Objective Bayesian Nets

Jon Williamson

Draft of July 14, 2005

Abstract

I present a formalism that combines two methodologies: *objective Bayesianism* and *Bayesian nets*. According to *objective Bayesianism*, an agent's degrees of belief (i) ought to satisfy the axioms of probability, (ii) ought to satisfy constraints imposed by background knowledge, and (iii) should otherwise be as non-committal as possible (i.e. have maximum entropy). *Bayesian nets* offer an efficient way of representing and updating probability functions. An *objective Bayesian net* is a Bayesian net representation of the maximum entropy probability function.

I show how objective Bayesian nets can be constructed, updated and combined, and how they can deal with cases in which the agent's background knowledge includes knowledge of qualitative *influence relationships*, e.g. causal influences. I then sketch a number of applications of the resulting formalism, showing how it can shed light on probability logic, causal modelling, logical reasoning, semantic reasoning, argumentation and recursive modelling.

Contents

§1
Introduction

Any theory of rationality must at some stage address the following key question:

Belief Representation What is the best way to represent an agent's rational belief state?

It is the aim of this paper to sketch a solution to the belief representation problem.

The proposed solution has two facets. First, *objective Bayesianism* tells us which degrees of belief an agent should adopt: she should adopt as her belief function a probability function, from all those that satisfy constraints imposed by her background knowledge, that is maximally non-committal, i.e. that maximises entropy (§2). Second, recent developments in probabilistic expert systems tell us how best to represent a probability function: a *Bayesian net* offers an efficient, clear and informative representation (§3).

Combining these two facets in §4, we use a Bayesian net to represent the agent's optimal belief function—such a Bayesian net will be called an *objective Bayesian net*.

The method for constructing an objective Bayesian net given in §4 requires that the agent's background knowledge be formulated as a set of quantitative constraints on her degrees of belief. However knowledge is often qualitative; the question arises as to how objective Bayesian nets can be constructed in the presence of such knowledge. In §5 we shall see that qualitative knowledge of influence relationships (e.g. causal influence) can be transformed into quantitative constraints on degrees of belief.

An objective Bayesian net is derived from background knowledge. Thus to understand how to perform an operation on an objective Bayesian net, one should perform the corresponding operation on background knowledge and derive the associated objective Bayesian net. For instance, when an objective Bayesian net needs to be updated, the updated net should be the same as the net generated by updated background knowledge (§6). The combination of two Bayesian nets should be the same as the net generated by the combination of their associated knowledge bases (§7).

Having presented the theory of objective Bayesian nets in Part I, we turn briefly to applications in Part II. We shall see that apart from their use in a general theory of rationality, objective Bayesian nets also shed light on a number of specific modes of reasoning. They can be used to perform inference in a probabilistic logic (§8), to justify the assumptions behind causal models (§9), to guide logical (§10) and semantic (§11) reasoning, and to develop a framework for argumentation (§12) and recursive modelling (§13).

Part I
Theory

§2
Objective Bayesianism

Suppose a patient has a high fever, a dry cough and appears confused—to what extent should one believe that he has Legionnaire's disease?

Bayesians hold that an agent's degrees of belief ought to satisfy the axioms of probability. Thus the above degree of belief has the form of a conditional probability statement, $p(l|fdc)$ where l signifies that the patient has Legionnaire's disease, f that he has a high fever, d that he has a dry cough and c that he appears confused. Subjective Bayesians stop there and consider an agent to be rational whatever probability function he adopts as her belief function. But objective Bayesians go further, insisting (i) that an agent's degrees of belief should also respect background knowledge—they should for example be calibrated with known frequencies (if she knows only the incidence rate of Legionnaire's disease in the population then $p(l)$ should match that rate)—and (ii) that the agent should commit to outcomes only to the extent warranted by background knowledge (e.g. if she knows nothing concerning l then she should not commit to l; instead she should equivocate between l and $\neg l$, i.e. set $p(l) = p(\neg l) = 1/2$).

More precisely, objective Bayesians suppose that an agent's background knowledge β delimits a set $\mathbb{P}_\beta$ of probability functions that are compatible with that knowledge, and that the agent should choose a function $p \in \mathbb{P}_\beta$ that maximises entropy as her belief function. The entropy of a probability function p is

$$H(p) = -\sum_{v \in \Omega} p(v) \log p(v), \tag{1}$$

where Ω is the space of all possible indivisible outcomes, e.g. $\Omega = \{\pm l \pm f \pm d \pm e\}$.[1] Entropy is interpreted as a measure of the uncertainty or lack of commitment of a probability function: the more middling the probabilities, the higher the entropy and the higher the uncertainty; the nearer the probabilities are to the extremes of 0 or 1, the lower the entropy and the more the probability function commits to certain outcomes.[2] A probability function in $\mathbb{P}_\beta$ that has maximum entropy is compatible with background knowledge but is maximally non-committal in other respects.[3] Such a probability function is to be desired as a representation of one's degrees of belief because it is guided by empirical information yet is on average maximally cautious when it comes to risky decisions, which tend to be embarked upon when one has more extreme degrees of belief.[4]

