From Certainty Factors to Belief Networks

David E. Heckerman* <dheck@sumex-aim.stanford.edu>

Edward H. Shortliffe <ehs@sumex-aim.stanford.edu>

Section on Medical Informatics
Stanford University School of Medicine
MSOB X-215, 300 Pasteur Drive
Stanford, CA 94305-5479

To appear in *Artificial Intelligence in Medicine,* 1992.

*Current address: Departments of Computer Science and Pathology, University of Southern California, HMR 204, 2025 Zonal Avenue, Los Angeles, CA 90033.

Abstract

The *certainty-factor* (CF) model is a commonly used method for managing uncertainty in rule-based systems. We review the history and mechanics of the CF model, and delineate precisely its theoretical and practical limitations. In addition, we examine the belief network, a representation that is similar to the CF model but that is grounded firmly in probability theory. We show that the belief-network representation overcomes many of the limitations of the CF model, and provides a promising approach to the practical construction of expert systems.

Keywords: certainty factor, probability, belief network, uncertain reasoning, expert systems

1 Introduction

In this issue of the journal, Dan and Dudeck provide a critique of the *certainty-factor* (CF) model, a method for managing uncertainty in rule-based systems. Shortliffe and Buchanan developed the CF model in the mid-1970s for MYCIN, an early expert system for the diagnosis and treatment of meningitis and bacteremia [38, 37]. Since then, the CF model has been widely adopted in a variety of expert-system shells and in individual rule-based systems that have had to reason under uncertainty. Although we concur with many of the observations in the article by Dan and Dudeck, we believe that the reasons for the development of the CF model must be placed in historical context, and that it is important to note that the AI research community has largely abandoned the use of CFs. In our laboratory, where the CF model was originally developed, we have not used CFs in our systems for over a decade. We accordingly welcome the opportunity to review the CF model, the reasons for its creation, and recent developments and analyses that have allowed us to turn in new directions for approaches to uncertainty management.

When the model was created, many artificial-intelligence (AI) researchers expressed concern about using Bayesian (or subjective) probability to represent uncertainty. Of these researchers, most were concerned about the practical limitations of using probability theory. In particular, builders of probabilistic diagnostic systems for medicine and other domains had largely been using the *simple-Bayes* model. This model included the assumptions that (1) faults or hypotheses were mutually exclusive and exhaustive, and (2) pieces of evidence were conditionally independent, given each fault or hypothesis (see Section 3). The assumptions were useful, because their adoption had made the construction of diagnostic systems practical. Unfortunately, the assumptions were often inaccurate in practice.

The reasons for creating the CF model were described in detailed in the original paper by Shortliffe and Buchanan [38]. The rule-based approach they were developing required a modular approach to uncertainty management, and efforts to use a Bayesian model in this context had been fraught with difficulties. Not only were they concerned about the assumptions used in the simple-Bayes model, but they wanted to avoid the cognitive complexity that followed from dealing with large numbers of conditional and prior probabilities. They also found that it was difficult to assess subjective probabilities from experts in a way that was internally consistent. Furthermore, initial interactions with their experts led them to believe that the numbers being assigned by the physicians with whom they were working were different in character from probabilities. Thus, the CF model was created for the domain of MYCIN as a practical approach to uncertainty management in rule-based systems. Indeed, in blinded evaluations of MYCIN, the CF model provided recommendations for treatment that were judged to be equivalent to, or better than, therapy plans provided by infectious disease experts for the same cases [47, 48].

Despite the success of the CF model for MYCIN, its developers warned researchers and knowledge engineers that the model had been designed for a domain with unusual characteristics, and that the model's performance might be sensitive to the domain of application. Clancey and Cooper's sensitivity analysis of the CF model [4, Chapter 10] demonstrated that MYCIN's therapy recommendations were remarkably insensitive to perturbations in the CF values assigned to rules in the system. MYCIN's diagnostic assessments, however, showed more rapid deterioration as CF values were altered. Since MYCIN was primarily a therapy advice system, and since antibiotic therapies often cover for many pathogens, variations in diagnostic hypotheses often had minimal effect on the therapy that was recommended; in certain cases where perturbations in CFs led to an incorrect diagnosis, the treatments recommended by the model were still appropriate. This observation suggests strongly that the CF model may be inadequate for diagnostic systems or in domains where appropriate recommendations of treatment are more sensitive to accurate diagnosis. Unfortunately, this point has been missed by many investigators who have built expert systems using CFs or have incorporated the CF model as the uncertainty management scheme for some expert-system shells.

In this article, we reexamine the CF model, illustrating both its theoretical and practical limitations. In agreement with Shortliffe and Buchanan's original view of the model, we see that CFs do not correspond to probabilities. We find, however, that CFs can be interpreted as measures of change in belief within the theory of probability. If one interprets the CF model in this way, we show that, in many circumstances, the model implicitly imposes assumptions that are stronger than those of the simple-Bayes model. We trace the flaws in the model to its imposition of the same sort of *modularity* on uncertain rules that we accept for logical rules, and we show that uncertain reasoning is inherently less modular than is logical reasoning. Also, we argue that the assessment of CFs is often more difficult and less reliable than is the assessment of conditional probabilities. Most important, we describe an alternative to the CF model for managing uncertainty in expert systems. In particular, we discuss the belief network, a graphical representation of beliefs in the probabilistic framework. We see that this representation overcomes many of the difficulties associated with both the simple-Bayes and CF models. In closing, we point to recent research that shows that diagnostic systems constructed using belief networks can be practically applied in real-world clinical settings.

2 The Mechanics of the Model

To understand how the CF model works, let us consider a simple example adapted from Kim and Pearl [25]:

> Mr. Holmes receives a telephone call from Eric, his neighbor. Eric notifies Mr. Holmes that he has heard a burglar alarm sound from the direction of Mr. Holmes' home. About a minute later, Mr. Holmes receives a telephone call from Cynthia, his other neighbor, who gives him the same news.

A miniature rule-based system for Mr. Holmes' situation contains the following rules:

R_1: if ERIC'S CALL then ALARM, $CF_1 = 0.8$

R_2: if CYNTHIA'S CALL then ALARM, $CF_2 = 0.9$

R_3: if ALARM then BURGLARY, $CF_3 = 0.7$

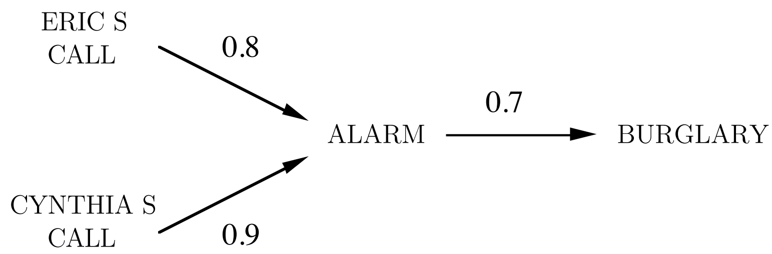


Figure 1: An inference network for Mr. Holmes' situation.

