On the Expressiveness of Rule-based Systems for Reasoning with Uncertainty

David E. Heckerman and Eric J. Horvitz

Medical Computer Science Group
Knowledge Systems Laboratory
Stanford University
Stanford, California 94305

Abstract

We demonstrate that classes of dependencies among beliefs held with uncertainty cannot be represented in rule-based systems in a natural or efficient manner. We trace these limitations to a fundamental difference between certain and uncertain reasoning. In particular, we show that beliefs held with certainty are more *modular* than uncertain beliefs. We argue that the limitations of the rule-based approach for expressing dependencies are a consequence of forcing non-modular knowledge into a representation scheme originally designed to represent modular beliefs. Finally, we describe a representation technique that is related to the rule-based framework yet is not limited in the types of dependencies that it can represent.

I Introduction

Original research on expert systems relied primarily on techniques for reasoning with propositional logic. Popular approaches included the rule-based and frame-based representation frameworks. As artificial intelligence researchers extended their focus beyond deterministic problems, the early representation methods were augmented with techniques for reasoning with uncertainty. Such extensions left the underlying structure of the representations largely intact.

In this paper, we examine the rule-based approach to reasoning with uncertainty. Within this context, we describe a fundamental difference between beliefs which are uncertain and beliefs which are held with certainty in the sense of monotonic propositional logic. In particular, we show that beliefs which are certain are more *modular* than uncertain beliefs. We demonstrate that because of this difference, simple augmentations to the rule-based approach are inadequate for reasoning with uncertainty.

We exhibit this inadequacy in the context of the MYCIN certainty factor model [Shortliffe 75], an adaptation to the rule-based approach for reasoning with uncertainty which has seen widespread use in expert systems research. We show that this adaptation does not have the *expressiveness* necessary to represent certain classes of dependencies that can exist among beliefs held with uncertainty. After demonstrating the limitations of the certainty factor model, we describe a representation technique called *belief networks* that is not similarly limited in its ability to express uncertain relationships among propositions.

II The MYCIN certainty factor model

In this section, we describe the aspects of the MYCIN certainty factor model that are central to our discussion. The knowledge in MYCIN is stored in rules of the form "IF E THEN H" where E denotes a piece of evidence for hypothesis H. A *certainty factor* is attached to each rule that represents the *change in belief* in the hypothesis given the piece of evidence for the hypothesis. Certainty factors range between -1 and 1. Positive numbers correspond to an *increase* in the belief in a hypothesis while negative quantities correspond to a *decrease* in belief. It is important to note that certainty factors do not correspond to measures of *absolute* belief This distinction, with respect to certainty factors as well as other measures of uncertainty, has often been overlooked in the artificial intelligence literature [Horvitz 86].

We will sometimes use the following notion to represent the rule "IF E THEN H":

$$E \xrightarrow{CF(H,E)} H$$

where CF(H,E) is the certainty factor for the rule. In the certainty factor model, as in any rule-based framework, multiple pieces of evidence may bear on the same hypothesis and a hypothesis may serve as evidence for yet another hypothesis. The result is a network of rules such as the one shown in Figure 1. This structure is called an *inference net* [Duda 76].

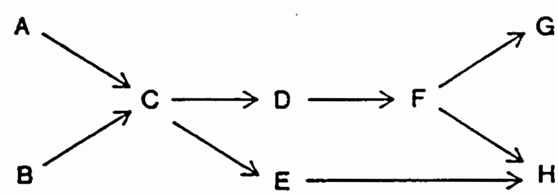


Figure 1: An inference net

The certainty factor model contains a prescription for propagating uncertainty through such an inference net. For example, the CF model can be used to compute the change in belief in hypotheses G and H when A and B are true (see Figure 1).

