Uncertainty, Bayesian Belief Nets, and Knowledge Management

Ulrich Metschl
TU München
Münchner Zentrum für Wissenschafts- und Technikgeschichte
ulrich.metschl@lrz.uni-muenchen.de

1 Uncertainty

When it comes to practical affairs, uncertainty is a notorious source of discomfort. Decision-making, whether private or public, would be straightforward if the outcomes of the choices which are available in a given situation could be foreseen with certainty. In the absence of reliable predictions beyond reasonable doubt, decision-making has become an art as much as a science.

This may seem trivial from a contemporary vantage point, but historically it was a long way to move from a liberation from the ‚quest for certainty' (John Dewey) to the acknowledgement of the inevitability of uncertainty. Along the way lay the inception of subjective (personalistic) probabilities and the invention of the modern concept of a cardinally measurable utility. While probabilities and utilities intuitively are independent of each other, they are nevertheless amenable to a unified treatment and can be simultaneously developed in a suitable formal framework. The joint derivation of probabilities and utilities from axioms for coherent, i.e. rational, preferences is the hallmark of Bayesianism, and has found its classical exposition in Leonard Savage's *Foundations of Statistics* (Savage 1957). For some, Bayesianism is the ultimate philosophical answer to the challenge of uncertainty, but this still leaves room for considerations on its scope. Thus, under one perspective Bayesianism is a doctrine about processing statistical or partial information, and its main components are the representation of partial information states by probability functions and the integration of new evidence or information -(‚updating') - by so-called conditionalization. Conditionalization here means that the (unconditional) probability function representing a new information state p_{new} after learning some piece of evidence e is equal to the conditional prior probability function representing the old information state, conditional, i.e., on the item of evidence: $p_{new}(A) = p_{old}(A|e)$.[1] Under a broader perspective, however, Bayesianism is an account of rational

[1] For more details see section 2.

choice, succinctly summed up in one recommendation: „Among the available options, choose one for which the expected utility is maximum". With the expected utility of an option A defined as:

$$EU(A) = \sum_{i=1}^{n} p(s_i)u(A(s_i))$$

where $s_1, \ldots, s_n$ are the relevant situations on which the agent is uncertain, $A(s_i)$ is the outcome of an option (‚act') A relative to the situation s_i, and u is a (subjective) utility function which is unique up to positive affine transformation, the Bayesian recommendation of maximizing expected utility generalizes the Bayesian approach as a method of statistical inference in an obvious manner.[2]

The mechanism of Bayesian deliberation, whose salient feature is the assessment of uncertainty in the context of choice, is nicely illustrated by the famous Monty Hall problem and similar probability puzzles.[3] In the Monty Hall problem, the correct choice, standing up to the requirements of rationality, requires the appropriate calculation of conditional probabilities. On the assumption that updating by new information conforms to conditionalization and in this sense refers to conditional probabilities, these can be calculated by the use of Bayes' theorem:

$$p(h|e) = \frac{p(h) \times p(e|h)}{p(e)}$$

However, while these examples are frequently seen as evidence that Bayesian inference is a powerful device, they indicate at the same time where it stands in need of improvement. In many cases, the unconditional probability of the evidence e is unknown and not open for direct assessment. What might be known, however, are probabilities for e, conditional on various hypotheses $h_1, \ldots, h_n$. Where these hypotheses $h_1, \ldots, h_n$ are mutually exclusive and jointly exhaustible (i.e. their probabilities add up to 1), $p(e)$ is, on basic probability postulates, equal to $p(e|h_1) \times p(h_1) + \ldots + p(e|h_n) \times p(h_n)$. As these conditional probabilities in turn may depend on other variables, the computation of a specific probability value may get quite involved.

[2]For philosophical purposes, it well deserves to be pointed out that the philosophy of pragmatism in the tradition of C.S. Peirce and John Dewey proclaims a priority of practice over theory, and thus sees beliefs and the cognitive standards for assessing the rationality of beliefs as dependent on the aim to make beliefs trustworthy for guiding actions, where actions may also include further inquiry.

