UNCERTAINTY AND PROBABILITY

Jane Terry Nutter
Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24060

ABSTRACT

This paper examines the relationship between probability theory and reasoning in uncertainty, and argues that (contra opposing views) probability theory does have a place in reasoning in uncertainty, but that its place is more restricted than many of its advocates claim. Two major theses are argued for. (1) Reasoning from probabilities works well in domains which permit a clear analysis in terms of events over outcome spaces and for which either large bodies of evidence or long periods of "training" are available; but such domains are relatively rare, and even there, care must be taken in interpreting probability results. (2) Some generalizations with which AI applications must concern themselves are not statistical in nature, in the sense that statistical generalizations neither capture their meanings nor even preserve their truth values. For these contexts, different models will be needed.

I INTRODUCTION

Probability estimates have been used to aid decision making in AI systems for over ten years now (Shortliffe and Buchanan 1975; Duda, Hart, and Nilsson 1976), and have been under fire even longer (McCarthy and Hayes 1969). Recently, the increased attention devoted to common sense reasoning and reasoning in contexts of uncertainty has fueled the debate, and clearly defined battle lines have emerged. Supporters point to a well-developed rigorous formalism for dealing with uncertainty (Cheeseman 1985; Ginsberg 1985). Opponents continue to object that applying Bayesian methods to calculate probabilities requires information that is not generally available ("where do the numbers come from?"), and that, insofar as it is available, is usually tainted (Kahneman, Slovic, and Tversky 1982). In addition, they charge that using probabilites suppresses important information (Cohen and Grinberg 1983; Sullivan and Cohen 1985); that statistical analyses fail to distinguish uncertainty from inherent "fuzziness" (Zadeh 1981); that judgements of typicality and normic generalizations underlie much reasoning from uncertainty and are not probabilistic (Rosch 1975; Rosch and Mervis 1975; Rosch, Mervis, Gray, Johnson and Boyes-Braem 1976; Scriven 1959; Scriven 1963; Nutter 1982); and so on.

Even supporters of statistical approaches separate into those who prefer straightforward Bayesian analysis (Cheeseman 1985), those who prefer the Dempster-Shafer (Dempster 1968; Shafer 1976) approach (Yen 1986; Ginsberg 1984; Ginsberg 1985; Strat 1984; Yu and Stephanou 1984); and those who prefer some other variant of the Bayesian approach (Snow 1986). Others are developing theones to permit logic-style inferences about probabilites, especially in the realm of reasoning about independence (Pearl 1986), while continuing to argue that probability provides an epistemically adequate and effective framework for the general problem of reasoning in uncertainty (Henrion 1986).

Both sides present their positions forcibly and plausibly, but not always with attention either to their opponents' points or to independent substantiation. As a consequence, arguments on both sides of the fence have tended to produce more heat than light. This paper aims to lower the temperature while illuminating the terrain. In the end, this paper argues neither for nor against statistical reasoning in AI. Rather it argues that statistical methods do apply, but only in some cases, and that care must be taken to identify those cases correctly, to fulfill requirements for reliable results, and to make sure that what is treated as a probability is indeed probabilistic in nature.

The body of this paper consists of three sections. The first describes elementary aspects of probability theory, to form a basis for discussion. The second characterizes some of the features an AI application and its domain must have for statistical analysis to be useful. The final section describes cases of uncertainty that cannot be represented as probability.

II PRELIMINARY REMARKS ON PROBABILITY

A. What arc Probabilities?

The philosophical nature of probabilities matters less for AI purposes than what kinds of phenomena classical and Bayesian probability analyses model. However, given the vehement disputes on this issue, a few observations may be useful. Statistics begins investigating probabilities in any particular instance by defining (at least loosely) a space of outcomes, that is, mutually exclusive observations of test results. Events are sets of outcomes from that space. When probability theorists refer to probabilities, they typically mean event probabilities, that is, the likelihood that the outcome of a particular test will belong to the set which defines the event. This likelihood is traditionally defined in terms of frequency: given a "sufficiently large" number of tests, what proportion of all outcomes fall in the event set?