In this paper, we will focus on two types of propagation, *parallel combination* and *divergent propagation*. Parallel combination occurs when two or more pieces of evidence impinge on a single hypothesis, as shown in Figure 2(a). In this case, the certainty factors on two rules are combined with the *parallel combination function* to generate a certainty factor for the hypothetical rule "IF E_1 AND E_2 THEN H."[1] Divergent propagation occurs when one piece of evidence

[1] The combination function is given in the EMYCIN manual [EMYCIN 79].

bears on two or more hypotheses as shown in Figure 2(b). In this case, the updating of each hypothesis occurs independently. More generally, if two sub-nets diverge from a common piece of evidence, uncertainty is propagated in each sub-net independently.

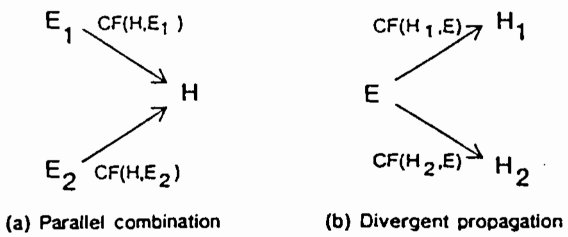


Figure 2: Two types of propagation in the CF model

III A fundamental difference

Let us now explore a fundamental difference between rules which represent deterministic relationships among propositions and rules which reflect uncertain relationships. Again consider the case of parallel combination as shown in Figure 2(a). Suppose the certainty factor for the rule involving E_1 is equal to 1. This corresponds to the situation where E_1 *proves* H with certainty. In this case, E_1 also proves H if E_2 is already known when E_1 is discovered. In other words, the certainty factor CF(H,E_1) does not depend on whether or not E_2 is known when the rule involving E_1 is invoked. Note we are assuming that deterministic beliefs are monotonic, a typical assumption of rule-based frameworks associated with schemes for reasoning with uncertainty. In contrast, suppose the certainty factor for this same rule lies between -1 and 1. This corresponds to the situation where E_1 potentially updates the belief in H but does not prove or disprove the hypothesis. In this case, it is reasonable to expect that the certainty factor for the rule may depend on the degree of belief assigned to E_2 when the rule is invoked.

The above is an instance of a fundamental difference between rules that are certain and those that are not. We say that deterministic or logical rules are *modular* while rules reflecting an uncertain relationship are *non-modular*. We use the term modular to emphasize that rules which are certain stand alone; the truth or validity of a deterministic rule does not depend upon beliefs about other propositions in the net. As mentioned above, the modularity of deterministic rules is a consequence of the assumption of monotonicity. We introduce the term modularity in lieu of monotonicity because we do not wish to confuse the notion of non-modularity, a concept we apply to uncertain beliefs, with non-monotonicity, a concept traditionally reserved for beliefs that are held with certainty.

That indeterministic rules are less modular than deterministic rules is also demonstrated in the case of divergent propagation (see Figure 2(b)). In particular, if E proves H_1 with certainty when the status of H_2 is unknown, then E will also prove H_1 with certainty when H_2 is known to be true or false. However, if E does not prove or disprove H_1 conclusively, the certainty factor for the rule involving H_1 may depend upon the belief assigned to H_2 at the time E is discovered.

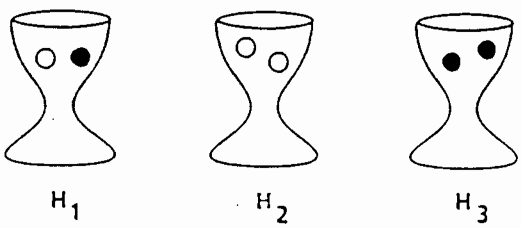
IV Limitations of the rule-based representation

As a result of this fundamental difference concerning modularity, there are certain classes of dependencies that cannot be represented in a natural or efficient manner within the rule-based framework.[2] In this section, we examine two such classes, termed mutual exclusivity and multiple causation, which occur commonly in real-world domains.

