EXPERIMENTAL INVESTIGATION OF UNCERTAINTY, STAKES, AND TIME IN PILOT DECISION MAKING

March 2000

Marvin S. Cohen, Leonard Adelman, and Bryan B. Thompson

Cognitive Technologies, Inc.
4200 Lorcom Lane
Arlington, VA 22207-3306
(703) 524-4331

Prepared for:

NASA-Ames Research Center
Moffett Field, CA 94035-1000

Contract No. NAS2-14075

ACKNOWLEDGEMENTS

We wish to acknowledge the advice and support of Judith Orasanu throughout this project. Captain Gil Coshland (United Airlines) provided invaluable service as a consultant. Jared Freeman and Margaret Will helped generate the experimental materials. Last but not least, we wish to thank the 64 pilots who graciously agreed to participate in our study.

INTRODUCTION AND OVERVIEW

This report describes: (1) a model of decision making in real-world environments, and its background in cognitive science; (2) an experimental test of implication of the model with active duty commercial airline pilots; and (3) a possible use of that model together with the experimental findings to develop pilot training. The three chapters of the report correspond to these objectives.

In a previous study (Freeman, Cohen, & Thompson, 1998; Freeman and Cohen, 1996), we tested other implications of the same model. In that study, as well as this one, we used active-duty commercial airline pilots in a low fidelity simulation, varying key features of an enroute scenario. In both studies, we examining pilot diversion decisions and information requests, and how they varied with pilot seniority. In the previous study, we examined how pilots dealt with different constraints on the time available for the decision. In the present study, we look at how pilots deal with different degrees of uncertainty about the necessity for a diversion, and different stakes, i.e., penalties for bad decisions.

Both studies have interesting implications for pilot training in what we might call *real-time critical thinking*. These are explored in the final chapter of the present report.

CHAPTER 1:
AN APPROACH TO DECISION MAKING AND CRITICAL THINKING

The purpose of this chapter is to describe the theoretical and empirical background of our approach to human decision making and to training critical thinking skills. This discussion has two parts. The second part will describe a cognitive model that provides the context for the experiment (Chapter 2) and the training recommendations (Chapter 3). First, however, we will describe the context of that model in the framework of other research on cognition, decision making, and problem solving.

BACKGROUND FOR CRITICAL THINKING AND DECISION MAKING

There are at present a variety of major competing conceptions of what decision making is. One way to classify them is in terms of whether their primary emphasis is on general-purpose strategies (i.e., weak methods), highly specialized routines or patterns (i.e., strong methods), intermediate strategies, or some combination that is contingent on the characteristics of the task, context, and decision maker. The most significant finding from the review reported below is the importance of medium level *strategies* for identifying and resolving different types of uncertainty. Of primary interest to us is the evolution of the notion of uncertainty-handling as a species of problem solving, the definition of a relatively small number of distinguishable strategies, and the specification of conditions under which they might be used.

Strategies: General, Specific, and Intermediate

General-purpose strategies: Decision Theory.

The dominant framework for study of decision making for many years, classical decision theory, remains a towering intellectual achievement that exerts a strong influence on work in inference and choice. The theory contains two main parts: *Bayesian probability theory* for drawing inferences about any situation in any domain, and *multi-attribute utility theory* for selecting an optimal action in any domain. These can be regarded either as procedures that people explicitly follow or as descriptive constraints that apply to their behavior, but of which they may not be explicitly aware. Bayesian probability theory requires that decision makers consider a set of mutually exclusive and exhaustive hypotheses, each of which is assigned a probability. Each potential observation that might bear on those hypotheses is assigned a diagnostic strength. Then, as new observations occur, beliefs in the hypotheses are appropriately updated. Multi-attribute utility theory is an analogous method for choice. Choices are made based on a combination of the probability of each uncertain state, the importance of each evaluative dimension, and the score of each action-state combination on every evaluative dimension.

We (along with others) have argued that decision theory is not in general *cognitively compatible* with the way experienced decision makers work (Cohen, 1993; Cohen & Freeman, 1996). Problems include the kinds of inputs it demands, the kind of processing it prescribes, and the outputs it produces. (1) By demanding a complete model up front, with fixed assessments of uncertainty and preference, decision theory overlooks the dynamic evolution of problem understanding through time, e.g., as new hypotheses, options, observations, outcomes, and even goals are discovered. (2) By reducing all uncertainty to a single measure (probability), decision theory obscures important qualitative differences in the way different types of uncertainty are handled, such as gaps, conflict, and unreliable assumptions. Decision theory, for example, treats conflicting evidence the same way that it treats congruent evidence, by essentially taking an average. Experienced decision makers, on the other hand, may use conflict as an opportunity for *problem solving*, i.e., to identify the faulty assumptions in the beliefs that produced the conflict (Cohen, 1986). Similarly, decision theory handles conflicting goals the same way it handles congruent goals, by calculating an overall score for each option that is an average of the different goals. Experienced decision makers, by contrast, may try to learn from the conflict, by creating a better option or a deeper understanding of their true objectives (Levi, 1986). (3) The output of a decision theoretic model is a statistical average – e.g., 70% chance hostile, 30% chance not hostile – rather than a single coherent picture of the situation. Decision makers cannot visualize, anticipate, or plan effectively for an abstract average. They often prefer to prepare against a specific, concrete possibility, while either accepting risk or hedging with respect to others.

Many researchers have claimed that under time stress, behavior no longer conforms to decision theoretic precepts. Janis (1972) attributes this to the irrationality induced by time stress, while others (Johnston & Payne, 1976) see it as a rational adjustment to the lack of time. As opposed to both of these positions, there is evidence that *even when time is available*, proficient decision makers do not typically use systematic methods, e.g., generating and considering a large number of options or outcomes (Cohen, 1993l; Klein, 1993)

Behavioral Decision Making: Heuristics and Biases.

