Statistical Science
1987, Vol. 2, No. 1, 3–44

Probability Judgment in Artificial Intelligence and Expert Systems

Glenn Shafer

Abstract. Historically, the study of artificial intelligence has emphasized symbolic rather than numerical computation. In recent years, however, the practical needs of expert systems have led to an interest in the use of numbers to encode partial confidence. There has been some effort to square the use of these numbers with Bayesian probability ideas, but in most applications not all the inputs required by Bayesian probability analyses are available. This difficulty has led to widespread interest in belief functions, which use probability in a looser way. It must be recognized, however, that even belief functions require more structure than is provided by pure production systems. The need for such structure is inherent in the nature of probability argument and cannot be evaded. Probability argument requires design as well as numerical inputs. The real challenge probability poses to artificial intelligence is to build systems that can design probability arguments. The real challenge artificial intelligence poses to statistics is to explain how statisticians design probability arguments.

Key words and phrases: Artificial intelligence, associative memory, Bayesian networks, belief functions, certainty factors, conditional independence, constructive probability, diagnostic trees, expert systems, production systems.

I have been asked to speak on the use of belief functions in artificial intelligence and expert systems. For the sake of perspective, I propose to address the broader topic indicated by my title. The theory of belief functions is part of the theory of probability judgment, and a general understanding of the role of probability judgment in artificial intelligence can help us understand the particular role of belief functions.

I will not attempt to evaluate all the ways in which probability has been used in artificial intelligence, nor even all the ways in which belief functions have been used. Instead, I will aim for some general insights into the interaction between probability ideas and artificial intelligence ideas. Many of my comments will be historical. I hope readers will forgive me for those cases where I belabor the obvious or repeat the well known; my excuse is that I hope to reach a dual audience—students of probability who may not know very much about artificial intelligence, and students of artificial intelligence who may not know very much about probability.

The first two sections of the paper are introductory in nature. Section 1 considers the reasons for the artificial intelligence community's initial disinterest in probability and its recent change of heart and outlines the paper's conclusions about how current expert systems fall short of putting probability judgment into artificial intelligence. Section 2 deals with probability judgment without reference to artificial intelligence; here I discuss the split between Bayesian and non-Bayesian methods and place the theory of belief functions in this historical context.

Section 3 reviews some strands of the development within artificial intelligence of ideas about using probability judgment in expert systems. Here we see how the general issues that separate the Bayesian and belief-function theories appear in the context of expert systems, and we gain some insight into why flexibility is harder to achieve with probability judgment than with other kinds of reasoning. Section 4 discusses the problem of giving an artificial intelligence a genuine capacity for probability judgment.

1. THE EMERGENCE OF PROBABILITY IN ARTIFICIAL INTELLIGENCE

Until recently, the artificial intelligence community showed relatively little interest in probability. There is little probability, for example, in the three-volume *Handbook of Artificial Intelligence* (Barr and

Glenn Shafer is a Professor in the School of Business, University of Kansas, Lawrence, Kansas 66045.

Feigenbaum, 1981, 1982; Cohen and Feigenbaum, 1982). During the past 4 or 5 years, however, probability and the management of uncertainty in intelligent systems has become a widely discussed topic. Why the initial disinterest, and why the change?

The reasons for the initial disinterest are clear. Probabilities are numbers, and number crunching is just what artificial intelligence was supposed not to be. When the artificial intelligence community was founded, computers were used mainly for number crunching. They were impressively good at this, but they were not intelligent. Intelligence seems to require more general kinds of symbol manipulation.

Moreover, when we begin to think about computer programs that will match the achievements of human intelligence, we find that we are thinking about programs with non-numerical inputs and outputs. What place is there for talk about numbers in the case of these programs? They are merely sets of rules for going from the inputs to the outputs, and while it might be possible to identify some intermediate steps that are analogous to operations on numerical probabilities, it seems pointless to do so. It seems better to tell what is really going on.

The prejudice against numbers in general and probabilities in particular has not entirely disappeared from artificial intelligence, and the argument sketched in the preceding paragraph is still made. This argument is part of the motivation for the continuing development within artificial intelligence of non-numerical methods for handling uncertainty. These include nonmonotonic logic (McCarthy, 1980; McDermott and Doyle, 1980; Reiter, 1980) and Paul Cohen's theory of endorsements (Cohen, 1985).

But the factors that caused this prejudice have substantially changed. The vague idea that artificial intelligence can be defined largely through the contrast with number crunching has been replaced by the equally vague but equally powerful idea that intelligence is produced by complexity and by access to large amounts of knowledge. Two specific openings have appeared for probability.

1. The ban on non-numerical inputs has been dropped in some cases. In addition to programs that try to match aspects of human intelligence, artificial intelligence is now also concerned with expert systems and other intelligent systems that interact with human users and can use numerical inputs supplied by these users.

2. The artificial intelligence community has absorbed David Marr's views on levels of explanation. In his work on vision, Marr convincingly made the point that full understanding of an intelligent system involves explanation at various levels. In addition to explanation at the level of implementation (what is really going on) we also need explanation at more abstract levels. "It's no use, for example, trying to understand the fast Fourier transform in terms of resistors as it runs on an IBM 370" (Marr, 1982, page 337). Understanding of this point takes the rhetorical force out of the argument that there is no place for probability ideas when inputs and outputs are non-numerical.

Most of the current interest in probability in artificial intelligence is the result of (1). In many cases it is impossible to build expert systems without the use of probability. But in the long run, (2) may be more important. Because of (2), we can now recognize the value to an artificial intelligence of an ability to design probability arguments and generate the numerical judgments they require.

The ban on numerical inputs in artificial intelligence was dropped because the artificial intelligence community became interested in expert systems. Why did this happen? The answer is that the community discovered ways of building expert systems that incorporated ideas that seemed to reflect important aspects of human intelligence. As I explain in Section 3, most of the expert systems developed within artificial intelligence have been production systems, relatively unstructured programs that have some of the flexibility in acquiring and using knowledge that is characteristic of intelligence.

I argue in this paper that the expert systems we can now build to use probability judgments do not have this kind of flexibility and hence fit awkwardly under the heading of artificial intelligence. The problem is that probability judgment requires an overall design and hence cannot be achieved by relatively unstructured methods of programming applied to individual numerical probabilities. I will argue in Section 4 that both the overall design of probability judgment and the determination of individual numerical probabilities can be achieved by an artificial intelligence only if it is equipped with a genuine associative memory.

As a result of the explosion of interest in expert systems, the field of artificial intelligence is now struggling to maintain its sense of identity. The idea of an expert system began in artificial intelligence, but any system with expert capabilities can justifiably claim the name, whether it is written in LISP or FORTRAN, and many systems developed outside of artificial intelligence have more impressive expert capabilities than those developed inside it. It is clear, therefore, that artificial intelligence must withdraw from its embrace of the whole field of expert systems in order to maintain intellectual coherence. But it is unclear just what parts of the field of expert systems will remain in the embrace. My suggestion here is that artificial intelligence will retain its newfound interest in probability but will look beyond the current expert systems to deeper uses of probability ideas.