The frequency view has been attacked for centuries; a recent criticism can be found in (Cheeseman, 1985). Probably the most pursuasive argument against the frequency view from an AI standpoint is that on that view, each event has exactly one correct probability. But for AI purposes, such a probability is neither attainable nor in some cases even interesting. Rather, we are interested in the probability of an hypothesis given the current evidence. Critics further object that the frequency theory "restricts probability to domains where repeated experiments (e.g. sampling) are possible, or at least conceivable" (Cheeseman, 1985). In addition, the concept of "long run frequency" has bothered people for centuries. How long? How do you know? Why should "large numbers" (how large?) have special properties?

These objections can be met without deserting a frequency-based approach. The probability of any hypothesis on the basis of the current evidence can be - and in normal statistical practice is - interpreted as the conditional probability of the hypothesis given the conjunction of events which that evidence reflects. In other words, in addition to a single, well-defined probability for every event over the space, the frequency view also provides a way to represent precisely the relativized probabilities we are most interested in (and these are exactly the probabilities that statisticians investigate).

Classical statistics texts also contain chapters on game theory and decision theory which describe techniques for estimating probabilities on the basis of very small samples (see e.g. Freund and Walpole 1980, Chapter 9, or almost any other freshman text). So not only does classical statistics recognize that this can be done, the theory instructs the interested in how to do it; only, it also warns not to place great faith in the accuracy of such estimates.

The hardest question to meet is the philosophical question of the significance of the Law of Large Numbers: what does it mean to talk about "long run" frequencies? Classical statistics provides some tests for whether an actual sample is large enough; but that cannot answer the philosophical question. The best that can be said here is that other approaches have their own philosophical questions that they cannot answer, but none of these philosophical questions seem to affect AI.

The classical alternative to the frequency view is the subjective probabilities view, which derives from the views of the 18th century English clergyman Thomas Bayes. On this approach, probabilities measure certainty levels. Two options here should be distinguished. The first is well-defined, and clearly subjective (as philosophers use the term): the probability of an event given the current evidence is the measure of the degree to which a particular specific "real live" individual believes that the event will occur on the basis of that evidence. The problem here is evident: people will believe all sorts of things, and different things at different times, for different reasons or none at all. There is no reason to suppose that one person's "probability" in this sense will match another's, and no grounds for a science of probability at all.

It is unlikely that many supporters of subjective probabilities ever meant that, though they often seem to say it:

> ... the following definition is put forward as one that withstands all previous criticisms: The (conditional) probability of a proposition given particular evidence is a real number between zero and one, that is a measure of an entity's belief in that proposition, given the evidence.
>
> (Cheeseman 1985; emphasis in original)

The alternative, and the view that is actually held, is that probabilities measure how much an ideal rational subject ought to believe that an event will occur, given the evidence. This option makes probabilities relative (to evidence), but not really subjective: no actual subjects are involved any more. This approach has two difficulties, both as obvious and as pressing as the problem the frequency theory has with understanding the long run. First, what makes someone an ideal rational subject? Probability cannot be considered well-defined on this view until that is spelled out. Second, how other than by measured frequencies can we establish the degree to which such a subject ought to believe that a given event will occur?

B. Htm do Probabilities Behave?

The mathematics for measuring probabilities is the same on both these competing definitions: Bayes's Theorem is a theorem of classical statistics, for example. The significant differences come in questions of when it is legitimate to apply the formulas, and what they can be taken as establishing. In this regard, it seems hat the frequency analysis has an advantage: designers of AI systems generally care less whether their systems "ought" to believe their answers than how often those answers are right. For systems whose judgements have practical consequences, we should measure and maximize that if we measure anything. But whatever philosophical view of probabilities we take, the mathematics always agrees with long run frequency expectations in all situations in which we can make sense of them.