Problems with the use of decision theory to describe behavior led to a counter-movement in cognitive psychology that focuses its attention on systematic deviations of performance from the constraints of decision theory, i.e., “biases” (e.g., Kahneman, Slovic, & Tversky, 1983). This work was, unfortunately, as limited in its own way as formal decision theory. (1) It focused on highly simplified questions, with no context, designed specifically to elicit errors. Such studies are not likely to be ecologically representative of the problems people deal with in real-world settings, or to shed much light on the processing strategies they use (Christensen-Szalanski, 1993). (2) In many cases, the experimenters assume one interpretation of the problem and define the normatively “correct” answer based on that interpretati0on, when it is not the only plausible one. If alternative interpretations are considered, the subjects’ responses often are seen to be reasonable rather than irrational (Smithson, 1989; L.J. Cohen, 1981; Cohen, 1993). (3) The processing theory adopted by Kahneman and Tversky focuses on “heuristics” that are defined and motivated by the way the behavior deviates from normative theory, rather than being integrated into a more systematic framework of human information processing. (4) Finally, the formal decision theoretic approach and its flip side, the heuristics and biases approach, share a common problem: They both regard decision theory as the final normative standard of decision making, though they differ regarding people’s ability to adhere to it. However, it can be argued that decision theoretic models, as they are typically applied, are not only descriptively inadequate, but normatively inadequate as well. Appropriate normative principles must capture the relevant *qualitative* features of the decision making process. If a normative standard is to be used to identify decision making errors, the standard must be close enough to actual performance for the discrepancies to be meaningful (Cohen, 1993). Other approaches, which define errors more naturalistically, may shed more light on the true strengths *and* weaknesses of decision making.

Specialized strategies: Pattern recognition.

An altogether different approach to decision making skill looks toward an extremely large number of acquired rules. It identifies expertise in general, and decision making skill in particular, with the accumulation through experience of a set of virtually automatic responses to recognized patterns. On this view, people do not make "decisions"; they simply recognize the situation and retrieve the response that is "typical" for that situation (Klein, 1993). This view has been popular in research on differences between experts and novices, beginning with Chase & Simon's (1973) work on chess.

Although pattern recognition is a key ingredient in proficient performance, it may not be the only one. A problem with the pattern recognition model as the sole explanation for expertise is that it abandons the effort to identify strategies that are general across different domains, or that can recur in different contexts in the same domain. Instead it resigns us to the identification of literally thousands of highly specific, narrowly applicable rules or patterns. In particular, it offers no response to questions such as: How is situation assessment accomplished in new and changing circumstances? How are conflicting and unreliable data dealt with? How do decision makers change their minds? When do they stop thinking and act? The response to all such questions is merely a domain-dependent list of patterns and responses.

This limitation of pure pattern matching approaches is shared with what would seem to be the diametrically opposite approach: the identification of highly general *elementary information processes*, or atoms of computation. This approach, like pattern recognition, responds to questions about what a decision maker did with a list of processing operations. It requires different theoretical tools to create a level of description that might shed some light on consistencies in the ways that people deal with uncertainty.

Intermediate strategies.

Based in part on such findings, an intermediate position has been gathering momentum in recent years. Proficient decision makers appear to use informal thinking strategies (such as, *make predictions and test them*; *look for reasons against your own position*; *look for analogies to previous problems*) that are not as general as decision theory claims to be, but not as particular as domain-specific patterns. A variety of thinking strategies have been identified in studies of expert performance, as well as in reflections of practitioners. Such strategies have been found in studies of self-regulation or metacognition (Metcalfe & Shimamura, 1994), expertise (Ericsson & Smith, 1991), everyday reasoning (Voss, Perkins, & Segal, 1991), and decision making (Cohen, Freeman, & Wolf, 1996). Proposed metacognitive strategies include: self-monitor while memorizing material, and form a hypothesis and test in reading comprehension.

Baron (1985) identifies a *general form of critical thinking strategy*: (i) Propose a statement; (ii) think of a counterargument to the statement (e.g., think of a counterexample to a general statement; think of an alternative explanation in scientific theorizing); and (iii) modify the statement so the criticism no longer applies. Halpern (1998) presents a similar framework in the form of a sequence of questions: what is the goal, what is known, which skills will get you to the goal, and have you achieved the goal. A similar intermediate-to-general strategy form is described in Cohen et al.'s (1996) Recognition / Metacognition framework, where strategies are characterized in general as cycles of identifying and filling gaps, identifying and resolving conflicts, and finding and evaluating assumptions in arguments, while monitoring the relative costs and benefits of continuing.

Some strategies have turned out to be weaker than suggested. For example, *lengthy search* (i.e., generating as many alternative solutions as possible, as suggested by de Bono), is not correlated with superior outcomes. Moreover, some have argued (e.g., Perkins, 1992) that there may be too many strategies for decision maker to remember, consider, select, and apply. On the other hand, Hayes (1985) notes that it requires 10 years to acquire the body of knowledge needed for proficiency in complex domains. He speculates that there may be several hundred different plausible strategies. It remains a challenge, however, to understand how decision makers *select* the appropriate strategy for a particular decision problem.

Problem solving strategies.

An appealing way to understand the selection of strategies is to view decision making as a special case of problem solving. Strategies may be the result of decision makers' generating subgoals to deal with impasses during search for a solution in a problem space (e.g., Newel, 1990; Anderson, 1983).

Unfortunately, problem-solving researchers have thus far not explicitly addressed the central role of *uncertainty* and *risk* in decision making (Fischhoff & Johnson, 1990). None of the classes of strategies that are studied (e.g., breadth-first versus depth-first search, backward versus forward reasoning, subgoal generation) shed any *specific* light on the way decision makers deal with uncertainty. This has primarily been left to researchers in other areas (e.g., non-Bayesian inference theory), where work has been done, for example, on strategies reflecting epistemic caution versus epistemic risk in inference, or worst-case strategies in choice (e.g, Levi, 1986; Gardenfors & Sahlin, 1982). This work, however, has not been linked to mainstream work on problem solving.

Another issue concerns the tendency in problem solving work to treat the relation between weak and strong methods as mutually exclusive, with strong methods replacing weak ones with growing experience, through a process of *chunking* (Newell & Rosenbloom, 1981) or *compiling* (Anderson, 1981). Explicit declarative knowledge, which is used by general-purpose strategies, is supplanted by relatively automatic domain-specific recognitional procedures. One problem with this viewpoint is that much, and perhaps most, recognitional knowledge is acquired directly, e.g., through associative and/or reinforcement learning, rather than by compiling initially declarative information or instructions (Berry & Broadbent, 1987). Another difficulty is that recognitional and reflective processes appear to *interact with and enhance* one another (Cohen et al., 1998). Reflective skills build on a base of recognitional knowledge, and in turn help people add to and make better use of their recognitional knowledge. In fact, it is this interaction, we believe, that holds the key to understanding how humans deal with uncertainty. A full-scale problem solving approach has not yet been applied to decision making under uncertainty and stress.