Each arc represents a rule. For example, the arc from ALARM to BURGLARY represents the rule R_3 ("if ALARM then BURGLARY"). The number above the arc is the CF for the rule. The CF of 0.7 indicates that burglar alarms can go off for reasons other than burglaries. The CFs of 0.8 and 0.9 indicate that Mr. Holmes finds Cynthia to be slightly more reliable than Eric. (Figure taken from D. Heckerman, The Certainty-Factor Model, S. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Second Edition*. Wiley, New York.)

In general, rule-based systems contain rules of the form "if e then h," where e denotes a piece of evidence for hypothesis h. Using the CF model, an expert represents his uncertainty in a rule by attaching a single CF to each rule.

Shortliffe and Buchanan intended a CF to represent a person's (usually, the expert's) *change* in belief in the hypothesis given the evidence. In particular, a CF between 0 and 1 means that the person's belief in h given e increases, whereas a CF between -1 and 0 means that the person's belief decreases. The developers of the model did not intend a CF to represent a person's *absolute* degree of belief in h given e, as does a probability [38]. We return to this point in Section 4.3.

Several implementations of rule-based representation of knowledge display a rule base in graphical form as an *inference network*. Figure 1 illustrates the inference network for Mr. Holmes' situation. Each arc in an inference network represents a rule; the number above the arc is the CF for the rule.

Using the CF model, we can compute the change in belief in any hypothesis in the network, given the observed evidence. We do so by applying simple *combination functions* to the CFs that lie between the evidence and the hypothesis in question. For example, in Mr. Holmes' situation, we are interested in computing the change in belief of BURGLARY, given that Mr. Holmes received both ERIC'S CALL and CYNTHIA'S CALL. We combine the CFs in two steps. First, we combine CF_1 and CF_2, the CFs for R_1 and R_2, to give the CF for the new composite rule R_4:

R_4: if ERIC'S CALL and CYNTHIA'S CALL then ALARM, CF_4

We combine CF_1 and CF_2 using the function

$$CF_4 = \begin{cases} CF_1 + CF_2 - CF_1 CF_2 & CF_1, CF_2 \geq 0 \\ CF_1 + CF_2 + CF_1 CF_2 & CF_1, CF_2 < 0 \\ \frac{CF_1 + CF_2}{1 - \min(|CF_1|, |CF_2|)} & \text{otherwise} \end{cases} \tag{1}$$

For $CF_1 = 0.8$ and $CF_2 = 0.9$, we have

$$CF_4 = 0.8 + 0.9 - (0.8)(0.9) = 0.98$$

Equation 1 may be called the *parallel-combination function*.

The earliest version of the CF model employed a parallel-combination function slightly different from Equation 1. There, positive CFs—called MBs—were combined as in Equation 1. Also, negative CFs—called MDs—were combined as in Equation 1. The final CF for the hypothesis, however, was given as the difference between MB and MD. This combination function has the following undesirable property. Let us suppose we have many strong pieces of evidence for a hypothesis. In particular, suppose that the combined certainty factor for the hypothesis has asymptotically approached 1. In addition, suppose that we have one weak piece of evidence against the same hypothesis, with a CF of -0.5. Using the original combination function, the net CF for the combined evidence would be approximately 0.5, which represents only weak evidence for the hypothesis. Shortliffe and Buchanan found unappealing this ability of a single, weak piece of negative evidence to overwhelm many pieces of positive evidence. Consequently, they and van Melle modified the function to that described by Equation 1 [44]. In this issue, Dan and Dudek argue that the original combination function is satisfactory. We find the argument made here more compelling then their argument.

Second, we combine CF_3 and CF_4, to give the CF for the new composite rule R_5:

R_5: if ERIC'S CALL and CYNTHIA'S CALL then BURGLARY, CF_5

The combination function is

$$CF_5 = \begin{cases} CF_3 CF_4 & CF_3 > 0 \\ 0 & CF_3 \leq 0 \end{cases} \tag{2}$$

In Mr. Holmes' case, we have

$$CF_5 = (0.98)(0.7) = 0.69$$

Equation 2 may be called the *serial-combination function*. The CF model prescribes this function to combine two rules where the hypothesis in the first rule is the evidence in the second rule (i.e., when the rules "chain" together).

If all evidence and hypotheses in a rule base are simple propositions, we need to use only the serial and parallel combination rules to combine CFs. The CF model, however, also incorporated combination functions to accommodate rules that contain conjunctions and disjunctions of evidence. For example, suppose we have the following rule in an expert system for diagnosing chest pain:

R_6: if CHEST PAIN *and*
SHORTNESS OF BREATH
then HEART ATTACK, $CF_6 = 0.9$

Further, suppose that we have rules that reflect indirect evidence for chest pain and shortness of breath:

R_7: if PATIENT GRIMACES then CHEST PAIN, $CF_7 = 0.7$

R_8: if PATIENT CLUTCHES THROAT then SHORTNESS OF BREATH, $CF_8 = 0.9$

We can combine CF_6, CF_7, and CF_8 to yield the CF for the new composite rule R_9:

R_9: if PATIENT GRIMACES *and*
PATIENT CLUTCHES THROAT
then HEART ATTACK, CF_9

The combination function is

$$CF_9 = CF_6 \min(CF_7, CF_8) = (0.9)(0.7) = 0.63 \quad (3)$$

That is, we compute the serial combination of CF_6 and the minimum of CF_7 and CF_8. We use the minimum of CF_7 and CF_8, because R_6 contains the conjunction of CHEST PAIN and SHORTNESS OF BREATH. In general, the CF model prescribes that we use the minimum of CFs for evidence in a conjunction, and the maximum of CFs for evidence in a disjunction.

There are many variations among the implementations of the CF model. For example, the original CF model used in MYCIN treats CFs less than 0.2 as though they were 0 in serial combination, to avoid the generation of unnecessary questions to the the user under its goal-directed reasoning scheme. For the sake of brevity, we will not describe other variations, but they are thoroughly outlined in [4].

3 The Simple-Bayes and CF Models

The simple-Bayes model is restrictive, in part, because it includes the assumption that pieces of evidence are conditionally independent, given each hypothesis. In general, propositions a and b are *independent*, if a person's probability (or belief) of a does not change once b becomes known. Propositions a and b are *conditionally independent*, given proposition c, if a and b are independent when a person assumes or knows that c is true. Thus, in using the simple-Bayes model, we assume that if we know which hypothesis is true, then observing one or more pieces of evidence does not change our probability that other pieces of evidence are true.

In the simple case of Mr. Holmes, the CF model is an improvement over the simple-Bayes model. In particular, ERIC'S CALL and CYNTHIA'S CALL are not conditionally independent, given BURGLARY, because even if Mr. Holmes knows that a burglary has occurred, receiving Eric's call increases Mr. Holmes belief that Cynthia will call. The lack of conditional independence is due to the triggering of calls by the sound of the alarm, and not by the burglary. In this example, the CF model represents accurately this lack of independence through the presence of ALARM in the inference network.

