Reasoning About Uncertainty: A Knowledge Representation Perspective

Paul R. Cohen and Thomas R. Gruber
Department of Computer and Information Science
University of Massachusetts, Amherst
COINS Technical Report 85-24

EKSL Report 3

This research is funded by a grant from the National Science Foundation IST 8409623. Published in *Pergamon Infotech State of the Art Report on Expert Systems*, (John Fox, Ed.), 1984.

Uncertainty is a vague concept. Nonetheless, it denotes a class of issues that must be faced if intelligent systems are to deal with real world problems adequately. This paper discusses some of the circumstances in which uncertainty arises, and thus clarifies what we mean by uncertainty. We will discuss important sources of uncertainty in rule-based inference, then present an organization of AI approaches to uncertainty, and finally propose a knowledge-representation view of reasoning about uncertainty.

1.0 INTRODUCTION

Uncertainty is a state of mind that arises in the reasoning process. Our approach is to ask what aspects of the process give rise to uncertainty. We emphasize the sources of uncertainty and its consequences, rather than uncertainty as a mental phenomenon. A second emphasis is how, in light of these sources and consequences, a system responds to uncertainty. When a system cannot change its behavior in response to uncertainty, which nonetheless has deleterious effects, we say it is reasoning *under* uncertainty. A system that incorporates in its problem-solving repertoire some kind of response to uncertainty is said to reason *with* uncertainty. A system that explicitly represents sources and consequences of uncertainty, and reasons about them to control its own behavior (e.g., by selecting problem-solving responses) is said to reason *about* uncertainty. Reasoning about uncertainty thus places the most responsibility for managing uncertainty on the system; reasoning with uncertainty is inflexible, because the system does not reason autonomously about *how* to manage its uncertainty; reasoning under uncertainty does not involve any management of uncertainty, autonomous or otherwise.

Our emphasis on the many sources of uncertainty has led us to a position we call the *composite* view of uncertainty, contrasting with the *one-dimensional* view. Consider a property of animals called "nastiness". We propose to rank animals on this one dimension by their nastiness: sharks and vipers are very, very nasty; shrews are a bit less nasty; and so on until we reach koala bears. The inquiring mind will look at our ranking and ask, "What features make one animal nastier than another?" because even though the ranking is on a single dimension, the features that contribute to nastiness are several. Those who must deal with nasty animals will want to know *why* their subjects are nasty -- their nasty characteristics -- not merely the extent of their nastiness. Just so with uncertainty. People and computers need to know why situations are uncertain, not merely the extent of their uncertainty. Thus, we believe that theories of uncertainty should emphasize the sources of uncertainty and its consequences.

2.0 SOURCES OF UNCERTAINTY

Uncertainty in rule-based inference has three general sources. It enters in *evidence*, which may be inaccurate or insufficient; it is implicit in any *model* of a domain (which is often encoded in production rules); and it is associated with the *beliefs* that result from inference. We discuss these in turn. Throughout the discussion we assume that the *environment* supplies evidence, which evokes *inferences*, which result in *beliefs*. Beliefs may be used as evidence for further inference.

2.1 Sources of Uncertainty About Evidence

Among the things we can say about evidence are that it is errorful, irrelevant, or insufficient. These are causes of uncertainty. In addition, we can say that a situation has a chance of being true; for example, it *might* rain, or Sally has an 80% chance of beating Fred´s poker hand. Because we want to understand the sources of uncertainty, we are unwilling to summarize with a number the argument that evidence is, say, irrelevant; since we would no longer be able to distinguish irrelevance from insufficiency or other causes of uncertainty. We will try to maintain this distinction, though it is easily muddled when probabilistic arguments are combined with other causes of uncertainty; for example, the evidence "Sally´s chances are 80%" may be insufficient, and evidence may have an 80% chance of being insufficient.

Errorful evidence is common in systems that rely on sensory information. For example, the tactile sensors of a robot may malfunction, resulting in an errorful interpretation of any evidence the sensor provides. Noise causes uncertainty about whether one´s evidence is relevant. Systems such as HEARSAY-II (Erman, Lesser, Hayes-Roth, and Reddy, 1980) and HASP/SIAP (Nii and Feigenbaum, 1977) contended with noise from their transducers. Before they could ask whether data from transducers was errorful, they had first to cope with uncertainty about its relevance – whether it was signal or noise. Many applications are uncertain because they need more data than is readily available, quite apart from the questions of whether the data that *is* available is errorful or relevant. In medicine, for example, some invasive tests are expensive or life-threatening, and so diagnosis might proceed on the basis of incomplete evidence. In other cases, the needed evidence will never be available; for example, pollsters necessarily make statistical inferences from small samples because it is impossible, or impractical, to query an entire population.