Several of the mathematical features of event probabilities and their measurement are counterintuitive enough to be worth mentioning. For example, experiments structured by statisticians always assume more than is known. Statistical experiments define a hypothesis about the likelihood of an event, and then compare actual observations against the predictions of the hypothesis. This fact has implications frequently overlooked in AI debates. One such implication relates to the controversy over the so-called assumption of maximum entropy: the policy of assuming all events independent unless a connection has been found (see Cheeseman 1985). Opponents claim that this practice involves assuming more than is known, since the events in question may be dependent; supporters respond that the assumption of maximum entropy provides "a neutral background against which any systematic (non-random) patterns can be observed.... [W]ithout this prediction, it is difficult to detect if the current information is incomplete, and thus to discover new information" (Cheeseman 1985). The truth is that any hypothesis provides a background for detecting deviation; and no experiment can be run without some hypothesis. The real question, then, is which hypotheses yield the best results without extensive "training"; this question must be answered by experiment.

Another, and for AI more serious, implication is that the outcome of an appropriate experiment must be observable independent of the statistical prediction. This is a problem for medical expert systems, for instance. The outcomes predicted by a system trying to solve problems at the level of "diagnose infectious disease" can not be straightforwardly observed: if they could, we wouldn't need the systems. This has serious consequences concerning the "trainability" of such systems; we will return to this later.

Finally, some simple properties of probabilities should be noted. For independent events the joint probability (probability that all events will occur) is the product of the probabilities of the events. (Events are independent provided that whether an outcome belongs to one does not affect how likely it is to belong to another.) Since all probabilities lie between zero and one, the joint probability of several independent events is always smaller than the probability of any one of them, unless all but one have probability one or at least one has probability zero. For dependent events, the joint probability is at most the maximum of the individual event probabilities, and it is that only if the corresponding event entails all the others. This has important consequences which we will return to later. Notice that the joint probability for dependent events may be zero even though none of the individual probabilities is, and it will always be so if at least two of the events are mutually exclusive. More subdy, the joint probability of, say, six events may be zero even though no two of them are mutually exclusive, if, say, five of them together exclude the sixth.

Similarly, the probability that an outcome will fall into at least one of several independent events is the sum of the probabilities of the events in question. If they are dependent, it is at least the maximum of the individual probabilities, and at most their sum. Notice that a false assumption of independence underestimates the probability of disjunctions *and* overestimates the probability of conjunctions. In a long chain of reasoning involving both, it is not at all clear that these offsetting errors would be easy to detect and isolate.

III REQUIREMENTS FOR PROBABILITY-BASED REASONING

Where decisions or predictions must be made on the basis of partial information, and where there is enough information to tell what outcomes are most likely given what is known, probability theory can be used to make these decisions and predictions accurately and responsibly. The mechanism is available, it is well-defined and well-understood, it gives good results, and it is the only mechanism we have with those properties. These facts alone show that probability has a place in reasoning in uncertainty, and henceforth I will take that as established. But probability-based reasoning takes more than arithmetic. This section looks at some of those needs and their consequences for AI systems.

A. Where Do the Numbers Come From?

Any application must consider where the system gets its data. There are two possiblities: a system may use known probability values and distributions, or it may begin with initial probability estimates that are refined in the light of further evidence (in Bayesian terminology, these are called "prior probabilities" or "priors"). The first choice gives better results, and is easier to implement. Unfortunately, it requires a depth of knowledge in the application domain that is almost never attainable, and so the second approach is more often taken.

"Prior probabilities" are not probabilities: they are guesses. (The only interpretation under which priors qualify as probabilities is the extreme subjective view above, on which anyone's level of commitment is *a fortiori a* probability, but not an interesting one, since on this view a science of probabilities is impossible and in any case would have nothing to do with what *is* or is not likely *to* happen.) If the priors are bad guesses, results based on them will be bad results, even if all other assumptions hold. Designers can deal with this problem two ways: use data and experiments to start with good prior probability estimates, or validate or train the system to improve bad ones.

