Should we use Probability in Uncertain Inference Systems?

Max Henrion

Department of Engineering and Public Policy,
and Department of Social and Decision Science,
Carnegie Mellon University.

Abstract

Criticisms of probability as being epistemologically inadequate as a basis for reasoning under uncertainty in AI and rule-based expert systems are largely misplaced. Probabilistic schemes appear to be the best way to deal with dependent evidence, and to properly combine diagnostic and predictive inference. Suggestions that expert systems should duplicate human inference strategies, with their documented biases, seem ill-advised. There is evidence that popular schemes perform quite poorly under some circumstances and there is an urgent need for careful study of when they can be relied upon. Some promising probabilistic alternatives are available, but they need to be demonstrated in realistic applications.

Introduction

Historically, probability has been by far the most widely used formalism for quantifying uncertainty and making inferences about it. However, for various conceptual and pragmatic reasons the majority of AI researchers have not, hitherto, found standard probabilistic techniques very appealing for use in rule-based, expert systems. Among the many alternatives they have used are the Certainty Factors used in Mycin (Shortliffe & Buchanan, 1975) and its descendants, Fuzzy Set Theory (Zadeh, 1984), the quasi-probabilistic scheme of Prospector (Duda et al, 1976), the Belief functions of Dempster-Shafer theory (Shafer, 1976), Paul Cohen's theory of endorsements (Cohen, 1985), Doyle's theory of reasoned assumptions (Doyle, 1983), and non-numerical, linguistic representations of uncertainty (Fox, 1986). We shall refer to both probabilistic and alternative methods, generically, as *uncertain inference schemes*, or UISs.

Each of these techniques has its partisans and its detractors, and discussion about their various merits and flaws seems to be heating up of late (Kanal & Lemmer, 1986, Gale, 1986). Much of the discussion hitherto has focussed on the theoretical issues of *epistemological adequacy* -- how well can each UIS represent the different aspects of reasoning with uncertainty? System developers have understandably been more concerned with the pragmatic issues of *heuristic adequacy* -- how easy is it to use and what are its computational demands?

The purpose of this paper is to give a personal view of some of the basic issues in evaluating and comparing UISs. Following a very brief account of the subjective or personalist view of probability and its applications in AI, I shall summarize the most common objections to probabilistic schemes, and attempts to rebut them that have appeared in the literature. The main focus will be on four issues that seem to have attracted less attention. The first concerns the treatment of correlated sources of evidence, and assumptions about dependence. The second is the issue of combining diagnostic and predictive reasoning, and the separation of inference rules from domain knowlege. The third is the vexed question of whether or not UISs for expert systems should try to to approximate human reasoning. The last is the question of whether it matters much which approach you use, and I shall argue the importance of more systematic comparison of alternative UISs to find out. The issue is not simply what can each UIS do or not do, but how much practical difference to the conclusions can it make which you use?

There has been considerable controversy on several of these topics. Researchers are operating under different paradigms with different programmatic goals, so I cannot expect agreement with all my arguments. What I hope is that it will be a contribution towards creating a more focussed debate, as a prerequisite for more cumulative science in this important area.

The appeal of probability

The probability of a proposition or a future event, according to the Bayesian or personalist view, is a measure of a person's degree of belief in it, given the information currently known to that person. The notion of probability may be derived from a set of simple axioms of rational decision-making under uncertainty, which form the basis of decision theory (Savage, 1954). The force of these axioms, and hence of the laws of probability derived from them, arises from the fact that a people who violate them and are willing to act on "incoherent" probabilities (for example, which do not satisfy Bayes' rule) are liable to demonstrable loss. Notably, an opponent could always design a "Dutch book", that is a combination of bets that they would be willing to accept, according to their professed beliefs, but which, in sum, would result in a guaranteed loss (de Finetti, 1974).

One advantage of being embedded in a theory of decision making is that it provides an *operational definition* for the probability of an event, in terms of the person's willingness to take bets based on the outcome of the event. Secondly, in combination with a utility model of preferences, it provides a clear, axiomatically-based approach for making decisions under uncertainty (Holtzman & Breese, 1985). Thirdly, it provides well-defined ways of using empirical data (Spiegelhalter, 1986), and evaluating the accuracy, resolution and calibration of UISs (Lichtenstein, Fischhoff & Phillips, 1982). No non-probabilistic measure of uncertainty offers these advantages. It has also been shown that for any reasonable scoring rule (which rewards a decision maker based on his or her probability assessments and the actual outcomes or truth of the propositions), any scalar measure of uncertainty is either worse than probability (produces a expected lower score) or is equivalent to it (Lindley, 1982).