Control over Strategies

Contingency models.

The problem solving approach can be seen as an instance of an even wider class of *contingency* models. Such models assert that strategies are selected based on properties of the task, the context, or the experience of the decision maker. For example, Klein (1993) argues that familiar situations are recognized quickly and the obvious response is implemented. In less familiar situations, on the other hand, another strategy prevails: The decision maker evaluates the most typical option by a process of mental simulation; if problems are found, the option is modified or rejected in favor of the next most typical reaction. Klein does not address issues of the cost of time required for mental simulation versus the potential benefits.

These issues are explicitly addressed by Payne, Bettman, & Johnson (1993), and Beach and Mitchell (1978). According to them, people adaptively adjust their decision making strategies in accordance with a cost-benefit balance between the demand for accuracy and the cost of being accurate. Payne et al., operationalized cost in terms of effort, defined as the number of elementary information operations required by a strategy. They picture the choice process as initially involving a set of "metacognitive productions that have as their actions the explicit (conscious) consideration of accuracy and error conditions…" Over time, these metacognitive processes become automatic, and are invoked directly by task features such as complexity, e.g., the number of options or the variance among probabilities and importance weights. These metacognitive choices can lead to the highly formal strategies dictated by normative models when accuracy is vital, or to highly approximate, abbreviated strategies, when time is more costly than errors. Unfortunately, as noted in Cohen (1993), this model does not tie either effort or accuracy to domain-specific knowledge, including recognitional patterns. It seems possible, for example, that experts might sometimes bypass the tradeoffs Payne et al, focus on: An immediate recognitional strategy could be *both* less costly *and* at least as accurate for an expert than more formal methods. Yet Payne et al.'s model does not permit this.

Hammond's cognitive continuum theory (1993) relates the choice of strategy type (in this case, *analytic* versus *intuitive*) to intrinsic properties of the task (e.g., redundancy and number of cues, continuous versus discrete distribution of cue values, linear versus nonlinear relation between cues and criterion, etc.) rather than personal familiarity or expertise (as in Klein's model). It might be possible, though Hammond does not do so, to formulate this model in terms of the Payne et al., framework as basing strategy selection on the relative effortfulness and likely accuracy of different strategies, as determined by the structure of the task stimuli.

A more long-range contingency hypothesis has been proposed by Holyoak (1991) in the area of expert problem solving. Holyoak argues that experts are not characterized by any specific processing strategy. For example, in some domains experts appear to use a recognitional strategy of working forward from the given to the goal (a strong method), while in other domains they use the more analytical strategy of working backwards from the goal to the given (regarded as a weak strategy). Experts adapt to the inherent constraints of the task, and perform it in whatever way is most efficient.

As we shall discuss shortly, Cohen et al. (1996) offer a model of contingent decision making which integrates features of the above models within a problem solving framework. In their Recognition / Metacognition model, the amount of time devoted to critical thinking about a recognitional response is a function of the familiarity of the situation (as in Klein's model), the amount and type of prior knowledge (as in Hammond's and Holyoak's approach), as well as the cost of errors and the cost of time (as in Payne et al.'s model).

The availability of alternative strategies, which are effective in different situations, implies an ability to choose either globally or locally. It seems plausible that persisting individual differences in the use of one or the other type of strategy might indicate differences in cognitive styles.

Cognitive style

Another factor that may influence the choice of strategy is an individual's cognitive style. Cognitive styles are regarded by Baron (1985) as stable, general dispositions to behave a certain way in mental tasks, and as the most general level of decision making skill that is learnable. Baron identifies two style parameters: (1) The amount of search for goals, possibilities, and evidence relative to the optimum range of the search processes. This dimension corresponds to the impulsivity (too little time spent searching) and reflectivity (too much time spent searching). (2) Whether the person is equally fair to possibilities that are already weak and strong in the search for additional evidence and in the use of that evidence. This corresponds to open-mindedness or flexibility versus a tendency to premature closure (Langer, 1989). According to Baron, these styles are usually under voluntary control (although they can be influenced by stress and other affective states). These parameter settings are affected by values, expectation, & habits, as well as by emotions and beliefs about one's self. They are also subject to long-term modification by learning. As a result of relatively persistent styles, decision making behavior should be correlated across moderately discrepant situations, and the styles themselves should be teachable in general form. Baron speculates that styles rather than strategies may account for observed differences in use of thinking strategies and for transfer effects in strategy training.

Epistemic attitudes, described by King and Kitchener (1994) as fundamental beliefs about the nature of knowledge, can be regarded as a variant of cognitive style. involves a sequence of qualitatively different stages of cognitive development, characterized by. Each stage of development is characterized by a different, coherent system including assumptions about what kind of knowledge is possible and corresponding justification strategies, e.g., a pre-reflective stage in which knowledge is either certain or derived from unquestioned authority, followed by a quasi-reflective stage in which all opinions are questioned and considered relative, followed by a reflective stage in which opinions can be evaluated and accepted, and subjected to reevaluation if necessary.

Stress and Workload

Stress is another likely influence on the choice of strategies. One view of stress's impact on decision making is that it disrupts "rational, logical" thought: the careful generation and evaluation of alternatives characteristic of analytical thinking (e.g., Janis, 1972). As we have seen, however, there is evidence that even unstressed decision makers do not evaluate options in the way required by normative models. This view appears to be supported by recent research on stress effects.

Pennebaker (1987) cites evidence for several effects of stress, which he combines under the idea of a *reduced level of thinking*. Stress (i) narrows the breadth of perspective, both in terms of time horizon and considering divergent information; (ii) makes people less self-aware, less likely to *reflect* on the causes and effects of their own actions, and less able to *self-regulate*; and (iii) makes people less aware of their own emotions. High-level thinking, by contrast, involves a broad perspective, self-reflective thoughts, and reference to emotions and moods. Most of these effects of stress appear to involve the disruption of reflective, self-regulative abilities.

Driskell & Johnston (1998; Mandler, 1982) present evidence for a model of stress that involves reflective processes in self-reinforcing cycles, which consume ever increasing amounts of the decision maker's cognitive resources. Novel and unpredictable situations cause people to lower their judgments of their own self-efficacy. These negative self-evaluations then lead to autonomic symptoms of stress, which seize attention. The symptoms may then be "overinterpreted" as suggesting incapacity, leading to even more stress. At the same time, the situation itself makes direct demands on attention because of its unpredictability, reducing resources for performing the task, leading to still lower judgments of self-efficacy, and more stress. Finally, attempts to remove the source of stress, or maladaptive responses such as worrying and negative self-evaluations, can consume even more attention.