Basing priors on data and experiments is straightforward and the most reliable course, when it can be done. Where data from reasonably representative samples already exist, those data are used. Where data dp not exist, classical experiments based on the outcome distribution and the predictions of a testable hypothesis are designed. For some AI domains, this procedure is feasible. The domain for PROSPECTOR, for instance, is small and reasonably well understood. We have fairly good information on the occurrence rates of different minerals, and it is easy to imagine, at least, what it would be like to have reliable estimates of joint and conditional probabilities over many of the relevant events. Since knowledge of conditional probabilities entails knowledge of which events are independent of each other, it also largely eliminates the need for assumptions like maximum entropy.

Unfortunately, domains in which this kind of information is available are rare. By contrast, if we consider a medical domain, the number of possible outcomes is huge, their distributions are less well known, their interactions are frequently unknown, and reliable data are notoriously hard to come by. The number and scope of experiments needed to establish such data are overwhelming. In cases such as this, some other course must be taken. The usual approach relies on expert opinions elicited formally or informally from individuals or from panels by any of a variety of strategies.

There are many reasons to doubt the accuracy of such estimates. First, people in general and trained scientists in particular are lousy at estimating probabilities. The classic studies showing this were published over a decade ago (Tversky and Kahneman 1971) (Tversky and Kahneman 1974) and many more have followed confirming and extending the results (Kahneman, Slovik, and Tversky 1982). Second, even supposing the experts on the panel avoid the most common kinds of error, they lack the information they would need to make accurate estimates. (The problem isn't that the experts refuse to give us this information; they don't have it cither.) The question is what to do about it.

Option One: Do nothing. It is a theorem that repeated application of Bayesian analysis to a given event yields a sequence of priors which converges on the frequency probability, however abysmal the original guess. So if we start with whatever priors we have and let the system refine them as it goes, its results will improve in the course of nature.

There are three problems with this, (a) The system can only improve its estimates if it has independent confirmation of the outcome. For a medical system, this independent information is often unavailable. (That a patient got well after treatment does not confirm the diagnosis, for instance, since many treatments help a broad range of problems, and anyhow most patients get well no matter what.) Without independent information on the outcome the system's estimates will not improve. Worse, they may *seem* confirmed because disconfirming instances, although present, go undetected, (b) Even if the system improves over time, it starts badly. If we care what answers we get, we may not want to tolerate this. (It is good if a medical expert system learns ro recommend the right treatment; it is less good if it learns by recommending wrong treatments that kill half of New York.) (c) Although the mathematics guarantees that estimates will converge, it doesn't guarantee that they will do so rapidly. As a general rule, estimates converge fairly slowly, and the more inaccurate one there are (and the more inaccurate they are), the slower the convergence. If the initial priors are bad enough, it may be a very long time before the system's predictions get much better.

Option One A: Do nothing, but report ignorance. Many researchers propose ranges instead of point probabilities to reflect "second order" uncertainty: wide ranges reflect shaky estimates, while narrow ranges reflect more solid ones. Dempster-Shafer theory gives way to calculate probabilities from these ranges. This approach has the advantage that a user who gets an answer with an attached probability of, say, [0.41, 0.99] can tell that the system really doesn't know, whereas a one who gets the same answer with an attached probability of 0.7 can not. But we have no more reliable guide for setting ranges than for setting priors; and a bad answer with a warning is still a bad answer. This approach can be combined with the steps below for improving the quality of the priors; whether and how much improvement will result remains to be seen.

Option Two: Validate the system's predictions. It may not be possible to check a system's judgements against actual outcomes, they can be checked against human judgements. Comparison against human performance may not give results as good as "true" priors would, but it meets any standard that could reasonably be expected. But in this context as in any other, care must be taken to ensure that validation is not tainted.

In particular, the system's performance should be compared with the performance of several experts (the more the better) on the same cases (not just ones that look similar). In addition, the experts in question should not know the system's design, its basic assumptions (including its priors), what questions it asked, what conclusions it reached, or how it reached them. In practice, this means that systems cannot be validated in the environments in which they would be used: if the considered opinions of several experts were routinely available, nobody would need the expert system. Finally, since the point of validation is to tune the system's priors, this phase must be pursued with the attitude that in case of disagreement, and in the absence of overwhelming evidence to the contrary, the experts are right and the system is wrong.