For a set $\mathbb{Q}$ of probability functions we shall write $p \uparrow \mathbb{Q}$ as shorthand for $p \in \{q \in \mathbb{Q} : H(q) \text{ is maximised}\}$. Objective Bayesians maintain then that one should take $p \uparrow \mathbb{P}_\beta$ as one's belief function, given background knowledge β. This principle is often called the *maximum entropy principle*; it considerably narrows

[1] We shall assume throughout that Ω is finite. The extension to the infinite is discussed in Williamson (2005c, §19).

[2] (Shannon, 1948)

[3] (Jaynes, 1957)

[4] (Williamson, 2005b)

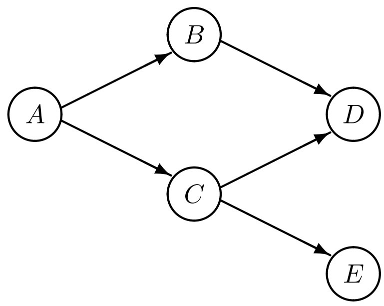

Figure 1: A directed acyclic graph.

down the values one can ascribe to $p(l|fdc)$.[5]

Two questions remain: how should one best represent the probability function $p \uparrow \mathbb{P}_\beta$? how should one calculate probabilities like $p(l|fdc)$? One can appeal to Bayesian nets to address these questions: a Bayesian net offers an efficient and perspicuous representation of a probability function and offers an efficient way to calculate conditional probabilities.

§3
Bayesian Nets

Consider a domain $V = \{A_1, A_2, \ldots, A_n\}$ of finitely many variables, each of which has finitely many possible values. Let $a_i@A_i$ signify that a_i is an assignment of a value to variable A_i. Associated with V there is set $\Omega_V = \{a_1 a_2 \cdots a_n : a_i@A_i, 1 \leqslant i \leqslant n\}$ of indivisible outcomes.

A *Bayesian net* $\mathfrak{b}_V = (G, S)$ contains

- a directed acyclic graph G whose nodes are variables in V (e.g. Fig. 1),
- a probability specification S which contains the probability distribution of each variable in V conditional on its parents in G (Table 1 contains an example distribution—B, C, and D each take two possible values, superscripted by 0 and 1).

A Bayesian net is also subject to an assumption, the *Markov Condition*, which holds that each variable A_i is probabilistically independent of its non-descendants in G conditional on its parents in G, written $A_i \perp\!\!\!\perp ND_i \mid Par_i$.

A Bayesian net determines a unique probability function p over Ω_V since the Markov Condition implies that

$$p(a_1 a_2 \cdots a_n) = \prod_{i=1}^{n} p(a_i | par_i),$$

where $par_i@Par_i$ is determined by $a_1 a_2 \cdots a_n$, and since the probabilities in this product are all contained in S.

[5]Plausibly $\mathbb{P}_\beta$ will be a closed convex set of probability functions, in which case $p \uparrow \mathbb{P}_\beta$ is uniquely determined—see Williamson (2005a, §5.3).

Table 1: The probability distribution of D conditional on B and C.

$p(d^0\|b^0c^0) = 0.7$	$p(d^1\|b^0c^0) = 0.3$
$p(d^0\|b^0c^1) = 0.9$	$p(d^1\|b^0c^1) = 0.1$
$p(d^0\|b^1c^0) = 0.2$	$p(d^1\|b^1c^0) = 0.8$
$p(d^0\|b^1c^1) = 0.4$	$p(d^1\|b^1c^1) = 0.6$

A Bayesian net $\mathfrak{p}$ offers an attractive representation of a probability function p for a number of reasons. First, $\mathfrak{p}$ perspicuously represents probabilistic independencies satisfied by p in the sense that one can simply read independencies off the graph: for $X, Y, Z \subseteq V$, $X \perp\!\!\!\perp Y \mid Z$ if Z *blocks* each path between X and Y, i.e., for each path between $A_i \in X$ and $A_j \in Y$, there is some node on the path in Z whose adjacent arrows meet head-to-tail or tail-to-tail, or there is a node on the path whose adjacent arrows meet head-to-head and Z contains neither that node any of its descendants. Second, $\mathfrak{p}$ is an efficient representation in the sense that relatively few probability specifiers $p(a_i|par_i)$ determine a large number of probabilities $p(a_1 a_2 \cdots a_n)$ (this depends on the structure of G: roughly speaking the sparser the graph G, the smaller the specification S). Third, $\mathfrak{p}$ admits efficient probabilistic inference: there are algorithms for quickly determining conditional probabilities from the Bayesian net (again, the efficiency of these algorithms depends on the structure of the graph).[6]

