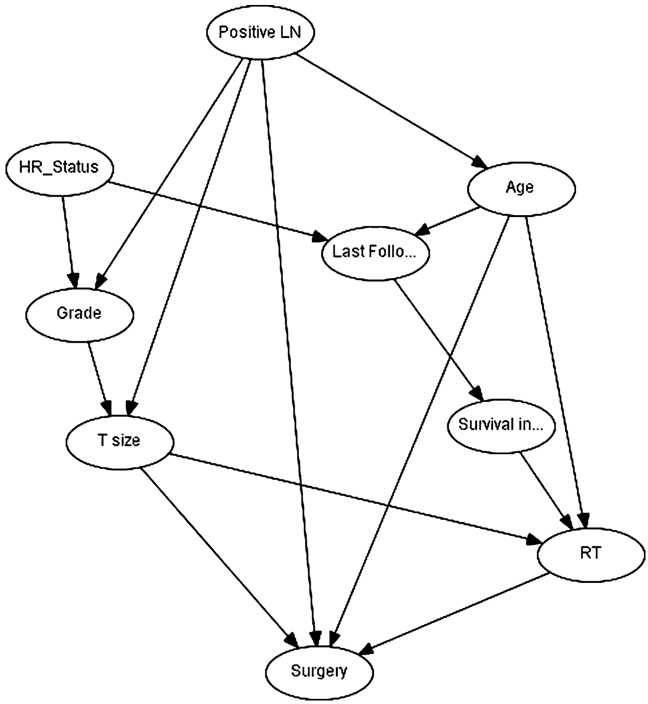

Fig. 13.11 Graph of a Bayesian net representing the empirical distribution of the clinical data

Table 13.1 Graph of a Bayesian net representing the empirical distribution of the clinical data

1p31	1p32	1p34	2q32	3q26	4q35	5q14	7p11	8q23	20p13	Xp11	Xq13
0	0	0	1	−1	0	0	1	0	0	0	−1
0	0	1	1	0	0	0	−1	−1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...

Table 13.2 Graph in a Bayesian net representation of a genomic dataset

Lymph Nodes	1q22	1q25	1q32	1q42	7q36	8p21	8p23	8q13	8q21	8q24
0	1	1	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...

The second source consists of genomic data from a Progenetix dataset, with 502 cases. A sample appears in Table 13.1.

The empirical distribution of this dataset is represented by a Bayesian net with the graph of Fig. 13.12. Again, from an objective Bayesian point of view, this data imposes the constraint that $P_{\mathcal{E}}(\theta) = P^*(\theta)$ for all θ involving just variables in the dataset.

The third source was a further genomic dataset (119 cases with clinical annotation) from the Progenetix database (Table 13.2).

The fourth source was a paper published study (Fridlyand et al. 2006), which contains causal and quantitative information concerning the probabilistic dependence between the variables HR_status and 22q12 – this provided a further bridge between clinical and genomic variables represented in Fig. 13.13.

The resulting objective Bayesian net has the graph depicted in Fig. 13.14. This kind of representation is attractive in that it involves both clinical and molecular variables, permitting inferences from one kind of variable to the other. Thus one can use molecular as well as clinical evidence to determine an appropriate prognosis. See Nagl et al. (2008) for a fuller discussion of the construction and uses of this objective Bayesian net.

13.8 Conclusion

Complexity of evidence is one kind of epistemic complexity. In this paper we have seen how objective Bayesian epistemology can begin to tackle this kind of epistemic complexity. Objective Bayesianism offers a unifying framework for integrating and interpreting not just evidence of empirical probability, but also evidence of causal, hierarchical and logical structure. Objective Bayesian probability can be defined over predicate languages as well as propositional languages, and the machinery of objective Bayesian nets can be used to represent and reason with objective Bayesian degrees of belief.