A set of hypotheses is said to be *mutually exclusive* and exhaustive when one and only one hypothesis from the set is true. We examine the case where two or more pieces of evidence are relevant to a set of three or more mutually exclusive and exhaustive hypotheses. We will show that parallel combination cannot be used to efficiently represent this situation. *Multiple causation* occurs when a piece of evidence has two or more independent causal or explanatory hypotheses. We will show that divergent propagation cannot be used to efficiently represent this situation. Rather than present proofs of these results which can be found elsewhere [Heckerman 86], we will present examples that facilitate an intuitive understanding of the limitations.

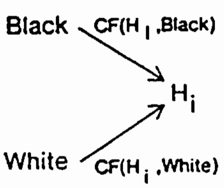
A. Mutual exclusivity

To illustrate a difficulty with representing mutual exclusivity in a rule-based framework, consider an example that harkens back to simpler days. Suppose you are given one of three opaque jars containing mixtures of black licorice and white peppermint jelly beans:



The first jar contains one white jelly bean and one black jelly bean, the second jar contains two white jelly beans, and the third jar contains two black jelly beans. You are not allowed to look inside the jar, but you are allowed to draw beans from the jar, one at a time, *with* replacement. That is, you must replace each jelly bean you draw before sampling another. Let H_i be the hypothesis that you are holding the *ith* jar. As you are told that the jars were selected at random, you believe each H_i is equally likely before you begin to draw jelly beans.

It seems natural to represent this situation with the following rules for each hypothesis H_i:



That is, each time a black jelly bean is observed, the belief in each hypothesis is revised using the certainty factors CF(H_i, Black) in conjunction with the parallel combination rule. Beliefs are similarly revised for each white jelly bean observed.

[2] We note that some classes of dependencies can be represented efficiently. In another paper [Heckerman 86], several of these classes are identified in a probabilistic context. It is shown that if relationships among propositions satisfy certain strong forms of *conditional independence*, then these relationships are accommodated naturally by the rule-based framework. Unfortunately, such conditions are rarely met in practice.

Unfortunately, such a representation is not possible because the modularity of rules imposed by parallel combination is too restrictive. To see this, suppose a black jelly bean is selected on the first draw. In this case, the belief in H_3 increases, the belief in H_2 decreases to complete falsity, while the belief in H_1 remains relatively unchanged. Thus, the certainty factor for the rule "IF Black THEN H_1" is close to zero.[3] In contrast, suppose a black jelly bean is selected following the draw of a white jelly bean. In this case, the certainty factor for the rule "IF Black THEN H_1" should be set to 1 as H_1 is confirmed with certainty. As only one certainty factor can be assigned to each rule, it is clear that the above representation fails to capture the dependencies among beliefs inherent in the problem.

This result can be generalized. It has been shown that parallel combination cannot be used to represent the situation where two or more pieces of evidence bear on a hypothesis which is a member of a set of 3 or more mutually exclusive and exhaustive hypotheses [Johnson 86, Heckerman 86].

We should mention that the above problem can be forced into the rule-based framework. For example, it can be shown that the following set of rules accurately represents the situation for H_1.

IF		1st draw Black	THEN	H_1. 0
IF		1st draw White	THEN	H_1. 0
IF	AND	1st draw Black Current draw Black	THEN	H_1. -.5
IF	AND	1st draw Black Current draw White	THEN	H_1. 1
IF	AND	1st draw White Current draw Black	THEN	H_1. 1
IF	AND	1st draw White Current draw White	THEN	H_1. -.5

Unfortunately, this representation is inefficient and awkward. The simplicity of the underlying structure of the problem is lost.

We note that there are even more pathological examples of inefficient representation. For example, if we add another white and black jelly bean to each jar in the above problem, it can be shown that the number of rules required to represent *N* draws is greater than *N*.