Unfortunately, the CF model cannot represent most real-world problems in a way that is both accurate and efficient. This limitation may not be serious in domains such as MYCIN's, but it does call into question the use of the CF model as a *general* method for managing uncertainty. In the next section, we shall see that the assumptions of conditional independence associated with the parallel-combination function are stronger (i.e., are less likely to be accurate) than are those associated with the simple-Bayes model.

4 Theoretical Problems with the CF Model

Rules that represent logical relationships satisfy the *principle of modularity*. That is, given the logical rule "if e then h," and given that e is true, we can assert that h is true (1) no matter how we established that e is true, and (2) no matter what else we know to be true. We call (1) and (2) the *principle of detachment* and the *principle of locality,* respectively. For example, given the rule

R_{10}: if L_1 and L_2 are parallel lines then L_1 and L_2 do not intersect

we can assert that L_1 and L_2 do not intersect once we know that L_1 and L_2 are parallel lines. This assertion depends on neither how we came to know that L_1 and L_2 are parallel (the principle of detachment), nor what else we know (the principle of locality).

The CF model employs the same principles of detachment and locality to belief updating. For example, given the rule

R_3: if ALARM then BURGLARY, $CF_3 = 0.7$

and given that we know ALARM, the CF model allows us to update Mr. Holmes' belief in BURGLARY by the amount corresponding to a CF of 0.7, no matter how Mr. Holmes established his belief in ALARM, and no matter what other facts he knows.

Unfortunately, uncertain reasoning often violates the principles of detachment and locality. Use of the CF model, therefore, often leads to errors in reasoning.[1] In the remainder of this section, we examine two classes of such errors.

4.1 Multiple Causes of the Same Effect

Let us consider a simple embellishment to Mr. Holmes' problem:

> As he is preparing to rush home, Mr. Holmes recalls that the previous sounding of his alarm was triggered by an earthquake. A moment later, he hears a radio newscast reporting an earthquake 200 miles from his house.

Figure 2 illustrates a possible inference network for his situation. To the original inference network of Figure 1, we have added the rules

R_{11}: if RADIO NEWSCAST then EARTHQUAKE, $CF_{11} = 0.9$

R_{12}: if EARTHQUAKE then ALARM, $CF_{12} = 0.6$

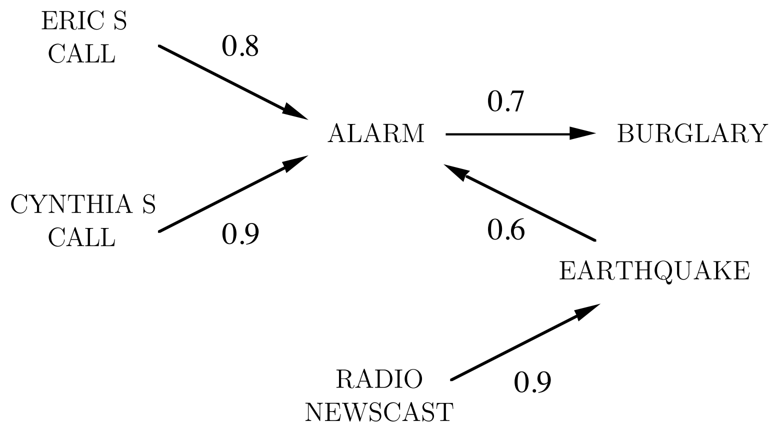


Figure 2: Another inference network for Mr. Holmes' situation.
In addition to the interactions in Figure 1, RADIO NEWSCAST increases the chance of EARTHQUAKE, and EARTHQUAKE increases the chance of ALARM. (Figure taken from D. Heckerman, The Certainty-Factor Model, S. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Second Edition*. Wiley, New York.)

The inference network does not capture an important interaction among the propositions. In particular, the modular rule R_3 ("if ALARM then BURGLARY") gives us permission to increase Mr. Holmes' belief in BURGLARY, when his belief in ALARM increases, no matter how Mr. Holmes increases his belief for ALARM. This modular license to update belief, however, is not consistent with common sense. When Mr. Holmes hears the radio newscast, he increases his belief that an earthquake has occurred. Therefore, he decreases his belief that there has been a burglary, because the occurrence of an earthquake would account for the alarm sound. Overall, Mr. Holmes' belief in ALARM increases, but his belief in BURGLARY decreases.

When the evidence for ALARM came from ERIC'S CALL and CYNTHIA'S CALL, we had no problem propagating this increase in belief through R_3 to BURGLARY. In contrast, when the evidence for ALARM came from EARTHQUAKE, we could not propagate this increase in belief through R_3. This difference illustrates a violation of the detachment principle in uncertain reasoning: the *source* of a belief update, in part, determines whether or not that update should be passed along to other propositions.

Pearl describes this phenomenon in detail [33, Chapter 1]. He divides uncertain rules into two types: diagnostic and predictive.[2] In a *diagnostic rule*, we change the belief in a cause, given an effect. All the rules in the inference network of Figure 2, except R_{12}, are of this form. In a *predictive rule*, we change the belief in an effect, given a cause. R_{12} is an example of such an rule. Pearl describes the interactions between the two types of rules. He notes that, if the belief in a proposition is increased by a diagnostic rule, then that increase can be

[1]Heckerman and Horvitz first noted the nonmodularity of uncertain reasoning, and the relationship of such nonmodularity to the limitations of the CF model [18, 17]. Pearl first decomposed the principle of modularity into the principles of detachment and locality [33, Chapter 1].

[2]Henrion also makes this distinction [20].

passed through to another diagnostic rule—just what we expect for the chain of inferences from ERIC'S CALL and CYNTHIA'S CALL to BURGLARY. On the other hand, if the belief in a proposition is increased by a predictive rule, then that belief should not be passed through a diagnostic rule. Moreover, when the belief in one cause of an observed effect increases, the beliefs in another cause should decrease—even when the two causes are not mutually exclusive. This interaction is just what we expect for the two causes of ALARM.

We might be tempted to repair the inference network in Figure 2, by adding the rule

$$R_{13}\text{: if EARTHQUAKE then BURGLARY, } CF_{13} = -0.7$$

Unfortunately, this addition leads to another problem. In particular, suppose that Mr. Holmes had never received the telephone calls. Then, the radio newscast should not affect his belief in a burglary. The modular rule R_{13}, however, gives us a license to decrease Mr. Holmes' belief in BURGLARY, whether or not he receives the phone calls. This problem illustrates that uncertain reasoning also can violate the principle of locality: The validity of an inference may depend on the truth of other propositions.

To represent accurately the case of Mr. Holmes, we must include a rule for every possible combination of observations:

if	ERIC'S CALL *and*
	CYNTHIA'S CALL *and*
	RADIO NEWSCAST
then	BURGLARY

if	NOT ERIC'S CALL *and*
	CYNTHIA'S CALL *and*
	RADIO NEWSCAST
then	BURGLARY

⋮

This representation is inefficient, is difficult to modify, and needlessly clusters propositions that are only remotely related. Ideally, we would like a representation that encodes only direct relationships among propositions, and that infers indirect relationships. In Section 6, we examine the belief network, a representation with such a capability.

