

THINKING NATURALLY ABOUT UNCERTAINTY

Marvin S. Cohen and Jared T. Freeman
Cognitive Technologies, Inc.
Arlington, VA

Methods for handling uncertainty should be evaluated in terms of their cognitive compatibility with real-world decision makers. Bayesian models of uncertainty demand precise up-front assessments of all problem elements and discourage the dynamic evolution of problem understanding. They handle missing or conflicting data by mathematical aggregation, while real-world decision makers regard gaps in knowledge and conflicting evidence as problems to be solved. Finally, they produce as output a statistical average rather than a coherent picture of the situation. Another approach to decision making, based on pattern-matching, does not address the ways in which situation pictures are evaluated and modified. A third approach, however, called the Recognition / Metacognition model, treats decision making under uncertainty as a problem-solving process that starts with the results of recognition, verifies them, and improves them where necessary. Critiquing strategies identify problems of incompleteness, conflict, and unreliability in situation models, and lead to correcting steps that retrieve or collect additional information or adopt assumptions. Training methods based on this model have been developed and tested with active-duty Naval officers.

INTRODUCTION

Is probabilistic reasoning the only way to handle uncertainty? It may seem to some that the answer must be *yes* by definition; uncertainty is defined as probabilities other than one or zero. To others, the hoary tradition of probability theory, stretching from Pascal and LaPlace to its elaboration in modern statistics and decision theory, provides convincing support for this answer. Yet few researchers are still prepared to argue, as many did three decades ago, that real-world decision makers correctly use probabilistic reasoning in real-world tasks. Naturalistic models, based on observation of experienced decision makers rather than laboratory studies of artificial tasks, have focused on mechanisms for situation assessment like pattern recognition or schema instantiation that match cues to stored knowledge. In doing so, they have turned attention away from uncertainty as such. Yet no one even vaguely familiar with domains such as combat, business, or medicine is prepared to argue that uncertainty is an unimportant or incidental feature of real-world decision making environments.

How do real world decision makers handle uncertainty, and how can they be trained or aided to handle it better? Must they choose between inappropriate “intuitive” approaches based on pattern matching, on the one hand, and technically difficult and “unnatural”

analytical methods, on the other? We contend that the study of naturalistic decision making is now mature enough to face uncertainty head on. Is there a *naturalistic* way to handle uncertainty? If so, it will be distinct from formal probabilistic reasoning, and it will go beyond pattern matching while nonetheless building on it. We shall first review two of the most popular current approaches, before describing a new model, training based on the model, and an empirical test of the training.

Probabilities, Patterns, or Problem-Solving?

The probabilistic approach starts with an abstract view of how uncertainty *ought* to be dealt with and provides an elaborate formal apparatus to do the job. By means of this apparatus, the unknown probability of a hypothesis can be expressed in terms of other probabilities that are more easily estimated or assessed. The decision maker generates a set of exhaustive and mutually exclusive hypotheses, makes numerical assessments of the probabilistic relationships between the hypotheses and evidence, and combines the numerical assessments to get probabilities for the hypotheses.

When compared with such normative models, real-world decision makers consistently fall short, revealing systematic errors, or “biases.” But the models themselves may also fall short in various ways as *intellectual tools*. The idea of an effective intellectual tool has to do with *cognitive compatibility* between the tool and the decision

maker: (1) Does the tool require inputs about which the decision maker has confident and precise intuitions? (2) Are the operations applied by the tool to the inputs understandable and plausible? (3) Does the tool produce as outputs judgments or choices that the decision maker cannot confidently make, but which he needs? In short, a good intellectual tool matches the pattern of a decision maker's knowledge and ignorance: it uses what he knows to generate what he needs to know, using reasoning processes that he trusts.

Decision analytic methods do not fair well when evaluated in terms of cognitive compatibility. (1) Can decision makers confidently and easily provide the inputs required by formal uncertainty models? Often not. The most obvious problem is sheer effort and time: Very large numbers of exact numerical probability judgments are demanded even by the simplest real-world problems. A more subtle problem, however, is that these methods require that the problem be completely modeled up front. They foreclose (or at least discourage) the dynamic process by which decision makers discover new hypotheses, new kinds of evidence, new options, and even new goals as a problem unfolds in time.