Probabilistic UISs

A set of m propositions, $\{A_1, A_2, \ldots A_m\}$, each of which may be true or false, gives rise to 2^m different possible elementary *events*, each being a particular combination of proposition values. e.g. $(A_1 \;\&\; \sim A_2 \;\&\; \ldots\; A_m)$. A complete joint probability distribution over these propositions specifies a probability for each event, and so requires specification of 2^m-1 parameters. The exponential complexity of this *complete* representation clearly rules it out as a viable approach for practical systems and so simplifying assumptions are essential. In all practical UISs, the evidential relationships are modelled as an *inference network*, in which each proposition (or variable) is directly related to only a few others. Each such link is represented by a *rule*, which provides evidence about a consequent proposition, C, based on the degree of belief in some logical combination of its *antecedents*, A_i, for example, $(A_1 \;\&\; \sim A_2) \dashrightarrow C_3$. A "strength" (one or more numbers) is associated with each rule, whose probabilistic interpretation varies according to the scheme. Each such UIS needs to provide functions for propagating the uncertainty measures through logical conjunction, disjunction, negation, and generalized *modus ponens*, as well as a function for combining the evidence from multiple rules that bear on one consequent. The best known schemes are Certainty Factors (CFs) developed for Mycin (Shortliffe & Buchanan, 1975) and the scheme used in Prospector (Duda et al, 1976), from which many variants have been derived. Both of these were originally intended as approximations to Bayesian inference.

Neither system is completely consistent with a complete probabilistic scheme, and the implied probability distributions are incoherent. Kim and Pearl have devised an ingenious scheme for representing and propagating probabilistic information over a kind of inference net they term *Bayes' networks* (Kim & Pearl, 1983). This can maintain global coherence over the network, using only an efficient local updating mechanism. These are similar in spirit to the *influence diagrams* developed for decision analysis (Shachter, 1985).

An alternative approach to dealing with a partially specified probability distribution is to estimate the full distribution using the *Maximum Entropy Principle.* This minimizes the additional information assumed in filling out the distribution, consistent with the specified constraints (usually marginal and conditional probabilities). While this approach has several desirable properties (Shore and Johnson, 1980), computation of maximum entropy distributions is, in general, prohibitively expensive with more than a few propositions, despite attempts to improve algorithms (Cheeseman, 1983). However, many popular probabilistic updating schemes, including , conditional independence assumption, Jeffrey's rule, and odds-ratio updating, are actually special cases of Maximum Entropy, and the related Minimum Cross Entropy update (Wise, 1986).

Objections to probability

Despite its attractions, probability has been under sustained attack as a viable scheme for representing uncertainty in AI, ever since McCarthy and Hayes dismissed it as "epistemologically inadequate". Among the criticisms have been the following:

1. Probability requires vast amounts of data or unreasonable numbers of expert judgments.
2. It can't express ignorance, vagueness or "second-order uncertainty".
3. It doesn't distinguish reasons for and against, or identify sources of uncertainty.
4. The inference process is hard to explain.
5. It can't express linguistic imprecision.
6. It requires unrealistic independence assumptions.
7. It is computationally intractable.
8. It is not how humans reason.
9. It doesn't make much difference what method you use anyway.

Several recent articles have assembled similar lists of objections, overlapping with the first five or six listed here; they have provided eloquent rebuttals (Spiegelhalter, 1986, Pearl, 1985a, Cheeseman, 1985). Below is an extremely brief summary of their conclusions, without attempt at explanation. The interested reader is referred to the original articles. The main focus of this article will be objections six to nine and some related issues of the *heuristic adequacy* of probabilistic schemes.

Summary of rebuttals

In evaluating the criticisms and rebuttals, it is important to distinguish claims about probabilistic inference in general from claims about specific quasi-probabilistic UISs incorporating various heuristic assumptions. These rebuttals have been primarily in defense of the theoretical possibilities of probability rather than particular UISs. Failure to keep in mind this distinction has sometimes resulted in misunderstanding and fruitless argument.

The belief that probabilistic representations require vast amounts of data seems to derive from frequentist interpretations of probability, and does not apply to the Bayesian or subjectivist interpretations usually advocated. Inordinate quantities of subjective judgments should not be necessary either, since humans are subject to analagous limitations to other UISs, and our intuitive knowledge of probabilistic dependencies is represented by relatively sparse networks (Bayes networks), where most variables are not directly dependent. The question of whether or not two variables are directly probabilistically dependent is a qualitative judgment which is relatively easy to make (Pearl, 1985a).