We prefer to characterize a poll as "accurate within a 2% margin of error," or a diagnosis as "lacking the evidence from a brain scan," since these characterizations guide us in dealing with our uncertainty. The more we know about the causes and consequences of uncertainty about evidence, the better we are able to cope with the uncertainty.

2.2 Sources of Uncertainty About the Model

Rule-based inference systems capture knowledge about the world in inference rules (which constitute a *world model*). Uncertainty is caused by the processes of constructing and using these rules. When constructing rules, uncertainty is an inevitable consequence of summarizing knowledge. We recognize that expert inference rules are *compilations* of dozens or hundreds of experiences, and that minor differences between the experiences are "smoothed out" in the rule. When using such rules, "relevant" features of a case – those mentioned in the condition of a rule – are attended to, but discrepant features are ignored.

A related source of uncertainty in rule-based inference is that rules are constructed with some purpose in mind, but the context in which rules are used does not necessarily correspond to the purpose for which they were intended. For example, imagine a simple rule that states "If it's raining, then take an umbrella." This rule assumes that one's purpose is to stay dry, when in fact one may want to be drenched. It doesn't work to add another conditional clause to the rule, specifying that one wants to stay dry, because other implicit assumptions are easily generated; for example, we are assuming that the umbrella works. One cannot escape the uncertainty caused by not knowing whether the implicit assumptions of an inference rule are met.

Uncertainty arises from limitations of the world model. In terms of rule-based inference systems, uncertainty is caused by not knowing whether one has rules for all situations that may arise. It is worth making this source of uncertainty explicit, because it makes an interesting qualification on one's conclusions. Expert knowledge may be relatively complete, so when the expert says "As far as I know, you are healthy," you can be pleased. But the knowledge of expert systems is usually less complete, so a clean bill of health from one of *them* is probably less encouraging. The expert system *should* say "As far as I know, . . .," because far from being a conversational nicety, it is an important qualification.

Note that "I'm pretty sure you're healthy" is not as informative as "As far as I know, you're healthy," since the latter states the cause of any uncertainty and the former does not. We re-emphasize the point we first made in connection with uncertainty about evidence: The more one knows about the sources of uncertainty about inference rules, the better one might cope with this uncertainty.

2.3 Sources of Uncertainty About Belief

Beliefs, in our simplified model of rule-based inference, are the conclusions of inferences. Thus, important sources of uncertainty in beliefs are uncertainty in evidence and inference rules. We will discuss how these sources of uncertainty are *propagated* to beliefs in later sections. Here, we concentrate on uncertainty that arises from one's beliefs independent of their derivation. The chief cause of uncertainty is that beliefs are sometimes *inconsistent*. For example, we believe that we pay too much money in taxes, but we also believe that taxation for social programs is a Good Thing. Inconsistent beliefs lead to uncertainty about our future conclusions and actions; it is not possible to predict with certainty whether we will vote for tax-cutting or tax-raising political candidates. Another source of uncertainty concerns how beliefs are accessed. In humans, at least, one can easily demonstrate *priming* and *availability* biases (e.g., Kahneman, Slovic, and Tversky, 1982). Briefly, people do not bring all beliefs to mind with equal facility, and we use apparent facility as evidence about the truth of statements. In the simplest case, if we cannot think of any examples of a proposition (e.g., that books can dance the polka) then we say the proposition is false. This is fair enough, but we often misjudge the likelihood of propositions by this same device. In AI inference systems, access can be interpreted as search, which may be bounded by resource considerations, or deduction, also susceptible to limits. Since the structure of the representations of belief can affect the efficiency of access, judgments based on the relative ease of access can introduce uncertainty about beliefs regardless of their content.