We find the same difficulties encountered in using CFs to represent Mr. Holmes' situation whenever we have multiple causes of a common effect. For example, if a friend tells us that our car will not start, we initially may suspect that either our battery is dead or the gas tank is empty. Once we find that our radio is dead, however, we decrease our belief that the tank is empty, because now it is more likely that our battery is dead. Here, the relationship between CAR WILL NOT START and TANK EMPTY is influenced by RADIO DEAD, just as the relationship between ALARM and BURGLARY is influenced by RADIO NEWSCAST. In general, when one effect shares more than one cause, we should expect violations of the principles of detachment and locality.

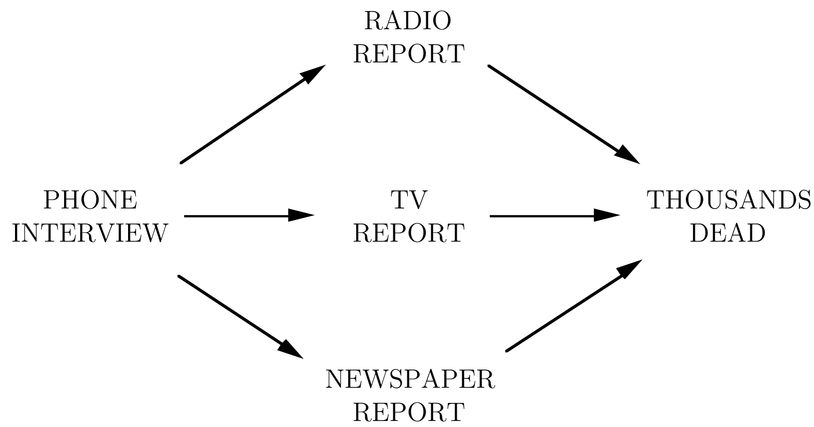


Figure 3: An inference network for the Chernobyl disaster (adapted from [19]).
When we combine CFs as modular belief updates, we overcount the chance of THOUSANDS DEAD.

4.2 Correlated Evidence

Figure 3 depicts an inference network for news reports about the Chernobyl disaster. On hearing radio, television, and newspaper reports that thousands of people have died of radioactive fallout, we increase substantially our belief that many people have died. When we learn that each of these reports originated from the same source, however, we decrease our belief. The CF model, however, treats both situations identically.

In this example, we see another violation of the principle of detachment in uncertain reasoning: The sources of a set of belief updates can strongly influence how we combine those updates. Because the CF model imposes the principle of detachment on the combination of belief updates, it overcounts evidence when the sources of that evidence are positively correlated, and it undercounts evidence when the sources of evidence are negatively correlated.

4.3 Probabilistic Interpretations for Certainty Factors

Although the developers of MYCIN and derivative systems used the CF model and combining functions without explicitly depending upon a specific interpretation of the numbers themselves, several researchers have assigned probabilistic interpretations to CFs. We can use these interpretations to understand better the limitations of the CF model. In the original work describing the model, Shortliffe and Buchanan proposed the following approximate interpretation:

$$CF(h \rightarrow e|\xi) = \begin{cases} \frac{p(h|e,\xi)-p(h|\xi)}{1-p(h|\xi)} & p(h|e,\xi) \geq p(h|\xi) \\ \frac{p(h|e,\xi)-p(h|\xi)}{p(h|\xi)} & p(h|e,\xi) < p(h|\xi) \end{cases} \tag{4}$$

where $CF(h \rightarrow e|\xi)$ is the CF for the rule "if e then h" given by an expert with background knowledge ξ; $p(h|\xi)$ is the expert's probability (degree of belief) for h given ξ; and $p(h|e,\xi)$ is the expert's probability for h given evidence e and ξ.[3] Adams examined this interpretation in detail [1]. He proved that the the parallel combination function—with the exception of the combination of CFs of mixed sign—is consistent with the rules of probability, provided (1) evidence is marginally independent and (2) evidence is conditionally independent, given h and NOT h.

Heckerman analyzed the model in a different way [12]. In particular, Heckerman showed that Shortliffe and Buchanan's probabilistic interpretation was inconsistent with the combination functions that were used by MYCIN and its descendants. For example, he showed that the probabilistic interpretation prescribes noncommutative parallel combination of evidence, even though the parallel combination function (Equation 1) is commutative.[4] Given this inconsistency, Heckerman argued that either the probabilistic interpretation, the combination functions, or both components of the model must be reformulated. Heckerman, in contrast to Adams, argued that Shortliffe and Buchanan's interpretation should be discarded, because he believed (as do Shortliffe and Buchanan) that the combination functions are the cornerstone of the CF model; their proposed definitions (Equation 4) were simply an attempt to show how the numbers used by MYCIN might be interpreted.

Heckerman went on to show that we can interpret a certainty factor for hypothesis h, given evidence e, as a monotonic increasing function of the likelihood ratio

$$\lambda(h, e) = \frac{p(e|h, \xi)}{p(e|\text{NOT } h, \xi)} \tag{5}$$

In particular, he showed that, if we make the identification

$$CF(h \to e|\xi) = \begin{cases} \frac{\lambda(h,e)-1}{\lambda(h,e)} & \lambda(h, e) \geq 1 \\ \lambda(h, e) - 1 & \lambda(h, e) < 1 \end{cases} \tag{6}$$

then the parallel-combination function used by MYCIN (Equation 1) follows exactly from the rules of probability. In addition, with the identification in Equation 6, the serial-combination function (Equation 2) and the combination functions for disjunction and conjunction are close approximations to the rules of probability. Using Bayes' theorem, we can write Equation 6 as

$$CF(h \to e|\xi) = \begin{cases} \frac{p(h|e,\xi)-p(h|\xi)}{(1-p(h|\xi))p(h|e,\xi)} & p(h|e, \xi) \geq p(h|\xi) \\ \frac{p(h|e,\xi)-p(h|\xi)}{p(h|\xi)(1-p(h|e,\xi))} & p(h|e, \xi) < p(h|\xi) \end{cases} \tag{7}$$

This probabilistic interpretation for CFs differs from Shortliffe and Buchanan's interpretation only by an additional term in the denominator of each case. These extra terms make the relationship between the probabilities $p(h|e, \xi)$ and $p(h|\xi)$ symmetric. It is this symmetry that makes this interpretation consistent with the combination functions.

[3]Shortliffe and Buchanan did not make the expert's background knowledge ξ explicit. Nonetheless, in the Bayesian interpretation of probability theory, a probability is always conditioned on the background knowledge of the person who assesses that probability.

[4]In this issue, Dan and Dudeck dispute this demonstration. They argue that we must combine evidence before applying the interpretation. However, we do not see any reason why this or any other probabilistic interpretation should be subjected to this limitation; a probabilistic interpretation should be limited only by the rules of probability.