B. Multiple causation

In discussing another limitation of the rule-based framework, let us move from the simple world of jelly beans to a more captivating situation. Consider the following story from Kim and Pearl [Kim 83]:

> Mr. Holmes received a telephone call from his neighbor notifying him that she heard a burglar alarm sound from the direction of his home. As he was preparing to rush home, Mr. Holmes recalled that last time the alarm had been triggered by an earthquake. On his way driving home, he heard a radio newscast reporting an earthquake 200 miles away.

[3] In another paper [Heckerman 87], a method for calculating numerical values for certainty factors is described. The results of this method are consistent with the intuitive results presented here.

It seems natural to represent this situation with the inference net shown in Figure 3. However, a problem arises in trying to assign a certainty factor to the rule "IF Alarm THEN Burglary." Had Mr. Holmes not heard the radio announcement, the alarm sound would have strongly supported the burglary hypothesis. However, since Mr. Holmes heard the announcement, support for the burglary hypothesis is diminished because the earthquake hypothesis tends to "explain away" the alarm sound. Thus, it is necessary to attach two certainty factors to the same rule; one for the case where Mr. Holmes hears the announcement and another for the case where he does not. As only one certainty factor can be assigned to each rule, the inference net in Figure 3 fails to capture the situation.

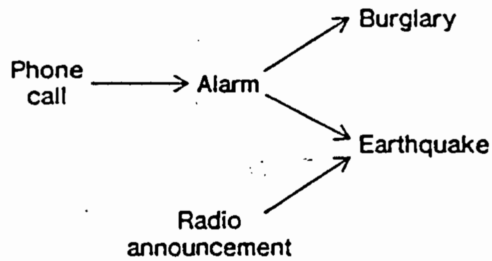


Figure 3: An inference net for Mr. Holmes' situation

The problem of Mr. Holmes can be generalized. It has been shown that divergent propagation cannot be used to represent the case where a single piece of evidence is *caused* by two explanatory hypotheses if either of these hypotheses can be updated with independent evidence [Heckerman 86].

As in the jelly bean problem, Mr. Holmes' situation can be forced into a rule-based representation. For example, the case can be represented by writing a rule for almost every possible combination of observations.

IF	AND	Phone call Announcement	THEN	Burglary, .1
IF	AND	Phone call No announcement	THEN	Burglary, .8
IF	AND	No phone call Announcement	THEN	Burglary, -.07
IF	AND	No phone call No announcement	THEN	Burglary, -.05
IF		Announcement	THEN	Earthquake, 1
IF	AND	Phone call No announcement	THEN	Earthquake, .01
IF	AND	No phone call No announcement	THEN	Earthquake, -.01

As in the previous problem, however, this representation is undesirable. In particular, the underlying causal relationships among the propositions are completely obscured. Moreover, the representation will become inefficient as the problem is modified to include additional pieces of evidence. For example, suppose the radio announcement is garbled and Mr. Holmes makes use of many small clues to infer that an earthquake is likely to have occurred. In this case, the number of rules required would be an exponential function of the number of clues considered.

V A more appropriate representation

We will now describe a representation technique that is closely related to the rule-based framework yet is not limited in the types of dependencies among propositions that it can represent. The representation, termed ***belief networks***, has recently become a focus of investigation within the artificial intelligence community [Pearl 86].[4] After briefly describing belief networks, we will show how the examples discussed above are represented within the methodology. We will then define a weaker notion of modularity that is more appropriate for uncertain knowledge in the context of belief networks. Finally, we will show how this weaker notion of modularity can facilitate efficient knowledge base maintenance.

A belief network is a two-level structure. The upper level of a belief network consists of a directed acyclic graph that represents the uncertain *variables* relevant to a problem as well as the relationships among the variables. Nodes (circles) are used to represent variables and directed arcs are used to represent dependencies among the variables. The bottom level represents all possible *values* or *outcomes* for each uncertain variable together with a *probability distribution* for each variable. The arcs in the upper level represent the notion of probabilistic conditioning. In particular, an arc from variable A to variable B means that the probability distribution for B may depend on the values of A. If there is no arc from A to B, the probability distribution for B is independent of the values of A.

