THE INCONSISTENT USE OF MEASURES OF CERTAINTY IN ARTIFICIAL INTELLIGENCE RESEARCH*

Eric Horvitz and David Heckerman

Medical Computer Science Group, Knowledge Systems Laboratory
Departments of Medicine and Computer Science, Stanford, California 94305

There has been a great deal of interest among artificial intelligence researchers in methodologies for using subjective measures of belief for reasoning about uncertain events. We emphasize the fundamental difference between the use of measures of *absolute* belief and measures of *change* in belief. We show that expert system investigators have elicited measures of absolute belief and have used them as though they were measures of change in belief. We argue that such use is counterintuitive and formally inconsistent. In our discussion of the inconsistency, we introduce a belief-updating paradigm used frequently in plausible reasoning. Finally, we attempt to characterize the potential impact of the inconsistency.

1. INTRODUCTION

Most work in reasoning about uncertainty has centered on the manipulation of measures of *absolute* belief. That is, most methods for managing uncertainty concern themselves with questions of the following form: Given current evidence, how certain is some hypothesis? However, there has also been investigation into the use of measures of *change* in belief. Here, questions are of the form: Given a piece of evidence, how has the certainty of some hypothesis changed? Expert knowledge about the uncertain relations among evidence and hypotheses can be elicited and encoded as absolute beliefs or as changes in belief. The distinction between reasoning with quantities representing changes in belief and reasoning with quantities representing absolute belief is significant in the design and characterization of methods for handling uncertainty in reasoning systems. Unfortunately, the fundamental difference between the nature of measures of absolute belief and change in belief tends to be overlooked in research on reasoning under uncertainty.

We shall focus on problems with the misuse of measures of change in belief or *belief updates*. We will first briefly review the historical background of inattention to the difference between absolute belief and belief updates. We will then introduce fundamental intuitive properties of the two measures and formally show that it is important to distinguish one from the other. Next, we will define and formalize the concept of *modularity* and describe a plausible reasoning paradigm used for reasoning under uncertainty in a number of expert systems. We shall examine the use of the reasoning scheme in two representative expert systems and argue that expert system engineers have elicited measures of absolute belief from experts and have used them as though they were modular belief updates. Finally, we explore the consequences of this misuse.

*This work was supported in part by the Josiah Macy, Jr. Foundation, the Henry J. Kaiser Family Foundation, and the Ford Aerospace Corporation. Computing facilities were provided by SUMEX-AIM under NIH grant RR-00785. Order of authorship based on a coin flip.

2. DISTINGUISHING BELIEF UPDATES FROM ABSOLUTE BELIEFS

Techniques for reasoning under uncertainty have become increasingly central as artificial intelligence researchers have focused their attention on tasks of real world complexity such as medicine. We believe that the fundamental difference between measures of absolute belief and measures of belief update has not been adequately stressed in research on reasoning under uncertainty in artificial intelligence. We suggest that inattention to the difference has contributed to the inconsistent use of one measure for another.

Inattention to the distinction between encoding knowledge about uncertainty as changes in belief as opposed to absolute belief has led to a widespread misuse of measures of belief, especially among users of commercial products [1]. As we shall describe below, expert system engineers often elicit measures of absolute beliefs and use them as if they were belief updates. The artificial intelligence literature has tended to reinforce rather than correct the problem. For example, presentations of measures of a belief update termed *certainty factors* [2], often describe the quantities as absolute beliefs in discussions of experimental and commercially-available expert systems.

3. HISTORICAL BLURRING OF BELIEF AND BELIEF UPDATE

Inattention to the distinction between reasoning with absolute beliefs and reasoning with belief updates is not unique to the artificial intelligence community. Signs of the confusion have arisen in a number of arenas. A striking example comes from the philosophy literature. Philosophers of science have referred to research on problems with belief as the study of *confirmation* [3]. Discussions of confirmation have tended to confuse the degree to which belief in a hypothesis is confirmed or refuted by evidence with the absolute belief in the hypothesis.

Problems surrounding the relationship of measures of absolute belief to belief updates surfaced in discussions about confirmation in the mid-twentieth century as researchers began to rediscover the subjective interpretation of probabilities and to seek general theories of belief in hypotheses within the subjectivist framework. Whereas Bayes and Laplace proposed probability as a personal measure of belief over two hundred years ago, the early twentieth century was marked by the rise of the interpretation of probabilities as frequencies or averages of the occurrence of events over numerous identical trials. Although the frequency interpretation of probability is still quite popular, the early 1930's saw a "rediscovery" of probability as a measure of personal belief [4]. Rediscoveries of the subjective interpretation and critical reactions to the arguments of the frequentists bolstered the popularity of the subjective interpretation. Among others, Ramsey, Savage, DiFenetti and Carnap led this rediscovery [5].

Confusion about the difference between absolute belief and belief update seems to have been amplified by problems with terminology as philosophers grappled for language to describe subjective belief, frequencies and changes in belief. For example, Carnap originally used the phrase, *degree of confirmation* in his classic work, *Logical Foundations of Probability* [6], primarily to emphasize the difference between the subjective interpretation of probabilities from the classical frequency interpretation. Unfortunately, Carnap and others began to use *degree of confirmation* to represent two very different concepts. Whereas he used *degree of confirmation* to denote what is now commonly called subjective probability, he also used the phrase to refer to changes in belief. Much confusion had already proliferated by the time Carnap acknowledged his wavering use of terminology in the second of edition of his book [7].