Ignorance, vagueness or second order uncertainty may be represented by a range of probabilities, or by a predictive distribution over a probability, expressing how the prior probability might change after consulting a specified information source (Cheeseman, 1985, Spiegelhalter, 1986). Although it is often sufficient to represent each probability by its mean value, unless decisions about gathering new information are being contemplated.

It is true that a single probability by itself doesn't distinguish the sources, type and effect of the pieces of evidence on which it is based, but it is certainly possible to retrieve and clearly express this information in probabilistic schemes (Pearl, 1985a). For example, the *evidence weight* (log likelihood ratio) provides a convenient additive measure of the relative importance of each piece of evidence for a conclusion. The weights of supporting evidence can be added to the prior weight, and weights of disconfirming evidence subtracted in a sort of "ledger sheet" to arrive at the total final weight (Spiegelhalter, 1986). Evidence weights are also useful in explaining probabilistic reasoning. As long as the underlying inference network is sparse, as Pearl argues it will be, the inference process should be explainable in simple, comprehensible steps (Pearl, 1985a).

The advantage often claimed for Fuzzy Set Theory over probability is that the former can model linguistic imprecision, whereas probabilities are only defined for unambiguously specified ("crisp") events or propositions (Bonissone, 1982). Indeed probabilists have generally not addressed the issue of linguistic imprecision, aside from studies of the correspondence between probability phrases and numbers (Beyth-Marom, 1982). There is plenty of experimental evidence that probabilistic inference is not a very good model for human linguistic reasoning (Kahneman, Slovic & Tversky, 1982). But there has been little experimental investigation of claims that alternative UISs offer better models. A study comparing human judgment to Fuzzy Set Theory found that subjects' judgment of the "plausibility" of the intersection of two fuzzy sets was better modelled by the multiplication of "plausibilities", analogous to the probabilistic rule for the intersection of two independent events, rather than by the minimum plausibility rule of Fuzzy Set Theory (Oden, 1977). A problem in such studies of non-probabilistic schemes is setting up a convincing comparison when the measure of uncertainty has no operational definition.

Assumptions about dependence

While probability can in theory cope perfectly with non-independent sources of evidence, most actual UISs cannot. Consider the following:

Chernobyl example: *The first radio news bulletin you hear on the accident at the Chernobyl nuclear power plant reports that the release of radioactive materials may have already killed several thousand people. Initially you place small credence in this, but as you start to hear similar reports from other radio and TV stations, and in the newspapers, you believe it more strongly. A couple of days later, you realize that the news reports were all based on the same wire-service report based on a single unconfirmed telephone interview from Moscow. Consequently, you greatly reduce your degree of belief again.*

Figure 1: Inference network for Chernobyl example

This illustrates how multiple, independent supporting sources of evidence increase the confirmation of a hypothesis, but the confirmation is reduced if they are correlated. Most of us seem quite capable of handling this kind of intuitive reasoning in practice, even if we don't have the terminology to describe it. However none of the better known UISs are actually capable of distinguishing between independent and correlated sources of evidence. They each make various arbitrary fixed assumptions about the presence or absence of dependence. So all are inherently incapable of performing this normal commonsense reasoning. For example, the Fuzzy Set operators for *and* and *or*, are equivalent in effect to probabilistic rules assuming *subsumption* among antecedents, i.e. where the least likely proposition logically implies the more likely one(s). This is equivalent to assuming the *maximum possible* correlation between input propositions. Prospector and

Mycin CFs use similar rules for *and* and *or*. On the other hand, Prospector, and Bayes networks assume *conditional independence* when combining evidence from different rules, as in the Chernobyl example. Figure 2 shows *Pr(A&B)* as a function of *Pr(A)* given *Pr(B)=0.6*, assuming minimum overlap between *A* and *B* (Minc), independence (Ind), or maximum overlap (MaxC), which is the Fuzzy Set assumption. It illustrates the range of results possible from alternative assumptions about correlation.