Option Three: Train the system. If outcomes are independently observable, they can be used to put a system through a training phase. This amounts to the "do nothing" approach, but pursued "off-line" and with careful supervision until there are signs of convergence. Given observable outcomes, this lets the mathematics work for the designers, while allowing intervention if thrashing values show that a particular estimate was so bad that it will take a long time to converge. It should be noted, however, that if many factors are involved, or if original estimates are very bad, this course may require very large amounts of data before responses become reliable enough for the system to "go on-line".

B. What Can the Numbers Tell You?

The most likely hypothesis is usually not the best explanation. The reason for this has to do with the intuitively paradoxical fact that a hypothesis which covers half the information and ignores (is indifferent to) the rest is more probable than one which gives exactly the same explanation of the first half of the data and then goes on to explain the rest. Recall that the probability of a conjunction is at most equal to the maximum of the probabilities of its conjuncts, and only reaches that limit when the conjunct with the highest probability entails all the others. It follows that more specific hypotheses are mathematically guaranteed to have lower probabilities than any less specific hypotheses they entail: adding information to a hypothesis reduces its probability. Always. Ten out of ten.

So, consider the following example, once again in the medical domain. Suppose (our data) that a patient has a fever, a sore throat, white spots on the tonsils, nausea, diarrhoea, and vomiting. Now consider two possible hypotheses: (A) the patient has strep throat; (B) the patient has strep throat and a gastro-intestinal virus. For the above reasons, A is necessarily the more probable hypothesis; but B is the better explanation.

In many cases, we really want the best explanation. This means something like, we want the hypothesis that best covers the facts while maintaining a reasonable probability. If the "answer space" has more than one level of granularity, this is different from the most probable hypothesis, because more specific explanations are better than more general ones, so long as they do not become intolerably unlikely. This is not to say that a system which is looking for the best explanation cannot use probability-based reasoning to advantage. But it needs a more sophisticated answer selection mechanism than "most probable hypothesis": whatever else goes on, at the end of the calculations, some reasoning takes place - to select the best explanation as opposed to most likely hypothesis - which is not probabilistic in nature. Hence, pace Cheeseman, even in situations which admit of a clear statistical analysis, it takes more than just probabilities to reason in uncertainty.

C. What do the Numbers Cost?

(Ginsberg 1985) proposes to use probabilities as a limiting threshhold to reduce the cost of inference. In particular, he suggests setting a threshhold so that once a conclusion's probability reaches that limit, it may be taken as established. For the sake of argument, say that the limit in question is 0.98. Suppose we know the following: anything which is A has a probability of 0.99 of also being B: x is A; anything which is C has a probability of 0 of also being B; x is C. Now: is x iB1 The answer, of course, is no. But if the system first finds the rule about As and the fact that x is A, giving a probability of 0.99 that x is B, it will cut off inference there. So the argument that using probabilities allows early termination should be taken with a grain of salt. The "best" (highest hit rate) hypothesis for any rare event is always the hypothesis that it never happens. That isn't useful if we are trying to predict, detect, and reason about rare events. Non-numerical inference mechanisms can stop inference early, for instance using resource limitation (see e.g. Donlon 1982); only, the system may miss an answer it would otherwise get. Likewise a system that threshholds on probabilities can stop early; but it may get the wrong answer. For more on threshholding problems, especially with regard to the Dempster-Shafer approach, see (Dubois and Prade 1985).

In addition, the training process may prove more expensive than it appears. (Ginsberg 1985) actually goes into detail on the process of getting tests to improve the quality of probability judgements, so the following remarks will be made in terms of his system's behavior. But these costs result from necessary steps if the system trains "on-line": any system that counts on this must either sacrifice accuracy or pay these prices. Each time an inference establishes a probability for a proposition, Ginsberg's system records the evidence that was used. Then every time new evidence changes a proposition's probability, the system retraces every inference that proposition was used in to update the conclusion's probability. Unfortunately, while the number of tests is small, none of the system's probabilities are reliable. So for most of its early life, the system has to recompute the probability for virtually all its inferences every time it sees anything. Using cut-offs makes this worse, by the way, since then the system really cannot just retrace the proofs that went through; it should also recheck those that were terminated early, of which no record exists. In effect, this means that rather than retracing a known path, it must perform the entire inference again from scratch.