To illustrate these concepts, consider once again the jelly bean problem. An belief network for this problem is shown in Figure 4.

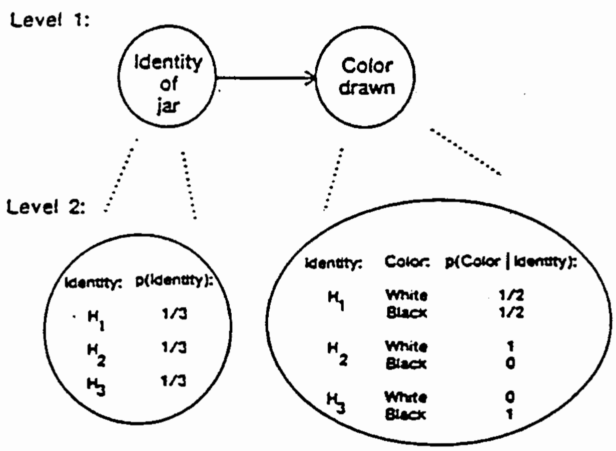


Figure 4: A belief network for the jelly bean problem

The two nodes labeled "Identity of jar" and "Color drawn" in the upper level of the belief network represent the uncertain variables relevant to the problem. The tables in the lower level list the possible values for each variable. The arc between the two nodes in the upper level means that the probability distribution for "Color drawn" depends on the value of "Identity of jar." Consequently, the probability distribution for "Color drawn" given in the second level of the network is conditioned on each of the three possible values of "Identity of jar": H_1, H_2, and H_3. Since there are no arcs into the "Identity of jar" node, an unconditional or *marginal* distribution for this variable is given.

Note that the same jar problem can be represented by a belief network with the arc *reversed*. In this case, an unconditional probability distribution would be assigned to "Color drawn" and a conditional probability distribution would be assigned to "Identity of jar." This highlights a distinction between inference nets and belief networks. Inference networks require dependencies to be represented in the evidence-to-hypothesis direction. In a belief network, dependencies may be represented in whatever direction the expert is most comfortable.[5]

As discussed earlier, it is difficult to represent this situation in an inference net because the three hypotheses reflecting the identity of the jar are mutually exclusive. In a belief network, however, this class of dependency is represented naturally. Rather than attempt to list each hypothesis in the upper level, these mutually exclusive hypotheses are moved to the second level and are considered together under the single variable, "Color drawn."

Now let us reexamine the story of Mr. Holmes to see how a belief network can be used to represent multiple causation. The upper level of a belief network for Mr. Holmes' situation is shown in Figure 5.

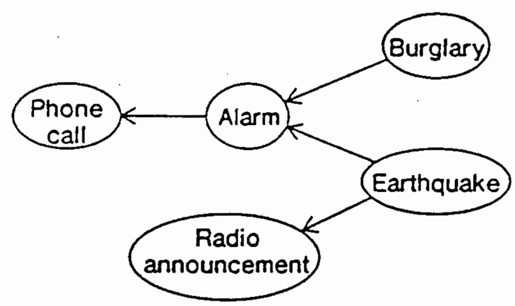


Figure 5: A belief network for Mr. Holmes' situation

The lower level of the belief network contains value lists and probability distributions as in the previous problem. For example, associated with the nodes "Phone call," "Alarm," "Burglary," and "Earthquake" are the value lists {Received, Not received}, {Sounded, Not sounded}, {True, False}, and {True, False} respectively. Associated with the node "Phone call" are the two probability distributions p(Phone call | Alarm = Sounded) and p(Phone call | Alarm = Not sounded).[6]

As mentioned earlier, an inference net cannot represent this situation in a straightforward manner because there are two causes affecting the same piece of evidence. However, this dependency is represented naturally in a belief network. In this example, the dependency is reflected in the probability distributions for "Alarm." In particular, a probability distribution for each combination of the values of the two variables "Burglary" and "Earthquake" is associated with the "Alarm" variable. That is, the following probability

[4] We note that the *influence diagrams* of Howard [Howard 81] and the *probabilistic causal networks* of Cooper [Cooper 84] are closely related to belief networks.