Some philosophers noticed the inattention to the difference between absolute belief and belief update early on. For example, philosopher of science Karl Popper was provoked by problems with this distinction and worked to refute "...all those theories of induction which identify the degree to which a statement is supported or confirmed or corroborated by empirical tests with its degree of probability" [8].

In his attempt to stress the clear difference between a change in belief and absolute belief, Popper presented clear examples distinguishing the two notions. For example, he argued that even though a piece of evidence may confirm one hypothesis and disconfirm another, the resulting degree of belief in the confirmed hypothesis may still be less than the degree of belief in the disconfirmed hypothesis. [8] That is, there can exist hypotheses H_1 and H_2, new evidence E and prior information e such that:

$$p(H_1|Ee) > p(H_1|e) \quad \text{and} \quad p(H_2|Ee) < p(H_2|e)$$

but

$$P(H_1|Ee) < P(H_2|Ee).$$

The resulting belief in the disconfirmed hypothesis can still overwhelm the belief in the confirmed hypothesis because of dominating prior information. Popper identified several cases where there were problems with the distinction, including Carnap's inconsistent use of the word confirmation [8]. Popper went on to introduce *corroboration*, his term for a well-defined measure of change in belief.

Confusion between measures of absolute belief and belief update can be found at the root of many of the historical paradoxes of confirmation [9]. Some of these paradoxes led many to the conclusion that probability could not capture the essential aspects of confirmation. For example, Carnap, Barker [10], Harre [11] and others, in viewing confirmation as analogous to absolute probabilities dwelled on what seemed to be puzzling aspects of the relation of confirmation of a hypothesis to the confirmation of the negation of the hypothesis. In particular, it was noted that it is counterintuitive that the confirmation of a hypothesis should be equal to one minus the confirmation of the negation of the hypothesis [12].

Popper and others have pointed out that the theory of probability allows for measures of belief update and should therefore not be abandoned. In support of this view, Good [13] showed that several probabilistic quantities satisfied a slightly modified version of Popper's axioms for measures of change in belief. We shall return to this work below.

Debate and confusion in the mid-twentieth century based in terminological and conceptual difficulties about belief and belief updates may have been contributing factors to more recent inconsistencies surrounding the use of belief updates. The historical problems also suggest a tendency to blur the fundamental difference between absolute belief and belief update.

4. INTUITIVE PROPERTIES OF MEASURES OF BELIEF

We could develop a discussion of the inconsistency in plausible reasoning methodologies without resorting to formal arguments. However, we can make more powerful claims by formalizing what we mean by measures of absolute belief and belief update. In this section, we will outline a set of simple properties that define the concept of an absolute belief. We will then discuss the consequences of this definition. In the following section, we will similarly examine the concept of a belief update.

The first property of a measure of absolute belief pertains to the hypothesis or proposition to which a degree of belief can be assigned. The nature of a proposition has at times been the source of debate about alternate formalisms for plausible reasoning. We believe that a desired property of a proposition to which one may assign a degree of belief is that it be defined precisely enough so that its ultimate truth value is indisputable. Furthermore, it seems reasonable that a degree of belief can be assigned to *any* proposition which is precisely defined.

Now we examine the factors upon which a degree of belief should depend. Clearly, an individual's degree of belief in some proposition or hypothesis H should depend on the particular hypothesis itself. In addition, the degree of belief in some hypothesis should depend upon the current state of information of the individual possessing the belief. Therefore, we will use the term H|e to represent the degree of absolute belief for hypothesis H by an individual with information e.

We also wish measures of degree of belief in the truth of a hypothesis to be able to vary continuously between values of certain truth and certain falsehood. We assume that a single real number is sufficient to represent a degree of belief and assume that larger numbers represent larger beliefs.

Now consider two hypotheses, H_1 and H_2 which are logically equivalent. That is, it is possible to show that H_1 logically implies H_2 and vice-versa. In this case, we require that $H_1|e = H_2|e$ for any prior information e. In other words, if two hypotheses have the same truth value, an individual should believe each of them with equal conviction. For example, if XY denotes the proposition "X and Y," XY|e should equal YX|e.

The next desired property is that the belief in the proposition XY should be related to the belief in X alone as well as to the belief in Y given that X is true. Formally, measures of absolute belief should have the property that there exists some function F such that

$$XY|e = F(X|e, Y|Xe).$$

This intuitive property is related to what is often termed hypothetical reasoning. Individuals commonly assign belief to events conditioned on the truth of another. It seems clear that in such cases, the belief in the event of interest should depend on one's degree of belief in the event given the truth of some conditioning event as well as the degree of belief in the conditioning event itself. A technical requirement is that F be continuous and monotonic increasing in both arguments when the other is held constant.[1]

Finally, we wish that one's belief in the negation of a hypothesis H, denoted ~H, should be determined by one's belief in the hypothesis itself. Formally, there should be a function G such that

$$\sim H|e = G(H|e).$$

Another technical point is that G should be continuous and monotonic decreasing.