2. BAYESIAN AND BELIEF-FUNCTION ARGUMENTS

In this section I review some general ideas about probability judgment, without reference to the particular problems of artificial intelligence. I begin by sketching a way of looking at the frequentist vs. Bayesian controversy, a controversy that has dominated discussions of probability judgment for more than a century. After developing a constructive understanding of the Bayesian theory, I introduce another constructive theory, the theory of belief functions. I argue that both theories should be thought of as languages for expressing probability judgments and constructing probability arguments.

2.1 Two Strategies for Probability Judgment

What we now call the mathematical theory of probability was originally called the theory of games of chance. Probability was an entirely different topic; something was probable when there was a good argument or good authority for it. When James Bernoulli and others began to use the word probability in connection with the theory of games of chance, they were expressing the ambition that this theory might provide a general framework for evaluating evidence and weighing arguments. But just how might this work? How can the theory of games of chance help us evaluate evidence?

In the nineteenth century, it became clear that there are two distinct strategies for relating evidence to the picture of chance. Today, these two strategies might be called the frequentist and Bayesian strategies, but in order to avoid some of the connotations of these names, let me call them, for the moment, the *direct probability* and *conditional probability* strategies.

The direct probability strategy relies on direct application of the idea that in life, as in games of chance, what happens most often is most likely to happen in a particular case under consideration. The ideal kind of evidence for this strategy is knowledge of the frequency of outcomes in similar cases. I assign a 98% probability to the prediction that a student who first appears 3 weeks after the beginning of my elementary statistics course will not be able to pass the course, because it has almost always turned out that way in the past.

The conditional probability strategy uses the picture of chance in a deeper way. It observes that games of chance unfold step by step, with the probabilities for different possible final outcomes changing at each step, and it suggests that the accumulation of evidence should change probabilities in a similar step by step way. Thus, my probability for whether the late-appearing student will pass my course should change when I learn more about his history and circumstances, just as my probability for whether two successive rolls of a die will add up to nine will change when I learn the result of the first roll. The conditional probability strategy usually leads to a more complicated argument than the direct probability strategy, since it involves construction of a probability measure over a more complicated frame and then the reduction of this measure and frame by conditioning.

In general, there is not, I believe, any *a priori* reason to prefer one of these two strategies to the other. We cannot say that it is normative to use one and irrational to use the other. They are both strategies for producing arguments, and it is the cogency of the arguments that must be evaluated. It may be most cogent to lump my new late-appearing student with all my past late-appearing students, on the grounds that particulars have not made much difference in the past. Or I may have had enough experience with late-appearing students like this one on some particulars that I can make a better direct probability argument by looking at the past frequency of success just for these late-appearing students. Or I may have the experience and insight needed to construct a probability measure that I can condition on the particulars. The issue cannot be settled in the abstract, without reference to the experience I bring to bear on the problem.

Moreover, neither of the two strategies is inherently more objective or subjective than the other. It is true that the direct probability strategy, since it tends to consider broader classes, is more likely to result in probability judgments based on actual frequency counts. But the objectivity of these frequencies must always be coupled with a subjective judgment of their relevance. And even with broad classes we most often have hunches and impressions rather than actual counts.

Historically, however, the direct probability strategy has come to be associated with claims to objectivity, whereas the conditional probability approach has come to be associated with claims to rationality. This fact seems to be a result of efforts to square the interpretation of probability with the empiricist and positivist philosophical trends of the late nineteenth and early twentieth centuries.

2.2 The Frequentist vs. Bayesian Deadlock

Laplace, writing at the beginning of the nineteenth century, was able to define numerical probability as the measure of the "reason we have to believe." But by the middle of the nineteenth century, many students of probability were looking for a more empirical definition. They found this definition in the idea of frequency, and they proceeded to reject those applications of probability theory that could not be based on observed frequencies. In particular, they rejected Laplace's method of calculating the probability of causes, which is a special case of the conditional probability strategy.

The frequentist philosophy severely restricted the domain of application of numerical probability, and those who wanted to use numerical probability more generally were forced to search for a philosophical foundation for the conditional probability strategy that would fit the positivist mind-set. Such a philosophical foundation was finally established in the twentieth century by Ramsey, de Finetti, and especially Savage. These authors conceived the idea that subjective probability should be given a behavioral and hence positivist interpretation—a person's probabilities should be derivable from his choices. They formulated postulates for what they called rational behavior, postulates that assure that a person's choices do determine numerical probabilities. And they argued that it is normative to follow these postulates and hence normative to have subjective probabilities.

During the past two decades, the philosophical foundation provided by Savage's postulates has led to a remarkable resurgence, both mathematical and practical, of the conditional probability strategy. The resulting body of theory has been called "Bayesian," because the conditional probability strategy often uses Bayes' theorem.

Although the new Bayesian philosophy has played a historically valuable role in rescuing the conditional probability strategy from its frequentist opponents, it has its own obvious shortcomings. Most important, perhaps, is its inability to explain how the quality of a probability analysis depends on the availability and quality of relevant evidence. Whereas the frequentist philosophy tries to limit applications of probability to models for which we have clearly relevant and objective frequency counts, there is nothing in the Bayesian philosophy to make our choice of a model depend in any way on the availability of relevant evidence. The postulates apply equally to any model.

We have, then, a deadlock between two inadequate philosophies of probability. On the one side, the frequentist philosophy, which recognizes the relevance of evidence but tries to justify claims to objectivity by limiting numerical probability judgment to cases where the evidence is of an ideal form; on the other side, the Bayesian philosophy, which recognizes the subjectivity of all probability judgment but ignores the quality of evidence and claims it is normative to force all probability judgment into one particular mold.

We have been caught in this deadlock for three decades. We have tired of it, and we are inclined to ask the two sides to compromise (see, e.g., Box, 1980). But we have not been able to find a philosophical foundation for probability judgment that can resolve the deadlock.

I believe that the way out of the deadlock is to back up and recognize that a positivist philosophical account of probability is no longer needed. Our intellectual culture has moved away from positivism and toward various sorts of pragmatism, and once we recognize this we will be free to discard both the frequentists' claims to objectivity and the Bayesians' claims to normativeness.

2.3 Constructive Probability

In several recent papers (especially Shafer, 1981; Shafer and Tversky, 1985) I have proposed the name "constructive probability" for the pragmatic, postpositivist foundation that I think we need for probability judgment. The idea is that numerical probability judgment involves fitting an actual problem to a scale of canonical examples. The canonical examples usually involve the picture of chance in some way, but different choices of canonical examples are possible, and these different choices provide different theories of subjective probability, or, if you will, different languages in which to express probability judgments. No matter what language is used, the judgments expressed are subjective; the subjectivity enters when we judge that the evidence in our actual problem matches in strength and significance the evidence in the canonical example.