[5] Typically, the direction of arcs in belief networks reflect causal relationships [Pearl 86, Shachter 87].

[6] Note that we are using a short-hand notation for probability distributions. For example, p(Phone call | Alarm = sounded) is an abbreviation for the two probabilities p(Phone call = Received | Alarm = sounded) and p(Phone call = Not received | Alarm = sounded).

distributions will be included in the lower level of the belief network:

p(Alarm | Burglary=False AND Earthquake=False)

p(Alarm | Burglary=False AND Earthquake=True)

p(Alarm | Burglary=True AND Earthquake=False)

p(Alarm | Burglary=True AND Earthquake=True).

The interaction between the "Burglary," "Earthquake," and "Alarm" variables is completely captured by these probability distributions.

The above example points out that the representation of dependencies in a belief network does not come without increased costs. In particular, additional probabilities must be assessed and computational costs will increase. However, these costs are *no greater* and typically less than costs incurred in attempting to represent the same dependencies within the rule-based representation. For example, in the case of the garbled radio announcement discussed above, the belief network approach will generally not suffer the same exponential blow-up which occurs in the inference net representation.

VI A weaker notion of modularity

Notice that many of the nodes in Figure 5 are not directly connected by arcs. The missing arcs are interpreted as statements of *conditional independence*. For example, the absence of a direct arc between "Burglary" and "Phone call" indicates "Burglary" influences "Phone call" only through its influence on "Alarm." In other words, "Burglary" and "Phone call" are conditionally independent given "Alarm." This would not be true if, for example, Mr. Holmes believed his neighbor might be the thief. Thus, belief networks provide a flexible means by which a knowledge engineer or expert can represent assertions of conditional independence.

The concept that a variable may depend on a subset of other variables in the network is the essence of a weaker notion of modularity more appropriate for representing uncertain relationships. In this section, we define this concept more formally.

To define weak modularity, we first need several auxiliary definitions:

1. A node j is a *direct predecessor* of node i if there is an arc from j to i.
2. A node k is a *successor* of node i if there is a directed path from i to k.
3. P_i is the set of all direct predecessors of i.
4. S_i is the set of all successors of i.
5. S'_i is the complement of the set of all successors of i excluding i.

For example, in the belief network for Mr. Holmes' situation,

$P_{\text{Phone call}}$ = {Alarm}

$S_{\text{Phone call}} = \emptyset$

$S'_{\text{Phone call}}$ = {Alarm, Burglary, Earthquake, Radio announcement}

Now consider a particular node i. The conditional independence assertion associated with this node is

$$p(i|S'_i) = p(i|P_i). \quad (1)$$

Relation (1) says that if the outcomes of the direct predecessors of a node i are known with certainty, then the probability distribution for node i is independent of all nodes that are not successors of node i. Thus, whenever an arc is omitted from a non-successor of node i to node i, an assertion of conditional independence is being made. It is important to remember that such assertions are under the control of the knowledge engineer or expert. For example, in the belief network for Mr. Holmes, arcs from "Burglary," "Earthquake," and "Radio announcement" to "Phone call" are omitted because it is believed that "Phone call" is independent of these variables once the status of "Alarm" is known.

We identify relation (1) as a weaker notion of modularity more appropriate for uncertain reasoning. Note that (1) is a *local* notion of modularity: assertions of conditional independence are made about each variable individually. This is in contrast with the modularity associated with inference nets where straightforward representation of uncertain relationships requires global assumptions of independence [Heckerman 86].

