Progress Report on the Theory of Endorsements: A Heurisitic Approach to Reasoning About Uncertainty.

Paul R. Cohen

COINS Technical Report 84-15

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

Abstract

This paper presents an approach to heuristic reasoning about uncertainty, called the *theory of endorsements*. In the first part of the paper, we describe an implementation of the theory – a program called SOLOMON. In the second half we concentrate on the problem of combining evidence in the theory of endorsements.

1. Introduction

In the theory of endorsements, uncertainty is represented, not as a numerical and passive comment on the credibility of beliefs, but as a body of richly structured knowledge *about* uncertain situations; including, but not limited to, *reasons* for believing and disbelieving. For example, we know whether evidence is currently available, whether it will become available and whether its "time of arrival" is predictable; we know when active seeking will produce evidence, and when passive waiting is most expeditious; we know the costs of obtaining evidence, and its diagnostic worth; we know what our evidence will be used for; we know the assumptions upon which the credibility of our evidence is based; we have a rich source of heuristic strategies (some summarized in the statistical literature on experimental design) for collecting and arranging evidence for maximum effect. Remarkably, in a world in which nothing is certain, we use our knowledge about uncertainty to behave *as if* almost nothing is uncertain.

We base the theory of endorsements on the assumption that numerical subjective degrees of belief (e.g., Shortliffe and Buchanan, 1975; Duda, Hart, and Nilsson, 1976; Shafer, 1976; Lowrance, 1982) are summaries of one's knowledge about uncertain situations: that they are *constructed* by intellectual effort from this knowledge (Shafer and Tversky, 1983). The fundamental contributions of the theory of endorsements are to *make explicit* this knowledge about uncertainty and evidence that would otherwise be summarized in a number, and to show how to reason with this knowledge *directly*, instead of indirectly through a numerical calculus. We believe that numbers are generally poor representations of our knowledge about uncertain situations, and that their semantics are often unclear. This position is argued at length in Cohen (1983); here we focus on an alternative approach.

2. SOLOMON -- An Implementation of the Theory of Endorsements.

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 (gleaned from a program called FOLIO; see Cohen and Lieberman, 1983). Our implementation of the theory of endorsements, a program called SOLOMON, reasoned about the uncertainty associated with these heuristics and their use. All such reasoning was mediated by structures called endorsements that represented reasons to believe and disbelieve their associated propositions. Endorsements are frame-like knowledge structures representing reasons to believe (*positive* endorsements) and disbelieve (*negative* endorsements). They are associated with propositions and inference rules at various times during reasoning. Five classes of endorsements appeared important for reasoning about uncertainty in rule-based systems:

Rule endorsements. Reasons to believe and disbelieve inference rules (e.g., a clause in a premise may be endorsed as *maybe-too-restrictive*, that is, the premise might occasionally fail due to this clause when the conclusion is in fact valid.)

Data endorsements. Reasons to believe and disbelieve raw data (e.g., a statement about one's own tolerance of risk is often *conservative*)

Task endorsements. Arguments about the evidence that executing tasks are likely to produce, used to schedule the tasks (e.g., a task is worth doing because it may produce a *corroborating* conclusion.)

Conclusion endorsements. Reasons to believe and disbelieve conclusions. These are combinations of a priori rule endorsements and detected relationships -- such as corroboration -- between conclusions (e.g., a *conservative* conclusion about one's risk tolerance is *corroborated* by other evidence.)

Resolution endorsements. Records of the application of methods to resolve uncertainty (e.g., no rules conclude a desired goal, but after eliminating a *maybe-too-restrictive* clause from a rule, we achieved the desired conclusion.)

The style of reasoning mediated by these endorsements is, by design, similar to the goal-directed reasoning of many expert systems: SOLOMON starts by trying to conclude a goal, usually the value of a parameter, such as risk-tolerance in the domain of investments. It then backchains through its rulebase, directed by this goal and its subgoals. As it proceeds, SOLOMON develops bodies of endorsements -- reasons to believe and disbelive its conclusions. These provide justifications for the conclusions, and also play a role in the control of SOLOMON's reasoning.