3.0 AI APPROACHES TO UNCERTAINTY

Given the diversity of sources of uncertainty, it is not surprising to find a plethora of responses in AI inference systems. The current approaches to uncertainty can be organized into three major groups. Systems constructed to circumvent the effects of uncertainty are of the *engineering approach*. Systems that control their behavior to avoid or reduce the effect of uncertainty use the *control approach*. Some systems divide the inference process into two separate subprocesses, one that performs inference as if there were no uncertainty, and another that associates representations of partial belief with the conclusions of the first process; this approach is called *parallel certainty inference* (Cohen, 1983). Although there is some overlap in these categories, they provide a useful organization to the discussion of current AI approaches to uncertainty.

3.1 The Engineering Approach

The designer of an inference system can anticipate some causes of uncertainty that effect the performance of a system, and then formulate the *problem* to eliminate any need to consider the uncertainty. For example, elementary textbook problems in physics ignore the effects of friction, relieving the student of the need to calculate the (uncertain) effect of this difficult-to-measure factor. It is common in AI inference systems to engineer the uncertainty out of problems, especially for prototype systems. Problems are frequently hard enough without considering noise or error, so the *clean data assumption* is often made to eliminate the effect of uncertainty introduced by the evidence. Of course, the same techniques that work with clean data must often be modified to cope with the problems of noise and error.

A second way to engineer uncertainty out of AI systems is to *assume relevance*. It is sometimes difficult to decide which features of the environment are relevant to a task, especially if one's world model is incomplete. Systems that are free of this uncertainty are conceptually simpler. For instance, Winston's (1975) "concept learning" program was presented with a set of training instances and inferred a "rule" to classify them. The program assumed that the teacher would supply typical and "near miss" cases of a special form. The problem was made tractable by assuming relevance, but the more difficult task of generating and evaluating training instances was finessed. Other learning systems (discussed in Dietterich, 1982; Michalski, Carbonell, and Mitchell, 1983) have made similar assumptions.

A third form of the engineering approach is a response to the kind of uncertainty that results from incomplete models of a domain. Since a system cannot know everything about its domain, it must make tentative decisions on the basis of uncertain beliefs. For instance, it is common to make the *closed world assumption* (Reiter, 1980) when working with a finite database of facts. The assumption asserts that something is false if it is not known to be true.[1] Thus, under the closed world assumption, to decide whether X is true, one checks if X is known; if it is not known, then X is assumed to be false. In a rule-based inference system, if no rule has asserted a proposition (even though it is possible that one might in the future), the proposition is false under this assumption. A system that makes the closed world assumption is freed from the need to have a complete model; it has removed one source of uncertainty - the uncertainty of the unknown - by hiding it. (However, not all systems ignore the uncertainty introduced by assumptions. See the discussions of dependency-directed backtracking and reason maintenance in the following two sections for techniques that recognize and reason with the uncertainty introduced by assumptions.)

3.2 Control Approaches

Control approaches to uncertainty recognize the characteristics of a domain that cause uncertainty, and utilize control strategies to reduce the effects of uncertainty or eliminate its sources. As an illustration, consider a control strategy for solving a jigsaw puzzle: build the borders first, and then work in towards the center. This strategy exploits the local constraints provided by the straight edges of the border pieces to reduce the number of pieces that could be fit. Border pieces are less unconstrained and should be placed first; then, any piece that looks as if it might extend the frame should be placed next. A control strategy that exploits domain constraints this way can sometimes minimize uncertainty or its effects. AI systems use knowledge about uncertainty in their control strategies in a variety of ways. By recognizing those points where uncertainty is introduced, a control strategy can provide a mechanism to retract errorful conclusions or mark problematic issues for careful analysis. A control strategy can also concentrate on hypotheses (partially

---

[1] Something is typically considered known if it is immediately available in a database or if it can be found by some limited inference. But in some logic-based paradigms, something is known if it can be proven -- deduced from the current set of assertions (Artificial Intelligence, 1980). See (Levesque, 1984) for a discussion of the difference.

A general assumption relating knowledge of a proposition to its truth is that X is true if and only if X is known. The contrapositive of implication in one direction (~known(x) -> ~true(x)) is the closed world assumption as commonly understood. The positive implication in one direction (true(x) -> known(x)) is the assumption made by (Collins, et. al., 1975) in plausible reasoning. The positive implication in the other direction (known(x) -> true(x)) is the assumption made by reasoning processes that ignore the effects of uncertainty in their beliefs, as in parallel certainty inference discussed in a later section.

supported belief) with especially high or low certainty, or modify action on the basis of characteristics of uncertain evidence.