[3]In the Monty Hall problem, a guest in a TV show is offered a choice between three doors, with the prize, a car, behind one of the doors, and a goat behind each of the other two doors. After having chosen one door, the show master opens one of the two doors that were *not* chosen by the guest, and one of the goats appears. The guest is then asked whether she wants to stick with her original choice or whether she wants to 'switch'. It turns out that switching increases the prospects for winning the prize to 2/3.

In general then, it is a considerable improvement, when coping with uncertainty, to find a joint probability distribution which provides all the necessary probability values. Joint probability distributions are, however, with an increasing number of variables complex to handle and require considerable computational power. In particular, the task of integrating new evidence, which amounts to updating all the values in the joint probability distribution, can be quite demanding. At this point Bayesian belief nets come into play.

2 Bayesian Belief Nets

Usually, probabilities are defined for propositions or for events. For our exposition, we will assume, however, that probabilities are defined for variables with a (finite) number of possible states, such that each variable is in exactly one state at one time. To each variable is assigned a probability distribution. It should be obvious that the variables stand for individual propositions (or for events, according to one's taste). A situation (in Savage's terminology a state), then, is defined jointly by the states of each of the variables. As in Savage's exposition, an event may also be defined as a set of situations (or Savage states). Where A is a variable such that its possible states are $a_1, \ldots, a_n$, a probability distribution $p(A)$ for A can be written as

$$p(A) = (x_1, \ldots, x_n)$$

such that $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 1$. As should be obvious, x_i denotes the probability that variable A is in state a_i.

Conditional probabilities are defined by the following equation:

$$p(A|B) = \frac{p(A, B)}{p(B)}$$

Bayes' Theorem follows from this definition straightforwardly, and it can be stated in a general form as:

$$p(A_i|E) = \frac{p(E|A_i) \times p(A_i)}{\sum_{j=1}^{n} p(E|A_j) \times p(A_j)}$$

Note that $\sum_{j=1}^{n} p(E|A_j) \times p(A_j)$ equals $p(E)$, which accounts for the general version of Bayes' theorem, once its standard version is established.

Suppose, we are concerned with a situation in which n variables $A_1, \ldots, A_n$ are involved. A *joint probability distribution* for the universe under consideration $p(U)$ is defined as

$$p(U) = p(A_1, \ldots, A_n)$$

By standard techniques like marginalization, $p(U)$ can be used to compute $p(A_i)$, or even $p(A_i|e)$, where e is some evidential input. However, when the number of variables is large enough (and it has not to be very large), the joint probability distribution $p(U)$ seems to get almost useless for practical purposes, as the complexity increases considerably with the number of variables and the number of the states for each variable.

However, there may exist dependencies between some of the variables, e.g. as causal relations, such that information on one variable affects the probability distribution for another variable. For example, in Jensen's car start problem (Jensen 2001) the filling of the gas tank results (with a certain degree of reliability) in a certain fuel meter reading, and thus the fuel meter reading provides information about the gas tank. If dirty spark plugs and lack of gas are identified as the potential causes for a failed car start, and the fuel meter reads 'full', then we will be willing to conclude that the problem is more likely to be due to the spark plugs.

Given a pattern of (causal) relations, the computation of a joint probability distribution is simplified by the so called chain rule. Let the relational pattern be represented by a non-cyclical graph whose edges are directed (representing the idea that the influence, whether causal or not, has a direction), and let $Pr(A)$ - the set of variables preceding A - stand for the set of variables which are connected with A by a directed edge such that they precede A in the graph. Then the chain rule states that

$$p(U) \quad = \quad \prod_i p(A_i|Pr(A_i))$$

where A_i is among the variables in the universe U. A proof of the chain rule basically exploits the fact that $p(A,B) = p(A|B) \times p(B)$ which follows immediately from the definition of conditional probabilities. In addition, it relies on information about what can be 'omitted' because of some sort of informational separability.[4] Due to the chain rule which is applicable in an Bayesian Belief Net (BBN), joint probability distributions are more readily tractable than they would be in the absence of information about influence patterns. A BBN in this context is formally defined as a directed, acyclic graph where each of the variables forming the nodes can be in one of a finite number of mutually exclusive states, and where to each variable A with predecessors $B_1, \ldots, B_n$ is assigned a probability distribution $p(A|B_1, \ldots, B_n)$.

