Two Endorsement-based Approaches to Reasoning About Uncertainty

Paul Cohen
Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

§1 ABSTRACT

Two approaches to reasoning about uncertainty are discussed. A parallel certainty inference model superimposes reasoning about the credibility of inferences on a deductive framework. A second approach identifies and implements representativeness as the general determinant of credibility in classification tasks. These approaches are steps in the evolution of endorsement-based reasoning, a view of uncertainty in terms of structured objects that represent characteristics of evidence.

§2 INTRODUCTION

This paper is about evidence. It is also about uncertainty, since uncertainty is a state of mind that arises when evidence is incomplete, inaccurate, irrelevant, and so on. Since these and other characteristics of evidence give rise to uncertainty, reasoning about uncertainty ought to be reasoning about characteristics of evidence. But this is difficult for Artificial Intelligence (AI) programs, which are not provided with explicit representations of characteristics of evidence. Most frequently, all knowledge about uncertainty is summarized in a single degree or range of belief.

Endorsements are explicit representations of characteristics of evidence. This paper describes two stages in the theory of reasoning with endorsements. It draws together results from several sources,[1] and is necessarily condensed. The endorsement-based view is prescriptive, in the sense of saying how reasoning about uncertainty *ought* to be done by intelligent computers. Some prescriptions have been implemented, others remain open research problems. One version of endorsement-based reasoning, described below, also purports to describe human reasoning under uncertainty.

Three prescriptions for intelligent reasoning about uncertainty guide the development of endorsement-based reasoning:

Orientation towards action. It is striking that humans, though uncertain about many or most decisions *act as if certain* about these decisions. AI programs are

[1] Cohen and Grinberg (1983); Cohen (1985); Sullivan and Cohen (1985); Cohen, Davis, Day, Greenberg, Kjeldsen, Lander and Loiselle (1985).

less able because they lack representations of *reasons* for uncertainty. Decision is trivialized by reliance on *degrees* of belief, exclusive of reasons, since decision strategies have access only to the weight of evidence, not its other characteristics. For example, given only the degree of belief in a hypothesis, one cannot decide to minimize uncertainty by attempting to prove its converse. Nor can one decide to "wait and see," in the expectation that more evidence is forthcoming. Both decision strategies require knowledge about *why* a hypothesis is uncertain: Is its converse more credible? is it currently unsupported but expected to gain support soon? An orientation to action under uncertainty mandates this kind of explicit knowledge about uncertainty.

Justification. Humans can justify their decisions. AI programs can generally explain how they derive factual conclusions, but not why they believe or disbelieve the conclusions. Once again, the lack of explicit representations of characteristics of evidence is at fault.

Advocacy and consensus. Humans can view the same evidence from different perspectives – arguing for and against a position to convince oneself or another. Avron Barr (1985) points out that, in addition to advocacy, humans may argue to find a consensus. The goal here is not to bludgeon an opposing position with the club of evidence, but to seek a new interpretation of the evidence that accommodates two apparently opposing views. Both advocacy and consensus focus on how evidence supports arguments. The weight of evidence is certainly important, but no more so than the source, cost, reliability, and other characteristics of evidence.

Guided by these considerations, we developed two forms of endorsement-based reasoning about evidence.

§3 PARALLEL CERTAINTY INFERENCE

The parallel certainty inference model divides reasoning under uncertainty into two parallel streams. In one, deductive inference is mediated by modus ponens; and in the other, representations of the credibility of deductive inferences are propagated. This split is necessary because deduction is not defined for propositions that are neither true nor false. Therefore, to reason deductively under uncertainty, one must add representations of uncertainty to logical (true or false) propositions. Degrees of belief in expert systems are the best-known example of parallel certainty inferences. For example, MYCIN uses modus ponens to backchain through a rulebase of implications, and uses quasi-Bayesian combining functions to propagate degrees of belief from one conclusion to the next (Buchanan and Shortliffe, 1984, Chapter 11). Since we believe that numbers are inadequate representations of characteristics of uncertainty, we developed a parallel certainty inference model in which numbers were replaced by endorsements.

Cohen (1984) and Sullivan and Cohen (1985) studied the role of endorsements in plan recognition. The problem is this: Given a set of plans, each composed of elementary actions, determine the plan a user has in mind by watching his actions. For example, imagine the given "library" contains only three plans, each composed of three elementary actions:

Plan	Steps
plan1	a b c
plan2	b d e
plan3	d f g

Further assume, for simplicity, that plan steps are not "shared": step **b**, for example, does not simultaneously continue plan 1 and start plan 2.