Bayesian nets are typically constructed in one of two ways. One is to employ a machine learning methodology to construct a net that represents the frequency distribution of a database of past observations of assignments to variables in V. The other is to elicit a graph and probability specifiers from an expert to construct a net that represents the expert's (subjective Bayesian) belief function. Here we are interested in objective Bayesian probability rather than frequency or subjective Bayesian probability—clearly neither of these two approaches are appropriate for representing an objective Bayesian belief function. We thus need a technique for constructing a Bayesian net that represents a probability function, from all those that satisfies constraints imposed by background knowledge, that maximises entropy.

§4
Objective Bayesian Nets

An *objective Bayesian net*, or *obnet* for short, is a Bayesian net that represents an objective Bayesian probability function p, i.e. a probability function that maximises entropy subject to constraints imposed by background knowledge β.

An objective Bayesian net can be constructed using the following strategy:

Step 1 determine conditional independencies that $p \uparrow \mathbb{P}_\beta$ must satisfy,

Step 2 represent these by a directed acyclic graph G that satisfies the Markov Condition with respect to p,

[6](Neapolitan, 1990)

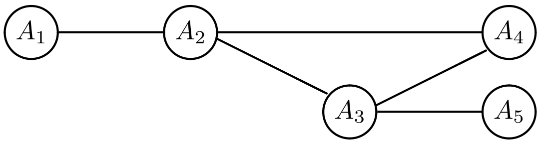

Figure 2: Constraint graph.

STEP 3 maximise entropy to calculate the numerical parameters $p(a_i|par_i)$ in the probability specification S.

We shall briefly run through each of these steps in turn—this procedure for constructing an objective Bayesian net is presented in more detail in Williamson (2005a, §§5.6–5.7).

Step 1: Determine Conditional Independencies

We shall suppose β can be construed as probabilistic constraints $\pi_1, \ldots, \pi_m$ on the probability function p. For example, π_1 might be $p(a_1|a_2) \geqslant 0.7$, and π_2 might be $p(a_2a_4) = p(a_3)^2p(a_2)$.

Construct an undirected graph, the *constraint graph*, by taking the variables in V as nodes and by connecting two nodes with an edge if they occur in the same constraint.

Suppose, for example, that $V = \{A_1, \ldots, A_5\}$, π_1 is a constraint involving A_1 and A_2, π_2 involves A_2, A_3, A_4, π_3 involves A_3, A_5, and π_4 involves A_4. Then the constraint graph is depicted in Fig. 2.

The constraint graph tells us about probabilistic independencies that the maximum entropy function will satisfy, since the following key property holds:[7] if Z separates X from Y in the constraint graph then $X \perp\!\!\!\perp Y \mid Z$ for $p \uparrow \mathbb{P}_\beta$.

In Fig. 2, for example, A_2 separates A_1 from A_3, A_4 and A_5, so we know that a maximum entropy function renders A_1 probabilistically independent of A_3, A_4 and A_5 conditional on A_2.

Step 2: Construct a Graph Satisfying the Markov Condition

One can transform the constraint graph into a directed acyclic graph G that satisfies the Markov Condition via the following algorithm:[8]

- triangulate the constraint graph,
- re-order V according to maximum cardinality search,
- let $D_1, \ldots, D_l$ be the cliques of the triangulated constraint graph ordered according to highest labelled node,
- set $E_j = D_j \cap (\bigcup_{i=1}^{j-1} D_i)$ for $j = 1, \ldots, l$,
- set $F_j = D_j \backslash E_j$ for $j = 1, \ldots, l$,
- take variables in V as the nodes of G,
- add an arrow from each vertex in E_j to each vertex in F_j $(j = 1, \ldots, l)$,
- ensure that there is an arrow between each pair of vertices in D_j $(j = 1, \ldots, l)$.

[7](Williamson, 2005a, Theorem 5.3)

[8]See Williamson (2005a, §5.7) for an explanation of the graph-theoretic terminology.

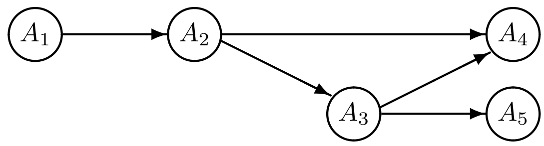

Figure 3: Graph satisfying the Markov Condition.

The resulting directed graph often looks much like the undirected constraint graph—in our example G is depicted in Fig. 3.