The effects of stress on decision making appear to be mediated in large degree through metacognition or reflective judgment. Entin (1990) lists three typical causes of stress in a decision making context: overload, conflict, and uncontrollability. Uncontrollability refers to a reflective belief that one does not have control over events. Overload requires the perception that task difficulty outstrips ability, whether because time is too limited or because standards of success are too high. Conflict requires the perception that all one's goals cannot be achieved by available options, or that competing interpretations of a situation cannot be resolved by accessible knowledge.

On a more optimistic note, a variety of training methods can be effective in breaking these vicious cycles (e.g., Driskell et al.). In addition, situations where decision makers experience moderate stress do not produce the pathologies described above, but lead to a reasonable adjustment to changes in *workload*. Decision makers have been observed by a number of researchers to adaptively adjust their workload under moderate stress. For example, Payne and his colleagues found that under time stress decision makers adopted more "attribute-based" information-search strategies: they tended to evaluate all options against the most important attribute first, then move on to the next most important attribute, and so on. They were thus assured of having some reasonably significant information about every option. Similarly, in an air defense identification-friend-or-foe context, Cohen, et al. (1988) found that high target density led operators to examine fewer classification cues per contact, while continuing to examine all contacts. Several studies have observed that time stress causes selective focusing on negative attributes or worst-case outcomes of options (Leddo, Chinnis, Cohen, and Marvin, 1987; Wright, 1974), which might be construed as the most critical attributes in a time-stressed choice problem. In some studies, time stress has caused subjects to select options that conservatively hedge against different possible enemy actions rather than seize an opportunity, since time is not available to resolve the uncertainty (Leddo et al., 1987; Ben Zur and Breznitz, 1981). Entin (1990) observed that subjects under time stress became more likely to select information in the form of a predigested recommendations than in the form of raw data.

In sum, the effects of stress and workload on decision making may involve a reasonable metacognitive adjustment of strategies to adapt to the lack of time or task difficulty. In more severe cases, however, they may involve more pathological effects of diminished cognitive capacity, to which metacognitive self-evaluation also contributes, in this case negatively.

Expertise

In addition to cognitive style and stress, a major determinant of strategy selection is degree of experience or expertise. Studies of expert-novice differences suggest that expertise develops along two paths over time, one leading to better performance in *familiar* situations, the other leading to improved ability to handle *unusual* situations. A considerable body of research has focused on the first path: Experts accumulate a large repertoire of patterns and associated responses, which they use to recognize and deal quickly with familiar situations (Chase &

Simon, 1973; Larkin, McDermott, Simon, & Simon, 1980; Klein, 1993). The difference between experts and novices, however, goes well beyond the quantity of patterns they draw on or the number of situations they regard as familiar.

A key hallmark of expertise is *goal-setting*, or intentional creation of novelty. In fields such as writing and historical or scientific research, for example, experts are more likely than novices to identify opportunities for original, productive work, establish their own goals, and create challenging tasks for themselves, which cannot be solved by pattern matching alone (Ericsson & Smith, 1991; Anzai, 1991; Holyoak, 1991). Novel ideas and strategies are also important in military and business environments.

When performing a challenging task, whether self-created or externally imposed, experts and novices differ in other ways that are not fully accounted for by pattern recognition. Scardamalia and Bereiter (1991) found that expert writers, compared to novice writers, discovered more problems with their own work and struggled longer to find solutions, revising both their goals and their methods more often than novices. Patel and Groen (1991) found that expert physicians spent more time *verifying* their diagnoses than did less experienced physicians. Physics experts are more likely than novices to verify the correctness of their method and result, and to actively change their representation of the problem until the solution becomes clear (Larkin, et al., 1980; Larkin, 1981; Chi, Glaser, and Rees, 1982). Expert programmers pay more attention to the goal structure of a task than novices, searching first for a global program design, while novices tend to be more "recognitional," plunging rapidly into a single solution (Adelson, 1984). In foreign policy problems, expert diplomats spent more time formulating their goals and representing the problem, while students primarily focused on the options (Voss, Wolf, Lawrence, & Engle, 1991). VanLehn (1998) found that less successful physics learners were more likely to solve new problems by analogy with old problems (a recognitional strategy), while more successful learners used general methods for solving new problems, drawing on analogies only when they reached an *impasse* or wished to *verify* a step in their solution. Chi, Bassok, Lewis, Reimann, & Glaser (1987) found that better performing physics students were more likely to generate self-explanations and self-monitoring statements than poor students. Glaser (1996) identifies effective self-evaluation and self-regulation as key components in the acquisition of expertise.

Tactical battlefield problems tend to be viewed differently by experts and by novices. Novices often regard them as puzzles, which have "school book" solutions, while more experienced officers view them in a more challenging light, acknowledging the possibility that the enemy may not succumb so readily to a predictable course of action. Serfaty, MacMillan, Entin, & Entin (1997) compared experienced Army planners to novice planners, and found that the experienced planners did not appear to use recognitional strategies; that is, they did not did not generate an initial plan more rapidly (e.g., based on similarities with prior situations), tended to see the situation as more *complex*, and felt the need for more *time* to think about their plan than novices. Among the distinguishing features of experts that Shanteau (1992) identified in his research was the ability to handle adversity, to identify exceptions, and to adapt to changing conditions (Shanteau, 1992).