The original interpretation of CFs (Equation 4) reflects Shortliffe and Buchanan's view that a CF represents a measure of change in belief (see the difference terms in the numerators). Nonetheless, in this interpretation, as $p(h|\xi)$ approaches 0 with $p(h|e,\xi)$ fixed, $CF(h \rightarrow e|\xi)$ approaches $p(h|e,\xi)$, a measure of absolute belief. The odd fact that a measure of change in belief can approach a measure of absolute belief led Shortliffe and Buchanan to emphasize the approximate nature of their probabilistic interpretation. In fact, they had speculated that CFs may not admit any probabilistic interpretation. Heckerman's interpretation, however, does not exhibit this unusual behavior. In particular, as $p(h|\xi)$ approaches 0 with $p(h|e,\xi)$ fixed, $CF(h \rightarrow e|\xi)$ approaches 1. This behavior is reasonable: If a hypothesis is extremely unlikely, we require a strong belief update to make that hypothesis at all likely. Thus, under the interpretation of the CF model as proposed by Heckerman, one can dispute previous claims that the CF model is fundamentally different from the theory of probability.

Heckerman's interpretation of CFs can help us to understand the limitations of the model. In developing this interpretation, Heckerman showed that the parallel and serial combination functions impose assumptions of conditional independence on the propositions involved in the combinations. In particular, when we use the parallel-combination function to combine CFs for the rules "if e_1 then h" and "if e_2 then h," we assume implicitly that e_1 and e_2 are conditionally independent, given h and NOT h. Similarly, when we use the serial-combination function to combine CFs for the rules "if a then b" and "if b then c," we assume implicitly that a and c are conditionally independent, given b and NOT b. Heckerman also showed that the combination functions for disjunction and conjunction impose specific forms of conditional dependence on the propositions involved in the combinations [13].

Overall, Heckerman's interpretation shows us that the independence assumptions imposed by the CF model make the model inappropriate for many—if not most—real-world domains. Indeed, the assumptions of the parallel-combination function are stronger than are those of the simple-Bayes model, the same model whose limitations motivated in part the development of the CF model. That is, when we use the simple-Bayes model, we assume that evidence is conditionally independent given each hypothesis. When we use the parallel-combination function, however, we assume that evidence is conditionally independent given each hypothesis and the *negation* of each hypothesis. Unless the space of hypotheses consists of a single proposition and the negation of that proposition, the parallel-combination assumptions are essentially impossible to satisfy, even when the simple-Bayes assumptions are satisfied [23].

In closing this section, we comment on Dan and Dudek's suggestion that $CF(h \rightarrow e|\xi)$ should be interpreted as a measure of the absolute belief in h, given e. In making this suggestion, they observe that most users of the model assess and use CFs as though they were absolute measure of belief. Moreover, they argue that we should not seek a probabilistic interpretation, because of the "inappropriate logical foundation" of probability theory. The authors of this article disagree strongly with this view. First, we believe that—from a theoretical perspective—probability theory is the most appropriate representation of uncertain beliefs. The theory is self consistent and well developed, and allows for the unambiguous representation of independence assumptions. In addition, psychologists and decision analysts have shown that the use of probability theory can help people avoid mistakes in reasoning [6, 42, 40, 24]. We know of no other representation for uncertain beliefs that have all of these benefits.

Second, if a $CF(h \rightarrow e|\xi)$ were to represent an expert's absolute belief in h, given e, then the parallel combination of CFs for multiple pieces of evidence would overcount the expert's initial or prior belief in h. For example, let us suppose that e_1 and e_2 are two pieces of evidence for h. Because both $CF(h \rightarrow e_1|\xi)$ and $CF(h \rightarrow e_2|\xi)$ incorporate the expert's prior belief in h, we doublecount this prior belief when we combine the CFs using the parallel combination function. The observation that most people interpret CFs as measures of absolute belief simply shows that most people are making this error; the observation is not an argument for how we *should* interpret CFs.

4.4 A Fundamental Difference

Under Heckerman's interpretation, we can identify precisely the problem with the CF representation of Mr. Holmes' situation. There, we use serial combination to combine CFs for the sequence of propositions EARTHQUAKE, ALARM, and BURGLARY. In doing so, we make the inaccurate assumption (among others) that EARTHQUAKE and BURGLARY are conditionally independent, given ALARM. No matter how we manipulate the arcs in the inference network of Figure 2, we generate inaccurate assumptions of conditional independence.

We can understand the problems with the CF model, however, at a more intuitive level. Logical relationships represent what we can observe directly. In contrast, uncertain relationships encode invisible influences: exceptions to that which is visible. For example, a burglary will not always trigger an alarm, because there are hidden mechanisms that may inhibit the sounding of the alarm. We summarize these hidden mechanisms in a probability for ALARM given BURGLARY. In the process of summarization, we lose information. Therefore, when we try to combine uncertain information, unexpected (nonmodular) interactions may occur. We should not expect that the CF model—or any modular belief updating scheme—will be able to handle such subtle interactions. Pearl provides a detailed discussion of this point [33, Chapter 1].

5 A Practical Problem with the CF Model

In addition to the theoretical difficulties of updating beliefs within the CF model, the model contains a serious practical problem. Specifically, the CF model requires that we encode rules in the direction in which they are used. That is, an inference network must trace a trail of rules from observable evidence to hypotheses.

Unfortunately, we often do not use rules in the same direction in which experts can most accurately and most comfortably assess the strength of the relationship. Kahneman and Tversky have shown that people are usually most comfortable when they assess the strength of relationship in predictive rules ("if CAUSE then EFFECT") rather than in diagnostic rules ("if EFFECT then CAUSE"). For example, expert physicians prefer to assess the likelihood of a finding, given a disease, rather than the likelihood (or belief update) of a disease, given a finding [43]. Henrion attributes this phenomenon to the nature of causality. In particular, he notes that a predictive probability (the likelihood of a finding, given a disease) reflects a stable property of that disease. In contrast, a diagnostic probability (the likelihood of a disease, given a finding) depends on the incidence rates of that disease and of other diseases that may cause the finding. Thus, predictive probabilities are a more useful and parsimonious way to represent uncertain relationships—at least in medical domains (see [21], pages 252–3). The developers of QMR, a diagnostic program for general internal medicine that uses ad hoc measures of uncertainty for both diagnostic and predictive rules, make a similar observation [28]. Indeed, the majority of medical literature (both textbooks and journal articles) describes predictive rules for a given disease, rather than diagnostic rules for a given finding.

Unfortunately for the CF model, effects are usually the observable pieces of evidence, and causes are the sought-after hypotheses. Thus, in using the CF model, we force experts to construct diagnostic rules. Consequently, we force experts to provide judgments of uncertainty in a direction that is more cognitively challenging. We thereby promote errors in assessment. In the next section, we examine the belief network, a representation that allows experts to represent knowledge in whatever direction they prefer.

6 Belief Networks: A Language of Dependencies

The examples in this article illustrate that we need a language that helps us to keep track of the sources of our belief, and that makes it easy for us to represent or infer the propositions on which each of our beliefs are dependent. The belief network is such a language.[5] Several researchers independently developed the representation—for example, Wright [46], Good [8], and Rousseau [34], and Pearl [31]. Howard and Matheson [22] developed the influence diagram, a generalization of the belief network in which we can represent decisions and the preferences of a decision maker.