From a purely theoretical perspective, BBNs may not look utterly impressive. But they are of some interest with regard to practical applications, and an increasing number of expert systems is based on the technique of BBNs. Prominent examples for recent applications comprise, amongst others, weather forecasting (Kennett/Korb/Nicholson, 200x), medical diagnosis for brain injuries (Sakellaropoulos/Nikiforidis 2000), electronic mail

[4]For details on the proof and the notion of separability as 'd-separation', see Jensen (2001), 21

delivery systems and diagnosis of printing problems (Microsoft), fraud detection (AT&T), failure detection in telecommunication networks (Nokia), and logistics in car manufacturing (Volkswagen). The general aim of all of these systems is an efficient handling of information about uncertain events in the form of hypotheses whose probability has to be assessed in the light of the available evidence. While it is plausible to claim that experts systems in general are still in their infancy, first comparisons seem to suggest that BBNs tend to fare better than most of their rivals, and outperform in particular rule-based systems that do not take account of probabilities as a measure of uncertainty.

But in spite of their technical success in these and other promising areas of application, Bayesian Belief Nets are not exempt from critical discussion. At present, their design as expert systems seems to be founded on the idea that the expert supplies the structure of the dependencies, or causal or informational influence - the model -, while the system calculates the probabilities on the basis of the model. But this situation is unsatisfactory for several reasons. First, despite a fully developed formalism for probabilistic reasoning, Bayesian Belief Nets, in relying on human expertise and intuition for the identification of causal relations, still depend, on this view, fundamentally on ‚tacit knowledge' that so far seems largely unassailable for scientific methods and a systematic treatment. Although it certainly would betray a form of scientific hubris if one would look with contempt at all areas of human activities which hitherto have withstood formalization and which are heavily based on tacit, non-declarative knowledge, this aspect nonetheless makes BBN's slightly less attractive as a technically powerful way for handling uncertainty. Progress in formal learning theory, however, may help to overcome the state where model building is largely a matter of individual ingenuity.[5]

However, more serious than this aspect is the (vaguely related) fact that the models on which the computations of joint probabilities are based are not necessarily unique. To be sure, in those cases where the domain of possibly uncertain events was explicitly designed, the uniqueness of the model and a thorough knowledge on the causal relations between possible events or variables can be expected. But in many areas of research, where the domain of uncertain events is not subject to human invention or intervention but instead is given in advance, the model itself stands in need of exploration. In those cases, therefore, it is not only conceivable but highly likely that different hypotheses about (causal) dependencies between the variables under consideration can be entertained, and as these lead to different models the probability calculations based on these models must be expected to differ.

[5]It is fair, however, to emphasize that efforts to formalize model building and formal learning about causal structures are currently under way. Therefore, one should not conclude from the present state of affairs that model building will remain beyond formalization for all times.

Although this problem is explicitly acknowledged in the BBN community, its consequences are not sufficiently embraced. Many workers in the field of BBN's seem to be content to take several models into account and to accept all of the probability results that can be deduced from these different models. Where the goal is to demonstrate the superiority of BBN's over other theories of modelling uncertainty, this equanimity is not surprising.[6] But from a more philosophical perspective this situation is unsatisfactory. In particular, when the aim is to estimate the likelihood of events in the context of choice, ambiguous probability assessments surely should be avoided, or otherwise the Bayesian recommendation for choice under uncertainty is of no help. Where, due to the incompleteness of our knowledge, more than one model representing causal dependencies is available, the situation is equal to one where several agents entertain different probability judgements and, hence, find themselves in a conflict concerning the likelihood of one or more hypotheses. The Bayesian approach in general is ill-suited for dealing with such conflicts for reasons to which we now turn.

3 Conflict and Knowledge Management

An important feature of Bayesianism in any of its versions is that it is a forward-looking doctrine in the sense that a new information state can only be the result of new informational input. Uncertainty, on this view, is due to the incompleteness of our knowledge and not to the unreliability or imperfection of given information.[7] There is much to said in favor of this conception, but it obviously ignores the everyday experience that sometimes new informational input requires adjustments or corrections with respect to previously held beliefs. The Bayesian postulates neglect the occasional need for a revision of information states as a result of, for example, the impacts of new evidence where this is incoherent with the given information.[8] But a need to revise or modify an information state may also be prompted by a desire to seek a consensus between agents whose assessments prove to be incompatible, and this is the kind of situation with which we were concerned at the end of the preceding section. Again, Bayesianism is at a loss when it comes to conflicting probability judgments, because, in some sense, conflict does not exist under a Bayesian perspective, or, to be more exacting, can

[6]See, e.g., Kennett/Korb/Nicholson (200x), where three Bayesian nets are shown to outperform considerably the hitherto established method of sea-breeze prediction in accuracy.