Dependency directed backtracking (Stallman and Sussman, 1977; Doyle 1979; London, 1978) is a method for efficiently recovering from errors in choices made with uncertainty. The behavior of a system can be seen as a tree, with each node representing a choice made under uncertainty. The power of backtracking is that the reasoning process assumes all nodes (choices) along a single branch of the tree are certain. When a choice is found to be wrong, the reasoning process reconsiders and makes an alternate choice at that point. An efficient method for redoing the choice leaves the bulk of the belief set unchanged. A related approach, which records the *reasons* for the uncertainty at each choice point, is discussed below.

Least-commitment planning (Sacerdoti, 1977; Stefik, 1980) is a strategy to manage the uncertainty introduced by not knowing the effects of actions (i.e., incompleteness of the domain model). The construction of plan steps introduces uncertainty because possible interactions with other plan steps are not known in advance. By delaying the commitment to these plan steps until more interactions are known, the uncertainty in the effects on other parts of the plan is reduced.

Opportunistic control, as in the HEARSAY-II speech understanding system, (Erman, et.al., 1980) directs the system to focus its attention to those hypotheses that are supported with the greatest certainty, that is, to follow the most promising leads. These *islands of certainty* are sources of local constraints that make it easier to propose and support new hypotheses. In the speech understanding domain, the effects of uncertainty were minimized by this opportunistic strategy, which relied on the redundancy of information in the speech signal. One can imagine cases in which an opportunistic strategy is not as well-suited to the characteristic uncertainty of a domain. The point is that for the control approach to work, the control strategy must be matched to the kinds of uncertainty that arise in a domain.

Heuristic search can also benefit from the consideration of uncertainty. The term "heuristic knowledge" implies that the knowledge is imperfect (uncertain) in some way. Understanding the limitations of heuristic knowledge can be a source of power in using it. For instance, many computer chess programs incorporate a *static evaluator*, a heuristic that estimates the worth of a board position. By searching a few moves ahead and applying the static evaluator at the terminal nodes of the search tree to compare the relative worth of each move, a chess program can choose a reasonable move. A difficulty called the *horizon effect* (Berliner, 1974) occurs when beneficial positions are missed because the static evaluator is applied at a uniform depth. Important positions are missed if they are just over the horizon of the evaluator's view. A control strategy can improve performance if it extends the search at points where the horizon effect is most likely.

In summary, the control approach to uncertainty recognizes *where* uncertainty arises and incorporates a control strategy to provide flexibility at those points.

3.3 Parallel Certainty Inference

The parallel certainty approach divides the reasoning process into two semi-independent processes. One proceeds as if there were no uncertainty in its conclusions. The other decides on the certainty of the conclusions derived by the first. This is convenient because it allows the first process to concentrate on the domain problem without considering difficulties introduced by uncertainty. The first process decides *what* to believe, and the second, *how much* or *why* to believe. Three broad categories of parallel certainty will be discussed: degrees of belief, reason maintenance, and the theory of endorsements.

3.3.1 Degrees of belief

The most popular parallel certainty methods represent uncertainty as a degree of belief, [2] an expression of *how much* something is to be believed. The canonical example is the *certainty factor* representation of MYCIN (Shortliffe and Buchanan, 1975) and PROSPECTOR (Duda, Hart, and Nilsson, 1976). A proposition is associated with a number between 0 and 1 that represents how much the system believes it. Inference rules are applied without regard to the certainty of their premises[3] (or conclusions, if the inference is backward-chaining), while degrees of belief are propagated from premises to conclusions via a rule of combination.

At least two sources of uncertainty are represented by certainty factors: certainty in inference rules and certainty in beliefs. A certainty factor for a rule represents the expert´s confidence in it, but it is not always clear what confidence means. For example, a rule that states that obesity implies illness might have a certainty factor of 0.8 associated with it. This number might represent the proposition that 80% of obese people get sick, or that the probability is .80 that a sick person is obese, or that the general rule that obesity causes sickness is more applicable than a rule with a certainty factor of 0.6. Whatever its meaning, the *effect* of the certainty factor on a rule is to weight the belief in its conclusions; the higher the rule´s cf, the higher the belief in conclusions from that rule (all things being equal). Certainty in

[2] The term is due to Shafer (1976).