(2) Do the reasoning processes applied to the inputs make sense? Again, there are serious problems. For example, laws of probability treat conflicting data the same way they treat congruent data, by quantifying the force of the evidence and then mathematically aggregating. From the point of view of a real-world decision maker, this resort to mathematical aggregation may often seem overhasty. It is not dealing with uncertainty so much as sweeping it under the rug. If two sources who are each regarded as highly reliable contradict one another, the decision maker is not likely to take an average. He usually wants to understand why the sources disagree, and may well learn (or assume) that one or both of them is not as reliable as he supposed. In other words, conflict is a problem that needs to be solved. It can be a symptom of erroneous assumptions about the evidence or its meaning. Finding and correcting mistaken beliefs can lead to a more plausible judgment in the present case and enduring lessons for future situations.

(3) Finally, are the outputs of formal analyses useful for decision makers? Again, the verdict is not wholly positive. The output is not a coherent picture of the situation but an unrealizable abstraction: an assignment of probabilities across propositions (e.g., 30% chance the enemy will attack; 70% chance the enemy will not attack). Such an abstraction cannot be visualized, anticipated, or planned for in the real world. In exchange

for this loss of concreteness, we should at least get a good depiction of uncertainty itself. Paradoxically, we do not. Bayesian models do not make allowances for varying degrees of knowledge underlying the conclusions. The probability of an event might be judged .5 because of complete ignorance regarding two possible hypotheses *or* because there is a mass of conflicting evidence pointing in both directions. The decision maker is expected to treat both cases the same way.

A second approach looks in a different direction for the nature of decision making skill, toward a set of virtually automatic responses to recognized patterns. This view has been popular in research on differences between experts and novices, beginning with Chase and Simon's work on chess. Experts are thought to be distinguished from novices by the accumulation of a large repertoire of patterns or schemas over the course of their experience in a specific domain. This point of view — with its emphasis on rapid, automatic, and domain-specific processes rather than slow, deliberative, and general-purpose ones — seems diametrically opposed to the decision theoretic approach. Unfortunately and ironically, normative methods and pattern recognition have some parallel shortcomings. They both give short shrift to the problem-solving processes that are often triggered in the real world by gaps and conflicts in data. What do real-world decision makers do when the elements of a pattern are incomplete, or when the data do not perfectly match any single pattern, but partially fit conflicting patterns instead? How do decision makers create a picture of the situation, evaluate it, then change their minds? How do they balance the need to act with the need to think more about a problem? Although recognition may be at the heart of proficient decision making, other processes are often crucial for success.

What is needed is a descriptively based account of the problem-solving strategies that experienced decision makers bring to bear on uncertainty. Such an account may in turn provide the foundation for cognitively compatible tools for training and decision aiding.

A Naturalistic Model of Critical Thinking

Our observations of decision making performance, in Naval anti-air warfare as well as other domains, suggest that recognition is supplemented by processes that verify and improve its results. Because of their function, we call these processes *meta-recognitional*. Meta-recognitional skills probe for flaws in recognized assessments and plans, try to patch up any weaknesses that are found, and evaluate the results. They are analogous to the *meta-comprehension* skills that

proficient readers use when they try to construct a mental model based on the information in a text. For example, according to researchers in this area, skilled readers continually test and evaluate the current state of their comprehension, and they adopt a variety of strategies for correcting problems that are found, such as inconsistencies or gaps in their understanding.

To reflect the complementary roles of recognition and metacognition in decision making, we have called this framework the Recognition / Metacognition (R/M) model. In the R/M model, the basic level of cognition is recognitional, including processes that activate assessments in response to internal and external cues. Assessments may in turn be associated with structures that organize actual and potential information into a situation model or plan. According to the R/M model, the integration of observations into situation models and plans often occurs under the influence of metacognitive control. Meta-recognitional processes include:

1. Identification of evidence-conclusion relationships (or *arguments*) within the evolving situation model and plan. This is simply an implicit or explicit awareness that cue A was *observed* on this occasion, while some conclusion (e.g., intent to attack) along with expectations of observing cue B were *inferred*. On some other occasion cue B might be observed and cue A inferred.

2. Processes of *critiquing* that identify problems in the arguments that support the situation model or plan. Critiquing can result in the discovery of three kinds of problems: *incompleteness*, *unreliability*, or *conflict*. A model or plan is incomplete if expected arguments are missing; that is, some expected information has not been obtained that would ordinarily confirm or disconfirm the assessment. An argument is unreliable if it depends on implausible assumptions. An argument in support of an assessment is conflictual if there are other arguments that support contradictory assessments.

3. Processes of *correcting* that respond to these problems. Correcting may instigate additional observation, additional information retrieval, revision of assumptions, and/or the generation of new hypotheses, options, or goals. These processes fill gaps in models or plans, resolve conflict among arguments, and substitute more reliable assumptions for less reliable ones.