Figure 4 shows a belief network for Mr. Holmes' situation. The belief network is a directed acyclic graph.[6] The nodes in the graph correspond to uncertain variables relevant to the problem. For Mr. Holmes, each uncertain variable represents a proposition and that proposition's negation. For example, node b in Figure 4 represents the propositions BURGLARY and NOT BURGLARY (denoted b_+ and b_-, respectively). In general, an uncertain variable can represent an arbitrary set of mutually exclusive and exhaustive propositions; we call each proposition an *instance* of the variable. In the remainder of the discussion, we make no distinction between the variable x and the node x that represents that variable.

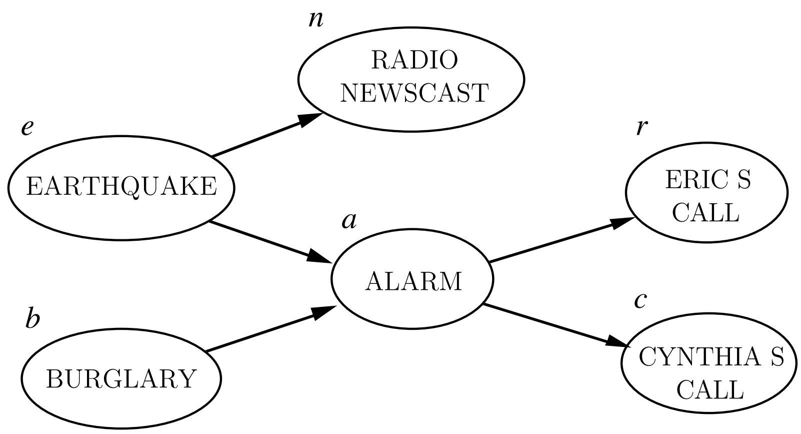
Each variable in a belief network is associated with a set of probability distributions.[7] In the Bayesian tradition, these distributions encode the knowledge provider's beliefs about the relationships among the variables. Mr. Holmes' probabilities appear below the belief network in Figure 4.

The arcs in the directed acyclic graph represent direct probabilistic dependencies among

[5]Other names for belief networks include probabilistic networks, causal networks, and Bayesian networks.

[6]A directed acyclic graph contains no directed cycles. That is, in a directed acyclic graph, we cannot travel from a node and return to that same node along a nontrivial directed path.

[7]A probability distribution is an assignment of a probability to each instance of a variable.



$p(b_+ | \xi) = 0.003$

$p(e_+ | \xi) = 0.001$

$p(r_+ | a_-, \xi) = 0.04$

$p(r_+ | a_+, \xi) = 0.2$

$p(c_+ | a_-, \xi) = 0.04$

$p(c_+ | a_+, \xi) = 0.4$

$p(n_+ | e_-, \xi) = 0.00002$

$p(n_+ | e_+, \xi) = 0.2$

$p(a_+ | b_-, e_-, \xi) = 0.0003$

$p(a_+ | b_+, e_-, \xi) = 0.6$

$p(a_+ | b_-, e_+, \xi) = 0.5$

$p(a_+ | b_+, e_+, \xi) = 0.8$

Figure 4: A belief network for Mr. Holmes' situation.

The nodes in the belief network represent the uncertain variables relevant to Mr. Holmes' situation. The arcs represent direct probabilistic dependencies among the variables, whereas the lack of arcs between nodes represents assertions of conditional independence. Each node in the belief network is associated with a set of probability distributions. These distributions appear below the graph. The variables in the probabilistic expressions correspond to the nodes that they label in the belief network. For example, $p(b_+|\xi)$ denotes the probability that a burglary has occurred, given Mr. Holmes' background information, ξ. The figure does not display the probabilities that the events failed to occur. We can compute these probabilities by subtracting from 1.0 the probabilities shown. (Figure taken from D. Heckerman, The Certainty-Factor Model, S. Shapiro, editor, *Encyclopedia of Artificial Intelligence, Second Edition*. Wiley, New York.)

the uncertain variables. In particular, an arc from node x to node y reflects an assertion by the builder of that network that the probability distribution for y may depend on the instance of the variable x; we say that *x conditions y*. Thus, a node has a probability distribution for every instance of its conditioning nodes. (An instance of a set of nodes is an assignment of an instance to each node in that set.) For example, in Figure 4, ALARM is conditioned by both EARTHQUAKE and BURGLARY. Consequently, there are four probability distributions for ALARM, corresponding to the instances where both EARTHQUAKE and BURGLARY occur, BURGLARY occurs alone, EARTHQUAKE occurs alone, and neither EARTHQUAKE nor BURGLARY occurs. In contrast, RADIO NEWSCAST, ERIC'S CALL, and CYNTHIA'S CALL are each conditioned by only one node. Thus, there are two probability distributions for each of these nodes. Finally, EARTHQUAKE and BURGLARY do not have any conditioning nodes, and hence each node has only one probability distribution.

The lack of arcs in a belief network reflects assertions of conditional independence. For example, there is no arc from BURGLARY to ERIC'S CALL in Figure 4. The lack of this arc encodes Mr. Holmes' belief that the probability of receiving Eric's telephone call from his neighbor does not depend on whether or not there was a burglary, provided Mr. Holmes knows whether or not the alarm sounded.

Pearl describes the exact semantics of missing arcs [33]. Here, it is important to recognize that, given any belief network, we can construct the joint probability distribution for the variables in any belief network from (1) the probability distributions associated with each node in the network, and (2) the assertions of conditional independence reflected by the lack of some arcs in the network. The *joint probability distribution* for a set of variables is the collection of probabilities for each instance of that set. The distribution for Mr. Holmes' situation is

$$p(e, b, a, n, r, c|\xi) = p(e|\xi)\ p(b|\xi)\ p(a|e, b, \xi)\ p(n|e, \xi)\ p(r|a, \xi)\ p(c|a, \xi) \tag{8}$$

The probability distributions on the right-hand side of Equation 8 are exactly those distributions associated with the nodes in the belief network.

6.1 Getting Answers from Belief Networks

Given a joint probability distribution over a set of variables, we can compute any conditional probability that involves those variables. In particular, we can compute the probability of any set of hypotheses, given any set of observations. For example, Mr. Holmes undoubtedly wants to determine the probability of BURGLARY (b_+) given RADIO NEWSCAST (n_+) and ERIC'S CALL (r_+) and CYNTHIA'S CALL (c_+). Applying the rules of probability[8] to the joint probability distribution for Mr. Holmes' situation, we obtain

$$p(b_+|n_+, r_+, c_+, \xi) = \frac{p(b_+, n_+, r_+, c_+|\xi)}{p(n_+, r_+, c_+|\xi)}$$

[8] See, for example, [41].

$$= \frac{\sum_{e_i,a_k} p\left(e_i, b_+, a_k, n_+, r_+, c_+|\xi\right)}{\sum_{e_i,b_j,a_k} p\left(e_i, b_j, a_k, n_+, r_+, c_+|\xi\right)}$$

where e_i, b_j, and a_k denote arbitrary instances of the variables e, b, and a, respectively.