If expertise develops along two paths, what is the nature of the second, non-recognitional path? One view distinguishes it sharply from the first path: Experts define and deal with challenging problems by substituting formal analytical methods for pattern matching. This is the general approach urged by decision analysts (e.g., Watson & Buede, 1987), who define *normative* methods that require breaking novel problems down into components parts (e.g., options, outcomes, goals), assessing them quantitatively, then recombining them in order to calculate a recommended decision. The research reviewed above, however, suggests that this characterization of the second path is wrong. Formal methods are both too time-consuming, and too divorced from the knowledge experts have accumulated (Cohen, 1993). Dreyfus (1997) puts it well: “Usually when experts have to make such decisions they are in a situation in which they have already had a great deal of experience. The expert, however, is not able to react intuitively, either because the situation is in some way unusual or because of the great risk and responsibility involved… the experts draw on their context-based intuitive understanding, but check and refine it to deal with the problematic situation…”

Instead of dropping pattern recognition in novel situations, experienced decision makers learn to pause and *think critically about the results of recognition*. For example, according to Baker (1985), skilled readers exercise *meta-comprehension* skills, by continually looking for problems, such as inconsistencies or gaps, in the current state of their comprehension, and adopting appropriate corrective response, such as referring back to earlier parts of the text or relating the text to information already known. In both reading comprehension and in situation assessment more generally, decision makers ask, in effect: “What in this situation conflicts with my expectations? How can I stretch the pattern, i.e., tell a new story, to make the pattern fit? What assumptions must I accept to believe this story? What information is missing that would clarify the assumptions? How plausible is the story? What alternative patterns might apply? What story must I tell to make one of these other patterns fit, and what assumptions does it require? Which story is more plausible?” Reflective processes of this kind amplify the power and flexibility of recognitional processes without altogether throwing away their advantage in rapid access to knowledge. Moreover, critical thinking can make itself unnecessary the next time round. Decision makers sometimes handle novel situations by identifying regularities underlying exceptions to known patterns. Mental models embodying these newly discovered regularities provide patterns that can be recognized in later situations (Chi et al., 1981; McKeithen et al., 1981; Adelson, 1984; Larkin et al., 1980; Thompson, Cohen, & Shastri, 1997).

Because their function is to monitor and regulate recognition, we call the reflective processes used in unusual situations *metarecognitional*.[1] and we call this framework the Recognition / Metacognition Model (Cohen, Freeman, & Wolf, 1996; Cohen, Freeman, & Thompson, 1998). The R / M model implies that the two paths along which expertise develops are intertwined. Reflection increases the power of recognition, but itself gains power as a base of recognitional knowledge is built. We will discuss it in more detail in chapter 4.

It is reasonable to suppose that expertise in *teamwork* evolves with increasing experience in a domain along the same two paths as expertise in *taskwork* (McIntyre & Salas, 1995). Yet Orasanu & Salas (1993) note that “most current team training aims at developing habits for routine situations… Habit and implicit coordination will carry people a long way in routine situations; we need to prepare them for the unusual.” In this report we will explore how the dual

---

[1] This name is by analogy to other so-called *metacognitive* skills, such as *meta-memory* (skills for monitoring and improving memory performance), *meta-attention* (skills for improving the control of attention), and *meta-comprehension* (skills for monitoring and improving the understanding of text). See Forrest-Pressley, MacKinnon, & Waller (1985); Metcalfe & Shimamura (1994); Nelson (1992).

nature of expertise sheds light on the tension between initiative and coordination in teamwork, and provides a framework within which both initiative and coordination can be trained.

Teamwork Strategies : Coordination and Initiative

The concept of initiative plays a key role in the theory of critical thinking processes, in the real-world practice of critical thinking, and in critical thinking training. To see why, we can start by distinguishing two advantages that teamwork may provide over an individual acting alone, and then look at why each of these advantages may fail to materialize: (1) The first advantage is based on *bringing together complementary inputs*, and derives from the coordination of multiple hands, eyes, heads, etc. to accomplish a complex task. Increased effectiveness comes from sharing of both physical and cognitive workload and through specialization of knowledge and skills.

However, there is another side of the coin. Increasing the size of an organization tends to reduce its overall efficiency unless there is also an increase in departmentalization and standardization of tasks (Blau, 1970). The latter features reduce flexibility of response in a changing or novel environment (Donaldson, 1995). A related problem is *goal displacement*, in which specialized units lose sight of the larger organizational purpose, and pursue their own goals as if they were fixed ends rather than means, which should be reevaluated when conditions change (Scott, 1998).

(2) The second advantage of teamwork is based on *choosing from among substitutable alternatives*, and derives from the diversity of competing solutions to the same problem that different members of a team can generate. Better decisions result if there is an effective organizational mechanism for selecting from, averaging, or mixing these diverse ideas to arrive at a single decision (e.g., Kerr, MacCoun, & Kramer, 1996).

But there is another side to this coin as well. Groups may be affected by socialization biases, such as "groupthink," which induce conformity rather than diversity of thought (Janus, 1972; March, 1996.). For this reason, group decisions tend to be better when individuals think about the problem independently before arriving at a group judgment (Castellan, 1993; Sniezek & Henry, 1990).

Both dangers –slowness of response to change and lack of innovative thinking – can be addressed by organizational structures that emphasize *decentralization*: granting individuals or subteams the autonomy to make decisions in their own spheres (Burns & Stalker, 1961; Van Creveld, 1985). The degree of appropriate autonomy varies. Decentralization and initiative are adaptive responses to specific organizational environments, and are not everywhere appropriate. Interdependency among team tasks, on the one hand, heightens the importance of coordination (Thompson, 1967), whether it is achieved implicitly on the basis of stable, shared knowledge of tasks, procedures, and other team members (Cannon-Bowers, Salas, and Converse, 1993; Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Kleinman & Serfaty, 1989), by contingency planning that begins when unexpected possibilities first become apparent (Orasanu, 1993), or by mutual monitoring, feedback, back-up, and closed-loop communication as the tasks are carried out (McIntyre & Salas, 1995). On the other hand, when the task environment is rapidly changing and uncertain, and especially when individuals or teams are *spatially dispersed*, decentralization and initiative gain in importance. In some cases, outcomes may be better when individual team members bypass standard procedures, question the accepted beliefs or practices of the group, and act on their own responsibility.

This is a not uncommon predicament in combat: Company E's job is to guard Company F's flank while Company F secures a bridge that the division intends to cross. Now, however, Company F appears to be stalled in a major firefight some distance from the bridge. Company E cannot raise either Company F or higher headquarters on the radio (and it will take too long for runners to find them and return). Should Company E sit tight until Company F is ready to seize the bridge or until communications are reestablished? Should it go help Company F in the firefight, at the risk of getting bogged down itself? Or should Company E take over Company F's task and attempt to seize the bridge now – a risky choice, but possibly the only way to accomplish the higher-level purpose of supporting the division in a timely manner?