The above properties have a striking logical relationship to probability theory. First, it is straightforward to show that any monotonic transformation of probability satisfies the above properties. That is, any monotonic transformation of p(H|e) qualifies as a measure of absolute belief. Moreover, it can be shown that the above properties are *only* satisfied by p(H|e) or some monotonic transformation of this quantity. It can be shown that the above properties are enough to prove that there exists some monotonic function B such that

$$0 \leq B(H|e) \leq 1 \tag{1}$$

$$B(TRUE|e) = 1 \tag{2}$$

$$B(H|e) + B(\sim H|e) = 1 \qquad \text{(sum rule)} \tag{3}$$

$$B(XY|e) = B(X|e) \cdot B(Y|Xe) \qquad \text{(product rule)} \tag{4}$$

These relations are just the axioms of probability theory as commonly stated. Given the simple properties stated above, only probabilities or monotonic transformations of probabilities can serve as measures of absolute belief.

These results were proved by a physicist named Cox in 1946 [14]. It is interesting to note that Cox's proof was the first to formally show that measures of belief satisfying the simple properties listed above must satisfy the axioms of probability. In particular, Cox was the first to "prove" Bayes' theorem in the subjective framework. Several investigators [15, 16] similarly interested in fundamental aspects of measures of belief have since enumerated variants of the Cox work. They have all elegantly proved that the axioms of probability theory follow from the small set of intuitive properties for measures of absolute belief presented above.

The significance of this work is apparently not yet appreciated by many in the artificial intelligence community. We recommend the full proof by Cox to the reader. We use his argument to justify framing our discussion of the relationship of absolute belief to belief updates in terms of probability theory.

5. PROPERTIES OF BELIEF UPDATES

We shall now formally define what we mean by a measure of change in belief. As in the previous section, we will assert a set of properties that should be satisfied by a belief update, and show that certain forms of belief update follow.

First, let us examine the factors that will influence a belief update. Clearly, a belief update depends on the hypothesis being updated and the evidence responsible for the new update. As in the case of absolute beliefs, an update may also depend on an individual's current state of information. Therefore, we will use the term U(H,E,e) to denote a belief update for a hypothesis H given evidence E and prior information e. We will assume that H, E, and e are precisely defined as in the case of absolute beliefs.

There are several other properties of belief updates which closely resemble the properties of absolute beliefs. One is that a belief update U(H,E,e) can be represented by a single real number. Another is that if H_1 is logically equivalent to H_2 and if E_1 is logically equivalent to E_2 it follows that the two updates $U(H_1,E_1,e)$ and $U(H_2,E_2,e)$ should be equal.

Another property of belief updates explicitly relates a belief update to measures of absolute belief. Consider a hypothesis H and an item of evidence E for the hypothesis. The belief update U(H,E,e) along with the *prior* belief in H, H|e, should be sufficient to determine the *posterior* belief in H, H|Ee. Formally, we say that there is some function[2] f such that

$$H|Ee = f(U(H,E,e),\ H|e).$$

In the context of the above discussion on the necessity of probability as a measure of belief we can write

$$p(H|Ee) = f(U(H,E,e),\ p(H|e)). \tag{5}$$

We will call equation (5) the fundamental property of belief update. This definition of a belief update should be contrasted with the more general notion of *belief updating*. The concept of belief updating refers to the fact that, in general, one's beliefs change as new information is received. The more specific notion underlying the definition above is that the effect of information upon one's beliefs can be captured by single real number updates.

There are two additional properties of belief updates which are related to the combination of evidence. We will discuss these properties below after we introduce the notion of *modularity*.

As in the case of absolute beliefs, it can be shown that the above set of weak properties, in conjunction with properties associated with the combination of evidence that we shall present below, imply that a belief update must have a very special form. In particular, it can be shown that any quantity which satisfies these properties must necessarily be some monotonic transformation of the likelihood ratio, $p(E|He)/p(E|\sim He)$, denoted $\lambda(H,E,e)$. A proof of this assertion, with minor differences in notation, can be found in the appendix of the paper in this publication discussing probabilistic interpretations for MYCIN's certainty factors [17]. This is not the first proof of the necessity of the likelihood ratio, $\lambda(H,E,e)$ as a measure of belief update. Good showed, over twenty-five years ago, that $\lambda(H,E,e)$ is the only quantity which satisfies Popper's desiderata for degree of change in belief [13]. Popper's desiderata are similar to the properties discussed above. We will not repeat proofs that a likelihood ratio $\lambda(H,E,e)$ is the *only* measure of change in belief which satisfies the properties discussed above. However, in the next section we will show that $\lambda(H,E,e)$ *satisfies* the fundamental property of a belief update (5).