[3] Actually, MYCIN did not fire rules whose conditions were believed with less than 0.2 cf, so it is not strictly a parallel inference method, since domain inferences are not kept entirely separate from inferences about uncertainty.

beliefs is also represented by numbers. Again, it is difficult to be clear about what the certainty factor of a belief *means*, other than to say that higher numbers mean stronger belief.

Belief is propagated across inferences. The propagation rules used by MYCIN and PROSPECTOR are variants of Bayes´ rule, which provides a mathematical method for updating the probability of a hypothesis given an observation of evidence. Bayesian methods are based on the axioms of probability theory, and have been applied in several ways to combine the degrees of belief for multiple hypotheses given evidence from multiple distinct sources that might disagree. They are quite general.

A related method of representing and reasoning with degrees of belief is the Shafer-Dempster method (Dempster, 1968; Shafer, 1976; Lowrance and Garvey, 1982). In contrast to the Bayesian approach, belief is represented by an interval between 0 and 1, rather than as a single point. The Shafer-Dempster method has a number of advantages over a strictly Bayesian approach, mainly because it makes weaker assumptions. (Bayesians require all hypothesis to be mutually exclusive, exhaustive, singletons). The two-number representation allows for ignorance (the inescapable result of incomplete knowledge), as well as degree of belief, whereas in Bayesian models ignorance is commonly *mis*represented as belief in the negation. The Shafer-Dempster representation can capture belief in sets of hypotheses, which is particularly useful when uncertainty about the relevance of evidence prevents the assignment of belief to individual (singleton) hypotheses.

Many objections can be raised to representing uncertainty with degrees of belief. First, the semantics of the numbers are not well defined. Some authors interpret the numbers as subjective probabilities, others as frequencies, and others entirely ignore the issue of what the numbers mean. An emphasis of recent research (Rich, 1983; Kim and Pearl, 1983; Wesley, 1983; Ginsberg, 1984; Strat, 1984) has been to make numerical degrees of belief represent an increasing variety of *kinds* of uncertainty, so the interpretation of the numbers is a bewildering task. We believe that numbers are not an adequate representation for everything one wishes to say about the causes and consequences of uncertainty. A second problem, which is a consequence of the representational inadequacy of numbers, is that the numbers are *used* to represent combinations of factors; for example, certainty factors in rule-based systems frequently account for *salience* and *utility* as well as degree of belief. A third and related problem is that if the components of a degree of belief are unknown, or if their relative contributions are unknown, then it is impossible to predict whether transformations of degrees of belief – such as combining functions – have any effects on the meanings of the numbers, since the meanings were obscure to begin with. A rule may be given a high certainty factor because it is important, or useful, even if it is not very accurate. What interpretation does one give a number produced by combining two such certainty factors? A fourth problem, again closely related, is succinctly put in the question: "where do the numbers come from?".

Salmon (1967) calls this the criterion of *ascerntainability*. How do we hope to effectively capture the knowledge of a human expert with numbers when the expert doesn´t reason that way?

3.3.2 Reason Maintenance

Reason Maintenance (Doyle, 1980, 1983a). was developed specifically to deal with uncertainty caused by incomplete knowledge. Often, the truth of a proposition cannot be determined, but one can proceed as if it were known. Reason maintenance, and the theory of "reasoned assumption" most recently developed by Doyle (1983b), calls for jumping to conclusions in the case where the truth of a proposition is not known but can be assumed. In making assumptions of various forms, the system *consciously* introduces uncertainty; it records the *reasons* for the assumption, and thus represents sources of uncertainty associated with it. In terms of the parallel certainty inference model, the first inference process proceeds as if it has confidence in assumed propositions, and the second provides a mechanism to carefully retract assumptions if they are found to be wrong. Thus, reasons for belief support sophisticated reasoning *about* uncertainty.