In any case, AI systems rarely perform controlled statistical experiments, testing selected observable variables against the predictions of a hypothesis. Instead, they face specific situations, get information, and reason from it. This means that to get the full benefit of their "tests", systems that train must analyze every new piece of information to find out which of the events they know about this datum may fit. Consider such a system when it first meets Fred the Female Flamingo. Not only is it meeting a bird that flies, it is meeting something pink that flies, something over three feet tall that is pink, a female named Fred, and so on and on and on. The computational cost of extracting events from data and of retracing past inferences should be clear.

IV WHERE PROBABILITIES DON'T BELONG

Pace Cheescman and many, many others, not all that is not universal is probabilistic. For instance: if, as Cheeseman claims, the by now tormented example "Birds fly" really means "Most birds fly", then birds don't fly in the spring. In nesting season, baby birds outnumber adults. Baby birds don't fly. Hence in nesting season, "Most birds fly" is false. By the way, we can do even better with "Birds lay eggs," which is out-and-out false year round of at least half the population (none of the males do). So if Cheeseman is right, anyone who says in the spring that birds fly or at any time that birds lay eggs is mistaken. This is nonsense.

"Birds fly" must be decoded with respect to typicallity. If typicallity can be modeled by any statistical concept, it is category cue validity, not probability (Rosen 1975, Rosch and Mervis 1975; Rosch, Mervis, Gray, Johnson, and Boyes-Braem 1976). "Birds lay eggs," on the other hand, is not statistical at all. It is shorthand for a genuine, accept-no-substitutes universal - but not for "For all x, if x is a bird, then x lays eggs". Instead, it is in a class with the non-universal generalizations "Mammals bear young alive" (duck-billed platypi lay eggs) and "Reptiles and fish lay eggs" (garter snakes and sharks bear live young). By the way, these generalizations cannot be translated straightforwardly into probability claims counting over species instead of individuals: *no* species *either* bears live young *or* lays eggs; only (female) individuals belonging to species do.

The typicality-based uncertainty involved in generalizations like "Birds fly" centers on whether an individual has a property typical of things of its kind. Another kind of uncertainty, also related to typicality, centers instead on the extent to which a given property applies to an individual. This is the issue of vagueness, and the kind of inferences justified on the basis of degree-of-applicability are different from the kinds based on either typicality or probability. The difference between measure-of-membership and typicality is subtle but real. Typical birds fly. But how typical a bird Tweety is does not measure how well Tweety flies or how even how likely Tweety is to fly (hummingbirds are atypical in many ways, but spectacularly good fliers).

More importantly, because more often confused, degree-of-applicability does not work like probability. Consider the following two claims about Oscar the Ostrich:

(i) Oscar is a (typical) bird at 0.6
(ii) Oscar is male at 0.5

Claim (i) says that Oscar is not very birdlike (ostriches aren't). It is the sort of claim fuzzy set theory was originally developed to handle; it tries to measure the extent to which an individual falls in the bounds established by a fuzzy concept. Claim (ii) is a probability claim, reflecting that the system doesn't know whether Oscar is male but does know that Oscar is a bird, and that half of all birds are male, making the chances that Oscar is male 50-50. That is *not* to say that Oscar is half male: the system can consistently hold (ii) and also hold that any given bird is either completely male or not at all. The claims look superficially alike, but they cannot be taken the same way: (i) says that Oscar is not a very typical bird; (ii) docs not say that Oscar is not a very typical male, (i) does not really reflect incomplete information at all: it reflects a fundamental fact about how Oscar relates to a vague concept. In case (ii), the information is incomplete and can be completed by a single experiment (look at Oscar and see). If no difference in representation reflects this basic difference in content, the system will reason incorrectly a good part of the time.