6. A PROBABILISTIC BELIEF UPDATE

The relationship between the likelihood ratio $\lambda(H,E,e)$ and absolute beliefs can be easily derived. First, consider Bayes' theorem for hypothesis H, evidence E, and prior evidence e:

$$p(H|Ee) = \frac{p(E|He)p(H|e)}{p(E|e)} . \tag{6}$$

If we divide the above equation by the corresponding formula for the negation of the hypothesis, ~H, we get

$$\frac{p(H|Ee)}{p(\sim H|Ee)} = \frac{p(E|He)}{p(E|\sim He)} \frac{p(H|e)}{p(\sim H|e)} . \tag{7}$$

In viewing the above equation from left to right, the first and last ratios are the posterior and prior *odds* respectively, written $O(H|E,e)$ and $O(H|e)$. We see that the second ratio is the likelihood ratio $\lambda(H,E,e)$. Therefore, we can rewrite (7) as

$$O(H|Ee) = \lambda(H,E,e)\ O(H|e). \tag{8}$$

This is called the *odds-likelihood* form of Bayes' theorem.

We see from equation (8) that the likelihood ratio $\lambda(H,E,e)$ relates the prior odds of H to the posterior odds. As the odds of an event is a monotonic transformation of the probability of an event, we see that the likelihood ratio λ satisfies equation (5), the fundamental property of updates. That is, the posterior belief in H is determined by λ and the prior belief.

7. INCONSISTENCY OF EQUATING ABSOLUTE BELIEFS WITH BELIEF UPDATES

Now that we have defined both absolute beliefs and belief updates and have examined the consequences of these definitions, we can discuss the general problem with blurring the distinction between absolute beliefs and belief updates. The odds-likelihood form of Bayes' theorem is helpful in analyzing the confusion. Again, consider equation (8):

$$O(H|Ee) = \lambda(H,E,e)\ O(H|e).$$

From this relation, we see that $O(H|Ee)$, a measure of absolute belief, can be viewed as containing two components: a prior belief and an update. Thus, if an absolute belief is mistakenly used instead of a belief update, the prior belief in hypothesis H is being implicitly ignored. The only time a belief update could be legitimately equated with an absolute belief is when the prior odds of the hypothesis in question happens to be 1. It is clear from (8) that it is possible to identify an absolute belief with a measure of update only in this particular situation.

8. EVIDENCE COMBINATION AND MODULARITY

Expert system plausible reasoning methodologies calculate measures of absolute belief in hypotheses by combining the impact of several pieces of evidence. In this section, we will consider the combination of the effects of multiple pieces of evidence. We will first discuss properties of updates in general with respect to combination. Our discussion of the combination will include an introduction of the notion of *modularity*, a property often assumed in the construction of plausible reasoning systems. Afterwards, we will focus on the likelihood ratio λ and show that it satisfies these properties.

Consider a single hypothesis H and a set ξ_H containing distinct pieces of evidence relevant to belief in hypothesis H. A belief update is associated with each item of evidence in this set. As an example, consider two pieces of evidence, E_1 and E_2, relevant to belief in hypothesis H. Furthermore, suppose that some subset e of ξ_H has already been used to update H.[3] That is, e represents prior evidence. If we first update with E_1, the update will take place given prior information e. If we then update with E_2, the update will take place given prior information e *and* E_1. Therefore, if we wish to combine updates, we must assert that there is some function g such that

$$U(H,E_1E_2,e) = g[U(H,E_1,e),\ U(H,E_2,E_1e)] \tag{9}$$

The only requirements on g is that it be continuous and that it be monotonic increasing in each argument when the other is held constant.[4]

Brief consideration of the above updating combination scheme uncovers a fundamental methodological problem with its use in general plausible reasoning: for complex applications, an intractable number of updates may have to be acquired and manipulated by an expert system as it is possible that a belief update associated with a hypothesis and a piece of evidence will be dependent on background evidence. That is

$$U(H,E,e_1) \neq U(H,E,e_2).$$

However, this intractability is removed if we assume that

$$U(H,E,e) = U(H,E,\emptyset) \tag{10}$$

for all subsets e of ξ_H. Note that, with this assumption, it is possible to write U as a function of only two arguments: U(H,E). We will say that updates with this property are *modular*; their value does not depend on prior evidence.

In reducing the complexity of inference and knowledge aquisition in expert system engineering, researchers have often implicitly assumed the modularity property; that is, *all* updates are considered modular.[5] The modularity property has often been assumed in the construction of expert systems which use plausible reasoning to avoid the consideration of possible complex interactions among evidence. In particular, the modularity assumption allows the addition and deletion of knowledge without detailed consideration of its impact on remaining portions of the knowledge base. Asserting the modularity property eases the task of knowledge acquisition and explanation, facilitates the construction and maintenance of knowledge bases and allows the use of relatively simple functions for belief updating.

The desirability of these features makes it quite tempting for a knowledge engineer to make use of the modularity assumption even when dependencies among events make its use inappropriate. Expert system builders have justified their assertions of modularity by suggesting that human experts may structure knowledge and manage complexity in particular domains by relying on or imposing modularity between events. Such simplification may indeed be supported by research in cognitive psychology, where severe limitations in the ability of humans to consider many elements and relations [18] in the short-term has been demonstrated. However, we fear that the need for managing complexity may apply more to the knowledge engineer grappling with an unfamiliar domain in the short-term than to the expert who has had long-term immersion within a discipline.