Step 3: Determine Probability Specification

Having found a graph G that satisfies the Markov Condition, to construct an objective Bayesian net it only remains to determine the probability distribution of each variable conditional on its parents.

Here it helps to rewrite the entropy equation as $H = \sum_{i=1}^{n} H_i$ where

$$H_i = -\sum \left(\prod_{A_j \in Anc_i} p(a_j|par_j) \right) \log p(a_i|par_i),$$

(Anc_i being the set of ancestors of A_i in G.)

One can then use numerical optimisation techniques or Lagrange multiplier methods to find the parameters $p(a_i|par_i)$ that maximise entropy. This entropy maximisation problem will in practice be a smaller problem than the original problem of maximising entropy over the whole domain (Equation 1) since there will in practice be far fewer parameters of the form $p(a_i|par_i)$ than there were of the form $p(v) = p(a_1 a_2, \ldots, a_n)$ (this is because in practice there tend to be few variables in each constraint in comparison with n, as n becomes large, so G tends to be sparse).

We see then that by pursuing this three-step procedure it is quite straightforward to construct an obnet, given a set of probabilistic constraints.

Quantitative probabilistic constraints are clearly required in order to apply the maximum entropy principle. However background knowledge does not always take the form of a set of quantitative constraints on degrees of belief—an agent may know of qualitative causal relationships, for instance. The task of converting qualitative constraints into quantitative constraints is a significant challenge for objective Bayesianism.[9] We shall see next how qualitative knowledge of influence relationships (e.g. causal influences) can be converted into quantitative constraints on an agent's belief function p.

§5
Influence Relations

We turn now to the question of how one can construct an objective Bayesian net when background knowledge includes qualitative knowledge of influence relationships.

[9] (Williamson, 2005c, §18)

We shall take the following to be the defining feature of the notion of *influence*: learning of the existence of new variables that are not influences of the other variables should not change degrees of belief concerning those other variables.[10]

The causal relation, for example, is an influence relation. If an agent learns of a new variable that is known not to be a cause of any the variables she already knew about, then this new information provides no reason for the agent to change her degrees of belief concerning those other variables. In the absence of any reason for change, her degrees of belief should stay the same. (In contrast, learning of new causes may motivate a change in degrees of belief: at first glance the flooding of glacial valleys in Kyrgyzstan and the insect population of southern England seem quite unrelated, but the knowledge that global warming affects both these variables may warrant an increase in the degree to which one believes insect populations will rise given that glacial flooding is increasing.) Causality and other examples of influence relationships will be discussed in Part II.

Given the above implicit definition of influence, it is straightforward to see that qualitative knowledge concerning influences can be transferred into quantitative constraints on degrees of belief. Suppose $V \supseteq U$ is a set of variables containing variables in U together with other variables that are known not to be influences of variables in U. As long as any other knowledge concerning variables in $V \backslash U$ does not itself warrant a change in degrees of belief on U, then $p^V_{\beta \downharpoonright U} = p^U_{\beta_U}$, i.e. one's belief function on the whole domain V formed on the basis of all one's background knowledge β, when restricted to U, should match the belief function one would have adopted on domain U given just the part β_U of one's knowledge involving U. Thus knowledge of influences is transferred into equality constraints on degrees of belief.

Once qualitative knowledge has been transferred into quantitative constraints on degrees of belief, the three-step procedure of §4 for constructing an objective Bayesian net can be directly applied. However, the fact that the new constraints are equality constraints leads to a simplification: these new constraints can be ignored in Step 1 of the process.[11] We thus have a slightly modified three-step procedure:

STEP 1 determine conditional independencies that $p \uparrow \mathbb{P}_\beta$ must satisfy from the constraint graph, *ignoring constraints yielded by knowledge of influences*,

STEP 2 represent these independencies by a directed acyclic graph G that satisfies the Markov Condition with respect to p,

STEP 3 maximise entropy to calculate the numerical parameters $p(a_i|par_i)$ in the probability specification S *(remembering to take equality constraints yielded by knowledge of influences into account)*.

Thus knowledge of influences does not add to the complexity of an objective Bayesian net, in the sense that the graph in the net is just as sparse as it would have been if there were no such knowledge.

A further simplification is possible in the case in which the agent knows all the influence relationships amongst the variables and has no quantitative

[10](Williamson, 2005a, §11.4)
[11](Williamson, 2005a, Theorem 5.6)

knowledge that overrides the equality constraints generated by these influence relationships (n.b. quantitative information regarding the strengths of the influence relationships will not override the equality constraints).[12] As long as the *influence graph*—i.e. the directed graph in which there is an arrow from variable A to variable B if and only if A directly influences B—is acyclic, we can go straight to Step 2: the influence graph itself satisfies the Markov Condition.[13] Step 3 is also simpler in this case: we can maximise entropy by maximising each component H_i of the modified entropy equation sequentially (rather than maximising their sum).[14] This breaks down the entropy maximisation problem into n smaller problems. In this case, then, the objective Bayesian net is just the influence graph plus sequentially-determined conditional probability distributions.