How can we interpret the user actions **a** followed by **b**? The first action suggests the user intends plan 1, since it is unique to plan 1. However, it might have been a mistake: the user might actually intend plan 2. The second action, **b** appears to support the plan 1 interpretation, since it is the next step in the plan. However, **b** is ambiguous: the first action might have been a mistake and the user might be starting plan 2. Both interpretations are uncertain. Sullivan and Cohen represent the uncertainty with mnemonic endorsements:

Step	Interpretation	Endorsements
1: **a**	(start plan1 **a**)	(**a** unique to plan 1 +) (**a** could be a mistake -)
2: **b**	(continue plan1 **b**)	(**a b** continuity is desirable +) (**b** ambiguous -) (**b** could be a mistake -)
b	(start plan2 **b**)	(**a b** discontinuity is undesirable -) (**b** ambiguous -) (**b** could be a mistake -)

Asked to argue that the actions **a, b** are evidence of plan 1, one can read the positive endorsements (that's what the + means) that **a** is unique to plan 1, and is continued by **b**. Asked to argue against this interpretation, one reads the negative endorsements: **a** might be a mistake, **b** is not unique to plan 1. The plan 2 interpretation of the actions accrues no positive endorsements, but the negative ones can be used to argue against it: the user prefers not to start one plan while in the middle of another (i.e., discontinuity is undesirable); **b** is ambiguous, and might have been a mistake.

These endorsements seem to explicitly capture the sources of uncertainty in the plan interpretation task. Yet they are only tokens: their meanings must be supplied from somewhere. "Discontinuity is undesirable" stands for the intuition that a user prefers to finish one task before starting another. Imagine replacing the mnemonic endorsements with arbitrary symbols, and you will see how important it is to specify the semantics of endorsements, and not rely on their mnemonic power. How much knowledge would a program need to reason about the symbol **x43** as well as humans reason about the possibility that an action is a mistake? This knowledge imparts meaning to the token "could be a mistake." Sullivan and Cohen found that the meaning of endorsements was adequately specified by their applicability conditions – when to add them to a proposition – and by rules specifying when to add and delete endorsements from propositions in the light of new evidence. The latter kind of knowledge played an important role in combining evidence.

The task of combining endorsements in light of evidence is best introduced with an example. Given the plans above, we can interpret action **a** as evidence for plan 1, subject to the qualification that it could be a mistake. Now, if the next action is **b**, is the qualification still valid? Might our uncertainty be tempered by the fact that **b** is consecutive with **a** under the plan 1 interpretation of both actions? If the third action is **c**, the concluding step in plan 1, can there be any doubt that **a** was not a mistake? If it is reasonable to diminish uncertainty due to a possible mistake given subsequent consistent evidence, then we might erase the negative endorsement "could be a mistake - " given such evidence. This is exactly the method used by Sullivan and Cohen. They specified *semantic combining rules*, so called because they defined the meaning of endorsements in terms of the way endorsements combine:

If (plan N: step i could be a mistake -) and
(plan N: steps i j continuity is desirable +)
Then erase (plan N: step i could be a mistake -)

(Though the implementation of this and other semantic combining rules literally erased endorsements, a better implementation would withdraw support for them. This reasoning, in the style of Doyle's reason maintenance (Doyle, 1983), keeps an important record of the fate of hypotheses as evidence is accrued.)

Semantic combining rules provide one mechanism for implementing advocate and consensus reasoning. Consider the justification offered by one advocate for the plan 1 interpretation of **a**, **b**: "The actions are consecutive – evidence that plan 1 is intended. True, **a** might be a mistake, and **b** is ambiguous, but, for me, the likelihood that **a** is a mistake given the consecutive action **b** is small." Another advocate uses the same evidence and endorsements but takes another view: "True, the actions are consecutive, but **a** might have been a mistake, and since **b** is ambiguous, I cannot argue that both belong to plan 1." We can model these different views in terms of semantic combining rules. The first advocate is governed by the rule shown above; the second by a more conservative rule:

If (plan N: step i could be a mistake -) and
(plan N: steps i j continuity is desirable +) and
(plan N: step j has only one interpretation +)
Then erase (plan N: step i could be a mistake -)

Consensus is more difficult to model, but involves comparing the semantic combining rules: The advocates disagree because the ambiguity of **b** is significant to one and not to the other. Several intermediate, consensus positions can be worked out; for example, the third clause in the previous rule might be rewritten "only one interpretation of step j currently has support," which would lead to agreement on the plan 1 interpretation of **a, b**, since all the endorsements of the plan 2 interpretation of this input are negative, as discussed above.