[7]Which is to say, given knowledge may be imperfect because it is incomplete but to the extent that it is available it is not defective. Thus, we may not know everything but what we know is free of errors.

[8]This and other limitations of Bayesianism were criticized by Isaac Levi for more than two decades, see, e.g., Levi (1980). Levi's own proposals for reasoning under uncertainty are the background of the present paper and will be sketched i this section.

only exist as a transitory state to be remedied by additional information.

There are, in fact, two results which, taken together, can be understood as a refutation of the possibility of genuine conflict under a Bayesian perspective. The first one, known as ‚washing out of priors', states that different probability distributions will converge under a sufficiently large series of updatings by additional evidence. The second result, Aumann's theorem on agreeing to disagree, states that the posterior probabilities of two ideally rational agents for an event must be equal if their priors were equal and the posterior probabilities are common knowledge, and irrespective of how they arrived at the posteriors. This means, in Aumann's own words, that „*people with the same priors cannot agree to disagree*".[9] Lack of information, then, is the only conceivable source of conflict among rational agents, and any conflict as to questions of fact must be overcome by additional evidence. But even if this is accepted, it still would be desirable to supplement the Bayesian account by criteria for a rational consensus in the absence of additional information. In the case of two agents whose beliefs are in conflict, this means that there exists at least one belief which should be withdrawn in order to identify those convictions which are shared by both agents and which mark common ground.

How does all this relate to Bayesian Belief Nets and the possibility of rival probability assessments by different BBNs? If, as suggested above, experts can disagree on the causal structure of a field under investigation and therefore propose different models for computing the probabilities of the events of interest, then it is not unlikely that additional information for the solution of the conflict is not available and will not be available within the time in which a decision has to be reached. A rational choice, then, must find a solution for these competing hypotheses and the (possibly different) choice recommendations based on them.

One option, of course, is to identify the ‚worst-case scenario'. This means, preference would be given to that model for which the outcomes of the events of interest are the most unfavorable, a strategy which equals the maximin criterion for choice under uncertainty. For example, insurance companies trying to estimate the potential effects of a global climate change on the basis of different models which lead to different conclusions might well give priority to more pessimistic estimations. But although risk-averse ‚maximining' is reasonable in many situations, it is far from clear that more optimistic assessments can always be dismissed so easily. When, in addition, the problem of mediating between disagreeing experts is understood as a case of collective decision-making where non-uniform probability judgements as well as different preferences have to be aggregated, then a rule that automatically gives priority to the worst-case scenario is certainly, in social choice terminology, dictatorial. For these reasons, a more flexible approach

[9] Aumann (1976), 1236; the emphasis is Aumann's.

for mediating between different and incompatible probability assessments by different Bayesian Belief Nets, or by different experts in general, is needed.

A promising proposal, due to Isaac Levi, introduces so called ‚indeterminate probabilities‘. Indeterminate probabilities were primarily designed to overcome the rigid Bayesian commitment to unique numerical probabilities. Instead of identifying probabilities with unique real numbers in the interval $[0, 1]$, indeterminate probabilities are intervals of real numbers in $[0, 1]$. Formally, a closed interval of reals is the same as the convex set defined by the greatest lower bound and the least upper bound of the interval. This, in turn, is based on the following motivation for indeterminate probabilities. Suppose, we face a situation in which two or more conflicting probability estimations are available, stemming either from different agents or from different models. A resolution of this conflict should not simply ignore or overrule any one of the conflicting positions. Rather, mediation should proceed by taking all available (and prima facie sufficiently reasonable) probability estimations into account, and also those which could be seen as a comprise between the given estimations. Formally, this is done by taking the set of the existing probability values and forming its convex hull. Because probability values represent a degree of confidence for a single event (proposition), they are something like points in an cognitive space. Proceeding on the assumption that an epistemic state is represented by a probability function, we achieve the full generality of indeterminate probabilities by working with nonempty, convex sets of (finitely additive) probability functions (Levi, 1974).