The combination of time stress, spatial separation, and uncertainty – along with varying degrees of task interdependency – can alter the nature of teamwork, overlaying a set of qualitatively different decision tasks on the traditional ones. For example:

- *Should we communicate?* When events unfold in an unanticipated manner (*uncertainty*), advance planning and shared task understanding may fail to bring about coordination. The obvious solution is to communicate in real time, as the unexpected events occur. Yet the *dynamic, time-stressed* character of the situation limits the time available for real-time communication. Moreover, *spatial separation* imposes a bandwidth limitation on communication, slowing it down drastically and exacerbating the impact of both uncertainty and time constraints.[2] The upshot is that real-time closed-loop communication can no longer be regarded as routine. When an unexpected, time-critical problem arises, team members or subteams must decide whether or not the potential benefits of communicating and/or waiting for a response are worth the delay.
- *What will other team members do?* In time-critical situations, subteams will sometimes be unable to communicate, or choose not to communicate, with one another. If their tasks are interdependent, however, the success of one will depend on coordination with the actions of another. In these cases, team members or subteams must make autonomous decisions that depend on plausible assumptions about concurrent decisions being made by other subteams in other locations. Shared task, team, and team member models may help support such predictions, but cannot be fully relied on in novel circumstances.
- *How good is the information?* Even when team members and subteams do decide to communicate, the combination of bandwidth and time constraints will prevent them from sharing information fully. Communications (e.g., reports, feedback, orders, or advice) from another subteam will have to be evaluated with incomplete understanding of the sources and assumptions behind them, and, conversely, with the benefit of other information that is available locally but not to the subteam that originated the message.

Evidence for the role of reflective processes is relatively pervasive in decision making contexts: in solving complex and novel problems; in electing strategies as a function of the task and situation; in generating stress and in abating stress; in the way individuals differ in the time

---

[2] In earlier historical periods, commanders could often see a large part, if not all, of the battlefield, and could both see and be seen by their subordinates. In this situation, the shared visual context provided a high-bandwidth channel of communication, which could be effectively supplemented by a few quick words and gestures. By contrast, the lethality and mobility of modern war has led to a high degree of dispersion, for which modern communications technologies, such as radio, and sensors do not fully compensate (Van Creveld, 1985).

they spend thinking about a problem, and in the strategies used by proficient or expert decision makers. We have also noted that there is comparable evidence for reflective, or metacognitive skills in other domains, which in some cases seem analogous to those exercised in decision making. Later in this report (chapters 3, 4, and 5), we will describe an empirically based theory that addresses skills of this kind. We will argue that the skills underlying initiative involve *critical thinking about mental models* of the task and the team. We then describe a training strategy that is based on the theory and which focuses on the mental models and critical thinking skills that underlie decisions about initiative (chapters 6 and 7). The value of such training should be quite general. Virtually every team is to some degree a *distributed* team. Even when team members are within plain sight and hearing of each other (e.g., in an emergency room, airline cockpit, or the combat information center of a cruiser), the high workload associated with uncertainty and time stress can be quite sufficient to limit the rate of communication (Kleinman & Serfaty, 1989) and make initiative essential.

A MODEL OF CRITICAL THINKING ABOUT RECOGNITION

In the remainder of this chapter, we describe a model of critical thinking in real-world settings that is based on the research described above, as well as on other finding. In successful recognition, perceptual inputs and goals rapidly converge within a decision maker's mind onto one, and only one, stable "intuitive" decision. The basis for decision making, more often than not, is recognition, and in ordinary circumstances, the recognitional responses of experienced decision makers are likely to be adequate (Klein, 1993). In more unusual situations, however, recognition needs to be supplemented by other processes. The model to be described addresses the question: What are these processes, and how do they work?

Recognitional learning enables humans (and other animals) to escape the speed limit imposed by natural selection, with its glacially slow shaping of inherited behavioral responses to recurring environmental situations. Instead, recognitional learning permits the acquisition of adaptive responses to environmental conditions that recur with some regularity during a *single lifetime*, even when they have not appeared at all in the previous history of the species. On the other hand, recognitional learning itself takes many years to produce expertise in a particular domain (Ericsson, 1996); how long it takes is likely to depend on the extent of the environmental variability or novelty that must be mastered. *Critical thinking* provides a further gain in flexibility in changing or novel environments, where recognitional learning also turns out to be too slow. Critical thinking enables decision makers to find discriminative, adaptive responses to even finer-grained environmental variations, which have not appeared in the previous experience of the decision maker. It does so by building a relatively simple layer of attentional control over the recognitional processing that is already taking place.[3] The simplicity of the required attentional control processes (described below), along with their power, lends plausibility to the

[3] This hypothesis regarding the evolutionary origin of metacognitive control is consistent with the views of Campbell (1974), Simon (1962), Heylighen (1991), Turchin (1977) and others. Knowledge systems in general evolve through a process of variation and selection, which favors changes that improve the system's ability to maintain itself in the presence of environmental variability. The complexity of the system increases along with the variety of different situations it can distinguish and responses it can produce. This increase in complexity is self-limiting, since it magnifies the time required to learn the appropriate situation-response connections. A solution is to increase the variety of potential responses indirectly, by varying higher-level parameters – in short, to introduce a system that varies the constraints on the original lower-level system. This higher-level system itself adapts through variation and selection, and thus explores a vast space of lower-level configurations without disrupting the operation of the lower level system.

hypothesis that such a second-order capability could have evolved, and that specific skills drawing on that capability could be shaped by experience.[4]

The Recognition / Metacognition Model of critical thinking has three main components:

- Meta-recognitional processes
- Mental Models
- Argument structure

Metarecognitional Processes

Critical thinking includes *meta-recognitional* processes that monitor and regulate recognition. As shown in Figure 1, the Recognition / Metacognition model distinguishes three functions that these processes perform:

(1) The *Quick Test*, which is a rapid assessment of the value of taking more time for critical thinking versus acting immediately on the current recognitional response;

(2) *Critiquing* the current results of recognition in order to identify three kinds of uncertainty: incompleteness in situation understanding or plans, conflict of goals or evidence, ; and explicit or implicit assumptions;

(3) *Correcting* those problems by influencing the operation of the recognition system, by inhibiting recognitional responding, shifting attention, and making assumptions.

[4] The hypothesis that meta-recognitional strategies can be learned through experience is being tested by experiments with a computational implementation of the Recognition / Metacognition model. The implementation utilizes a connectionist architecture with a backpropagation learning algorithm, and employs temporal synchrony of firings for consistency of object reference in relational reasoning (Thompson, Cohen, & Shastri, 1997).