It is important that endorsements should affect control of processing in SOLOMON, because the theory of endorsements is oriented towards the effects of uncertainty on behavior. In SOLOMON these effects were two: 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 is similar to setting a threshold on the numeric degree of belief that a conclusion must accrue in a backchaining system (e.g., MYCIN set a global threshold of 0.2.) However, the "threshold" is determined dynamically for each goal and applied to its subgoals' endorsements; and the threshold is not a quantity but a boolean combination of desirable and undesirable endorsements. Importantly, a proposition that is not certain enough for one task may serve for another; 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 is achieved, in SOLOMON, by *resolution tasks*. The principle of these tasks is that negative endorsements are viewed as problems to be solved. SOLOMON will attempt to improve the endorsement of an important proposition. It has available general and domain-specific rules for resolving uncertainty. For example, when it is unable to derive a desired conclusion from its available rules, it can make small modifications to the premises of the rules, such as dropping clauses. Clauses to drop are selected by their endorsements; SOLOMON will not drop clauses endorsed as *criterial*. Dropping clauses results in additional endorsements noting the uncertainty that it introduces (see Cohen, 1983, pp. 148-158, for a detailed example).

In addition to rules to decide when a proposition is certain enough for a task, and rules for resolving uncertainty, SOLOMON had a simple rule to combine endorsements and propagate them over inferences. This was that a conclusion inherits all endorsements of its premise, plus any that result from posting the conclusion (such as a contradiction between the conclusion and another). In fact, this rule was doubly flawed: First, reasons to believe or disbelieve a premise are *not* always endorsements of the conclusion; and, second, the rule led to large bodies of endorsements after only a few inferences. The remainder of this paper reports recent work on the problem of combining endorsements.

3. Combining Endorsements

Combining evidence is something that numerical approaches to uncertainty do very well, because they represent uncertainty as a quantity increased or diminished by evidence. We do not represent uncertainty as a quantity: We represent it in terms of knowledge about evidence, and we do not summarize this knowledge in a degree of belief. Thus, it is not as easy to combine evidence in the theory of endorsements as it is in quantitative theories. If there is evidence from more than one source for a proposition, we must "calculate" a body of endorsements for the proposition by combining the endorsements of each piece of evidence. Simple syntactic union of the endorsements leads to the problems mentioned above: Large bodies of endorsements result, and not all endorsements remain relevant for all uses of their associated propositions. We are exploring *semantic combining rules* for endorsements -- so called because the combination of endorsements is mediated by rules that reflect what the endorsements mean.

A related problem is *ranking* endorsements. Again, quantitative approaches can rank the credibility of hypotheses easily, and again, it is more difficult with endorsements. However, endorsements can be ranked on an ordinal scale, if not an interval one, and so schemes for ranking endorsements can be designed. This is the subject of a research note in preparation.

The domain in which we are exploring issues of combining endorsements is plan recognition. In this task, a person types instructions to accomplish plan steps, and we try to determine which of several known plans the person has in mind, given knowledge about the plans and about the person. (A more sophisticated plan recognition system, called POISE, is the model for the the tasks we describe here. It is discussed in Carver, Lesser, and McCue, 1984.) Plan recognition is uncertain because an instruction may suggest more than one plan, and it may be a mistake. For example, imagine only two simple plans, each composed of just three steps:

plan 1: a b c
plan 2: b d e

If the first input is a, this results in a strong reason to believe that the most likely explanation (MLE) of **a** is plan 1, namely, there is no other explanation of a. Note, however, that **a** might have been a mistake. The endorsements of the statement MLE(a, plan1) are thus

MLE(a, plan 1)
1. no other explanation of **a** – positive
2. a may be a mistake – negative.

The next input is **b**. A possible explanation (PE)[1] is that **b** starts plan 2. However, our system knows some rudimentary facts about people that it uses as endorsements of interpretations of user actions. One such fact is that people prefer

1 We use the predicate PE, instead of MLE, to indicate that we haven't yet decided which is the most likely explanation of the input.

to do one thing at a time. This is an argument against the plan 2 explanation of **b**, because plan 1 is already believed to be underway. Note that the same fact is used as a *positive* endorsement of the plan 1 explanation of **b**:

PE(**b**, plan 1)
1. people prefer to do one thing at a time – positive
2. there is another explanation of **b** – negative
3. **b** may be a mistake – negative

PE(**b**, plan 2)
1. people prefer to do one thing at a time – negative
2. there is another explanation of **b** – negative
3. **b** may be a mistake – negative

Since **b** can continue plan 1, it is *further evidence* that the user has plan 1 in mind. We now consider how to combine the endorsements of plan 1 thus far, with the endorsements of the plan 1 explanation of **b**. The former endorsements are

MLE(**a**, plan 1)
1. no other explanation of **a** – positive
2. **a** may be a mistake – negative.

We invoke a *semantic combining rule*:

SCR-1: If an explanation of a step is negatively
endorsed by "may be a mistake," and
the successor of the step is the next input
Then drop the endorsement.