Having discussed the construction of obnets, we now turn to how they might be updated (§6) and combined (§7).

§6
Updating

An objective Bayesian net represents the degrees of belief that an agent should adopt and these rational degrees of belief are determined by the agent's background knowledge. So when her background knowledge changes, so too should the obnet. The extent to which the net changes will depend on the extent to which background knowledge changes.

If the new knowledge consists of an observation o of the values of some of the variables, then the new probability function $p' \uparrow \mathbb{P}_{\beta \cup \{o\}}$ is just the old function conditional on the observation, i.e. $p' = p(\cdot|o)$ where $p \uparrow \mathbb{P}_\beta$.[15] This type of update is known as *Bayesian conditionalisation*. It is simple to modify a Bayesian net to represent its Bayesian conditionalisation update: the graph in the net remains the same but the probability specification gets updated using standard propagation algorithms.[16]

More generally, when the new knowledge consists of new constraints on the agent's degrees of belief that are consistent with the old constraints, the new probability function $p' \uparrow \mathbb{P}_{\beta'}$ is the probability function satisfying β' that is *closest* to the old function $p \uparrow \mathbb{P}_\beta$ in the sense that it minimises the *cross entropy distance* to p, $d(p', p) = \sum_v p'(v) \log p'(v)/p(v)$.[17] Thus we need to modify the objective Bayesian net $\mathfrak{p}$ representing $p \uparrow \mathbb{P}_\beta$ to form its *cross entropy update* $\mathfrak{p}'$ that represents the $p' \in \mathbb{P}_{\beta'}$ which minimises $d(p', p)$. This involves reconstructing the part of the graph of $\mathfrak{p}$ that involves variables in the new constraints and their ancestors in the graph and updating the associated conditional probability distributions; the rest of the net stays the same.[18]

[12](Williamson, 2005a, pp. 99–100)
[13](Williamson, 2005a, Theorem 5.7)
[14](Williamson, 2005a, Theorem 5.8)
[15](Williams, 1980, pp. 134–135)
[16](Neapolitan, 1990, Chapters 6–7)
[17]N.b. $\beta \subseteq \beta'$. Here $p' \uparrow \mathbb{P}_{\beta'}$ is the function minimising $d(p', c)$, where c is the *central function* that gives the same probability to each elementary outcome (Paris, 1994, p. 120). As long as constraints are all *affine*, this as the same function as that found by minimising $d(p, c)$ first and then minimising $d(p', p)$—see Williams (1980, pp. 139–140).
[18](Williamson, 2005a, §12.11)

In other cases, the whole net may need be reconstructed. If the new constraints are inconsistent with the old, background knowledge can not be simply augmented, it must change: some element of background knowledge must be repealed to eradicate the inconsistency. In this case a new objective Bayesian net must be constructed around the changed constraints, via the three-step procedure of §4 and §5. Similarly if the new knowledge consists of knowledge of new variables as well as new constraints, a reconstruction of the net will be required, unless the new knowledge does not warrant a change of degrees of belief involving the old variables (in particular if the new variables are known not to be influences of the old). In this latter case one can just augment the old net by adding the new variables to the graph, adding arrows to the new variables from old variables that occur in the same constraints (and amongst new variables that occur in the same constraints) and adding the probability distributions of new variables conditional on their parents.

We see then that the updating of an objective Bayesian net hinges on the updating of background knowledge. This yields a *foundational* approach to updating—the warrant for degrees of belief, background knowledge, is the crucial determinant of those degrees of belief; one does not update by cohering with past degrees of belief but by satisfying constraints imposed by this knowledge.

§7
Combining

In certain circumstances it is useful to consider the combination $\mathfrak{p}$ of two objective Bayesian nets $\mathfrak{q}$ and $\mathfrak{r}$, written $\mathfrak{p} = \mathfrak{q} \star \mathfrak{r}$. (More generally, given a set $\mathfrak{Q}$ of obnets we can denote their combination by $\mathfrak{p} = \star\mathfrak{Q}$.) For example, two or more agents may need to come to some consensus and act as one agent, and the question arises as to which belief function this group agent should adopt.[19]

From the foundational point of view, the combination of a set of obnets should be determined from the combination of the set of background knowledge bases that underpin the respective obnets: $\mathfrak{p} = \mathfrak{q} \star \mathfrak{r}$ should represent $p \uparrow \mathbb{P}_\beta$ where $\beta = \gamma \star \delta$, the combination of the knowledge base γ that determines $\mathfrak{q}$ and the knowledge base δ that determines $\mathfrak{r}$.