VII Knowledge maintenance in a belief network

As a result of (1) and the fact that the graph component of a belief network is acyclic, it is not difficult to show that the probability distributions found at the second level of the belief network are all that is needed to construct the full *joint probability distribution* of the variables in the network [Shachter 86]. Formally, if a belief network consists of *n* uncertain variables $i_1, i_2, \ldots, $ and i_n, then

$$p(i_1 \text{ AND } i_2 \ldots i_n) = \prod_m p(i_m|P_{i_m}) \quad (2)$$

where $p(i_m|P_{i_m})$ is the probability distribution associated with node i_m at the second level of the belief network.

As the joint probability distribution for a given problem implicitly encodes *all* information relevant to the problem, property (2) can be used to simplify the task of modifying a belief network. To see this, imagine that a proposition is added to a belief network. When this occurs, the expert must first reassess the dependency structure of the network. For example, the new node may be influenced by other nodes, may itself influence other nodes, or may introduce conditional independencies or conditional dependencies among nodes already in the network. Then, in order to construct the new joint probability distribution, the expert need only reassess the probability distribution for each node which had its *incoming arcs* modified. Given (2), there is no need to reassess the probability distributions for the nodes in the network whose incoming arcs were not modified. Similarly, if a proposition is deleted from a belief network, the expert must first reassess dependencies in the network and then reassess only the probability distributions for those nodes which had their incoming arcs modified.

To illustrate this point, consider the following modification to Mr. Holmes' dilemma:

> Shortly after hearing the radio announcement, Mr. Holmes realizes that it is April first. He then recalls the April fools prank he perpetrated on his neighbor the previous year and reconsiders the nature of the phone call.

With this new information, an "April fools" node should be added to the belief network and a conditioning arc should be added from the new node to "Phone call" (see Figure 6). No other arcs need be added. For example, the arc from "April fools" to "Radio announcement" is absent reflecting the belief that radio announcers take their jobs somewhat seriously. The arc from "April fools" to "Burglary" is absent because it is assumed that burglars don't observe this holiday. The absence of an arc from "April fools" to "Earthquake" reflects certain beliefs about the supernatural.

Given the new graph, we see from (2) that the following probability distributions are needed to construct the new joint probability distribution:

```
p(Phone call | April fools AND Alarm)

p(Alarm | Burglary AND Earthquake)

p(Radio announcement | Earthquake)

p(Burglary)

p(Earthquake)
```

Fortunately, all nodes except for "Phone call" have retained the same predecessor nodes and so, by (2), the probability distributions corresponding to these nodes are available from the old belief network (see Figure 5). Only the probability distribution for "Phone call" needs to be reassessed.

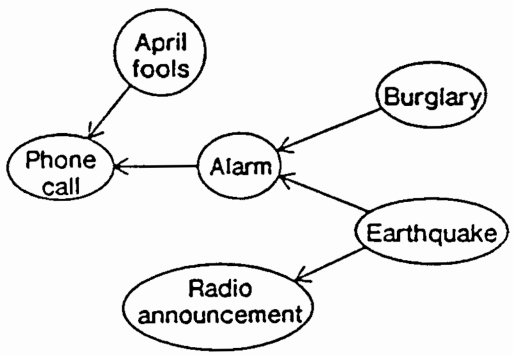


Figure 6: Mr. Holmes revisited

The modification above should be compared with the modification required in a rule-based framework. Because divergent propagation cannot be used to represent multiple causation in this framework, we are limited to an unnatural representation such as constructing a rule for each possible combination of observations. In this representation, modification to include the consideration of April fools results in a *doubling* in the number of rules. Thus, it is clear that the local modularity property associated with belief networks can help to reduce the burden of knowledge base maintenance.