Let us explore modularity further in the context of the likelihood ratio $\lambda(H,E,e)$. It is straightforward to show that $\lambda(H,E,e)$ satisfies the combination property of updates, (9). Consider Bayes' theorem for two pieces of evidence:

$$\frac{p(H|E_1E_2e)}{p(\sim H|E_1E_2e)} = \frac{p(E_1E_2|He)}{p(E_1E_2|\sim He)}\frac{p(H|e)}{p(\sim H|e)} = \frac{p(E_2|HE_1e)}{p(E_2|\sim HE_1e)}\frac{p(E_1|He)}{p(E_1|\sim He)}\frac{p(H|e)}{p(\sim H|e)} \quad (11)$$

From the definition of $\lambda(H,E,e)$, it follows that

$$\lambda(H,E_1E_2,e) = \lambda(H,E_2,E_1e)\ \lambda(H,E_1,e). \quad (12)$$

Therefore, $\lambda(H,E,e)$ satisfies the combination property where combination is achieved by simple multiplication.

Now suppose

$$p(E|He) = p(E|H) \quad \text{and} \quad p(E|\sim He) = p(E|\sim H) \quad (13)$$

where e is any subset of ξ_H and where $E \notin e$. That is, the belief that evidence E will is occur is not influenced by the presence of other evidence if either H or ~H is known to be true. When (13) holds, it is said that the items of evidence in the set ξ_H are *conditionally independent*. With this assumption, it immediately follows from the definition of $\lambda(H,E,e)$ that

$$\lambda(H,E,e) = \lambda(H,E,\emptyset). \quad (14)$$

That is, when (13) holds, the likelihood ratio $\lambda(H,E,e)$ satisfies the modularity property.

The relationship between global modularity and conditional independence is quite intuitive. Suppose that there are conditionally dependent items of evidence for H. If the modularity property were to be satisfied, then these complex dependencies could be calculated with a simple function. It is unreasonable to expect such "something for nothing" behavior. Considering the contrapositive, it is reasonable to assert that the modularity property, (10), implies conditional independence of the form (13).

Earlier, we mentioned that the properties of belief updates discussed in section 5 and two additional assumptions concerning the combination of evidence imply that *any* belief update must be some monotonic transformation of λ. These additional assumptions are the combination property, (9) and the requirement that modularity imply conditional independence.

9. THE MODULAR UPDATING PARADIGM

When the modularity property holds, we saw that it is possible to define updates of only two arguments. In addition, when modularity holds, updates of two arguments can be combined in a simple manner. To see this, again consider the likelihood ratio λ. It follows from (14) and (12) that

$$\lambda(H,E_1 \ldots E_n) = \Pi_i \lambda(H,E_i). \tag{15}$$

That is, the update corresponding to the combined evidence is just the *product* of the individual updates.

As all measures of belief update must be some monotonic transformation of λ, it follows that any combination scheme for modular belief updates must either be identical to (15) or isomorphic to this combination rule. We shall refer to this methodology for belief the combination of evidence as the *modular updating paradigm*. This paradigm has seen widespread use in expert systems. In the remainder of the paper, we will focus on this updating paradigm.

One interesting isomorphic mapping of (15) is known as the "weight of evidence" scheme. Taking the logarithm of (15), we get

$$\log[\lambda(H,E_1 \ldots E_n)] = \Sigma_i \log[\lambda(H,E_i)]. \tag{16}$$

That is, the update corresponding to the combined evidence is just the *sum* of individual updates. Peirce identified this combination scheme in 1878. He referred to these modular updates as "weights of evidence" because they are added just as weights on a scale [19]. Peirce argued that $\log[\lambda]$ is a natural quantity for acquiring and making inferences with uncertainty. Several other researchers independently discovered this simple but powerful measure of change in belief. They include Turing [5], Good [20, 13], and Minsky and Selfridge [21].

10. INCONSISTENT USE OF THE MODULAR UPDATING PARADIGM

We now describe a pattern of misuse of the modular updating methodology and examine consequences of this misuse. Namely, it has been common in expert system engineering to elicit measure of absolute belief and combine them using the modular updating paradigm described above. The result is then interpreted as an absolute belief in the hypothesis given the combined evidence.[6]

Let us examine, in general, the consequence of this inconsistent use of belief measures. Suppose items of evidence $E_1 \ldots E_n$ have an impact on hypothesis H and the absolute beliefs $\log[O(H|E_i)]$ or some monotonic transformation of these quantities are elicited from an expert. If we combine these quantities as though they were the updates $\log[\lambda(H,E_i)]$ as in equation (16) and then interpret the combined result as an absolute quantity, we get

$$\log[O(H|E_1 \ldots E_n)] = \Sigma_i \log[O(H|E_i)].$$

Using the fact that

$$\log[O(H|E)] = \log[O(H)] + \log[\lambda(H,E)]$$

we get

$$\begin{aligned}\log[O(H|E_1 \ldots E_n)] &= \Sigma_i \log[O(H)] + \Sigma_i \log[\lambda(H,E_i)] \\ &= n \cdot \log[O(H)] + \Sigma_i \log[\lambda(H,E_i)].\end{aligned} \tag{17}$$

However, the expansion of $\log[O(H|E_1...E_n)]$ consistent with Bayes' theorem, assuming modularity, is

$$\log[O(H|E_1...E_n)] = \log[O(H)] + \Sigma_i \log[\lambda(H,E_i)]. \tag{18}$$

Comparing (17) and (18), it is easy to see that the inconsistent use of measures of belief has the consequence that the prior belief in H is overcounted n-1 times in the final belief of H. The posterior probability of a hypotheses will diverge from its correct value as updating proceeds. The posterior belief assigned to a hypothesis with a large prior belief will be amplified while the posterior belief assigned to a hypothesis with a small prior will be attenuated.