Within a given language of probability judgment, there can be different strategies for fitting the actual problem to the scale of canonical examples. The direct and conditional probability strategies described above live, I think, in the same probability language, the language in which evidence about actual questions is fit to canonical examples where answers are determined by known chances. We may call this language the Bayesian language. (For a more detailed account of different strategies that are available within the Bayesian language, see Shafer and Tversky (1985). The distinction between the direct and conditional probability strategies corresponds to the distinction that is made there between total-evidence and conditioning designs.)

The constructive viewpoint tells us that when we work within the Bayesian language we must make a judgment about how far to take the conditional probability strategy in each particular problem. We make this judgment on the basis of the availability of evidence to support the conditional and unconditional probability judgments that are required.

It may be useful to elaborate on this point. Suppose we want to make probability judgments about a frame of discernment S. (A *frame of discernment* is a list of possible answers to a question; we want to make probability judgments about which answer is correct.) We reflect on our evidence, and we produce a list $E_1, \ldots, E_n$ of facts that seem to summarize this evidence adequately. The conditional probability strategy amounts to standing back from our knowledge of these n facts, pretending that we did not yet know them, and constructing a probability measure over a frame that considers not only the question considered by S but also the question whether $E_1, \ldots, E_n$ are or are not true; typically we construct this measure by making probability judgments $P(s)$ and $P(E_1 \& \ldots \& E_n \mid s)$ for each s in S. The problem with this strategy is that we now need to look for further evidence on which to base all these probability judgments. We have used our best evidence up, as it were, but now we have an even larger judgmental task than before. According to the behaviorist Bayesian theory, there is no problem—it is normative to have the requisite probabilities, whether we can identify relevant evidence or not. But according to the constructive viewpoint, there is a problem, a problem that limits how far we want to go. We may want to apply the conditional probability strategy to some of the E_i, but we may want to reserve the others to help us make the probability judgments (see Shafer and Tversky, 1985).

2.4 The Language of Belief Functions

Whereas the Bayesian probability language uses canonical examples in which known chances are attached directly to the possible answers to the question asked, the language of belief functions uses canonical examples in which known chances may be attached only to the possible answers to a related question.

Suppose S and T denote the sets of possible answers to two distinct but related questions. When we say that these questions are related, we mean that a given answer to one of the questions may fail to be compatible with some of the possible answers to the other. Let us write "$s\mathrm{C}t$" when s is an element of S, t is an element of T, and s and t are compatible. Given a probability measure P over S (assume for simplicity that P is defined for all subsets of S), we may define a function Bel on subsets of T by setting

(1) $$\mathrm{Bel}(B) = P\{s \mid \text{if } s\mathrm{C}t, \text{ then } t \text{ is in } B\}$$

for each subset B of T. The right-hand side of (1) is the total probability that P gives to those answers to the question considered by S that require the answer to the question considered by T to be in B; the idea behind (1) is that this probability should be counted as reason to believe that the latter answer is in B. We might, of course, have more direct evidence about the question considered by T, but if we do not, or if we want to leave other evidence aside for the moment, then we may call Bel(B) a measure of the reason we have to believe B based just on P.

The function Bel given by (1) is the *belief function* obtained by extending P from S to T. A probability measure P is a special kind of belief function; this is just the case where (i) $S = T$ and (ii) $s\mathrm{C}t$ if and only if $s = t$. Thus the language of belief functions is a generalization of the Bayesian language.

All the usual devices of probability are available to the language of belief functions, but in general we use them in the background, at the level of S, before we move to degrees of belief on T, the frame of interest.

Like other non-Bayesian approaches to probability judgment, the language of belief functions countenances the use of probability models that are less complete than Bayesian models. In order to obtain a belief function over T, we begin with a probability measure over S alone, and we use observed facts to create a compatibility relation C between S and T. A Bayesian conditional probability argument that used the frames S and T would extend the probability measure over S to a complete probability measure over $S \times T$, and it would then use the compatibility relation to condition this measure.

I have studied the language of belief functions in detail in earlier work—see especially Shafer (1976, 1986a). Here I will use some examples of (1) to illustrate the language and to contrast it with the Bayesian language.

Example 1. Is Fred, who is about to speak to me, going to speak truthfully, or is he, as he sometimes does, going to speak carelessly, saying whatever comes into his mind? Let S denote the possible answers to this question; $S = \{\text{truthful, careless}\}$. Suppose I know from experience that Fred's announcements are truthful reports on what he knows 80% of the time and are careless statements the other 20% of the time. Then I have a probability measure P over S: $P\{\text{truthful}\} = .8$, $P\{\text{careless}\} = .2$.

Are the streets outside slippery? Let T denote the possible answers to this question; $T = \{\text{yes, no}\}$. And suppose Fred's announcement turns out to be, "The streets outside are slippery." Taking account of this, I have a compatibility relation between S and T; truthful is compatible with yes but not with no, while careless is compatible with both yes and no. Applying (1), I find

(2) $$\mathrm{Bel}(\{\text{yes}\}) = .8 \text{ and } \mathrm{Bel}(\{\text{no}\}) = 0;$$

Fred's announcement gives me an 80% reason to believe the streets are slippery, and no reason to believe they are not.

How might a Bayesian argument using this evidence go? A Bayesian direct probability argument would use all my evidence, Fred's announcement included, to make a direct probability judgment about whether the streets are slippery. The judgment that Fred is 80% reliable need not appear explicitly in such an argument. On the other hand, I can construct a Bayesian conditional probability argument using this judgment as one ingredient. I need two other judgments as well: (i) A prior probability, say p, for the proposition that the streets are slippery; this will be a judgment based on evidence other than Fred's announcement. (ii) A conditional probability, say q, that Fred's announcement will be accurate even though it is careless. I can construct a probability measure from these judgments, and I can condition this measure on the content of Fred's announcement.

The probability measure constructed in this conditional argument is formally a measure over $S \times T$, where T is still the set of answers to the question whether the streets are slippery,

$$T = \{\text{yes, no}\},$$

but where S now tells us not only whether Fred is truthful or careless but also whether he is accidentally telling the truth in case he is careless,

$$S = \{\text{truthful, careless but accurate, careless and inaccurate}\}.$$

My probabilities for T are p for yes and $1 - p$ for no. My probabilities for S are .8 for truthful, $.2q$ for careless but accurate, and $.2(1 - q)$ for careless and inaccurate. Assuming probabilistic independence between the state of the streets and Fred's behavior, I multiply these numbers to obtain the product probability measure on $S \times T$, given in the second column of Table 1. Conditioning this measure on the content of Fred's announcement means eliminating the three possibilities marked with an × in the table; since Fred said the streets are slippery, he cannot be truthful or accurate if the answer to T is no, and he cannot be inaccurate if the answer to T is yes. Having eliminated these three possibilities, I renormalize the probabilities for the other three so that they add to one; this means multiplying each probability by K, where $K = 1/(.8p + .2qp + .2(1 - q)(1 - p))$. This results in the posterior probabilities given in the third column in Table 1. Adding the first two nonzero probabilities in this column, I obtain my total posterior probability that Fred's announcement that the streets are slippery is true:

$$\frac{.8p + .2qp}{.8p + .2qp + .2(1 - q)(1 - p)}. \qquad (3)$$

Is the Bayesian argument (3) better than the belief-function argument (2)? This depends on whether I have the evidence required. If I do have evidence to support the judgments p and q—if, that is to say, my situation really is quite like a situation where the streets and Fred are governed by these known chances, then (3) is a cogent argument, and it is better than (2) because it takes more evidence into account. But if the evidence on which I base p and q is of much lower quality than the evidence on which I base the number 80%, then (2) will be the better argument.