So the combination of obnets boils down to the combination of knowledge bases. How should knowledge bases be combined? This is a rather subtle question that turns on the origins of the constraints in the knowledge bases. Consider an example. Suppose Quentin's background knowledge γ contains the constraint $q(a) = 0.7$, while Ronette's background knowledge δ contains $r(a) = 0.8$. Clearly these are incompatible assignments of probability if reinterpreted as constraints on p. But the way this inconsistency is resolved depends on the origins of the these constraints. Suppose that both constraints originated from observed frequencies: for Quentin a falls under a reference class which has observed frequency 0.7 of a-type outcomes, while for Ronette a falls under a reference class which has observed frequency 0.8 of a-type outcomes. If Ronette's reference class is *narrower* than Quentin's, then her constraint should override Quentin's, and only the constraint $p(a) = 0.8$ should appear in the combined knowledge base $\beta = \gamma \star \delta$. On the other hand, if neither reference class is

[19] (Gillies, 1991)

narrower than the other then neither constraint is defeated by the other and the best one can do is include the constraint $p(a) \in [0.7, 0.8]$ in β.[20] In general, we can say that $\mathbb{P}_\beta$ is the smallest closed convex set of probability functions generated by *undefeated* constraints in $\gamma \cup \delta$.

In sum then, a combination of objective Bayesian nets will depend on defeasibility relationships amongst constraints in the associated knowledge bases. If one agent's knowledge is better than all the others' then the group obnet should match that agent's obnet. Typically though the combined obnet will need to be constructed afresh from the combined background knowledge.

Part II
Applications

We shall now quickly run through some applications of the theory developed above. A more detailed treatment is given in Williamson (2005a).

§8
Probability Logic

The simplest probability logics are concerned with questions of the form: given premiss sentences and their probabilities, what probability should attach to a conclusion sentence?

For instance,

$$a_1 \wedge \neg a_2,^{0.9} (\neg a_4 \vee a_3) \rightarrow a_2,^{0.2} a_5 \vee a_3,^{0.3} a_4.^{0.7} \models a_5 \rightarrow a_1.^{?}$$

is short for the following question: given that $a_1 \wedge \neg a_2$ has probability 0.9, $(\neg a_4 \vee a_3) \rightarrow a_2$ has probability 0.2, $a_5 \vee a_3$ has probability 0.3 and a_4 has probability 0.7, what probability should $a_5 \rightarrow a_1$ have?

Such questions can be given an objective Bayesian interpretation: supposing background knowledge consists of the constraints $p(a_1 \wedge \neg a_2) = 0.9, p((\neg a_4 \vee a_3) \rightarrow a_2) = 0.2, p(a_5 \vee a_3) = 0.3, p(a_4) = 0.7$, what degree of belief should be awarded to $a_5 \rightarrow a_1$?

An objective Bayesian net can be constructed to answer this question. The first step is to determine conditional independencies that must be satisfied by the probability function, out of all those that satisfy these constraints, that maximises entropy. To do this we link variables that occur in the same constraint, as in Fig. 2; separation in this graph determines conditional independencies. The second step is to transform this graph into a directed acyclic graph satisfying the Markov Condition, such as Fig. 3. The third step is to maximise entropy to determine the probability distribution of each variable conditional on its parents in the directed graph. This yields a Bayesian net. Finally we use the net to calculate the probability of the conclusion

$$\begin{aligned} p(a_5 \rightarrow a_1) &= p(\neg a_5 \wedge a_1) + p(a_5 \wedge a_1) + p(\neg a_5 \wedge \neg a_1) \\ &= p(a_1) + p(\neg a_5 | \neg a_1)(1 - p(a_1)) \end{aligned}$$

[20](Williamson, 2005a, §5.3)

Thus we must calculate $p(a_1)$ and $p(\neg a_5|\neg a_1)$ from the net, which can be done using standard algorithms.

This application of obnets to probability logic is quite straightforward because background knowledge is quantitative. Other applications use the apparatus of §5 to exploit qualitative knowledge, as we shall now see.

§9
Causal Modelling

Many types of causal model (e.g. structural equation models) consist of information about the qualitative causal relationships amongst a set of variables together with the quantitative strengths of these causal relationships.

In order to easily infer a causal model from data, a number of fundamental assumptions are made about connections between causal relationships and empirical phenomena. Perhaps the key assumption is the following:

Causal Markov Condition (CMC) each variable is probabilistically independent of its non-effects conditional on its direct causes.