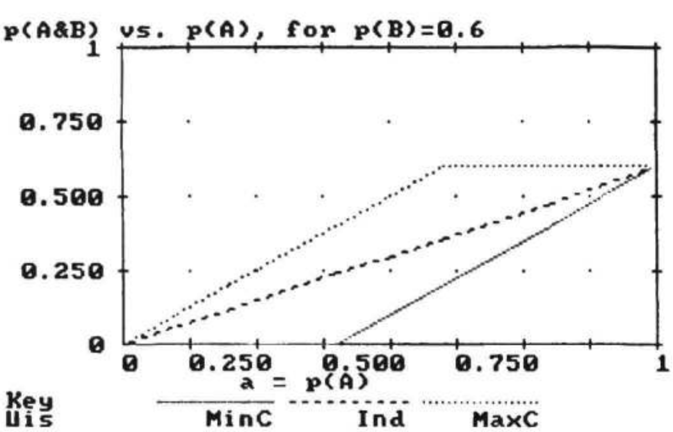

Figure 2: The effect of assumptions about correlation, (Wise & Henrion, 1986)

It has sometimes been claimed as an advantage of some non-probabilistic UISs, including Fuzzy Set Theory (Bonissone, 1986), that they avoid having to make *any* assumptions about dependencies. But in fact, as we have seen, the Fuzzy Set combination functions are equivalent, at least in effect, to specific probabilistic assumptions. It is true that non-probabilistic languages for uncertainty do not provide a general framework for modelling correlated evidence, since they do not provide a well-defined language for expressing the ideas. But to claim that they can therefore avoid making unsupported assumptions about correlations is akin to claiming that a new settler in Alaska can deal with the winter precipitation by adopting the language of an equatorial tribe with no word for snow.

Unfortunately to deal completely with probabilistic dependences is inherently complex (exponential in the number of evidence sources), and no inference network system which represents uncertainty in each proposition by one or a few parameters can deal with its full complexity. An alternative approach to is to represent uncertainty by a range of two probabilities and to compute both the largest and smallest probabilities compatible with the ranges of the antecedents. This does avoid making any specific, unsupported assumptions , although there is a danger of ending up with vacuous results (probability limits of 0 and 1).

The original Bayes' Network scheme of Kim and Pearl is restricted to Chow Trees i.e. singly connected graphs, so that conditional independence between convergent sources of evidence can be preserved (Kim & Pearl, 1983). However Pearl has suggested a method of removing the cycles, either by conditioning on variables in the cycle, or by adding extra nodes (hidden variables) that allows restructuring the probabilistic dependencies to avoid cycles (Pearl, 1985b).

Another approach is to represent the uncertainty in each proposition by a sample of truth values representing a random sample of possible worlds. These can be combined and propagating using the usual mechanisms of deterministic logic. Correlations due to multiple paths in the inference network or dependencies specified between inputs are handled correctly without special mechanisms. This *incidence calculus* (Bundy, 1986) or *logic sampling* (Henrion, 1986) involves a form of Monte Carlo simulation. Its accuracy depends on the sample size chosen. This approach can be reasonably efficient (it is linear in the network size) and seems promising, but its full potential remains to be explored.

Diagnostic and predictive inference

Diagnostic inference involves reasoning from observable manifestations to hypotheses about what may be causing them, for example reasoning from symptoms to diseases. *Predictive or causal inference* involves reasoning from causes (or causal influences, such as genetic or environmental factors that might increase susceptibility to a disease) to possible manifestations (Tversky & Kahneman, 1982). Consider the following:

> ***The sneeze example:*** *Suppose you find yourself sneezing unexpectedly in the house of an acquaintance. It might either be due to an incipient cold or your allergy to cats. You then observe animal paw marks, which increases your judged probability of a cat in the vicinity (diagnostic inference), which, in turn, increases the probability that you are having an allergic reaction (predictive inference). This also explains away the sneezing, and so decreases the probability you are getting a cold.*

Notice that this reasoning involves a mixture of both diagnostic and predictive inference. Having rules that allow reasoning backwards and forwards like this creates a danger of vicious circles, where, say, the probability of a cat would increase the probability of the allergic reaction, and vice versa. To avoid this, it seems necessary to keep a record of the sources of different uncertain evidence for each variable, so that you can avoid possible double counting. Pearl's scheme for Bayes' Networks keeps the flows of diagnostic and predictive evidence separate to avoid such cycling, and only combines them to calculate the aggregate degree of belief in each node (Pearl, 1985b). Figure 3 shows a Bayes Network representation of propositions mentioned.