The traditional debate between the frequentist and Bayesian views has centered on the quality of the evidence for prior probabilities. It is worth remarking, therefore, that q, rather than p, may well be the weak point in the argument (3). I probably will have some other evidence about whether it is slippery outside, but I may have no idea about how likely it is that Fred's careless remarks will accidentally be true.

A critic of the belief-function argument (2) might be tempted to claim that the Bayesian argument (3) shows (2) to be wrong even if I do lack the evidence needed to supply p and q. Formula (3) gives the correct probability for whether the street is slippery, the critic might contend, even if I cannot say what this probability is, and it is almost certain to differ from (2). This criticism is fundamentally misguided. In order to say that (3) gives the "correct" probability, I must be able to convincingly compare my situation to the picture of chance. And my inability to model Fred when he is being careless is not just a matter of not knowing the chances—it is a matter of not being able to fit him into a chance picture at all.

Example 2. Suppose I do have some other evidence about whether the streets are slippery: my trusty indoor–outdoor thermometer says that the

TABLE 1

	(s, t)	Probability of (s, t) Initial	Posterior
	(Truthful, yes)	$.8p$	$.8pK$
×	(Truthful, no)	$.8(1-p)$	0
	(Careless but accurate, yes)	$.2qp$	$.2qpK$
×	(Careless but accurate, no)	$.2q(1-p)$	0
×	(Careless and inaccurate, yes)	$.2(1-q)p$	0
	(Careless and inaccurate, no)	$.2(1-q)(1-p)$	$.2(1-q)(1-p)K$

TABLE 2

s	Probability of s		Elements of T compatible with s
	Initial	Posterior	
(Truthful, working)	.792	0	
(Truthful, not)	.008	.04	Yes
(Careless, working)	.198	.95	No
(Careless, not)	.002	.01	Yes, no

temperature is 31° Fahrenheit, and I know that because of the traffic ice could not form on the streets at this temperature.

My thermometer could be wrong. It has been very accurate in the past, but such devices do not last forever. Suppose I judge that there is a 99% chance that the thermometer is working properly, and I also judge that Fred's behavior is independent of whether it is working properly or not. (For one thing, he has not been close enough to my desk this morning to see it.) Then I have determined probabilities for the four possible answers to the question, "Is Fred being truthful or careless, and is the thermometer working properly or not?" For example, I have determined the probability $.8 \times .99 = .792$ for the answer "Fred is being truthful, and the thermometer is working properly." All four possible answers, together with their probabilities, are shown in the first two columns of Table 2. I will now construct a belief function over T by using these four answers as my frame S.

Taking into account what Fred and the thermometer have said, I obtain the compatibility relation between S and T given in the last column of the Table 2. (Recall that T considers whether the streets are slippery; $T = \{\text{yes, no}\}$.) The element (truthful, working) of S is ruled out by this compatibility relation (since Fred and the thermometer are contradicting each other, they cannot both be on the level); hence, I condition the initial probabilities by eliminating the probability for (truthful, working) and renormalizing the three others. The resulting posterior probabilities on S are given in the third column of the Table 2.

Finally, applying (1) with these posterior probabilities on S, I obtain the degrees of belief

(4) $\quad \text{Bel}(\{\text{yes}\}) = .04$ and $\text{Bel}(\{\text{no}\}) = .95$.

This result reflects that fact that I put much more trust in the thermometer than in Fred.

The preceding calculation is an example of *Dempster's rule of combination* for belief functions. Dempster's rule combines two or more belief functions defined on the same frame but based on independent arguments or items of evidence; the result is a belief function based on the pooled evidence. In this case the belief function given by (2), which is based on Fred's testimony alone, is being combined with the belief function given by

(5) $\quad \text{Bel}(\{\text{yes}\}) = 0$ and $\text{Bel}(\{\text{no}\}) = .99$,

which is based on the evidence of the thermometer alone. In general, as in this example, Dempster's rule corresponds to the formation and subsequent conditioning of a product measure in the background. See Shafer (1986a) for a precise account of the independence conditions needed for Dempster's rule.

Example 3. Dempster's rule applies only when two items of evidence are independent, but belief functions can also be derived from models for dependent evidence.

Suppose, for example, that I do not judge Fred's testimony to be independent of the evidence provided by the thermometer. I exclude the possibility that Fred has tampered with the thermometer and also the possibility that there are common factors affecting both Fred's truthfulness and the thermometer's accuracy. But suppose now that Fred does have regular access to the thermometer, and I think that he would likely know if it were not working. And I know from experience that it is in situations where something is awry that Fred tends to let his fancy run free.

In this case, I would not assign the elements of S the probabilities given in the second column of Table 2. Instead, I might assign the probabilities given in the second column of Table 3. These probabilities follow from my judgment that Fred is truthful 80% of the time and that the thermometer has a 99% chance of working, together with the further judgment that Fred has a 90% chance of being careless if the thermometer is not working.

When I apply (1) with the posterior probabilities given in Table 3, I obtain the degrees of belief

$$\text{Bel}(\{\text{yes}\}) = .005 \text{ and } \text{Bel}(\{\text{no}\}) = .95.$$

These differ from (4), even though the belief functions based on the separate items of evidence will still be given by (2) and (5).

In this example, the combination of two belief functions (2) and (5) departed from Dempster's rule in that the probability measure constructed over the joint probability space in the background was not a product measure. This is just one of the ways the language of belief functions can take dependence into account.

TABLE 3

s	Probability of s		Elements of T compatible with s
	Initial	Posterior	
(Truthful, working)	.799	0	
(Truthful, not)	.001	.005	Yes
(Careless, working)	.191	.950	No
(Careless, not)	.009	.045	Yes, no

Another way is to modify the compatibility relation between the joint probability space and the frame T (Shafer, 1986a). Another is to rework the way the evidence is broken up, so that different items of evidence better correspond to independent uncertainties (Shafer, 1984).