Translations of common generalizations into probabilities do not preserve truth values, and translations of degree-of-applicability claims do not preserve inferences. So neither preserves meaning. Hence wherever else they can be used, probabilities cannot be used to understand generalizations and expressions of uncertainty in understanding natural language or in any system which gets its data in natural language. Natural language understanding requires inference in contexts of uncertainty all over the place, including inference from previously processed non-statistical generalizations. Hence there are instances of inference in contexts of uncertainty which are not amenable to analysis as probabilities.

V CONCLUSIONS

Probability theory is an important tool for many AI applications involving uncertainty. Where outcome likelihood is at stake, and where the necessary data are available, it is the best known tool. But it is also *hard*. It requires a detailed analysis and understanding of the domain, and either a great deal of data or an extensive validation procedure, if the answers obtained are to be reliable. There are no short cuts.

In any case, statistics cannot provide a panacea for all problems of uncertainty, generality, vagueness, and ignorance. No mathematical model however rigorous can be expected to give reasonable answers unless the situations to which it is applied conform to its underlying assumptions. The existence and prevalence of non-statistical generalizations shows that non- statistical models uncertainty must also be investigated. The existence of a well-defined, long-studied, clearly articulated theory argues strongly for its use — but not for its extension, willy-nilly, into other fields. Sometimes one theory is developed before another because the first makes sense and the second doesn't; other times, as for instance with physics and biology, we make progress on the easier problem first.

Ultimately, AI systems which reason in uncertainty need to combine these modes of uncertainty. In particular, we need to distinguish representations of probabilities, fuzzy membership, and generalizations based on typicality. One such approach would represent the first two as functions which take properties and yield either numbers or second order relations, and represent the third using some form of default reasoning (I have argued elsewhere for a simple monotonic extension to first order logic; see Nutter 1983). Axioms and rules can then make use of information when and as it is available, without misrepresenting that information and so making wrong inferences from it. Obviously this scheme is Utopian; so what can we do meanwhile? Some application domains are particularly amenable to one of these forms of inference, and can do without the others; in these cases, we make choices, hopefully understanding the limitations and trade-offs. On the science front, we can develop as many models as we can, with as close attention to the phenomena to be modeled as possible. And meanwhile, we can remember that given the diverse kinds of reasoning involved, anyone who claims to have the one and only key to reasoning in uncertainty is almost certainly wrong.

ACKNOWLEDGEMENTS

My thanks to Dr. Janet C. Rice of the Department of Biostatistics and Epidemiology, School of Tropical Medicine and Public Health, Tulane University, for her patience in looking over early versions of this paper to check for gross statistical bloomers. As always, credit goes to her only for accuracy; any errors lie solely at the author's door.

REFERENCES

1 Cheeseman, P. "In Defense of Probability." In *Proc. IJCAI-85.* Los Angeles, California, August, 1985, pp. 1002-1009.

2 Cohen, P.R. and M.R. Grinberg "A Framework for Heuristic Reasoning about Uncertainty." In *Proc. IJCAI-83.* Karlsruhe, Federal Republic of Germany, August, 1983, pp. 355-357.

3 Dempster, A.P. "A Generalization of Bayesian Inference." J. *Royal Statistical Society* 30 (Series B) (1958) 205-247.

4 Donlon, G. "Using Resource Limited Inference in SNePS." SNeRG Technical Note 10, Department of Computer Science, SUNY at Buffalo, Buffalo, New York, 1982.

5 Dubois, D. and H. Prade "Combination and Propagation of Uncertainty with Belief Functions - a Reexamination." In *Proc. IJCAI-85.* Los Angeles, California, August, 1985, pp. 111-113.

6 Duda, R.O., P.E. Hart and NJ. Nilsson "Subjective Bayesian Methods for Rule-Based Inference Systems." In *Proc. 1976 N.C.C.* AFIPS Conf. Proc. Vol. 45 (New York) 1976, pp. 1075-1082.

7 Freund, J.E. and R.E. Walpole *Mathematical Statistics,* third edition, Prentice-Hall, Inc. (Englewood Cliffs, New Jersey) 1980.