§4 REPRESENTATIVENESS

Our second approach to endorsement-based reasoning addressed two related issues: However informative our endorsements appeared, they were still comments added post hoc to deductive inferences. This was an inevitable consequence of the parallel certainty inferences, but it leaves open the questions "where do endorsements come from, and what do they mean?" At the same time, we became interested in associative, as opposed to deductive, reasoning. Expert and commonsense reasoning derives its power from empirical associations – propositions that "go together" in the mind. We thought that the manner in which the propositions go together – the associations between them – could tell us something about the credibility of inferences based on those associations. Unfortunately, the nature of an association (be it causal, temporal, etc.) is lost when it is represented as an implication; and most empirical associations are represented this way, as production rules. The failure of this approach, it seemed, was that implications are interpreted deductively – as logical forms – and thus the parallel certainty inference model is inescapable. Yet, if the associations that underlie inferences are not discarded in the translation from empirical association to implication, then they might be used to assess the credibility of the inferences. By focusing on associative reasoning, we eliminate the parallel certainty inference model, and the questions it raises about the meaning of endorsements. At the same time, we gain a set of associations that are a natural basis for statements about the credibility of interpretations – endorsements.

We addressed a single associative task and its inherent uncertainty. The task was classification, prevalent in AI systems. Partial matching occurs whenever the criteria for a classification are not met exactly. For example, given the empirical association *a person with queasiness, fatigue, aching limbs, and a fever has flu*, what can we say of an individual with a poor appetite, headache, marginal fever, and a twitch in the left eye? Flu might be considered if poor appetite is evidence of queasiness, headache is evidence of aching limbs, and so on, although no reasonable interpretation of the twitch seems possible. The partial matching problem has two forms: Not all the criteria for a classification are met, and those that are may be met approximately. The individual above did not complain of fatigue – one of the criteria for flu – and did not complain of other criteria exactly as stated. Yet all complaints but the twitch could be interpreted as evidence for flu.

One form of the partial matching problem – inexact matches between evidence and classification criteria – reminded us of Tversky and Kahneman's notion of representativeness:

> Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that object A belongs to class B? What is the probability that event A originates from B? What is the probability that process B will generate event A? In answering these questions, people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. (Tversky and Kahneman, 1982, p.4)

One's certainty in a diagnosis of flu depends on the degree to which the symptoms are representative of flu. Returning to the previous example, is poor appetite (a symptom) representative of queasiness (a criterion for flu)? Is a headache representative of aching limbs? We accepted the idea that representativeness mediates credibility in classification tasks, and set about implementing judgments of representativeness. What follows is a necessarily abbreviated discussion of the work presented in Cohen et al. (1985).

Our approach is to represent propositions as structured objects (frames) and to measure representativeness in terms of the associations that hold between structures. For example, consider this structure for the concept *tobacco*:

<u>Tobacco</u>

PART-OF cigarettes
CAUSE cancer

Is the inference *cigarettes cause cancer* credible? If so, it must be justified by the PART-OF association between tobacco and cigarettes. Consider a similar case:

<u>Industrial emission</u>

HAS-PART: carbon dioxide
CAUSE : acid rain

Is it credible to infer that carbon dioxide causes acid rain? If not, it must be because carbon dioxide is only one part of industrial emissions, and thus cannot credibly be implicated as a cause, given other possible causes. The credibility of the first example is due to the PART-OF relation, but the HAS-PART relation does not support causal arguments credibly. We say that the one relation *preserves representativeness* and the other does not. More formally:

$$\frac{\begin{array}{l}\text{CAUSE(A,B)}\\ \text{PART-OF(A,C)}\end{array}}{\text{CAUSE(C,B)}} \qquad \text{but,} \qquad \frac{\begin{array}{l}\text{CAUSE(A,B)}\\ \text{HAS-PART(A,C)}\end{array}}{** \ \text{CAUSE(C,B)}}$$

The ** indicates an unrepresentative, or less credible, conclusion. This style of reasoning has much in common with Collins' plausible reasoning (1978). The central idea is that the credibility of a conclusion given evidence is determined by the associations that relate evidence and conclusion in a network of concepts. Representativeness, or credibility, is a function of associations. In the previous example, cigarettes may be considered to cause cancer, given the evidence that tobacco causes cancer, because the agent in the evidence (tobacco) is a PART-OF the agent in the conclusion (cigarettes).

We have developed a program to reason associatively in a classification task, and to qualify its conclusions as representative or unrepresentative based on the associations that underlie classifications. The domain of the program, called GRANT, is research funding. A proposal (P) states a researcher's interests. GRANT finds a set of funding agencies (A) that are representative of the proposal. The agencies judged representative of a proposal by GRANT were judged in turn by an expert to be those most likely to fund the proposal. I will briefly describe the structure, function, and testing of GRANT.