2.5 Conclusion

I would like to emphasize that nothing in the philosophy of constructive probability or the language of belief functions requires us to deny the fact that Bayesian arguments are often valuable and convincing. The examples I have just discussed were designed to convince the reader that belief-function arguments are sometimes more convincing than Bayesian arguments, but I am not claiming that this is always or even usually the case. What the language of belief functions does require us to reject is the philosophy according to which use of the Bayesian language is normative.

From a technical point of view, the language of belief functions is a generalization of the Bayesian language. But as our examples illustrate, the spirit of the language of belief functions can be distinguished from the spirit of the Bayesian language by saying that a belief-function argument involves a probability model for the evidence bearing on a question, whereas a Bayesian argument involves a probability model for the answer to the question.

Of course, the Bayesian language can also model evidence. The probability judgments made in a belief-function argument can usually be extended to a Bayesian argument that models both the answer to the question and the evidence for it by assessing prior probabilities for the answer and conditional probabilities for the evidence given the answer. The only problem is that we may lack the evidence needed to make all the judgments required by this Bayesian argument convincing. The advantage gained by the belief-function generalization of the Bayesian language is the ability to use certain kinds of incomplete probability models.

3. THE ATTEMPT TO USE PROBABILITY IN PRODUCTION SYSTEMS

The field of expert systems developed within artificial intelligence from efforts to apply systems of production rules to practical problems. The current interest in probability judgment in artificial intelligence began with efforts to incorporate probability judgments into production rules. In this section I review these efforts and relate them to what we learned in the preceding section about the Bayesian and belief-function languages.

A production rule is simply an if-then statement, interpreted as an instruction for modifying the contents of a data base. When the rule is applied, the action specified by its right-hand side is taken if the condition on its left-hand side is found in the data base. A production system is a collection of production rules, which are repeatedly applied to the data base either in the same predetermined order or else in an order determined by some relatively simple principle. Production systems were used in programming languages in the early 1960s, and they were advanced as cognitive models by Newell and Simon in the late 1960s and early 1970s (Newell and Simon, 1965; Newell, 1973). These systems are attractive models for intelligence because their knowledge is represented in a modular way and is readily available for use. Each rule represents a discrete chunk of knowledge that can be added to or removed from the system without disrupting its ability to use the other chunks, and the system regularly checks all the chunks for their relevance to the problem at hand (Davis and King, 1984).

When artificial intelligence workers undertook, in the 1970s, to cast various bodies of practical knowledge in the form of production rules, they found that in many fields knowledge cannot be encoded in the form of unqualified if–then statements. Instead, probability statements seem to be required: "If $E_1, E_2, \ldots E_n$, then probably (or usually or almost certainly) H." So these workers found themselves trying to use production systems to manipulate probability judgments.

Many tacks were taken in the effort to use probability in production systems, but I would like to emphasize two lines of development. One of these begins with PROSPECTOR and leads to Pearl and Kim's elegant work on the propagation of Bayesian probability judgments in causal trees, while the other begins with the certainty factors of MYCIN and leads to the use of belief functions in diagnostic trees. I will review these two lines of development in turn.

As it turns out, the results of both lines of development can be unified in a general scheme for propagating belief functions in trees (Shenoy and Shafer, 1986). I will briefly describe this general scheme.

3.1 Bayesian Networks

The artificial intelligence workers at SRI who developed the PROSPECTOR system for geological exploration in the middle 1970s thought of production rules as a means for propagating probabilities through a network going from evidence to hypotheses. Figure 1, taken from Duda, Hart and Nilsson (1976), gives an example of such a network; here, E_i denotes an item of evidence, and H_i denotes a hypothesis. The idea is that the user of the system should specify that

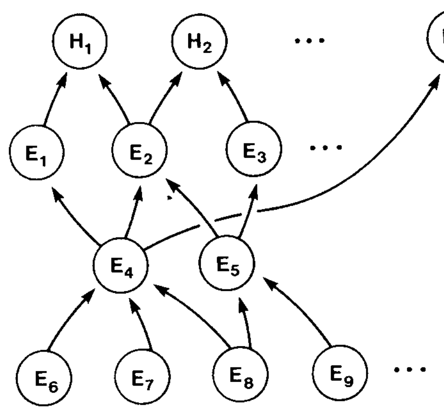

FIG. 1. *PROSPECTOR's inference network.*

some of the E_i at the bottom of the network are true and some are false, or should make probability judgments about them, and the production rules, corresponding to conditional probabilities for the links in the network, should propagate these probability judgments through the network to produce judgments of the probabilities of the hypotheses.

Unfortunately, the introduction of probabilities into production rules does not square well with the modularity we want these rules to have. The PROSPECTOR workers wanted to be able to elicit from a geologist statements of the form, "If E_i and E_j and . . . , then E_r, with probability p," and they wanted to allow the geologist to make each of these statements independently. But this led to problems in putting the statements together into a calculation of the probabilities of the hypotheses. For example: (1) The conditional probabilities elicited may not be sufficient to determine a joint probability measure over all the E's and H's. The geologist might give rules corresponding to $P(E_5 \mid E_8)$ and $P(E_5 \mid E_9)$ in Figure 1 but neglect or feel unable to give a rule corresponding to $P(E_5 \mid E_8 \,\&\, E_9)$. (2) The conditional probabilities that are given may be inconsistent. (3) The network may have cycles, which will cause trouble when propagation is attempted.

These problems were handled in PROSPECTOR in relatively ad hoc ways. Problem (1) was handled partly by independence assumptions and partly by maximum-minimum rules reminiscent of the theory of fuzzy sets. Problem (2) was handled by formulating rules of propagation which did not always accord with the rules of probability but which were insensitive to some kinds of inconsistencies. Problem (3) was handled by arbitrarily rejecting new production rules when they would introduce cycles into the network already constructed.

PROSPECTOR was only modestly successful, but it was very influential in the questions it raised. The PROSPECTOR workers subscribed to Bayesian principles, and they were conscious of their failure to follow those principles completely. Is it possible to do better? Can probability judgments be treated modularly within the Bayesian language? To what extent is the propagation of probabilities possible within this language?

The best work that has been done in response to these questions is that of Judea Pearl and his students at UCLA (Pearl, 1982, 1986; Kim, 1983; Kim and Pearl, 1983). Pearl has shown that we can make sense of the independence assumptions needed to construct a probability measure over a network from simple conditional probabilities and we can propagate updated probabilities through the network in a simple and elegant way provided that the network has a causal interpretation and a relatively simple form; it must be a simple directed tree or else a more general type of directed tree that we may call a *Kim tree.*

Recall that a tree is a graph in which there are no cycles. A simple directed tree is a tree in which the links are assigned directions that all run outward (or downward, if we want) from a single initial node, as in Figure 2a. A Kim tree is a tree in which the links are assigned arbitrary directions. Such a tree can always be laid out so that the directions are downward, as in Figure 2b. In Pearl's work, the nodes of a tree correspond to random variables, and the directions of the links are interpreted as directions of causation. Thus each variable is influenced by the variables above it in the graph and influences the variables below it. An observation of the value of one variable is diagnostic evidence about the value of a higher variable and causal evidence about the value of a lower variable.