A fundamental problem facing proponents of causal modelling is the question of the justification of the Causal Markov Condition. One approach—taken by Pearl (2000) for example—is to make a number of other assumptions that are collectively stronger than the CMC and which together imply CMC. For example Pearl assumes universal determinism, that variables are functions of just their direct causes and error terms that are not in the variable set, and that error terms are probabilistically independent.

Objective Bayesian nets offer a less drastic solution to this conundrum. The components of the causal model can be thought of as an agent's background knowledge β. As we saw in §5, causality is an influence relation, and if β contains just causal relationships and their strengths then the graph in the obnet generated by β is just the causal graph. By construction, the Markov Condition is guaranteed to hold for this graph. But the Markov Condition for the causal graph is just the Causal Markov Condition. Thus the Causal Markov Condition must hold, where the probabilities that CMC talks about are interpreted as the degrees of belief that an agent ought to adopt if all she knows is the causal model.

Thus objective Bayesian nets offer a framework for causal reasoning. But obnets can also be applied to other influence relations. We shall turn to other examples of influence relations now.

§10
Logical Reasoning

A sentence a is a *logical influence* of sentence b if either a or $\neg a$ is a necessary component of some set of sentences that logically imply either b or $\neg b$, i.e. $\pm a, d \models \pm b$ for some sentence d, and $d \not\models \pm b$.

By analogy with causal influence, logical influence is plausibly an influence relation: learning of variables that are not logical influences of the others provides no reason to change one's degrees of belief concerning those other variables.

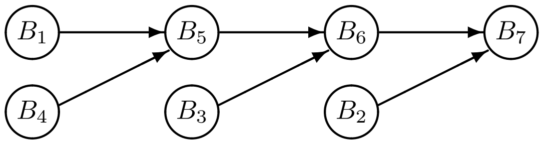

Figure 4: A logical influence graph.

Hence objective Bayesian nets can be used to represent an agent's degrees of belief in sentences given qualitative knowledge of logical influence relationships.

For example, suppose β consists of the following proof:
1: $\phi \rightarrow \psi$ [hypothesis]
2: $\theta \rightarrow \phi$ [hypothesis]
3: $(\theta \rightarrow (\phi \rightarrow \psi)) \rightarrow ((\theta \rightarrow \phi) \rightarrow (\theta \rightarrow \psi))$ [axiom]
4: $(\phi \rightarrow \psi) \rightarrow (\theta \rightarrow (\phi \rightarrow \psi))$ [axiom]
5: $\theta \rightarrow (\phi \rightarrow \psi)$ [by 1, 4]
6: $(\theta \rightarrow \phi) \rightarrow (\theta \rightarrow \psi)$ [3, 5]
7: $\theta \rightarrow \psi$ [2, 6]

This proof provides yields not only qualitative knowledge of logical influences but also quantitative constraints, namely $p(b_5|b_1b_4) = 1, p(b_6|b_3b_5) = 1, p(b_7|b_2b_6) = 1$, where variable B_i takes assignment b_i (respectively $\neg b_i$) just when the sentence on line i of the proof is true (respectively false). Then the graph in the obnet generated by β maps the structure of the proof, as in Fig. 4. The probability specification in the obnet contains the probabilities yielded by the quantitative constraints $p(b_5|b_1b_4) = 1, p(b_6|b_3b_5) = 1, p(b_7|b_2b_6) = 1, p(\neg b_5|b_1b_4) = 0, p(\neg b_6|b_3b_5) = 0, p(\neg b_7|b_2b_6) = 0$; all other probabilities in the specification, e.g. $p(b_6|\neg b_3b_5)$, will be set to $\frac{1}{2}$ by maximising entropy. This net can be used to calculate arbitrary probabilities, e.g. $p(b_1|\neg b_7)$.

§11
Semantic Reasoning

A concept a is a *semantic influence* of concept b if a (or its complement) is a b (or its complement). For example, 'flu is a semantic influence of virus, because 'flu is a virus.

Plausibly, semantic influence is an influence relation. Learning that 'flu and herpes are both viruses provides no reason to change degrees of belief involving 'flu and herpes: one's degree of belief that a patient has herpes given that he has 'flu and that they are both viruses should be the same as it would be in the absence of the knowledge that they are both viruses. Thus learning of non-semantic-influences should not change degrees of belief over other variables. (On the other hand, learning of semantic influences may warrant a change in degrees of belief: learning of 'flu and that 'flu is a short-term illness and a virus may increase one's degree of belief that a patient has virus given that he has a short-term illness.)