11. MODULAR BELIEF UPDATING IN MYCIN

We now turn our attention to two particular modular updating schemes that have used measures of belief inconsistently as described above. We shall first examine this problem in the MYCIN expert system for diagnosis of bacteremia and then look at the INTERNIST-1 expert system for diagnosis within internal medicine.

MYCIN, EMYCIN, and its descendants have reasoned with measures of changes in belief called certainty factors (CF's). MYCIN's knowledge is represented as *rules* in the form of "IF E THEN H" where H is a hypothesis and E is evidence having relevance to the hypothesis. A certainty factor is associated with each MYCIN rule. CF's were designed as measures of *change* in belief about a hypothesis given some evidence [22]. The quantities range between -1 and 1; positive numbers correspond to an *increase* in belief in a hypothesis while negative quantities correspond to a *decrease* in belief when a piece of evidence becomes available.

When more than one piece of evidence bears on a hypothesis, the MYCIN certainty factor model prescribes a method for combining updates. For two items of evidence with certainty factors x and y, the combined certainty factor, z, is given by the function

$$z = \begin{cases} x + y - xy & x,y \geq 0 \\ \dfrac{x + y}{1 - \min(|x|,|y|)} & x,y \text{ of opposite sign} \\ x + y + xy & x,y < 0 \end{cases} \tag{19}$$

With a little algebra, [17] it can be shown that the above combination function is isomorphic to the combination scheme for $\lambda(H,E)$ (15). In particular, the mapping

$$CF(H,E) = \begin{cases} (\lambda(H,E)-1) \;/\; \lambda(H,E) & \lambda \geq 1 \\ \lambda(H,E)-1 & \lambda < 1 \end{cases} \tag{20}$$

in conjunction with (15) generates the combination function (19).

While certainty factors were intended to represent measures of belief update, they were elicited from experts as absolute beliefs. In particular certainty factors were elicited from experts with the phrase "On a scale of 1 to 10, how much certainty do you affix to this conclusion?" [2] Therefore, the updating scheme used by MYCIN is equivalent to the inconsistent scheme presented in the previous section. In particular, the updating procedure in MYCIN is overcounting priors.

We should note that the creators of the certainty factor model provided a justification for the elicitation of certainty factors as absolute beliefs. This justification was based on their definition of the certainty factor quantity:

$$CF(H,E) = \begin{cases} \dfrac{p(H|E) - p(H)}{1 - p(H)} & p(H|E) > p(H) \\ \\ \dfrac{p(H|E) - p(H)}{p(H)} & p(H) > p(H|E) \end{cases} \qquad (21)$$

This definition was developed in part to stress that certainty factors were belief updates. The creators of certainty factors noted that when p(H) approaches 0 and when p(H|E) is not too close to 0, CF(H,E) is approximately equal to p(H|E) given the above definition.

In light of previous discussions, the above justification must be flawed. The flaw is based on the inadequacy of the original CF definition for representing a valid measure of belief update. The creators of certainty factors assumed that the definition of certainty factors (21) is consistent with the combination function (19). Several arguments can be used to show the inconsistency. For example, the definition prescribes non-commutative combination of evidence although the combination function is commutative. More fundamentally, it can be shown that the definition (21) is not equal to any monotonic transformation of the likelihood ratio λ.

The MYCIN system designers were not especially concerned with the theoretical soundness of their methodology for plausible reasoning. The group stressed the pragmatics of designing a system that performed well within the targeted domain. Unfortunately, versions of the original certainty factor model for reasoning under uncertainty have been used in a wide variety of domains in which their performance has not been validated.

12. MODULAR BELIEF UPDATING IN INTERNIST-1

INTERNIST-1 is the result of a major research project to build an expert system for internal medicine diagnosis [23]. The system uses an *ad hoc* methodology for reasoning under uncertainty. As the measures of belief and plausible reasoning methodology used in INTERNIST-1 have been defined less-formally than the MYCIN certainty factor model, it is more difficult to make strong claims about inconsistencies within the system. Nonetheless, it seems that the method for managing uncertainty in INTERNIST-1 is associated with difficulties similar to MYCIN's. In particular, the creators of INTERNIST-1 elicited measures of absolute belief but combined them using the modular updating paradigm.

The absolute measure of belief to which we refer is termed an *evoking strength* by the INTERNIST-1 researchers. The evoking strength ES(H,E) was informally defined to be the answer to the question: Given a patient with manifestation E, how strongly should I consider diagnosis H to be its explanation? [23] Evoking strengths can take on one of 6 values. The meaning of each value is given in table 1 below. These informal definitions suggest that ES(H,E) is a measure of absolute belief. Indeed, evoking strengths have been described as being "somewhat analogous to a posterior probability." [24]

However, evoking strengths for a disease hypothesis are combined using the modular updating paradigm. In particular, evoking strengths for a particular hypothesis are combined through addition. The resulting combined evoking strength determines, in part, the "score" for a disease. The score is interpreted as a measure of absolute belief that the disease is present.