8 Ginsberg, M.L. "Non-Monotonic Reasoning using Dempster's Rule." In *Proc. AAAI-84.* Austin, Texas, August, 1984, pp. 126-129.

9 Ginsberg, M.L "Does Probability Have a Place in Non-Montonic Reasoning?" In *Proc. IJCAI-85.* Los Angeles, California, August, 1985, pp. 107-110.

10 Henrion, M. "Uncertainty in AI: Is Probability Epistemologically and Heuristically Adequate?" Seminar presented at Massachusetts Institute of Technology, Boston, Massachusetts, 10 November 1986.

11 Kahneman, D., P. Slovic and A. Tversky, editors *Judgement under Uncertainty: Heuristics and Biases.* Cambridge University Press (New York) 1982.

12 McCarthy, J. and P. Hayes "Some Philosophical Problems from the Standpoint of Artificial Intelligence." In B. Meltzer and D. Michie, editors, *Machine Intelligence 4,* Edinburgh University Press (Edinburgh) 1969, pp. 463-502.

13 Nutter, J.T. "Defaults Revisited, or 'Tell Me if You're Guessing'." *Proc. 4th Ann. Cog. Sci.* Ann Arbor, MI, June, 1982, pp. 67-69.

14 Pearl, J. "On the Logic of Probabilistic Dependencies," *Proc. AAAI-86.* Philadelphia, Pennsylvania, August, 1986, pp. 339-343.

15 Rosch, E. "Cognitive Representations of Semantic Categories." /. *Exp. Psych.: General* 104 (1975) 192-233.

16 Rosch, E. and C. B. Mervis "Family Resemblances: Studies in the Internal Structure of Categories." *Cog. Psych.* 7 (1975) 573-605.

17 Rosch, E., C.B. Mervis, W.D. Gray, D.B. Johnson and P. Boyes-Braem "Basic Objects in Natural Categories." *Cog. Psych.* 8 (1976) 382-439.

18 Scriven, M. "Truisms as the Grounds for Historical Explanations." In P. Gardiner, editor, *Theories of History,* Free Press (New York) 1959.

19 Scriven, M. "New Issues in the Logic of Explanation." In 5. Hook, editor, *Philosophy and History,* New York University Press (New York) 1963.

20 Shafer, G. *A Mathematical Theory of Evidence,* Princeton University Press (Princeton, New Jersey) 1976.

21 Shortliffe, E.H. and B.G. Buchanan "A Model of Inexact Reasoning in Medicine," *Math. Biosciences* 23 (1975) 351-379.

22 Snow, P. "Bayesian Inference without Point Estimates." In Proc. AAAI-86. Philadelphia, Pennsylvania, August, 1986, pp. 233-237.

23 Strat, T.M. "Continuous Belief Functions for Evidential Reasoning." In Proc. AAAI-84. Austin, Texas, August, 1984, pp. 308-313.

24 Sullivan, M. and P.R. Cohen "An Endorsement-Based Plan Recognition Program" In Proc. IJCAI-85. Los Angeles, California, August, 1985, pp. 475-479.

25 Tversky, A. and D. Kahneman, D. "Belief in the Law of Small Numbers", Psych. Bull. 2 (1971) 105-110.

26 Tversky, A. and Kahneman, D., "Judgement Under Uncertainty: Heuristics and Biases." Science 185 (1974) 1124-1131.

27 Yen, J. "A Reasoning Model Based on an Extended Dempster-Shafer Theory." In Proc. AAAI-86. Philadelphia, Pennsylvania, August, 1986, pp. 125-131.

28 Yu, S.Y. and H.E. Stephanou "A Set Theoretic Framework for the Processing of Uncertain Knowledge." In Proce. AAAI-84. Austin, Texas, August, 1984, pp. 216-221.

29 Zadeh, L.A. "Possibility Theory and Soft Data Analysis." In L.M. Cobb and R.M. Thrall, eds, Mathematical Frontiers of the Social and Policy Sciences, AAAS Selected Symposium 54, Westvicw Press (Boulder, Colorado) 1981,69-129.