Once a Kim tree is constructed for a problem, the construction of a probability measure over it and the updating of the measure are straightforward. Given the independence conditions of Pearl and Kim, which are reasonable in the causal context, a measure over the tree can be constructed from prior probabilities for the topmost nodes and conditional probabilities for all the links. Moreover, this construction is straightforward; there are no complicated consistency conditions that the conditional probabilities must meet. Once construction is completed, the measure can be stored and updated locally. At each node we store information about the conditional probabilities corresponding to incoming and outgoing links, the current probability measures for the variable at the node and the variables at neighboring higher nodes, and likelihood-type information from neighboring lower nodes. When the value of a variable is then observed, this information can be propagated through the network to update the entire probability measure in one pass. All computations are made locally, with

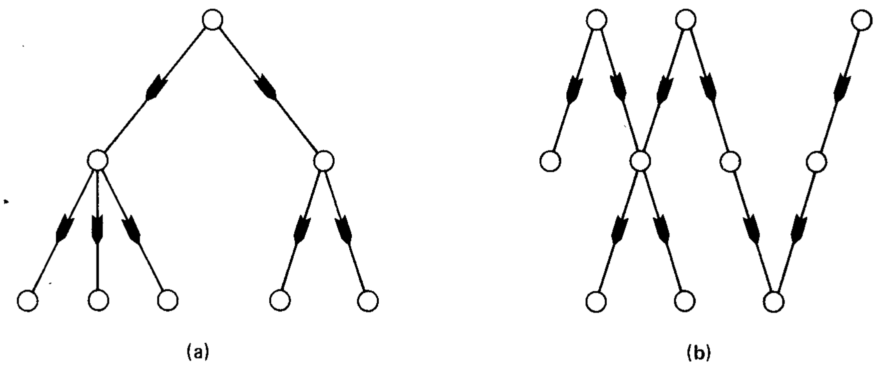

FIG. 2. *Pearl's causal tree.*

each node communicating only updated local information to its neighbors.

An obvious shortcoming of this elegant scheme is its restriction to Kim trees. In few problems will the causal relations that we think important take so simple a form. Kim (1983) and Pearl (1986) have shown how such trees might be used to approximate more realistic models; they propose first using a more general graph to elicit a probability measure from an expert, and then approximating this measure with a Kim tree. This solution does not seem very satisfactory, however. It is not clear that the approximation will be satisfactory, and more importantly, the constructive nature of the initial probability measure is put into question. In a Kim tree the initial probability measure can be constructed from probability and conditional probability judgments without concerns about consistency, but in a more general graph consistency conditions will be so complicated that it will be impossible for us to hope they will be met unless we pretend that we are indeed eliciting a measure instead of constructing one.

Another obvious shortcoming is the restriction to thoroughly causal models. In a sense, of course, all evidence is causal. With sufficient complication, we can always construct a model that relates the facts we observe to deeper causes and also relates these causes to the questions that interest us. But we may lack the evidence needed to make good probability judgments relative to such a model.

3.2 Certainty Factors and Belief Functions

The work on the MYCIN system for medical diagnosis began earlier and has been more extensive than the work on PROSPECTOR. It has also had more effect on subsequent expert systems; various versions of EMYCIN, the expert system shell that was abstracted from MYCIN, are now being widely used. The story of the MYCIN effort has been told in a recent book (Buchanan and Shortliffe, 1984), which includes extensive discussion of the certainty factors that were used by MYCIN and the similarities of these certainty factors to the values of belief functions.

MYCIN departed from the pure production system picture by using a backward-chaining strategy to select production rules to apply. This means that it selected rules by comparing their right-hand sides to goals instead of comparing their left-hand sides to statements already accepted. If the right-hand side of a rule matched a goal, its left-hand side was then established as a goal, so that there was a step by step process backward from conclusions to the knowledge needed to establish them.

MYCIN also differed from PROSPECTOR in that the MYCIN workers rejected at the outset the idea that the numerical probability judgments associated with the rules could or should be understood in Bayesian terms. They emphasized this point by calling these numbers "certainty factors" rather than probabilities. And they formulated their own rules for combining these certainty factors.

In spirit, and to a considerable extent in form, these rules agree with Dempster's rule for combining independent belief functions. I would explain this coincidence by saying that in developing their calculus for certainty factors, Shortliffe and Buchanan were trying to model the probabilistic nature of evidence while avoiding the complete probability models needed for Bayesian arguments.

In recent work (Gordon and Shortliffe, 1984, 1985), some of the MYCIN workers have taken a close look at the similarity between the calculus of certainty factors and the language of belief functions and have asked how belief functions can contribute further to the MYCIN project. They have drawn two main conclusions. First, it is sensible to modify some of the rules for certainty factors to put these rules into more exact agreement with the rules for belief functions. Second, the diagnosis problem that was central to MYCIN can be understood more clearly in terms of belief functions if it is explicitly expressed as a problem involving hierarchical hypotheses.

The term "hierarchical hypotheses" refers to the fact that the items of evidence in a diagnostic problem tend to support directly only certain subsets of the frame of discernment, subsets which can be arranged in a tree. Figure 3, taken from Gordon and Shortliffe (1984), illustrates the point. The four nodes at the bottom of this tree represent four distinct causes of cholestatic jaundice; they form the frame of discernment for the diagnostic problem. Some items of evidence may directly support (or directly refute) one of these causes for a particular patient's jaundice. Other evidence may be less specific. There may, for example, be evidence that the jaundice is due to an intrinsic liver problem, either hepatitis or cirrhosis. On the other hand, it is hard to imagine a single item of medical evidence supporting the subset {cirrhosis, gallstone} without supporting one of these more directly; this is reflected by the fact that this subset does not correspond to an intermediate node of the tree.

This picture suggests that a belief-function argument based on such medical evidence may involve combining many belief functions by Dempster's rule, where each belief function is a simple support function focused on a subset in the tree or its complement. (A simple support function is a belief function obtained from (1) when S has only two elements and one of these is compatible with all the elements of T.)

Two concerns can be raised about this use of Dempster's rule. First, there is the issue of computational complexity. Since the computational complexity of Dempster's rule increases exponentially with the size of the frame, it might not be feasible to implement the rule for a large diagnostic tree. Second, there is the issue of dependence. Will the items of evidence bearing on different nodes of the tree all be independent?

As it turns out, computational complexity is not a problem. By taking advantage of the tree structure, we can devise remarkably efficient algorithms for implementing Dempster's rule (Shafer and Logan, 1985).

Violations of the independence assumptions needed for Dempster's rule pose a more worrisome problem. It seems unlikely that the uncertainties involved in a very large number of items of medical evidence will all be independent. This does not mean that a belief-function analysis will be impossible or unsatisfactory, but it does mean that a satisfactory belief-function analysis may require modeling dependencies in the evidence.