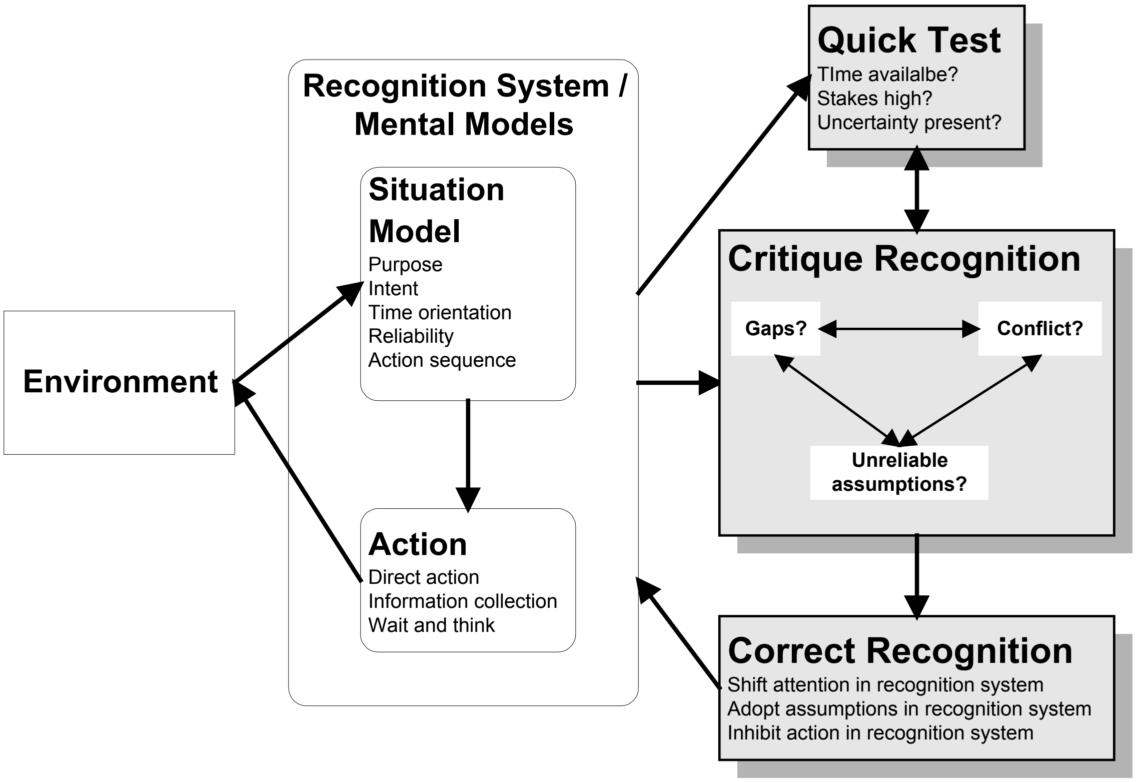


Figure 1. Basic components of the Recognition / Metacognition model. Shaded components are meta-recognitional, i.e., the reflective subsystem.

Meta-recognitional processes are general skill components that are effective across different tasks and domains. Their successful application, however, requires extensive domain-specific knowledge, such as mental models that describe causal relationships among events in the domain. We will discuss how meta-recognitional processes work in more detail later in this section and in chapter 11. Previous descriptions of the R/M model may be found in Cohen, Freeman, & Wolf (1996) and Cohen, Freeman, & Thompson (1998; see also Cohen, Parasuraman, Serfaty, & Andes, 1997).

Mental Models

Mental models are sets of correlated concepts and the causal (or other) structural relationships among them. Many mental models are highly specific to a domain, but some types of models are generalizable at least to some degree. For example, a large number of domains utilize mental models of intent with elements corresponding to motive, opportunity, and capability; and the proactive / predictive / predictive-reactive structure of initiative that discussed in chapter 3 is also widely relevant. In such domains, there is a distinction between: (i) mental models that support action based on predictions of future events (including the actions of other agents), (ii) mental models that support action designed to influence future events, and (iii) mental models of actions that are contingent on the specific future events that actually occur. Both domain specific and general mental models support meta-recognitional processes of verifying and improving situation understanding and plans. For example:

- Critiquing and correcting incompleteness: In predicting enemy plans, have I considered all the factors that might influence enemy intent? If I am unsure about a *prediction* of future enemy action, is there something I can do proactively to *influence* the enemy to act in a way that is advantageous to me? If my predictions, or my attempts to influence the enemy fail, what is my backup contingency plan?
- Critiquing and correcting conflict: Is the evidence that underlies my prediction of enemy actions consistent, or do some indicators point in opposing directions? Can I simultaneously attack where it will do the most harm to future enemy capabilities (i.e., be proactive), and attack where the enemy is currently the weakest (i.e., be predictive)? If I use artillery fires to reduce an enemy's strength prior to an attack, do I sacrifice the element of surprise? Which is more important in this particular case?
- Critiquing and correcting the reliability of assumptions: Do my predictions of enemy action or my plans depend on covert assumptions, for example, about enemy capabilities, the weather, or the passability of terrain? Have I assumed correctly that the enemy will panic in the face of a bold attack, rather than resist effectively? Have I assumed that I can implement a contingency plan or branch more quickly than is in fact possible?

Arguments

On the one hand, meta-recognitional skill is acquired in the process of gaining expertise in a particular domain. On the other hand, the skills that are in fact acquired need not be entirely domain-specific. We have already noted that meta-recognitional processes (identifying and correcting gaps, conflicts, and unreliable assumptions) are largely general across domains, and that some important mental models are also somewhat general. An additional source of generality is *argument structure*. Through experience in a domain, decision makers may learn to distinguish different *roles* that beliefs can play in any process of reasoning (e.g., the roles of *evidence*, *conclusion*, and *assumption*). By identifying the specific beliefs that play those roles in a particular case, decision makers can generalize the critiquing and correcting strategies that they have acquired in specific contexts.

A simplified version an argument structure (based on Toulmin, 1958) is shown in Figure 2. A *Claim* is any conclusion whose merits we are seeking to establish. It may be a assessment, e.g., about enemy intent, or part of a friendly plan, e.g., the time of an attack. The Claim is supported by *Grounds*, or evidence, e.g., considerations of likely enemy purpose, capabilities, and opportunities. Possible *Rebuttals* are condition under which the link between Grounds and Claim would not hold. Rebuttals are equivalent to implicit or explicit assumptions, that is, beliefs that are assumed true until shown to be false, and whose falsity would undermine the validity of the argument.