In other words, since we got **b** immediately after **a**, we no longer believe that **a** could have been a mistake. The endorsements of plan 1 are thus

plan 1
1. no other explanation of **a** – positive
2. people prefer to do one thing at a time – positive
3. there is another explanation of **b** – negative
4. **b** may be a mistake – negative

Consider what happens if the next input is **d**. Now it appears that plan 1 has been suspended or was never intended (i.e., a was a mistake). Since d supports the interpretation that the user intends plan 2, we must ask how the endorsements of the plan 2 interpretation of b combine with the endorsements of d. These endorsements are:

PE(b, plan 2)
1. people prefer to do one thing at a time – negative
2. there is another explanation of b – negative
3. b may be a mistake – negative

MLE(d, plan 2)
1. no other explanation of d – positive
2. people prefer to do one thing at a time – positive
3. d may be a mistake – negative

To combine these endorsements we invoke SCR-1 to eliminate the concern that b might be a mistake. We also use a similar rule to eliminate the "people prefer to do one thing at a time" endorsement:

SCR-2: If an explanation of a step is negatively endorsed
by "people prefer to do one thing at a time," and
the successor of the step is the next input
Then drop the endorsement.

That is, we may have disbelieved that b started plan 2, since plan 1 was already open and people prefer to do one thing at a time. But since d continues plan 2, we believe that plan 2 was the intent of b, and the negative endorsement is erased. Thus, the combined endorsement of plan 2 is:

plan 2
1. there is another explanation of b – negative
2. no other explanation of d – positive
3. people prefer to do one thing at a time – positive
4. d may be a mistake – negative

We have not yet considered how an input such as d could constitute evidence against the interpretation that the user intends plan 1. Evidence for one plan is, in this domain, evidence against another, but we are eager to see whether endorsements of opposing interpretations can be combined. This is for the all-important reason that we want to reason about evidence in domains where the hypotheses are not necessarily mutually exclusive and exhaustive. Consider this example of how such reasoning might proceed: One of the positive endorsements of plan 1 is that people prefer to do one thing at a time; this results from the interpretation of **b** as following **a** in plan 1. But if plan 2 is, in fact, the intended plan, then **b** was intended as part of it, and the positive endorsement just mentioned is invalid. Thus, the strength of the positive endorsement depends on how much we believe plan 2 was intended. And so, the endorsement of plan 1 *can* be changed, but only by combination with the endorsements of the competing interpretation. It is this condition – that of including endorsements of competing interpretations – that makes calculating the *negative* import of evidence more complicated that calculating its positive impact.

Finally, we note that semantic combining rules that drop endorsements will occasionally produce curious bodies of endorsements. SCR-1 dropped the possibility that **a** was a mistake, which, if it still existed, we could use as further support for the plan 2 interpretation of the user's actions. We could "tune" the combining rules to avoid this kind of situation, but we prefer not to: Our goal is a plausible theory of heuristic reasoning about uncertainty, and so we are just as interested in plausible errors as other behaviors.

4. Summary

The theory of endorsements represents uncertainty in terms of statements about evidence, particularly reasons to believe and disbelieve evidence. These statements, called endorsements, can be used to justify beliefs, and to control problem solving. The latter is desirable because the theory is intended as a model of how uncertainty affects behavior. A program, SOLOMON, explored some aspects of the theory of endorsements. More recently, we have explored the problem of combining endoresements in the plan recognition task.

References

Carver, N., Lesser, V., and McCue, D. Focusing in Plan Recognition, AAAI-84, in press.

Cohen, P. R. , 1983. Heuristic reasoning about uncertainty: An AI approach. PhD Dissertation. Stanford University. Report Number STAN-CS-83-986

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

Cohen, P. R., and Lieberman, M. D . A report on FOLIO: An expert assistant for portfolio managers. IJCAI-8, 212-215

Duda, R. O ., Hart, P .E ., and Nilsson, N., 1976. Subjective Bayesian methods for rule-based inference systems. Technical note 124, AI Center, SRI International, Menlo Park, CA.

Lowrance, J. 1982. Dependency-graph models of evidential support. PhD dissertation, Department of Computer and Information Science, University of Massachusetts, Amherst, MA.

Shafer, G. 1976. A Mathematical Theory of Evidence. Princeton University Press, Princeton, New Jersey.

Shafer, G. and Tversky, A., 1983. Weighing evidence: The design and comparison of probability thought experiments. Under review.

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