3.3 Propagating Belief Functions in Trees

It turns out that Pearl's method of propagating Bayesian probabilities in causal trees and Shafer and Logan's method of combining simple support functions in diagnostic trees are both special cases of a general scheme for propagating belief functions in qualitative Markov trees. The following comments on this general scheme are relatively technical but may be of interest to some readers. For more detail, see Shenoy and Shafer (1986).

The idea of a qualitative Markov tree is based on the idea of qualitative conditional independence. We say that two partitions $\mathbf{P}_1$ and $\mathbf{P}_2$ of a frame S are *conditionally independent* given a third partition $\mathbf{P}$ if $P \cap P_1 \cap P_2 \neq \emptyset$ whenever $P_1 \in \mathbf{P}_1$, $P_2 \in \mathbf{P}_2$, $P \cap P_1 \neq \emptyset$, *and* $P \cap P_2 \neq \emptyset$. This means that once we know which element of $\mathbf{P}$ contains the truth, knowledge of which element of $\mathbf{P}_1$ contains the truth tells us nothing more about which element of $\mathbf{P}_2$ contains the truth. Qualitative conditional independence is important for belief functions, because it is legitimate, when $\mathbf{P}_1$ and $\mathbf{P}_2$ are conditionally independent given $\mathbf{P}$, and we want to combine a belief function on $\mathbf{P}_1$ with a belief function on $\mathbf{P}_2$, to first simplify both to belief functions on $\mathbf{P}$. This can be helpful if $\mathbf{P}$ is a relatively coarse partition, for then the combination is easier to think about and computationally more feasible.

A qualitative Markov tree is a tree of partitions with the property that the disconnected branches that

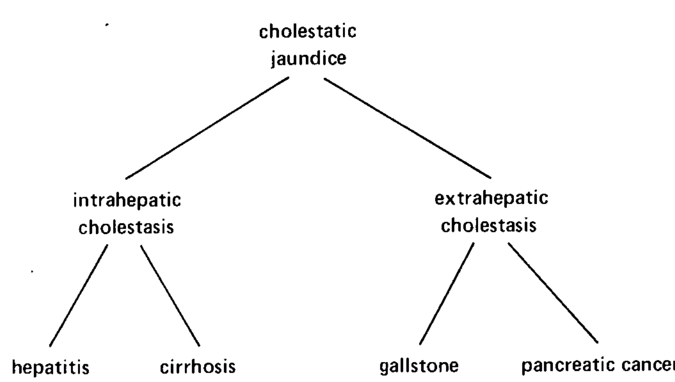

FIG. 3. *A diagnostic tree.*

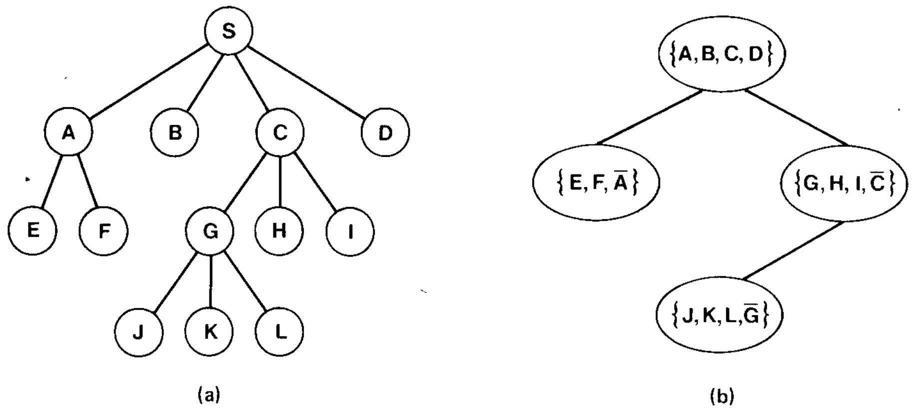

FIG. 4. *The tree of partitions. (b) derived from a diagnostic tree (a).*

result from the removal of a partition **P** are always conditionally independent given **P**. We obtain a qualitative Markov tree if we replace each random variable in a Bayesian causal tree with the partition of the sample space it induces. We can also construct a qualitative Markov tree from a diagnostic tree; for each mother node in the diagnostic tree, we form a partition whose elements are the daughters of the mother and the complement of the mother. Figure 4b shows the qualitative Markov tree obtained in this way from the diagnostic tree of Figure 4a.

Suppose we wish to combine belief functions defined on various partitions in a qualitative Markov tree. It is legitimate to do so in a stepwise way, simplifying the belief function on one partition to a belief function on its neighbor, combining all the belief functions projected to the neighbor in this way, and then projecting to the next neighbor. The schemes of Pearl and Shafer and Logan both turn out to be special cases of this simple general idea.

In addition to generalizing Pearl and Shafer and Logan, this scheme for propagating belief functions in trees promises to be useful as a general framework for designing probability arguments. Independent items of evidence often bear on different but related partitions (or questions, or variables), and a qualitative Markov tree provides a way of keeping track of the relations.

3.4 Conclusion

The preceding look at attempts to use probability judgment in expert systems justifies at least one general conclusion: probability judgment in expert systems is very much like probability judgment everywhere else. The general issues about probability judgment that we identified in Section 2 all reappear in the expert systems work. In expert systems, as elsewhere, probability judgment is constructive and requires an overall design. It is sometimes possible to provide such a design within the Bayesian language, but Bayesian designs often demand judgments for which we do not have adequate evidence. And belief-function analyses often require models for dependent evidence.

Production systems were attractive to the artificial intelligence community because these systems seemed to have the flexibility in acquiring and using knowledge that seems characteristic of intelligence. But it seems fair to say that the attempt to incorporate probability judgment into production systems has failed. The most successful production systems are still those, like R1 and DART, that do not attempt to use numerical measures of uncertainty. Many expert systems have recently been built using the EMYCIN shells, but more often than not the builders of these systems ignore the "certainty factor" capacities of the shells.

It appears that probability judgment simply does not have the extremely modular character that made production systems so attractive. Almost always, probability judgment involves not only individual numerical judgments but also judgments about how these can be put together. This is because probability judgment consists, in the final analysis, of a comparison of an actual problem to a scale of canonical examples.

I believe that progress will be made over the next few years in using probability in expert systems. But these systems will be intensely interactive. They will depend on the human user to design the probability argument for the particular evidence at hand: they will be able at most to help the user construct his or her causal, diagnostic, or qualitative Markov tree. And they will also depend on the human user to supply individual numerical probability judgments.

4. THE CONSTRUCTION OF ARGUMENTS

A genuine capacity for probability judgment in an artificial intelligence would involve both the ability to generate numerical probability judgments and the ability to design probability arguments. How might these abilities be programmed? We do not have an answer, but we should start thinking about the question.