In general, given a belief network, we can compute any set of probabilities from the joint distribution implied by that network. We also can compute probabilities of interest directly within a belief network. In doing so, we can take advantage of the assertions of conditional independence reflected by the lack of arcs in the network: Fewer arcs lead to less computation. Several researchers have developed an algorithm in which we reverse arcs in the belief network, applying Bayes' theorem to each reversal, until we have derived the probabilities of interest [22, 30, 35]. Pearl has developed a message-passing scheme that updates the probability distributions for each node in a belief network in response to observations of one or more variables [32]. Lauritzen and Spiegelhalter have created an algorithm that first builds an undirected graph from the belief network [26]. The algorithm then exploits several mathematical properties of undirected graphs to perform probabilistic inference. Most recently, Cooper has developed an inference algorithm that recursively bisects a belief network, solves the inference subproblems, and reassembles the component solutions into a global solution [5].

6.2 Belief Networks for Knowledge Acquisition

A belief network simplifies knowledge acquisition by exploiting a fundamental observation about the ability of people to assess probabilities. Namely, a belief network takes advantage of the fact that people can make assertions of conditional independence much more easily than they can assess numerical probabilities [22, 32]. In using a belief network, a person first builds the graph that reflects his assertions of conditional independence; only then does he assess the probabilities underlying the graph. Thus, a belief network helps a person to *decompose* the construction of a joint probability distribution into the construction of a set of smaller probability distributions.

6.3 Advantages of the Belief Network over the CF Model

The example of Mr. Holmes illustrates the advantages of the belief network over the CF model. First, we can avoid the practical problem of the CF model that we discussed in Section 5; namely, using a belief network, a knowledge provider can choose the order in which he prefers to assess probability distributions. For example, in Figure 4, all arcs point from cause to effect, showing that Mr. Holmes prefers to assess the probability of observing an effect, given one or more causes. If, however, Mr. Holmes wanted to specify the probabilities of—say—EARTHQUAKE given RADIO NEWSCAST and of EARTHQUAKE given NOT RADIO NEWSCAST, he simply would reverse the arc from RADIO NEWSCAST to EARTHQUAKE in Figure 4. Regardless of the direction in which Mr. Holmes assesses the conditional distributions, we can use one of the algorithms mentioned in Section 6.1 to reveal the conditional probabilities of interest, if the need arises. (See [36], for a detailed discussion of this point.)

Second, using a belief network, the knowledge provider can *control* the assertions of conditional independence that are encoded in the representation. In contrast, the use of the combination functions in the CF model forces a person to adopt assertions of conditional independence that may be incorrect. For example, as we discussed in Section 4.3, the inference network in Figure 2 dictates the erroneous assertion that EARTHQUAKE and BURGLARY are conditionally independent, given ALARM.

Third, and most important, a knowledge provider does not have to assess indirect independencies, using a belief network. Such independencies reveal themselves in the course of probabilistic computations within the network.[9] Such computations can tell us—for example—that BURGLARY and RADIO NEWSCAST are normally independent, but become dependent, given ERIC'S CALL, CYNTHIA'S CALL, or both.

Thus, the belief network helps us to tame the inherently nonmodular properties of uncertain reasoning. Uncertain knowledge encoded in a belief network is not as modular as is knowledge about logical relationships. Nonetheless, representing uncertain knowledge in a belief network is a great improvement over encoding all relationships among a set of variables.

6.4 Belief Networks in Real-World Applications

Despite the strong theoretical arguments that favor the use of belief networks for representing uncertainty in medical decision-support systems, it is appropriate to ask whether there are practical technologic approaches to their adoption. Indeed, the belief-network representation recently has facilitated the construction of several real-world expert systems. For example, researchers at Stanford University and the University of Southern California used a belief network to construct Pathfinder, an expert system that assists pathologists with the diagnosis of lymph-node diseases [9, 11]. The program reasons about over 60 diseases (25 benign diseases, 9 Hodgkin's lymphomas, 18 non-Hodgkin's lymphomas, and 10 metastatic diseases) and over 140 features of disease, including morphologic, clinical, laboratory, immunological, and molecular biological findings. The belief network for Pathfinder was constructed with the aid of a similarity network, an extension of the belief-network representation that permits the incremental construction of extremely large belief networks from cognitively manageable subproblems that involve the comparison of two diseases and their distinguishing features [14, 15, 16]. A formal evaluation of Pathfinder has demonstrated that its diagnostic accuracy is at least as good as that of the program's expert [10]. Currently, the program is undergoing clinical trials that will compare the diagnostic accuracy of general pathologists who have access to Pathfinder to that of pathologists who do not have such access.

Other medical expert systems have been built with belief networks. These systems include Munin, a program for the diagnosis of muscular disorders [2], Alarm, a program that assists physicians with ventilator management [3], Sleep-It, a program for the diagnosis of sleep disorders [29], and QMR-DT, a probabilistic version of QMR [39, 27].

In addition to the development of the belief network, there are several changes in com-

[9] In fact, we are not even required to perform numerical computations to derive such indirect independencies. An efficient algorithm exists that uses only the structure of the belief network to tell us about these dependencies [7].

puting environments which have been instrumental in making it practical to return to formal probabilistic models. In particular, the dramatic increase in raw computing power make it reasonable to consider using search algorithms that would have brought reasoning systems to a halt 20 years ago. Also, the graphical environments that are now routinely available allow us to address the issues of cognitive complexity that would have limited attempts to use belief networks in earlier decades.

7 Conclusions

The widespread adoption of the CF model in the late 1970s and early 1980s is clear evidence of the importance of practical and simple methods for managing uncertain reasoning in expert systems. As we have seen, however, the simplicity of the CF model was achieved only with frequently unrealistic assumptions and with persistent confusion about the *meaning* of the numbers being used.

Fortunately, the belief-network representation overcomes many of the limitations of the CF model, and provides a promising approach to the practical construction of expert systems. We hope that our discussion will inspire investigators to develop belief-network inference algorithms and extensions to the representation that will simplify further the construction and use of probabilistic expert systems. We believe that the time is right for the development of such systems in medical domains.

Acknowledgments

We thank Bruce Buchanan and Eric Horvitz for many valuable discussions about the CF model and its limitations. Mark Peot and Jack Breese provided useful comments on an earlier version of this manuscript. This work was supported by the National Library of Medicine under the SUMEX-AIM Grant LM05208 and the Pathfinder Grant RO1LM04529, and by the National Cancer Institute under the Pathfinder Grant RO1CA51729-01A1.

References

[1] J.B. Adams. A probability model of medical reasoning and the MYCIN model. *Mathematical Biosciences*, 32:177–186, 1976.

[2] S. Andreassen, M. Woldbye, B. Falck, and S.K. Andersen. MUNIN: A causal probabilistic network for interpretation of electromyographic findings. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* Milan, Italy, pages 366–372. Morgan Kaufmann, San Mateo, CA, August 1987.

[3] I.A. Beinlich, H.J. Suermondt, R.M. Chavez, and G.F. Cooper. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine,* London. Springer Verlag, Berlin, August 1989.

[4] B.G. Buchanan and E.H. Shortliffe, editors. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison–Wesley, Reading, MA, 1984.