For a demonstration of the advantages of indeterminate probabilities in the context of choice, Levi points occasionally to the well-known Ellsberg problem. The Ellsberg problem consists in a pair of games of chance where a ball is drawn randomly from an urn containing ninety balls of three colors. Thirty balls are known to be red and each of the remaining sixty balls is either black or yellow, in unknown proportion. The payoff matrices for the two games are as follows.

Game I

	30	60	
	red	black	yellow
A	\$100	\$0	\$0
B	\$0	\$100	\$0

Game II

	30	60	
	red	black	yellow
C	\$100	\$0	\$100
D	\$0	\$100	\$100

Empirical studies suggest that many agents (in most experiments a clear majority) prefer option A in Game I and option D in Game II. This kind of behavior, however, is at odds with Bayesian recommendations, because no utility function, assigning the same utilities to the same outcomes in both games, makes this choice behavior coherent. The force of the Ellsberg problem however, is not just to show that Bayesian decision theory is descriptively inadequate. Even more important is the point that in the absence of unique probabilities on which the outcomes depend, agents can plausibly justify their choice behavior that contradicts Bayesian prescriptions. On the basis of indeterminate probabilities, the following account can be given. Option A in Game I comes with a prospect of winning \$100 with a probability of 1/3, while in option B the chances of winning the same amount of money range from 0 to 2/3. No estimation is possible, however, where, within this range, the true chance will lie, because the proportion of black and yellow balls is supposed to be unknown. In Game II, on the other side, it is option D that comes with a fixed prospect, now of 2/3, while the chances for option C range from 1/3 to 1, and no information is available where it will lie exactly. Therefore, the chances for winning in option B could be better than in option A, but they also might be worse, just as in Game II option C might be better than option D but also might be worse. As the agents are unable to decide which of these possibilities really holds, they have reasons, by deliberations similar to ‚maximining‘, to resort to those options which offer a fixed security level, namely option A in Game I and option D in Game II.

Indeterminate probabilities were presented above as unconditional probabilities. It is more convenient to define indeterminate probabilities, as Levi himself has done in his original exposition (Levi 1974) as non-empty, convex sets $P_e(A)$ of (finitely additive) probability functions $p(A|e)$, for propositions e consistent with the background knowledge, and all propositions A. Usually, joint probability distributions are defined by reference to conditional probabilities, as $p(A, B) = p(A|B) \times p(B)$. Multiplying indeterminate probabilities ‚pointwise‘ yields joint probability distributions for indeterminate probabilities.

Without going into technical details, we concentrate instead on the obvious implications for knowledge management. The Ellsberg problem indicates that indeterminate probabilities overcome some of the deficiencies that affect the Bayesian paradigm with its insistence on unique numerical probabilities. In point of fact, Bayesian decision theory is a special case of the decision theory based on indeterminate probabilities, as is, under favorable conditions, the maximin strategy for choice under (severe) uncertainty.[10] But in contrast to the restricted methods of Bayesian deliberation, indeterminate probabilities accommodate to the possibility of conflicting assessments of

[10]For details see Levi (1974).

uncertain events. To the extent that different estimations about the probabilities of the possible situations pertinent to some context of choice give rise to different recommendations concerning the option to be chosen, group decision may be expected to take place under conflict. In order to overcome an impasse that may result from such a situation, indeterminate probabilities offer a model of mediation for conflicts under uncertainty in the following ways. First, by substituting unique numerical probabilities with probability intervals a weaker epistemic position is maintained. Therefore, the commitment to a certain degree of confidence in some hypothesis is less strict. By weakening their endorsement of a certain belief and by suspending judgment on the exact degree of uncertainty, agents are able to retreat to a weaker position that can be shared by others who initially held different beliefs concerning the likelihood of a hypothesis under dispute. In this way, they may find a mediating position that is sufficiently broad to be embraced by the agents with conflicting opinions, and yet encapsulates their joint beliefs as something like the ‚greatest lower bound'. Second, a partial suspense of judgment will in some cases imply a suspense of choice. This is due to the fact that in the absence of unique estimations for situations which are uncertain, calculations of expected utility also will typically fail to be unique. Different probabilities for an uncertain event will lead to different utility expectations. The obvious consequence is that different recommendations can be entertained how or by which action to maximize expected utility. In the lofty realms of scientific reasoning, states of conflicting assessments may be taken as a transient nuisance. Further information or evidence will have to be awaited for the resolution of such a conflict, and inquiries will be undertaken with the aim to procure the required information.