As the result of the work by psychologists during the past decade, especially the work of Kahneman and Tversky (see Kahneman, Slovic, and Tversky, 1982), we do have some ideas about how people generate numerical probability judgments. They conduct internal sampling experiments, they make similarity judgments, they construct causal models and perform mental simulations with these models, they consider typical values and discount or adjust these, and so on. An obvious and appropriate strategy for artificial intelligence is to try to implement these heuristics.

The heuristics sometimes lead to systematic mistakes or biases, and it is by demonstrating these biases that the psychologists have convinced us that people use them. There is a tendency, therefore, to think that people are doing something suboptimal or unnormative when they use them. Indeed, proponents of the Bayesian philosophy frequently assert that the psychological work only demonstrates what people do and is irrelevant to what people should do. When we face up to the artificial intelligence problem, however, we see that the heuristics are really all we have. People have to use such heuristics if they are to make quick probability judgments about questions they have not previously considered, and our programs will also have to use them if they are going to be equally flexible. The challenge is to figure out how to use the heuristics well enough that using them will not usually cause mistakes.

It is more difficult to say anything about how we might build the ability to design probability arguments. The lesson from Section 3 is clear, though: the chunks that we try to fit together when we search for a convincing argument must be larger than the chunks represented by production rules. It is also clear that the ability to construct cogent probability arguments must include an ability to evaluate whether a probability argument is cogent.

I believe that our ability to build systems with human-like capabilities in designing probability arguments and generating numerical probability judgments will ultimately depend on our ability to build associative memories. With a genuine associative memory, we could retrieve stored experiences that approximately match any arbitrary new situation, not just those that match a relatively few situations we might specify in advance. The retrieval of such stored experiences on a fine scale would permit us to calculate frequencies that could serve as numerical probability judgments, and the comparison to other problems on a coarser scale could give hints for the design of a probability argument. Associative memory is currently an active and exciting field of research in artificial intelligence (Hinton and Anderson, 1981; Hopfield, 1982; Kohonen, 1984). It is a field where statisticians should be making a greater contribution than they are.

The entire field of artificial intelligence poses a challenge to students of probability. I believe that probability judgment will turn out to be possible and important in artificial intelligence, but the extent of its ultimate usefulness cannot be taken for granted; it must be demonstrated.

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under Grants IST-8405210 and IST-8610293.

REFERENCES

BARR, A. and FEIGENBAUM, E. A., eds. (1981, 1982). *The Handbook of Artificial Intelligence*, **1**, **2**. William Kaufmann, Los Altos, Calif.

BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *J. Roy. Statist. Soc. Ser. A* **143** 383–430.

BUCHANAN, B. G. and SHORTLIFFE, E. H., eds. (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, Reading, Mass.

COHEN, P. R. (1985). *Heuristic Reasoning about Uncertainty: An Artificial Intelligence Approach.* Pitman, Boston.

COHEN, P. R. and FEIGENBAUM, E. A., eds. (1982). *The Handbook of Artificial Intelligence*, **3**. William Kaufmann, Los Altos, Calif.

DAVIS, R. and KING, J. J. (1984). The origin of rule-based systems in AI. In Buchanan and Shortliffe (1984), 20–52.

DUDA, R. O., HART, P. E. and NILSSON, N. J. (1976). Subjective Bayesian methods for rule-based inference systems. *AFIPS Conf. Proc.* **45** 1075–1082. (Reprinted in *Readings in Artificial Intelligence*, edited by B. L. Webber and N. J. Nilsson. Tioga Publishing, Palo Alto, Calif., 1981, pages 192–199.)

GORDON, J. and SHORTLIFFE, E. H. (1984). The Dempster-Shafer theory of evidence. In Buchanan and Shortliffe (1984), 272–292.

GORDON, J. and SHORTLIFFE, E. H. (1985). A method for managing evidential reasoning in hierarchical hypothesis spaces. *Artificial Intelligence* **26** 323–358.

HINTON, G. E. and ANDERSON, J. A., eds. (1981). *Parallel Models of Associative Memory.* Erlbaum, Hillsdale, N. J.

HOPFIELD, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci. U.S.A.* **79** 2554–2558.

KAHNEMAN, D., SLOVIC, P. and TVERSKY, A. (1979). *Judgments under Uncertainty: Heuristics and Biases.* Cambridge Univ. Press, Cambridge.

KIM, J. H. (1983). CONVINCE: A conversational inference consolidation engine. Ph.D. dissertation, Dept. of Computer Science, Univ. California, Los Angeles.

KIM, J. H. and PEARL, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. *Proc. of the Eighth International Joint Conference on Artificial Intelligence* 190–193. William Kaufmann, Los Altos, Calif.

KOHONEN, T. (1984). *Self-Organization and Associative Memory.* Springer, New York.

MARR, D. (1982). *Vision.* Freeman, San Francisco.

McCARTHY, J. (1980). Circumspection—a form of non-monotonic reasoning. *Artificial Intelligence* **13** 27–39.

McDERMOTT, D. and DOYLE, J. (1980). Non-monotonic logic I. *Artificial Intelligence* **13** 41–72.

NEWELL, A. (1973). Production systems: models of control structures. In *Visual Information Processing* (W. G. Chase, ed.) 463–526. Academic, New York.

NEWELL, A. and SIMON, H. A. (1965). An example of human chess play in the light of chess playing programs. In *Progress in Biocybernetics* (N. Wiener and J. P. Schade, eds.). Elsevier, Amsterdam.

PEARL, J. (1982). Reverend Bayes on inference engines: a distributed hierarchical approach. *Proc. of the Second National Conference on Artificial Intelligence, American Association for Artificial Intelligence* 133–136. William Kaufmann, Los Altos, Calif.

PEARL, J. (1986). Fusion, propagation, and structuring in belief networks. *Artificial Intelligence* **29** 241–288.

REITER, R. (1980). A logic for default reasoning. *Artificial Intelligence* **13** 81–132.

SHAFER, G. (1976). *A Mathematical Theory of Evidence.* Princeton Univ. Press, Princeton, N. J.

SHAFER, G. (1981). Constructive probability. *Synthese* **48** 1–60.

SHAFER, G. (1984). The problem of dependent evidence. Working paper no. 164, School of Business, Univ. Kansas.

SHAFER, G. (1986a). Belief functions and possibility measures. To appear in *The Analysis of Fuzzy Information* (J. C. Bezdek, ed.) **1**. CRC Press.

SHAFER, G. and LOGAN, R. (1985). Implementing Dempster's rule for hierarchical evidence. Working paper no. 174, School of Business, Univ. Kansas. To appear in *Artificial Intelligence.*

SHAFER, G. and TVERSKY, A. (1985). Languages and designs for probability judgment. *Cognitive Sci.* **9** 309–339.

SHENOY, P. and SHAFER, G. (1986). Propagating belief functions with local computations. *IEEE Expert.* **1** 43–52.