3.3.3 The Theory of Endorsements

The parallel certainty inference approach divides reasoning under uncertainty into two "streams"; one is a stream of logical inferences, typically the inferences that are needed to solve a problem. The other is a stream of inferences about the certainty of conclusions produced in the first stream. We have considered numerical inferences, based on Bayes´ and Dempster´s rules, and also reason maintenance -- the recording and maintenance of dependencies among conclusions. A third approach is to record arbitrarily complex messages (which we call *endorsements*) in the second stream. These messages record causes and remedies of uncertainty; for example, we might note that evidence is produced by an occasionally faulty sensor, or that a newspaper reporter considers a wide range of sources before filing a report, or that the margin of error on a poll is 5%, or that a recommendation comes from someone who doesn´t know his subject, and so on. The fundamental assumption of the *theory of endorsements* (Cohen, 1983) is that subjective degrees of belief, usually represented as numbers, are *composites* of reasons to believe and disbelieve. We suggest that, for many tasks, one needs to know more than simply the *extent* of one´s belief; one also needs to know the *causes* of belief. The theory of endorsements is concerned with how to represent and reason with this knowledge.

We misrepresent the theory of endorsements somewhat by grouping it with parallel certainty inference approaches. One advantage of knowing why a proposition is uncertain is the ability to "take evasive action" to eliminate the cause, or the effects, of uncertainty. For example, if one knows that the cause of uncertainty is the absence of attainable knowledge, then one might eliminate the uncertainty by simply asking for the missing information, or by searching for it. On the other hand, if the missing knowledge isn't obtainable, then one cannot eliminate the cause of uncertainty but one may minimize its effects. For example, hedging minimizes uncertainty arising from unattainable knowledge. The key to such evasive action is knowledge about uncertainty. The search for missing evidence, for example, depends on knowledge about its source. If the source produces evidence intermittently, like a volcanic fault, then one must sit around and wait. We adopt one strategy for a feedback-directed search for evidence (e.g., we have found the right book, then the right section, and finally the desired sentence), and another for evidence that just "pops up" without warning (e.g., waiting for a bus that may or may not be running). Thus, the key to making a system responsive to its uncertainty is knowledge about the causes of uncertainty; or, conversely, parallel certainty inference approaches aren't responsive to uncertainty because they know nothing about it except its extent.

The theory of endorsements was initially developed in the context of rule-based systems, and was tested with expert heuristics from the domain of portfolio management (from a program called FOLIO; see Cohen and Lieberman, 1983). Cohen (1983) built a program called SOLOMON to reason about the uncertainty associated with these heuristics. Endorsements -- frame-like knowledge structures representing reasons to believe (*positive* endorsements) and disbelieve (*negative* endorsements) -- were associated with propositions and inference rules at various times during reasoning. SOLOMON was intended originally as a tool for building expert systems to reason about uncertainty, and so its "built-in" endorsements were very general. The style of reasoning mediated by these endorsements was, by design, similar to the goal-directed reasoning of many expert systems: SOLOMON backward-chained through its rule base of domain-specific heuristics, accruing endorsements -- reasons to believe and disbelieve -- to intermediate conclusions.

Two useful effects of endorsements on *control* were implemented: First, SOLOMON used endorsements to decide whether a proposition was *certain enough for the task at hand*. It would ask whether the endorsements of a subgoal conclusion were good enough to warrant using the conclusion to assert its parent goal. This device made concrete the intuition that a body of evidence will suffice for some purposes but not others; for example, the word of a used-car salesman might barely suffice if you want to know who won last night's football game, but is perilously untrustworthy where the salesman's self-interest is concerned.

The second effect of uncertainty on behavior was achieved, in SOLOMON, by *resolution tasks*. Negative endorsements were viewed as problems to be solved: SOLOMON attempted to improve the endorsement of an important proposition. It had available general and domain-specific rules for resolving uncertainty. A general method was to drop clauses from the premises of rules. The effect was to make inferences that would otherwise have been prevented due to unmet preconditions, but to propagate uncertainty about the applicability of a rule to its conclusion. A second, powerful, method was to search for corroboration for uncertain conclusions (see Cohen, 1983, pp. 148-158, for a detailed example).

A serious problem for the theory of endorsements is how to combine endorsements. As long as uncertainty is represented by quantities (e.g., degrees of belief), mechanisms for combining evidence are easily constructed (even if the interpretations of the results are unclear). But endorsements represent aspects of evidence other than its weight. They are not quantities and thus are not easily combined. We doubt whether the meanings of numerical degrees of belief (whatever they are) are preserved by arithmetic rules of combination, and thus we believe that the ease of combining numeric representations of uncertainty is bought at an unacceptable cost. But the fact remains that reasoning under uncertainty requires the ability to combine evidence, and the theory of endorsements lacks simple methods for this end.