Evoking strength:

0: *Nonspecific; manifestation occurs too commonly to be used to construct a differential diagnosis.*

1: *The diagnosis is a rare or unusual cause of the listed manifestation.*

2: *The diagnosis causes a substantial minority of instances of the listed manifestation.*

3: *The diagnosis is the most common but not the overwhelming cause of the listed manifestation.*

4: *The diagnosis is the overwhelming cause of the listed manifestation.*

5: *The listed manifestation is pathognomonic for (implicates) the disease.*

Table 1: Definitions of INTERNIST-1 Evoking Strengths

It appears that the INTERNIST-1 reasoning methodology follows the general inconsistent scheme described above. That is, measures of absolute belief are combined as modular belief updates. It follows that there will be an overcounting of prior belief.

13. CONSEQUENCES OF THE INCONSISTENCY

We have identified a fundamental inconsistency in the use of absolute measures as updates in two expert systems. We realize that the methodologies for managing uncertainty in MYCIN and INTERNIST-1 were created as *heuristic* plausible reasoning strategies in response to perceived problems of complexity with using probability theory. As such, properties of the strategies may make them acceptable or useful for reasoning within their intended domains even with the identified inconsistencies.

Nevertheless, as the presence of inconsistencies in complex formal systems can make system behavior unpredictable, the identification of an inconsistency should stimulate vigorous attempts to make a system consistent. At times, clear approaches for removing inconsistencies in a reasoning system can be determined. If, for some reason it is difficult or undesirable to remove an identified inconsistency, a characterization of the possible impact of the inconsistency on system performance is indicated. In the case where an inconsistency has been identified in a system that performs well, it would be useful to develop an understanding as to the reasons that the particular inconsistency is so innocuous.

In the context of the inconsistent use of measures of belief, it is clear that removing the inconsistency based on the use of absolute beliefs as belief updates entails the elicitation of appropriate measures of change in belief from experts. The consequences of the use of measures of change of belief in lieu of belief updates in INTERNIST-1 and MYCIN can depend, among other things, on the sensitivity of the system to the methodology for reasoning under uncertainty and on the specifics surrounding the ultimate use of the measures of certainty. Indeed MYCIN and INTERNIST-1 perform impressively despite the identified inconsistency, indicating that the systems are insensitive to the nuances of the methodologies for plausible reasoning. In fact, a formal study demonstrated that MYCIN was quite insensitive to the precision of the numbers used for certainty factors [12].

Unfortunately, such serendipitous insensitivity may not always be the case. The sensitivity of a system's performance to inconsistency, assumptions of modularity, or to the use of inaccurate measures of belief will depend to a great extent on the application area. In some cases, the inconsistency regarding the use of absolute belief in systems structured to manipulate belief updates might be as inappropriate as the use of the position of a projectile as an approximation for its velocity in physics kinematics equations. Validation studies showing that MYCIN works well or is insensitive to the precision of measures of belief update does not demonstrate that similar systems, reasoning in other domains, will operate well or be similarly insensitive to problems in the use of measures of belief.

14. RELEVANCE OF BIASES IN THE ELICITATION OF BELIEF

Before concluding, we believe it is important to discuss an aspect of the belief elicitation process that may have relevance to the inconsistency resulting when researchers elicit measures of absolute belief and combine them in an updating scheme as if they were belief updates. We believe that a phenomenon detected by psychologists studying biases in human judgment may be significant in the the impact of that the misuse of belief measures may have in expert systems.

Psychologists studying possible heuristics and biases in human judgment have found that human decision makers tend to disregard prior information when assessing absolute beliefs given new evidence [25]. That is, human decision makers often overlook the absolute prior belief and make judgments based only on updates. The psychologists propose that the phenomenon is founded in the human use of particular heuristics for generating probabilities.

In light of this research, it is possible that experts may actually report belief updates when asked by expert system engineers to assess an absolute belief. If this were the case, the incorrect assessment would tend to cancel the effects of the inconsistent use of the measures of belief. Needless to say, even if the effect of the inconsistency might be diminished by the human bias, it seems expert system engineers should not rely on a potentially unpredictable phenomenon.

In addition to its relevance in the belief elicitation process, the bias discovered by cognitive psychologists may have more fundamental significance to the inconsistent use of measures of belief. It is possible that the phenomenon may have actually played a role in the historical and current inattention to the difference between measures of absolute belief and belief update.

15. SUMMARY

We have cited a current and historical failure to distinguish belief updates from absolute belief. In our discussion, we focused on formal properties of absolute belief and belief update. We introduced the notion of modularity and the commonly used modular belief updating paradigm and described an inconsistent use of measures of belief. In particular, we have argued that two representative expert systems have used measures of absolute belief in an updating scheme that requires measures of belief update. We finally attempted to characterize the nature of this inconsistency and reviewed problems with the elicitation of modular belief updates.

We believe that it is important that measures of belief or belief update are combined in a way consistent with their elicitation from experts lest serious flaws in reasoning may result. We believe that future systems would benefit from explicit consideration of the properties and requirements associated with the use of subjective measures of belief in plausible reasoning.