4. A control process called the *quick test*, which regulates critiquing and correcting. The quick test considers the available time, the costs of an error, and the degree of uncertainty or novelty in the situation. If conditions are appropriate, the quick test inhibits recognitional responding and interposes a process of

critical thinking. If conditions are inappropriate, it allows immediate action based on the current recognitional response.

The R / M model explains how experienced decision makers are able to exploit their experience in a domain and at the same time handle uncertainty and novelty. They construct and manipulate concrete, visualizable models of the situation, not abstract aggregations (such as 70% chance of hostile intent, 30% chance non-hostile). Uncertainty is represented explicitly at the metacognitive level, by “annotating” the situation model or plan to highlight points of incompleteness, conflict, and unreliability. In response to specific problems of this kind, metacognitive strategies try to improve the current situation model and plan or find better ones. Such strategies are highly dynamic and iterative: For example, gaps may be filled by collecting further data or by making assumptions (e.g., the worst case). The resulting arguments may then be found to conflict with other arguments. Such conflict may be resolved by evaluating the reliability of the assumptions in the conflicting arguments. This process stops when the quick test indicates that immediate action on the basis of the current best model is called for. The output is a coherent, consistent model of the situation together with an understanding of its strengths and weaknesses.

Our hypothesis is that decision aiding and training based on the R/M model will be cognitively compatible with the way real-world decision makers think about uncertainty. In terms of inputs, such tools begin with a recognitional response rather than a complete decomposition of the problem; they require judgments only in response to specifically identified problems in the recognitional response and only as long as time and stakes permit; and they encourage a dynamic evolution of understanding. In terms of processes, they rely on problem solving rather than mathematical aggregation, stimulating collection and retrieval of new information, adjustment of assumptions, or generation of new hypotheses, options, or goals. Finally, their outputs combine the concreteness of a coherent situation picture with an explicit recognition of qualitatively different types of uncertainty.

Training to Handle Uncertainty

Each definition of decision making skill has its own implications for training. From the formal, normative point of view, training should convey a set of general purpose methods for structuring and quantifying evidence, hypotheses, options, and outcomes. Examples of applying the techniques are useful only incidentally,

for practice and to demonstrate the generality of the methods. By contrast, examples are central from the recognitional point of view, in which decision training focuses intensely on a particular application area. Training can accelerate the accumulation of experiences with the characteristic patterns of the relevant domain by using realistic simulations and outcome feedback. From the point of view of the R/M model, the focus of training is neither a small set of general-purpose methods nor a vast quantity of specialized patterns. Our focus is a moderately sized set of strategies for critical thinking, which are general but at the same time *build on the recognitional skills of a particular domain*.

We have developed and tested training methods based on the R/M framework in Army battlefield situation assessment (Cohen & Freeman, 1995) and in Navy anti-air warfare (Cohen, Freeman, Wolf, & Militello, 1995). The Navy training focuses on the situation in which an air or surface platform with unknown intent is approaching own ship. The most recent version of the Navy training begins with an overview of the process of building and improving situation models, and then, in three additional units, focuses on particular aspects of that process.

We call the overview of the critical thinking process *STEP*. *STEP* consists of a cycle of four strategies: (1) Building a Story based on an assessment and filling the gaps in that story. If an officer has assessed the intent of an approaching platform, in order to take that assessment seriously, he must consider the events in the past and future that it implies. Such a story consists of events that could have motivated the intent and other events that would effectively achieve the intent. (2) Testing the story to identify data or knowledge that conflict with it, and attempting to revise the story to explain the conflicting information. Decision makers do not simply drop an assessment in the face of conflict; if they did, they would be paralyzed in situations where no familiar pattern fits all the data. Rather, they explore assumptions that would be sufficient to explain the conflicting evidence. (3) Evaluating the story and the assumptions on which it is based. Simply being able to construct a story around an assessment doesn't mean the assessment is true. Decision makers now step back and ask if the story makes sense. Can the assumptions be confirmed by other data, or accepted as the most plausible possibilities? If not, the decision maker may consider another assessment, and attempt to construct, test, and evaluate a new story. (4) Formulating contingency Plans to protect against unreliable assumptions in the current story. At any given

time, the decision maker has a best available assessment along with an appreciation of its weaknesses. Contingency plans guard against the possibility that assumptions in the story turn out to be false.

After receiving an overview and practice with *STEP*, in the second unit of training officers study and apply a particularly important kind of story, based on hostile intent. Training helps them identify issues that experienced decision makers typically consider in assessing whether intent is hostile. A complete hostile intent story includes information or assumptions about why the country involved would want to attack, why they chose the relevant platform as the attack vehicle, why they chose a particular asset as the target of attack, how the platform was able to localize the target, how the platform is managing to protect itself while arriving at a position for engagement, and what the platform must do in order to execute an engagement.