GRANT has two kinds of knowledge about the world. One is a network of about 1000 research concepts, and the other is a set of rules, like those above, for finding representative connections between concepts. Research proposals and the interests of funding agencies are represented as structures attached to the network of research concepts. For example, an agency that wants to study the effects of *diet* on *cardiovascular disease* is linked to the network by these topics. A researcher may propose to study the effects of dietary sodium on atherosclerosis. Is this representative of the agency? That is, is the agency likely to fund the proposal? The answer depends on the associative relations between the topics subsumed by the agency and the proposal. As it happens, dietary sodium is PART-OF diet, and atherosclerosis ISA cardiovascular disease. Both relations have been shown to preserve representativeness, so the proposal is representative of the agency's interests.

Most of the knowledge engineering effort on the GRANT project has been devoted to discovering, from a human expert in the domain of funding, the associations that link research concepts, and the rules that state whether the associations and their combinations preserve representativeness. Combinations of associations are called *path endorsements*. A set of *traversal rules* restricts GRANT to consider the associative pathways that are adequately endorsed. Given the rules, a program can find sources of funding for which a proposal is representative, simply by starting in the associative network of research concepts at the proposal and spreading activation through the network over representative associations until it "bumps into" topics associated with one or more funding agencies.

The use of traversal rules constrains blind spreading activation, and finds the agencies most likely to fund a research proposal. We tested this claim with 23 research proposals. For each, we ran GRANT in two modes. One, called *minimum-distance search* (MD for short) spread activation without regard for representativeness until it found, on average, 15 agencies. The other mode, called *endorsement constrained* (EC for short) searched until it found all representative agencies. For each proposal, the agencies found by MD search were given to an expert, who was asked to select the agencies most likely to fund the proposal. For the 23 proposals, the expert selected an average of 2 agencies from the average of 15 found by MD search. EC search found, on average, 80% of the agencies judged good by the expert. However, about 32% of the agencies found by EC search were not judged good by the expert. Thus EC search has a respectable hit rate (80%) and an adequate false positive rate (32%). Its false positive rate, compared with that of MD search under a variety of stopping conditions, is much superior. This suggests that the rules that govern EC search do, in fact, discriminate between representative and unrepresentative alternatives. Moreover, they effectively capture the determinants of the likelihood that an agency will fund a proposal.

Path endorsements have not yet been shown to mediate explantation and advocacy and consensus reasoning. Endorsements and combining rules of the kind discussed in Sullivan and Cohen are better suited to these purposes. However, the semantics of path endorsements is much clearer than those added to a parallel certainty inference model. Current research is devoted to attaining advocacy and consensus reasoning in GRANT.

§5 REFERENCES

Barr, Avron. 1985. Systems that don't understand. In the proceedings of Cognitiva: Artificial Intelligence and Neuroscience. Centre d'Etudes des Systemes et des Technologies Avancees, Paris.

Buchanan, B.G. and Shortliffe, E.H. 1984. Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project. Addison Wesley, 1984.

Collins, A. M. Fragments of a theory of human plausible reasoning. TINLAP-2. p.194-201.

Cohen, P.R., 1984. Progress report on the theory of endorsements: a heuristic approach to reasoning about uncertainty. IEEE Workshop on Principles of Knowledge-Based Systems. p. 139.

Cohen, P.R., 1985. Heuristic reasoning about uncertainty: An artificial intelligence approach. Pitman Research Notes. London. (Also published as Stanford University Technical Report 83-986, 1983)

Cohen, P.R., Davis, A., Day, D., Greenberg, M., Kjeldsen, R., Lander, S., and Loiselle, C. 1985. Representativeness and uncertainty in classification systems. AI Magazine. Fall 1985.

Cohen, P.R., and Grinberg, M.R. 1983. A theory of heuristic reasoning about uncertainty. AI Magazine, Summer, 1983.

Doyle, J. Some theories of reasoned assumptions: Essays in rational psychology. CMU-CS-83-125. Carnegie-Mellon University.

Sullivan M., and Cohen, P.R. 1985. An endorsement-based plan recognition program. IJCAI 1985, forthcoming.

Tversky, A., and Kahneman, D. 1982. Judgment under uncertainty: Heuristics and Biases. In Judgment Under Uncertainty: Heuristics and Biases, Kahneman, D. Slovic, P. and Tversky, A. (Eds.) Cambridge University Press.