The methods we *have* developed add to or delete from a body of endorsements associated with a conclusion. For example, one rule of combination says "If a user action might have been a mistake (a negative endorsement), and the next action is consistent with it (new evidence), then erase the negative endorsement." We call these rules of combination *semantic combining rules* because we think they preserve the commonsense meaning of endorsements. For example, the reasoning underlying the previous rule is that if consecutive are actions are consistent, then they *might* both be mistakes, but more likely neither are. Thus, the second action has the effect of erasing uncertainty about the first. The main advantage to this approach to combining endorsements is that the meaning of the endorsements is preserved. The main disadvantage is the need for many semantic combining rules.

4.0 CONCLUSIONS

To effectively reason *under* uncertainty, in the long run, intelligent systems must reason *about* uncertainty. This means specifying representations, thinking carefully about what they mean, developing operations for *combining* and *propagating* them, and considering what properties of uncertainty the operations preserve. Early work in reasoning with uncertainty concentrated on whether there was uncertainty and how much. This is adequate for some purposes, but the intelligent reasoning systems of the future will need richer representations for a more sophisticated approach to uncertainty. Some of the purposes to which sophisticated reasoning about uncertainty must be applied are explanation, evaluation, and control.

Explanation. We want to know "why" an agent believes something, not just "how much" it is believed. Early inference systems such as TEIRESIAS, (Davis, 1976) explained their behavior by displaying the chain of rules leading to a conclusion; they didn't explain why those particular rules fired. In particular, they failed to explain the basis for partial support (i.e., certainty factors). It is not clear how a degree of belief summarizes the reasoning under uncertainty that produced it, and yet, it is precisely in conditions of uncertainty that good explanations are most beneficial.

Evaluation. AI systems cannot be evaluated as black boxes. Proper validation requires a consideration of the structure and content of internal belief. For example, Lenat's (1976) AM program discovers fundamental concepts in mathematics. That's the black box view. Only after several years of experiment did anyone (including Lenat) really understand why and how AM worked. (Lenat and Brown, 1983) That analysis, which demystified the original program and provided valuable insights into the nature of learning, was based on experiments with the structure and content of AM's representations. Smilarly, we cannot hope to understand how our systems reason under uncertainty unless we "open up" the black box representations of uncertainty. As with AM, we can say that our systems "work." But they do not currently give us any insight into the sources and consequences of uncertainty.

Control. Most expert systems use relatively simple control strategies. Processing is data-driven or goal-driven, or the two may be mixed in an opportunistic manner. Focus of attention in opportunistic systems is managed by numerically weighing, in empirically derived equations, alternative actions (e.g., Erman and Lesser, 1980). Unfortunately these numeric assessments hide the reasons for performing one action over another. We propose that flexible control strategies for reasoning in uncertain domains must be sensitive to the causes and consequences of uncertainty. Only if these are represented explicitly, can a system tailor its actions to minimize uncertainty or its consequences.

In conclusion, sophisticated reasoning about uncertainty will require adequate representations of knowledge about the causes and consequences of uncertainty, and adequate mechanisms for weighing, combining, and selecting actions, based on these representations.

5.0 REFERENCES

Berliner, H. Chess as problem solving: the development of a tactics analyzer. Department of Computer Science, Carnegie-Mellon University, 1974.

Collins, A., Warnock, E., Aiello, N., and Miller, M. Reasoning from incomplete knowledge. in *Representation and Understanding*, D. G. Bobrow and A. Collins (Eds.), New York: Academic Press, 1975, 383-415.

Cohen, P. R. Heuristic reasoning about uncertainty: An artificial intelligence approach. Doctoral dissertation, STAN-CS-83-986, Stanford University, 1983; Also in press by Pitman Publishers Ltd. London.

Cohen, P. R. , and Lieberman, M. D. FOLIO: An expert assistant for portfolio managers. IJCAI 8. 1983.

Cohen, P. R. Progress report on the theory of endorsements. Proceedings of the IEEE workshop on Expert Systems. Forthcoming.