The third unit focuses on explaining conflict and generating alternative hypotheses. It presents a variant of the devil's advocate technique that forces officers to generate alternative interpretations of the evidence.

The fourth and final unit of training provides guidance regarding when critical thinking is appropriate (i.e., the quick test). It provides strategies for assessing the available time, the stakes, and the uncertainty in a situation in order to decide when it is necessary to commit to an action.

In all four units, training consists of presentation and discussion of the R/M concepts followed by practice and feedback with a simulated anti-air warfare scenario.

EMPIRICAL TESTS

Method

The training methods were tested in two experiments at United States Navy training facilities, one involving 60 officers (study 1) and the other involving 35 officers (study 2). In study 1, we were able to utilize a control group (which was exposed to a general discussion of knowledge representation and problem solving-strategies with practice examples from the participants' skill areas), as well as a pretest-posttest comparison. In that study, however, only 90 minutes were available for training, and practice utilized paper-and-pencil examples rather than automated simulations. In study 2, four hours were available for training and more realistic practice was possible due to the availability of automated simulations. In this study, however, there was no control group, and only the pretest-posttest comparison was available. Both

studies counterbalanced two scenario sequences for the pretest and posttest.

The evaluation examined the effects of training on both performance (i.e., assessments and actions) and on critical thinking processes in realistically simulated anti-air warfare scenarios. In both studies, each pretest and posttest scenario had three breaks, at which participants responded to questions. These questions probed for the likely intent of a designated track, degree of confidence in that assessment, and reasons for it, other possible intents, reasons in favor of one of the non-accepted intents, data that conflict with a designated intent, possible explanations of the conflicting data, and actions the participant would take or plan for. Dependent measures included the number of arguments generated for a hypothesis, the amount of conflicting evidence identified, success in explaining conflicting evidence, the number of alternative hypotheses generated, the accuracy of the accepted assessments, appropriateness of actions, and participants' own evaluations of the training (in a subsequent debriefing).

Summary of Results

Trained participants identified more evidence in favor of a hypothesis than untrained participants ($p=.001$ for the pretest-posttest comparison in study 2; $p=.092$ for the treatment-control group comparison in study 1). Perhaps more interestingly, trained participants identified significantly more evidence that conflicted with a designated hypothesis than untrained participants ($p<.001$ for study 2, $p=.015$ for study 1). Training also improved participants' ability to patch up stories by constructing explanations for the conflicting evidence ($p<.001$ for study 1, ns for study 2). Trained participants also generated more alternative assessments of possible intent than untrained participants ($p<.001$ for study 2, ns for study 1).

In both studies, training had a significant effect on assessments ($p=.001$ in study 1, $p=.002$ in study 2). In two of the four test scenarios (one of the two used in each study), training significantly increased agreement with the subject matter expert who had designed the scenarios ($p=.034$ in study 2, $p=.015$ in study 1). Actions taken by participants significantly varied in a way that reflected these assessments. Training did not reduce participants' confidence in the assessments that they accepted. It did, however, increase contingency planning in case assumptions proved wrong. Trained participants in study 1 (who were more likely to regard the designated contact as non-hostile) were more likely to make contingency plans for engagement of the contact than untrained

participants ($p=.005$ in study 1, not yet analyzed in study 2).

Finally, participants generally evaluated the training favorably both in numerical evaluations and in qualitative comments.

Conclusion

Trained participants were significantly better than untrained participants in all four aspects of the STEP method: building stories, testing and patching up stories, evaluating stories, and making contingency plans in case stories were wrong. This study provides a preliminary demonstration that meta-recognitional skills can be taught effectively, that decision makers will use them in relatively realistic tactical situations, and that use of such skills will improve outcomes. This is a promising avenue for the development of cognitively compatible tools for handling uncertainty based on naturalistic models of decision making.

ACKNOWLEDGMENTS

This work was supported by contracts N61339-92-C-0092 and N61339-95-C-0107 with the Naval Air Warfare Center / Training Systems Division. Klein Associates, Inc., served as a subcontractor in these efforts. Sonalysts, Inc., provided support in setting up the experimental simulations.

REFERENCES

- Cohen, Marvin S. and Freeman, Jared T. (1995). *Methods for Training Cognitive Skills in Battlefield Situation Assessment*. Arlington, VA: Cognitive Technologies, Inc.
- Cohen, Marvin S., Freeman, Jared T., Wolf, Steve, and Militello, L. (1995). *Training Metacognitive Skills in Naval Combat Decision Making*. Arlington, VA: Cognitive Technologies, Inc.