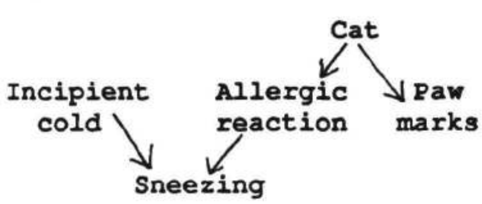

Figure 3: Bayes' network for sneeze example

Independent evidence for the allergy helps to explain the sneezing and so reduces the probability of a cold. Alternatively, the observation of a mild fever might increase the probability that it was a cold and so decrease the probability of the allergy. Thus the presence of sneezing induces a negative correlation between the cold and the allergy, which would otherwise be independent. This kind of reasoning, which we may term *intercausal*, is a natural consequence of the simple logical relation that sneezing can be caused either by a cold or an allergy. Pearl's propagation scheme for Bayes Networks models this correctly, but other rule-based schemes have a very hard time with it.

Consider the following rule for medical diagnostic inference which performs intercausal reasoning (Clancey, 1983):

> *If the patient has a petechial rash and does not have leukemia, then neisseria may be present.*

This reflects the medical fact that a petechial rash can be caused either by neisseria or by leukemia, and so the rash is evidence for neisseria unless it has been explained by leukemia.

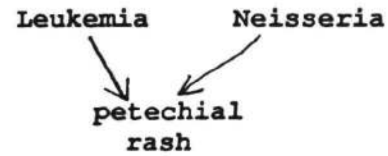

Figure 4: Bayes' network for rash example

If the system was also intended to help diagnose leukemia, it would need an additional rule:

> *If the patient has a petechial rash and does not have neisseria, then leukemia may be present.*

With both rules the inference network would contain a cycle (from leukemia to neisseria and back again), which is liable to cause updating difficulties, at least sensitivity to the sequence in which evidence arrives. The underlying problem is that these rules embody general knowledge about inference under uncertainty as well as specific medical knowledge.

It would be far better to be able to specify the essential medical knowledge in causal form, that "a petechial rash can be caused either by neisseria or by leukemia" (with specified conditional probabilities if the relationships are uncertain). The UIS should then be able use this to make the uncertain diagnostic inferences implied by either rule, or even reason predictively from the diseases to the symptom, according to the demands of the situation. Pearl's scheme can do this effectively and consistently, while maintaining a clear separation between the inference methods and domain knowledge. Clancey in a critique of Mycin has emphasized the desirability of separating the representation of inference strategy from domain knowledge (Clancey, 1983). But it does not appear that schemes, like Mycin, or Prospector representing knowledge primarily as diagnostic rules, rather than probabilistic causal relations, are capable of this.

Should UISs emulate humans?

This objection to probability has been forcefully stated by Paul Cohen:

> *"[it is] puzzling that AI retains models of reasoning under uncertainty that are derived from normative theories..., because the assumptions of the normative approaches are frequently violated, and because the probabilistic interpretation -- and numerical representation -- of uncertainty summarizes and fails to discriminate among reasons for believing and disbelieving. ... models of humans as perfect processors of information are not only inaccurate, but also unlikely to lead to efficient and intelligent reasoning." (Cohen, 1985), p.9*

Cohen here advocates the strategy, which appears to have been successful elsewhere in AI research, of adopting heuristic approaches based on human intuitive reasoning, rather than theoretically optimal, but computationally intractable schemes. Cognitive psychologists have indeed provided us with ample evidence that human inference under uncertainty is not accurately modelled by Bayesian decision theory (Kahneman, Slovic & Tversky, 1982). But there is little experimental evidence that proposed non-probabilistic UISs are better models. Very likely there is considerable variation between tasks and between individuals. It is an important and challenging task for cognitive psychologists to build better models of judgment under uncertainty, but it seems quixotic for those primarily interested in developing better expert systems to seriously attempt to emulate human judgment. That is not to say that evidence about human reasoning including introspection may not give us excellent ideas for devising new and better UISs, but the criterion for judging their usefulness should be the quality of their performance, rather than how well they simulate human thought processes.

One feature of human judgment observed by psychologists has been termed the *representativeness heuristic*: When asked the probability that object A belongs to class B, people typically evaluate it by the degree to which A is representative of B, that is by the degree to which A resembles B (Kahneman, Slovic & Tversky, 1982). This leads to judgments which are insensitive to the prior probability of A, and contrary to Bayes' rule. Cohen and colleagues explicitly adopt the representativeness heuristic for representing uncertainty in the classification system, GRANT, (Cohen *et al*, 1985), which deliberately ignores prior information. Other UISs also explicitly exclude prior probabilities, including Mycin (Buchanan & Shortliffe, 1984). The rationale has been that prior probabilities are too hard to estimate, and it is better to avoid them. However, for both GRANT and Mycin (and its derivatives), ignoring priors is functionally equivalent to assuming equal priors (for the probabilities of agencies funding a proposal, or the probabilities of disease organisms). Occasionally such flat priors may be appropriate, but more often it means ignoring important information about differing frequencies.