But in everyday situations as well as in the world of business affairs, the prevailing constraints on time or budget often will prevent that these conflict will be resolved by additional information. Decisions, then, have to be made with the available resources and on the given informational basis.

Entrepreneurial decision very often are group decisions. Except in rare cases of strong homogeneity (which in some groups, of course, may be brought about by compulsion), opinions on the likelihood of relevant situations will vary. With their commitment to economic success, managerial decisions find a preeminent task in the settlement of dispute in order to embark on a clear business strategy. Insurance companies have to fix premiums, stock brokers have to buy, hold, or sell assets, and so on. However, under conflicting judgments on what is the case or how risky some business may be, disagreement on proposed strategies is not entirely unlikely. One role, then, that knowledge management in situations like that can fruitfully play is that of mediation between conflicting assessments where these stand in the way of consensual decision making. Levi's account of indeterminate probabilities is a promising proposal for rational mediation between conflicting assessments. Indeterminate probabilities represent epistemic states of a more flexible sort than ordinary probability functions in their commitment to uniqueness. Where agents disagree on the likelihood of events under consideration a mediating proposal comprises the full range of conflicting assessments, and thereby calls for a suspense of judgment over the precise degree credence that should be assigned to the uncertain events under dispute. On the mediating position, less is taken for granted and decision making will proceed in a more cautious manner. Levi's decision theory, to the extent that it is based on indeterminate probabilities, is tied to a conception of stratified choice behavior. In a first step, those options bearing maximum expected utility *under some probability estimation* are identified, and in a second step - assuming that choice can be suspended - those of the options selected in step 1 which bear the highest security level, i.e. which under the least favorable probability estimation have the highest expected utility, are chosen.[11] Under this process, decision making, in particular when done by groups, will put some emphasis on security. This conforms to the intuition that less daring decisions, decisions which are risk-averse, are easier to justify than maximum risk-bearing choices.

Uncertainty marks an epistemic challenge for groups as well as individuals. Indeterminate probabilities were presented as a method for overcoming a notorious obstacle for deliberate decision making. Knowledge management should profit from this proposal to the extent that it offers a concept of mediation where groups have to reach decisions under conflicting factual assessments.

4 Literature

Aumann, Robert (1976): Agreeing to Disagree; *Annals of Statistics* **4**, 1236-1239

Cooke, Roger M. (1991): *Experts in Uncertainty. Opinion and Subjective Probability in Science*; Oxford University Press

Jensen, Finn V.(2001): *Bayesian Networks and Decision Graphs*; Springer, New York et al.

Kennett/Korb/Nicholson (200x): Seabreeze Prediction Using Bayesian Networks; Report School of Computer Science and Software Engineering, Monash University, Australia

Levi, Isaac (1980): *The Enterprise of Knowledge*; MIT Press, Cambridge (Mass.)

Levi, Isaac (1974): On Indeterminate Probabilities; *Journal of Philosophy* 71 (1974), 391-418. Reprinted in Levi (1997), 117-144

[11] This process was illustrated by Levi's solution for the Ellsberg problem above.

Levi, Isaac (1985): Consensus as Shared Agreement and Outcome of Inquiry; *Synthese* **62**, 3-11. Reprinted in Levi (1997), 145-153

Levi, Isaac (1986): *Hard Choice. Decision making under unresolved conflict*; Cambridge University Press

Levi, Isaac (1997): *The Covenant of Reason*; Cambridge University Press

Pearl, Judea (1988): *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*; San Francisco, Morgan Kaufmann Publishers, Inc.

Sakellaropoulos/Nikiforidis (2000): Prognostic performance of two expert systems based on Bayesian belief networks; *Decision Support Systems*, Vol. 27, January 2000, 431-442

Savage, Leonard J. (1954): *The Foundations of Statistics*; New York, Dover, 1972