[5] G.F. Cooper. Bayesian belief-network inference using recursive decomposition. Technical Report KSL-90-05, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, January 1990.

[6] W. Edwards, ed. Special issue on probabilistic inference. *IEEE Transactions on Human Factors in Electronics*, 7, 1956.

[7] D. Geiger, T. Verma, and J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.

[8] I.J. Good. A causal calculus (I). *British Journal of Philosophy of Science*, 11:305–318, 1961. Also in I.J. Good, *Good Thinking: The Foundations of Probability and Its Applications,* pages 197–217. University of Minnesota Press, Minneapolis, MN, 1983.

[9] D. Heckerman, E. Horvitz, and B. Nathwani. Toward normative expert systems: The Pathfinder project. *Methods of Information in Medicine*, 1991. In press.

[10] D. Heckerman and B. Nathwani. An evaluation of the diagnostic accuracy of Pathfinder. *Computers and Biomedical Research*, 1991. In press.

[11] D. Heckerman and B. Nathwani. Toward normative expert systems: Probability-based representations for efficient knowledge acquisition. *Methods of Information in Medicine*, 1991. In press.

[12] D.E. Heckerman. Probabilistic interpretations for MYCIN's certainty factors. In *Proceedings of the Workshop on Uncertainty and Probability in Artificial Intelligence,* Los Angeles, CA, pages 9–20. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1985. Also in Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence,* pages 167–196. North-Holland, New York, 1986.

[13] D.E. Heckerman. Formalizing heuristic methods for reasoning with uncertainty. Technical Report KSL-88-07, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, May 1987.

[14] D.E. Heckerman. Probabilistic similarity networks. *Networks*, 20:607–636, 1990.

[15] D.E. Heckerman. *Probabilistic Similarity Networks*. PhD thesis, Program in Medical Information Sciences, Stanford University, Stanford, CA, June 1990. Report STAN-CS-90-1316.

[16] D.E. Heckerman. *Probabilistic Similarity Networks*. MIT Press, Cambridge, MA, 1991.

[17] D.E. Heckerman and E.J. Horvitz. The myth of modularity in rule-based systems. In *Proceedings of the Second Workshop on Uncertainty in Artificial Intelligence,* Philadelphia, PA, pages 115–121. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1986. Also in Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 2,* pages 23–34. North-Holland, New York, 1988.

[18] D.E. Heckerman and E.J. Horvitz. On the expressiveness of rule-based systems for reasoning under uncertainty. In *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence,* Seattle, WA, pages 121–126. AAAI Press, Menlo Park, CA, July 1987.

[19] M. Henrion. Should we use probability in uncertain inference systems? In *Proceedings of the Cognitive Science Society Meeting,* Amherst, PA. Carnegie–Mellon, August 1986.

[20] M. Henrion. Uncertainty in artificial intelligence: Is probability epistemologically and heuristically adequate? In J.L. Mumpower, editor, *Expert Judgment and Expert Systems*, pages 105–130. Springer-Verlag, Berlin, Heidelberg, 1987.

[21] E.J. Horvitz, J.S. Breese, and M. Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2:247–302, 1988.

[22] R.A. Howard and J.E. Matheson. Influence diagrams. In R.A. Howard and J.E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 721–762. Strategic Decisions Group, Menlo Park, CA, 1981.

[23] R. Johnson. Independence and Bayesian updating methods. In *Proceedings of the Workshop on Uncertainty and Probability in Artificial Intelligence,* Los Angeles, CA, pages 28–30. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1985. Also in Kanal, L. and Lemmer J., editors, *Uncertainty in Artificial Intelligence,* pages 197–201. North-Holland, New York, 1986.

[24] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, 1982.

[25] J.H. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference engines. In *Proceedings Eighth International Joint Conference on Artificial Intelligence,* Karlsruhe, West Germany, pages 190–193. International Joint Conference on Artificial Intelligence, ?? 1983.

[26] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society B*, 50:157–224, 1988.

[27] B. Middleton, M. Shwe, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base—part 2: Evaluation of diagnostic performance. *Methods in Information and Medicine*, 1991. In press.

[28] R.A. Miller, E.P. Pople, and J.D. Myers. INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, 307:476–486, 1982.

[29] G. Nino-Murcia and M. Shwe. An expert system for diagnosis of sleep disorders. In M. Chase, R. Lydic, and C. O'Connor, editors, *Sleep Research*, volume 20, page 433. Brain Information Service, Los Angeles, CA, 1991.

[30] S.M. Olmsted. *On Representing and Solving Decision Problems*. PhD thesis, Department of Engineering-Economic Systems, Stanford University, Stanford, CA, December 1983.

[31] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings AAAI-82 Second National Conference on Artificial Intelligence,* Pittsburgh, PA, pages 133–136. AAAI Press, Menlo Park, CA, August 1982.

[32] J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[33] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[34] W.F. Rousseau. A method for computing probabilities in complex situations. Technical Report 6252-2, Center for Systems Research, Stanford University, Stanford, CA, May 1968.

[35] R.D. Shachter. Probabilistic inference and influence diagrams. *Operations Research*, 36:589–604, 1988.

[36] R.D. Shachter and D.E. Heckerman. Thinking backward for knowledge acquisition. *AI Magazine*, 8:55–63, 1987.

[37] E.H. Shortliffe. *Computer-based Medical Consultations: MYCIN*. North-Holland, New York, 1976.

[38] E.H. Shortliffe and B.G. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[39] M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base—part 1: The probabilistic model and inference algorithms. *Methods in Information and Medicine*, 1991. In press.

[40] C.S. Spetzler and C.S. Stael von Holstein. Probability encoding in decision analysis. *Management Science*, 22:340–358, 1975.

[41] M. Tribus. *Rational Descriptions, Decisions, and Designs*. Pergamon Press, New York, 1969.

[42] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[43] A. Tversky and D. Kahneman. Causal schemata in judgments under uncertainty. In D. Kahneman, P. Slovic, and A. Tversky, editors, *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, 1982.

[44] W. van Melle. *A domain-independent system that aids in constructing knowledge-based consultation programs*. PhD thesis, Department of Computer Science, Stanford University, 1980. Report STAN-CS-80-820 and HPP-80-22. Reprinted as van Melle, 1981.

[45] W. van Melle. *System Aids in Constructing Consultation Programs*. UMI Research Press, Ann Arbor, MI, 1981.

[46] S. Wright. Correlation and causation. *Journal of Agricultural Research*, 20:557–85, 1921.

[47] V.L. Yu, B.G. Buchanan, E.H. Shortliffe, S.M. Wraith, R. Davis, A.C. Scott, and S.N. Cohen. Evaluating the performance of a computer-based consultant. *Computer Programs in Biomedicine*, 9:95–102, 1979.

[48] V.L. Yu, L.M. Fagan, S.M. Wraith, W.J. Clancey, A.C. Scott, J.F. Hannigan, R.L. Blum, B.G. Buchanan, and S.N. Cohen. Antimicrobial selection by a computer: A blinded evaluation by infectious disease experts. *Journal of the American Medical Association*, 242:1279–1282, 1979.