The following example points up the dangers of this approach:

> **Blood test example:** *James is engaged to be married, and takes the routine pre-marital blood test required by the state. To his horror, the test comes back positive for syphilis. His physician explains to him that the test is very reliable, having an false positive rate of 1%, and so the chance he has the disease is 99%. Aghast, James wonders what to tell his fiancee.*

Most physicians will give the same advice as James' one does. Like other people, they are poor intuitive Bayesians (Kahneman, Slovic & Tversky, 1982) and tend to ignore prior or base-rate information. Using the representativeness heuristic, the chance that James has VD is judged by the degree to which he (having a positive blood test) is representative of people with VD. In this, and many similar cases, the heuristic leads to a conclusion that is badly wrong:

> *Fortunately, James' fiancee, Alice, is not only understanding, but a Bayesian statistician. She finds out from the physician that the prevalence of syphilis among men from James' background is about 1 in 10,000. Based on this, she concludes that the probability he actually has the disease is about 1%, and decides to go ahead with the wedding.*

For those of us not so lucky as to be marrying a Bayesian, would we rather consult a physician or an expert system modelled on normal human judgment, or would we prefer one based on normative Bayesian principles?

Does it matter which you use?

Even if one accepts the arguments that probability is epistemologically adequate to represent uncertainty, it is clear that only approximations to it are computationally tractable for real systems. Despite the theoretical differences between systems, does it really make much difference to the conclusions of a rule-based expert system which scheme you use? There has been a common perception in the AI community that the performance of systems is relatively insensitive to the choice of UIS; that the important differences are to do with qualitative knowledge rather than quantitative uncertain inference. This may be true, at least for some domains. But so far, belief in this insensitivity seems to have been based primarily on ideology, since there has been little systematic analysis or experimental evidence published.

One early piece of evidence was a comparison of Mycin's method for combining evidence from different rules with a probabilistic model (Shortliffe & Buchanan, 1975). This showed a pronounced tendency to under-respond. On average, strong aggregate evidence for or against a conclusion was computed to be as about half as strong (half the CF) as it should be. This in itself, may not have mattered much, since Mycin used relative CFs for making decisions. But in 25% of the cases the system responded in the wrong direction. Confirming evidence actually reduced the CF or vice versa. The developers of Mycin suggest that Certainty Factors are satisfactory for the initial application domain (selecting antibiotic therapy), but that "We would need to perform additional experiments to determine the breadth of the model's applicability", (Buchanan & Shortliffe, 1984) p. 700. However CFs and related UISs are now being used for many other applications, apparently without the benefit of such experiments.

Recently there have been a few comparative studies of UISs. Tong and colleagues have compared 12 variants of the Fuzzy Set Rules for *and*, *or*, and *modus ponens* combinations in terms of their performance in a fixed rule base (Tong, 1985). They found that the performance of all rules with smooth response (i.e. not discontinuous) did reasonably well in their example. Vaughan and colleagues have done a comparison of the Prospector scheme with odds-ratio updating (Vaughan, 1986) for a systematic range of single rules. They found that Prospector did well in many cases, but that there are some situations in which it performs poorly. Wise has argued that the appropriate standard for comparison is a system using Maximum Entropy to fit a complete prior to specified input probabilities and rule strengths, and Minimum Cross Entropy for updating it (Wise & Henrion, 1986, Wise, 1986). This ME/MXE approach is actually a generalization of the odds-ratio approach used by Vaughan *et al.* Wise has performed comparisons of six UISs, including CFs, Fuzzy Set Theory, and a probabilistic scheme with Conditional Independence, against the ME/MXE scheme, for individual rules, and small assemblies of 2 or 3 rules, 30 cases in all, each with all input probabilities systematically varied (Wise, 1986). For purposes of comparison, the degree of membership of a Fuzzy Set was equated to probability. The performance of the UISs varied considerably over the different situations. All worked well in at least some cases, and none worked well in all cases. There were some situations in which some UISs were worse than random guessing.

It is not hard to construct examples in which CFs (and other widely-used UISs) produce results that disagree badly with a complete probabilistic analysis, even having the wrong qualitative sensitivities. Experienced knowledge engineers may be aware of at least some of the problems inherent in the UIS they use, and may know how to modify rule-sets to mitigate the undesirable behavior, at least for some anticipated situations. However, some of the problems are quite subtle, even though their effects can be severe. In any case it seems dangerous to rely on the ability of the knowledge engineer to "program around" such problems, particularly given our sketchy understanding of what all the problems are.