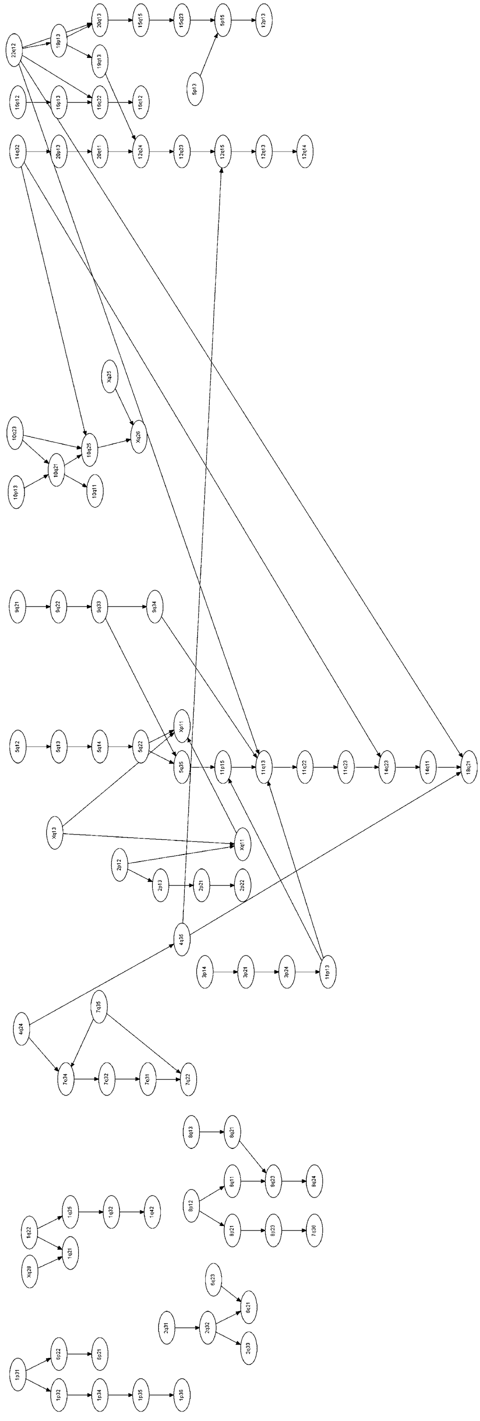

Fig. 13.12 Graph in a Bayesian net representation of a genomic dataset

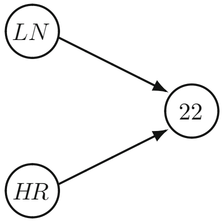

Fig. 13.13 Lymph nodes status, hormone receptor status and 22q12

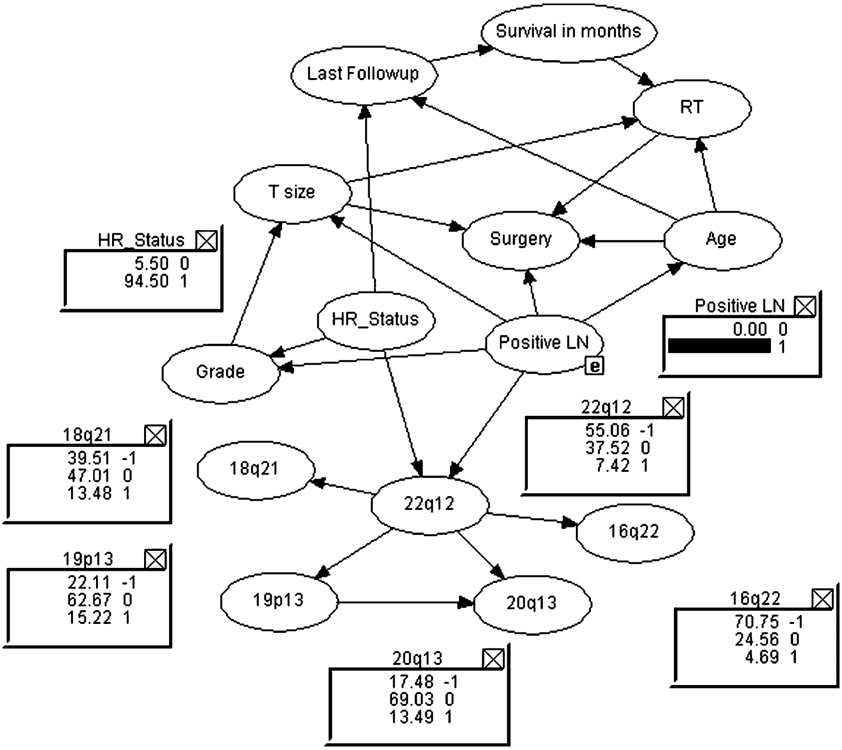

Fig. 13.14 Graph of the objective Bayesian net

Acknowledgment This research was carried out as a part of the project *progicnet: Probabilistic logic and probabilistic networks*, supported by the Leverhulme Trust.

References

Fridlyand J, Snijders A, Ylstra B, Li H, Olshen A, Segraves R, Dairkee S, Tokuyasu T, Ljung B, Jain A, McLennan J, Ziegler J, Chin K, Devries S, Feiler H, Gray J, Waldman F, Pinkel D, Albertson D (2006) Breast tumor copy number aberration phenotypes and genomic instability. BMC Cancer 6:96

Haenni R, Romeijn J-W, Wheeler G, Williamson J (2010) Probabilistic logic and probabilistic networks. Springer

Howson C, Urbach P (1989) Scientific reasoning: The Bayesian approach, 2nd edn. Open Court, Chicago, IL

Nagl S, Williams M, Williamson J (2008) Objective Bayesian nets for systems modelling and prognosis in breast cancer. In: Holmes D, Jain L, (eds) Innovations in Bayesian networks: Theory and applications. Springer, Berlin

Paris JB (1994) The uncertain reasoner's companion. Cambridge University Press, Cambridge

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

Williamson J (2007) Motivating objective Bayesianism: From empirical constraints to objective probabilities. In: Harper WL, Wheeler GR (eds) Probability and inference: Essays in honour of Henry E. Kyburg Jr., College Publications, London, 151–179

Williamson J (2008) Objective Bayesianism, Bayesian conditionalisation and voluntarism Synthese DOI 10.1007/s11229-009-9515-7

Williamson J (2009) Aggregating judgements by merging evidence. J Logic Comput 19(3): 461–473