VIII Summary

In this paper, we demonstrated that particular classes of dependencies among uncertain beliefs cannot be represented in the certainty factor model in an efficient or natural manner. We should emphasize that, to our knowledge, all uncertainty mechanisms designed as incremental extensions to the rule-based approach suffer similar limitations. Also, we identified a fundamental difference between reasoning with beliefs that are certain and reasoning with beliefs that are uncertain. We demonstrated that rules representing deterministic relationships between evidence and hypothesis are more *modular* than rules reflecting uncertain relationships. We showed that the limitations of the rule-based approach for representing uncertainty is a consequence of forcing non-modular knowledge into a representation scheme designed to represent modular beliefs. Finally, we described belief networks, a representation scheme related to the rule-based approach. We believe artificial intelligence researchers will find belief networks an expressive representation for capturing the complex dependencies associated with uncertain knowledge.

Acknowledgements

We wish to thank Judea Pearl and Peter Hart for discussions concerning divergence. We also thank Greg Cooper, Michael Wellman, Curt Langlotz, Edward Shortliffe, and Larry Fagan for their comments. Support for this work was provided by the NASA-Ames, the National Library of Medicine under grant R01-LM04529, the Josiah Macy, Jr. Foundation, the Henry J. Kaiser Family Foundation, and the Ford Aerospace Corporation. Computing facilities were provided by the SUMEX-AIM resource under NIH grant RR-00785.

References

[Cooper 84] Cooper, G. F. *NESTOR: A Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge.* Ph.D. Th., Computer Science Department, Stanford University, Nov. 1984. Rep. No. STAN-CS-84-48. Also numbered HPP-84-48.

[Duda 76] Duda, R., Hart, P., and Nilsson, N. Subjective Bayesian Methods for Rule-based Inference Systems. Proceedings 1976 National Computer Conference, AFIPS, 1976, pp. 1075-1082.

[EMYCIN 79] Van Melle, W. EMYCIN Manual. Stanford, 1979.

[Heckerman 86] Heckerman, D.E. Probabilistic Interpretations for MYCIN's Certainty Factors. In *Uncertainty in Artificial Intelligence*, Kanal, L. and Lemmer, J., Eds., North Holland, 1986.

[Heckerman 87] Heckerman, D.E., and Horvitz, E. J. The Myth of Modularity in Rule-Based Systems. In *Uncertainty in Artificial Intelligence*, Lemmer, J. and Kanal, L., Eds., North Holland, 1987.

[Horvitz 86] Horvitz, E. J., and Heckerman, D. E. The inconsistent Use of Measures of Certainty in Artificial Intelligence Research. In *Uncertainty in Artificial Intelligence*, Kanal, L. and Lemmer, J., Eds., North Holland, 1986.

[Howard 81] Howard, R. A., Matheson, J. E. Influence Diagrams. In *Readings on the Principles and Applications of Decision Analysis*, Howard, R. A., Matheson, J. E., Eds., Strategic Decisions Group, Menlo Park, CA, 1981, ch. 37, pp. 721-762.

[Johnson 86] Johnson, R. Independence and Bayesian updating methods. In *Uncertainty in Artificial Intelligence*. Kanal, L. and Lemmer, J., Eds., North Holland, 1986.

[Kim 83] Kim, J.H., and Pearl, J. A computational model for causal and diagnostic reasoning in inference engines. Proceedings 8th international joint conference on artificial intelligence, IJCAI, 1983, pp. 190-193.

[Pearl 86] Pearl, J. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence 29*, 3 (September 1986), 241-288.

[Shachter 86] Shachter, R.D. Intelligent probabilistic inference. In *Uncertainty in Artificial Intelligence*, Kanal, L. and Lemmer, J., Eds., North Holland, 1986.

[Shachter 87] Shachter, R., and Heckerman, D. A backwards view for assessment. In *Uncertainty in Artificial Intelligence*, Lemmer, J. and Kanal, L., Eds., North Holland, 1987.

[Shortliffe 75] Shortliffe, E. H. and Buchanan, B. G. A model of inexact reasoning in medicine. *Mathematical Biosciences 23* (1975), 351-379.