Davis, R. Applications of meta-level knowledge to the construction, maintenance, and use of large knowledge bases. (Doctoral Dissertation), Computer Science Department Report STAN-CS-76-552, Stanford University, 1976.

Dempster, A. P. A generalization of Bayesian inference. *Journal of the Royal Statistical Society, series B*, **30**, 205-247.

Dietterich, T. G. Learning. in P. Cohen and E. Feigenbaum (Eds.) *The Handbook of Artificial Intelligence, Volume III*. Los Altos, CA.: William Kaufmann, Inc., 1982, 323-513.

Doyle, J. A truth maintenance system. *Artificial Intelligence*, **13**, 1979, 81-132.

Doyle, J. The ins and outs of reason maintenance. *Proceedings IJCAI-83*, August 1983, 349-351. (a)

Doyle, J. Some theories of reasoned assumptions: Some essays in rational psychology. CMU-CS-83-125. Carnegie-Mellon University, 1983 (b).

Duda, R. O., Hart, P. E., and Nilsson, N. Subjective Bayesian methods for rule-based inference systems. Technical Note 124, Artificial Intelligence Center, SRI International, Menlo Park, California.

Erman, L. D., Hayes-Roth, F., Lesser, V. R., and Reddy, D. R. The HEARSAY-II speech understanding system: integrating knowledge to resolve uncertainty.

Computing Surveys, 12, 1980, 213-253.

Ginsberg, M. L. Non-monotonic reasoning using Dempster's Rule. *Proceedings AAAI-84*, August, 1984, 126-129.

Kahneman, D., Slovic, P., and Tversky, A. Judgment under uncertainty: Heuristics and biases. Cambridge University Press, 1982.

Kim, J. M. and Pearl, J. A computational model for causal and diagnostic reasoning in inference systems. *Proceedings IJCAI-83*, August 1983, 190-193.

Lenat, D. B. AM: an artificial intelligence approach to discovery in mathematics as heuristic search. (Doctoral Dissertation), Department of Computer Science report STAN-CS-76-570, Stanford University, 1976.

Lenat, D. B. and Brown, J. S. Why AM and Eurisko appear to work. *Proceedings AAAI-83*, August 1983, 236-240.

Levesque, H. J. A logic of implicit and explicit belief. *Proceedings AAAI-84*, August, 1984, 198-202.

London, P. E. Dependency networks as a representation for modeling in general problem solver. (Doctoral Dissertation), Department of Computer Science Technical Report 698, University of Maryland, 1978.

Lowrance, J. P., and Garvey, T. P. Evidential reasoning: a developing concept. *Proceedings of the International Conference on Cybernetics and Society*, October 1982, 6-9.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. *Machine Learning: An Artificial Intelligence Approach*, Palo Alto, CA.: Tioga, 1983.

Nii, H. P., and Feigenbaum, E. A. Rule-based understanding of signals. Memo STAN-CS-77-612. Stanford University. 1977

Reiter, R. On closed world data bases. in H. Gaillaire and J. Minicer (Eds.), *Logic and Data Bases*. New York: Plenum Press, 1978, 55-76.

Reiter, R. A logic for default reasoning. *Artificial Intelligence*, 13, 1980, 81-132.

Rich, Elaine. Default reasoning as likelihood reasoning. *Proceedings AAAI-83*, August 1983, 348-351.

Sacerdoti, E. D. A structure for plans and behavior. New York: American Elsevier. (Originally published in 1975)

Salmon, W. C. *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press, 1967.

Shafer, G. *A Mathematical Model of Evidence*. Princeton: Princeton University Press, 1976.

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

Stallman, R. M. and Sussman, G. J. Forward reasoning and dependency-directed backtracking in a system for computer-aided circuit analysis. *Artificial Intelligence*, **9**, 1977, 81-132.

Stefik, M. J. Planning with constraints. (Doctoral Dissertation), Department of Computer Science Report 80-784, Stanford University, 1980.

Strat, T. M. Continuous belief functions for evidential reasoning. *Proceedings AAAI-84*, August, 1984, 308-313.

Wesley, L. P. Reasoning about control: the investigation of an evidential approach. *Proceedings IJCAI-83*, August 1983, 203-206.

Winston, P. H. (Ed.). *The Psychology of Computer Vision*. New York: McGraw Hill, 1975.