ACKNOWLEDGEMENTS

We thank Judea Pearl, Benjamin Grosof, Peter Cheeseman, Curt Langlotz, Peter Hart, Ted Shortliffe, and Mark Musen for useful discussions.

NOTES

[1]Actually, we don't want to require that F is monotonic increasing everywhere. In particular, when either argument takes on an extreme value, we only require that F be non-decreasing in the other argument. For example, if X is believed to be false with certainty, then XY should also believed to be false independent of the belief in Y given that X is true.

[2]The only requirements on the function f is that it be continuous and that it be monotonic increasing in each argument when the other is held constant. The function need only be monotonic non-decreasing when either argument takes on an extreme value.

[3]For simplicity, we will assume that prior information is included in ξ_H.

[4]As in previous cases, the function need only be monotonic non-decreasing when either argument takes on an extreme value.

[5]This property of modularity can be viewed as the plausible reasoning analog of a related notion of modularity in logical deduction systems. The term modularity in such systems refers to the property of being able to add or delete rules without the need to consider interactions among rules.

[6]In addition to the problem with the use of absolute beliefs as belief updates, such updating schemes often have another inconsistency. It is common in such systems for the same item of evidence to be used to update the belief assigned to several hypotheses. It is easy to show that the assumption of modularity is inconsistent with the updating of more than one hypothesis with the same item of evidence. This inconsistency can have a major impact on the calculation of final belief. We shall not dwell on this additional inconsistency as it is our intent to focus in this paper solely on problems based in the use of measures of absolute belief in methodologies structured to reason with belief updates.

REFERENCES

[1] Teknowledge, Inc., "M.1 Reference Manual", p. 3-5

[2] Shortliffe, E. H., "Computer-Based Medical Consultations: MYCIN", Elsevier/North Holland, New York, 1976.

[3] Hempel, C.G., Studies in the logic of confirmation, in "Aspects of Scientific Explanation and Other Essays in the Philosophy of Science", Free Press, New York, 1965, pp. 3-51.

[4] Jaynes, E.T., Where do we stand on maximum entropy?, in "The Maximum Entropy Formalism", M.I.T. Press, 1979, Levine and Tribus, eds.

[5] Good, I.J., The Bayesian influence, or how to sweep subjectivism under the carpet, in "Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Vol. II", D. Reidel Publishing Company, 1976, pp. 125-174.

[6] Carnap, R., "Logical Foundations of Probability", Chicago University Press, 1950.

[7] Carnap, R., "Logical Foundations of Probability: Second Edition", Chicago University Press, 1962.

[8] Popper, K.R., Corroboration, the weight of evidence, in "The Logic of Scientific Discovery", Scientific Editions, New York, 1959, pp. 387-419.

[9] Salmon, W.C., Confirmation, *Scientific American*, (225), (1973), pp. 75-83.

[10] Barker, S. F., "Induction and hypothesis: A study of the logic of confirmation", Cornell University Press, Ithaca, NY, 1957.

[11] Harre, R., "The principles of scientific thinking", University of Chicago Press, Chicago, Ill, 1970.

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

[13] Good, I.J., Weight of evidence, corroboration, explanatory power, information and the utility of experiments, *J. R. Statist. Soc. B*, (22), (1960), pp. 319-331.

[14] Cox, R., Probability, Frequency and Reasonable Expectation, *American Journal of Physics*, 14 (1), January-February (1946), pp. 1-13.

[15] Jaynes, E.T., "Plausible reasoning", Chapter in unpublished course notes

[16] Tribus, M., What do we mean by rational?, in "Rational Descriptions, Decisions and Designs", Pergamon Press, New York, 1969, ch. 1.

[17] Heckerman, D.E., Probabilistic interpretations for MYCIN's certainty factors, in "Uncertainty in Artificial Intelligence", North Holland, New York, 1986, Also in Proceedings of AAAI Workshop on Uncertainty and Probabiliity in Artificial Intelligence, August 14-16, 1985, Los Angeles, California

[18] Miller, G.A., The magical number seven, plus or minus two, *Psychological Review*, 63 (1956), pp. 81-97.

[19] Peirce, C.S., The probability of induction, in "The World of Mathematics, V.2", Simon and Shuster, 1956, pp. 1341-1354.

[20] Good, I.J., "Probability and the weighing of evidence", Hafners, New York, 1950.

[21] Minsky, M., and Selfridge, O.G., Learning in random nets, in "Information Theory", Butterworths, London, 1961, pp. 335-347.

[22] Shortliffe, E. H. and Buchanan, B. G., A Model of Inexact Reasoning in Medicine, *Mathematical Biosciences*, 23 (1975), pp. 351-379, Reprinted as Chapter 11 in (Buchanan, 1984)

[23] Miller, R. A., Pople, H. E., and Myers, J. D., INTERNIST-1, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine, *New England Journal of Medicine*, 307 (8), (1982), pp. 468-476.

[24] Pople, H., Heuristic methods for imposing structure on ill-structured problems: the structuring of medical diagnostics, in "Artificial Intelligence In Medicine", (P. Szolovits, ed.), Westview Press, 1982.

[25] Tversky, A., Kahneman, D., Judgment Under Uncertainty: Heuristics and Biases, *Science*, 185 (1974), pp. 1124-1131.