Conclusions

Probability has often been criticized as epistemologically inadequate for representing uncertainty in AI, but many of these criticisms have stemmed from incomplete understanding of probabilistic inference. In this paper, I have focussed on a number of important advantages that probabilistic representations have over other proposed measures of uncertainty, which have not loomed so large in the debate hitherto. Personal probability has an unambiguous operational definition, and it is embedded in a rational theory of decision-making under uncertainty -- we know what it means, and we know how to make decisions using it. Probabilistic inference is epistemologically adequate to perform three important kinds of reasoning that humans are capable of: (a) taking into account non-independence between sources of evidence, (b) engaging in mixed diagnostic and predictive inference, and (c) inter-causal inference, between alternative causes of an event, as in "explaining away". Non-probabilistic rule-based UISs may be able to simulate these in particular cases, at least qualitatively, but only by confounding general knowledge about uncertain inference with the domain specific knowledge in the rules.

These types of inference are important, and we can learn a great deal from studying human reasoning. But it is not necessarily desirable that a UIS should duplicate *all* features of human judgment under uncertainty, including such strategies as the representativeness heuristic that can lead to severe biases, as in the blood test example. Where cognitive limitations cause human judgments to diverge from the results of normative theory, surely it is better to use the latter when expert systems are advising on important decisions, as in medical or defense applications.

If one accepts the arguments for the epistemological adequacy, or even superiority, of probability, serious questions may still be raised about its heuristic adequacy -- can practical, computationally efficient implementations be built? The Bayes' Network approach seems very promising, but work still remains to be done to deal conveniently and generally with multiply connected networks (i.e. dependent sources of evidence). Monte Carlo logic sampling seems to offer possibilities here, both as a practical implementation and as an intellectual link to deterministic logic. Although several probabilistic methods for dealing with second-order uncertainty, distinguishing the effect of different sources of evidence, and explaining probabilistic reasoning have been suggested, there remains considerable work to be done to develop implementations and experimental study of their acceptance and usefulness to system builders and users. Whatever the theoretical merits of probabilistic representations, the AI community has a venerable tradition of pragmatism, and many will understandably remain unconvinced until these more sophisticated probabilistic schemes have demonstrated success in large scale applications. On the other hand, disturbing evidence is emerging about the performance of the most popular UISs, and complacency would be inappropriate as they are applied to new tasks with major potential consequences. There is an urgent need for more rigorous experimental evaluations of UISs for a range of realistic rule-bases to find out under what circumstances they can be relied on, and when they may be seriously wrong.

Acknowledgements

These ideas have benefitted from discussions with Ben Wise and Chris Elsaesser. This work was supported in part by the National Science Foundation under grants PRA-8413097 and IST-8603493.

References

Beyth-Marom, R. How probable is probable? A numerical taxonomy translation of verbal probability expressions. *J. of Forecasting*, 1982, *1*, 257-69.

Bonissone, P.P. A Fuzzy Sets based linguistic approach: Theory and applications. In M.M. Gupta & E. Sanchez (Eds.), *Approximate Reasoning in Decision Analysis*. North-Holland, 1982.

Bonissone, P. P. Plausible Reasoning: Coping with Uncertainty in Expert Systems. In *Encyclopedia of AI*. John Wiley & Sons, 1986.

Buchanan, B. G. & Shortliffe, E. H. *Rule-based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, Mass, 1984.

Bundy, A. Incidence Calculus: A mechanism for probabilistic reasoning. In Kanal, L.N. & Lemmer, J. (Eds.), *Uncertainty in AI.* North-Holland: Amsterdam, 1986.

Cheeseman, P. A methods of computing generalized Bayesian probability values for expert systems. In *Proceedings of 8th International Joint Conference on AI.* Karlsruhe, W. Germany: , 1983.

Cheeseman, P. In Defense of Probability. In *Proceedings of 9th International Joint Conference on AI.* Los Angeles, Ca: , 1985.

Clancey, W.J. The Epistemology of a Rule-Based Expert System - A Framework for Explanation. *AI*, 1983, *20*, 215-251.

Cohen, Paul R. *Heuristic Reasoning about Uncertainty: An AI Approach.* Pitman: Boston, 1985.