Since semantic influence is an influence relation, objective Bayesian nets can be used to represent an agent's degrees of belief given qualitative semantic knowledge. Suppose the agent's background knowledge β consists of the follow-

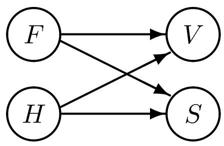

Figure 5: A semantic influence graph.

ing semantic knowledge:
- ◦ 'flu is a virus,
- ◦ herpes is a virus,
- ◦ 'flu is a short-term illness,
- ◦ herpes is not a short-term illness.

This consists of qualitative semantic knowledge, but also imposes the constraints $p(v|fx) = 1, p(v|hx) = 1, p(s|fx) = 1, p(\neg s|hx) = 1$, where v signifies virus, f 'flu, h herpes, s short-term illness and x is an arbitrary assignment. The resulting obnet will consist of the semantic graph Fig. 5 (a semantic graph is sometimes called a *semantic network* in AI) together with the entropy maximising probability specifiers e.g. $p(v|fh) = 1, p(v|f\neg h) = 1, p(v|\neg fh) = 1, p(v|\neg f \neg h) = 1/2$. One can use the obnet to calculate probabilities such as $p(h|vs)$.

§12
Argumentation

A proposition a is an *argumentative influence* of proposition b if a, or its negation, is an argument in favour of, or against, b. Plausibly, this is another example of an influence relationship: learning of propositions that are not argumentative influences of other propositions does not warrant a change in one's degrees of belief involving the other propositions.

If so, an objective Bayesian net can be used to represent an agent's degrees of belief given knowledge of argumentation structure. As before, given full knowledge of argumentation structure, the obnet will consist of an argument graph together with probability specifiers that maximise entropy.

§13
Recursive Modelling

In a *recursive model* the values that variables take may themselves be structured, containing further variables. Such models can be used to represent nested relationships. For example, the fact that smoking causes cancer causes governments to restrict tobacco advertising. This can be represented by a recursive model of the form $SC \longrightarrow A$ where SC is a variable taking value $S \longrightarrow C$ or value $S \not\longrightarrow C$, S represents smoking, C cancer and A advertising, the latter three variables just take the value true or false, and the arrow represents causal connection. Fig. 4 can also be thought of as a recursive model if each variable B_i takes as one value the sentence on line i of the proof used to generate the graph.

A variable A is *superior* to variable B if B occurs at a lower level to A. In the above causal model, SC is superior to S and C, but not to A. Arguably, superiority is an influence relation: learning of more structure at lower levels does not warrant a change in degrees of belief concerning higher levels. Full knowledge of superiority relationships leads to an obnet which contains arrows from superiors to their direct inferiors.

In fact a recursive model soon looks quite complicated if all these superiority arrows are included in the model. But one can eliminate them from the model if one imposes a new Markov Condition, called the *Recursive Markov Condition*, which holds that each variable is probabilistically independent of those other variables that are neither its inferiors nor at the same level, conditional on its direct superiors. This yields a *recursive Bayesian net*, a formalism that is explored in some detail in Williamson and Gabbay (2005).

§14
Concluding Remarks

We have explored a new, third way of constructing a Bayesian net: like a subjectively elicited Bayesian net, an objective Bayesian net represents an agent's degrees of belief; like a Bayesian net learned from a frequency distribution, an obnet is objectively determined from data. Objective Bayesian nets combine the best aspects of the other two methods: an obnet can make use of frequency information where available, but can also incorporate qualitative knowledge that is not reflected in frequencies.

A theory of rationality must tell us about knowledge (how it should be gleaned, updated, combined, and so on), about belief, and about decision-making, and must also offer a practical framework for their integration. Objective Bayesian nets provide the belief module: given knowledge, an obnet can be constructed to represent the agent's degrees of belief; given an obnet, a decision theory can advise the agent as to which decisions to make on the basis of her beliefs. Objective Bayesian nets are thus a crucial component of our normative toolkit.

References

Gillies, D. (1991). Intersubjective probability and confirmation theory. *British Journal for the Philosophy of Science*, 42:513–533.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Neapolitan, R. E. (1990). *Probabilistic reasoning in expert systems: theory and algorithms*. Wiley, New York.

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

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423 and 623–656.

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

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

Williamson, J. (2005b). Motivating objective Bayesianism: from empirical constraints to objective probabilities. In Harper, W. L. and Wheeler, G. R., editors, *Probability and Inference: Essays in Honour of Henry E. Kyburg Jr.* King's College Publications, London.

Williamson, J. (2005c). Philosophies of probability: objective Bayesianism and its challenges. In Irvine, A., editor, *Handbook of the philosophy of mathematics*. Elsevier. Handbook of the Philosophy of Science volume 9.

Williamson, J. and Gabbay, D. (2005). Recursive causality in Bayesian networks and self-fibring networks. In Gillies, D., editor, *Laws and models in the sciences*, pages 173–221. King's College Publications, London. With comments pp. 223–245.