Cohen, P., Davis, A., Day, D., Greenberg, M., Kjeldson, R., Lander, S., & Loiselle, C. Representativeness and Uncertainty in Classification Systems. *The AI Magazine*, Fall 1985, pp. 139-149.

de Finetti, Bruno. *Theory of probability.* New York: Wiley, 1974.

Doyle, J. *The Ins and Outs of Reason Maintenance* (Tech. Rep. CMU-CS-83-126). Carnegie Mellon University, 1983.

Duda, R. O., Hart, P. E., and Nilsson, N. J. *Subjective Bayesian Methods for Rule-Based Inference Systems* (Tech. Rep. SRI Technical Note 124). SRI International, 1976.

Fox, J. Knowledge, decision making and uncertainty. In W. Gale (Ed.), *AI and Statistics.* Addison-Wesley: Reading, MA, 1986.

Gale, W. (ed.). *AI and Statistics.* Addison-Wesley: Reading, MA, 1986. (In press).

Henrion, M. *Propagation of Uncertainty by Logic Sampling in Bayes' Networks* (Tech. Rep.). Department of Engineering and Public Policy, Carnegie Mellon University June 1986.

S. Holtzman & J. Breese. Exact reasoning about uncertainty. In *Workshop on Uncertainty and Probability in AI.* AAAI 1985.

Kahneman, D.; Slovic, P. & Tversky, A. *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press, 1982.

Kanal, L.N. & Lemmer, J. (eds.). *Uncertainty in AI.* North-Holland: Amsterdam, 1986.

Kim, J. & Pearl, J. A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems. In *Proceedings of IJCAI-83.* , 1983.

Lichtenstein,S., Fischhoff, B. & Phillips, L.D. Calibration of probabilities: The state of the art to 1980. In Kahneman, D.; Slovic, P. & Tversky, A. (Eds.), *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press, 1982.

Lindley, D.V. Scoring rules and the inevitability of probability. *international Statistical Review*, 1982(50), pp. 1-16. (with commentary).

Oden, G.C. Integration of Fuzzy Logical Information. *J. Experimental Psychology: Human Perception and Performance*, 1977, *3*(4), 565-75.

Pearl, J. *How to do with probabilities what people say you can't* (Tech. Rep. CSD-850031). Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles, CA, 1985.

Pearl, J. *Fusion, Propagation, and Structuring in Bayesian Networks* (Tech. Rep. CSD-850022). Cognitive Systems Laboratory, Computer Science Department, University of California, Los Angeles, CA, 1985.

Savage, LJ. *The Foundations of Statistics.* New York: John Wiley & Sons, 1954.

R.D. Shachter. Intelligent Probabilistic Inference. In *Workshop on Uncertainty and Probability in AI.* AAAI, 1985.

Shafer, G. *A Mathematical Theory of Evidence.*

Princeton University Press, 1976.

Shore, J. E., and Johnson, R. W. An Axiomatic Derivation of the the Maximum Entropy Principle. *IEEE Transactions on Information Theory*, January 1980, pp. .

E. H. Shortliffe & B. G. Buchanan. A model of Inexact Reasoning in Medicine. *Mathematical Biosciences*, 1975(23), pp. 351-379.

Spiegelhalter, D. J. A statistical view of uncertainty in expert systems. In W. Gale (Ed.), *AI and Statistics*. Addison-Wesley: Reading, MA, 1986. (in press).

Tong, R. M., Shapiro, D. G. Experimental Investigations of Uncertainty in a Rule-based System for Information Retrieval. *International Journal of Man-Machine Studies*, 1985.

Tversky, A. & Kahneman, D. Causal Schemata in Judgments under Uncertainty. In Kahneman, D.; Slovic, P. & Tversky, A. (Eds.), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982.

Vaughan, D. S., Perrin, B. M., Yadrick, R. M., Holden, P. D., Kempf, K. G. An Odds Ratio Based Inference Engine. In Kanal, L.N. & Lemmer, J. (Eds.), *Uncertainty in AI*. North-Holland: Amsterdam, 1986.

Wise, B.P. *An experimental comparison of uncertain inference systems*. Doctoral dissertation, Carnegie Mellon University, Department of Engineering and Public Policy, 1986.

Wise, B.P. & Henrion, M. A Framework for Comparing Uncertain Inference Systems to Probability. In Kanal, L.N. & Lemmer, J. (Eds.), *Uncertainty in AI*. North-Holland: Amsterdam, 1986.

Zadeh, L. A. Making Computers Think Like People. *IEEE Spectrum*, August 1984, pp. .