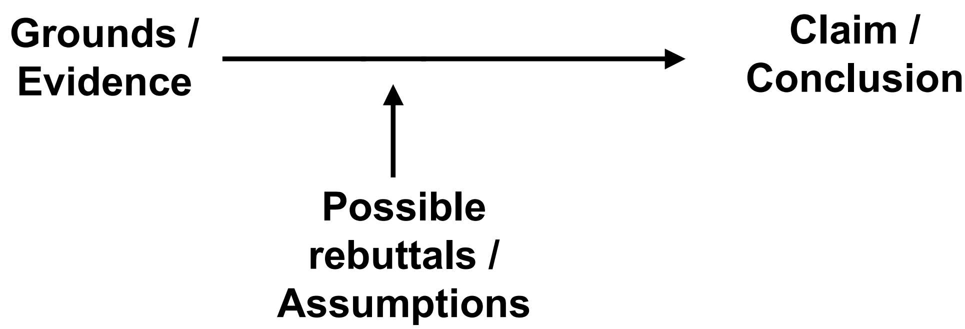


Figure 2. Toulmin's model of argument. The structure can be read: Grounds, so Qualified Claim, unless Rebuttal, since Warrant, on account of Backing.

An argument is typically *based on*, but not identical with, an underlying knowledge representation or mental model. For example, observations or analyses of enemy capability may provide grounds for conclusions about intent, since it is one its causes. Similarly, conclusions about intent may provide grounds for conclusions about the effects, i.e., actions the enemy is likely to take to achieve the intent. However, evidence-conclusion relationships do not always run from cause to effect. For example, observations about enemy actions lead to inferences about the intent behind the actions. Inferences or information about the enemy's intent can lead to inferences about enemy capability. Distinguishing grounds from conclusion must be a real-time discrimination (Kuhn, Amsel, & O'Loughlin, 1988), because the same event may serve as evidence in one situation and as a conclusion in another. The relationship between grounds, conclusion, and assumptions on a particular occasion is an *argumen*t, which may or may not be convincing.

Critiquing and correcting in terms of arguments is a more general skill than critiquing and correcting in terms of domain-specific mental models. It can take more time and be less effective than the corresponding specialized skill. However, in relatively unfamiliar domains, or novel situations, it may be the only available approach to resolving uncertainty. Table 3 provides examples of specialized skills, on the one hand, and more general versions of those skills, on the other.

Table 3. Specialized meta-recognitional skills and general meta-recognitional skills compared to one another, in how they deal with incompleteness, conflict, and unreliability.

	Domain-specific skill (based on mental models)	**General skill (based on roles in argument structures)**
Identifying and resolving incompleteness	Enemy intent may be to attack in the south or the north. Let me compare enemy capabilities in the north and the south and look at current enemy actions.	There are no grounds either for or against this conclusion, so both the conclusion and its negation are possible. What kinds of evidence are relevant (either as causes or effects of the conclusion)?
Identifying and resolving conflict	Enemy engineer capability is better in the south, but leadership is superior in the north. Does the enemy really need engineers for the terrain in the north? Is the leadership in the south better than we have supposed?	There are grounds both for and against this conclusion, so neither the conclusion nor its negation appear possible. What kinds of assumptions underlie my interpretation of the conflicting evidence?
Identifying and resolving unreliable assumptions	I have assumed that engineers will serve in a specialized engineers unit, as they have in the past. Perhaps the enemy has decided to integrate engineers in with other units. Maybe that's why we observed no engineers in the north.	There is evidence for one conclusion, but there are rebuttals that could neutralize it. The argument depends on generalizations that do not take into account the specifics of this situation, or which may be too limited, and so may not remain stable as I acquire more detailed information about this situation.

Correcting Processes

Critical thinking addresses these problems by removing one major limitation on recognitional learning: that the situation and the response retrieved to handle it must have been closely associated in the individual's previous experience. The mechanism that overcome this limitation involve relatively simple processes of controlled attention.[5] One important meta-

[5] The classic account of attentional control processes is in Atkinson & Shiffrin (1968), although more subtle models are now available.

recognitional correcting step involves shifting attention from cues in the situation to selected elements of the current recognitional conclusion. The result is activation of potentially relevant knowledge in long-term memory that has not played a role in the present argument because it is too distantly related to the situational cues. Activation of this new information may lead, via recognitional processes, to activation of still more indirectly related knowledge, to which attention may then be shifted, and so on.[6] Such attention shifting is equivalent to *posing queries* about the acceptability of the currently active situation model and plan (Shastri & Ajjanagadde, 1993; Thompson, Cohen & Shastri, 1997). A computational model of such processes is described in chapters 10 and 11 below.

As-if reasoning can be regarded as a more directive variant of attention shifting: i.e., to *persistently* attend to a *hypothetical* or *counterfactual* action or event. Persistent attention to such a possibility is equivalent to assuming or imagining that it is true, and posing a query about what *would* happen if the hypothesized action or event were the case (Ellis, 1995). This strategy extends the reach of recognitional processing even further, by activating relevant knowledge that is not closely associated either with cues in the actual situation or with the recognitional conclusion.

The result of attention shifting strategies of either kind is usually to increase the amount of knowledge brought to bear on a problem (assuming that conclusions can be retained and integrated across cycles of attention shifting).[7] Attention shifting, however, operates in different ways and has different consequences in response to different types of uncertainty. Experienced decision makers learn meta-recognitional strategies that respond differently to different types of uncertainty. Moreover, the solution to one problem may (but need not) lead to the creation or discovery of new problems. Figure 3 summarizes a variety of ways in which critiquing and correcting interact. We will explain these interactions in the following sections.

---

[6] It is plausible, but speculative to distinguish attention from consciousness. According to one view, consciousness results from a positive feedback process which recruits major parts of the brain into resonating activation cycles (Ellis, 1995). This is consistent with view that consciousness requires the, and also with Shruti's use of temporal synchronization to unify activities in different parts of the brain (chapter 10 below). Focused attention, on the other hand, involves querying a specific subset of the contents of perceptual or long term memory. Activation from this query may spread to other relevant contents, and return a signal to the queried node, creating a resonating cycle (Shastri and Ajjanagadde, 1993). Shifting attention, in order to query additional nodes, may result in recruitment of additional parts of the brain. Focal attention is thus one of the *causes* of consciousness.

[7] Priming and integration mechanisms were addressed in work on the Shruti system for this project. See chapter 10.

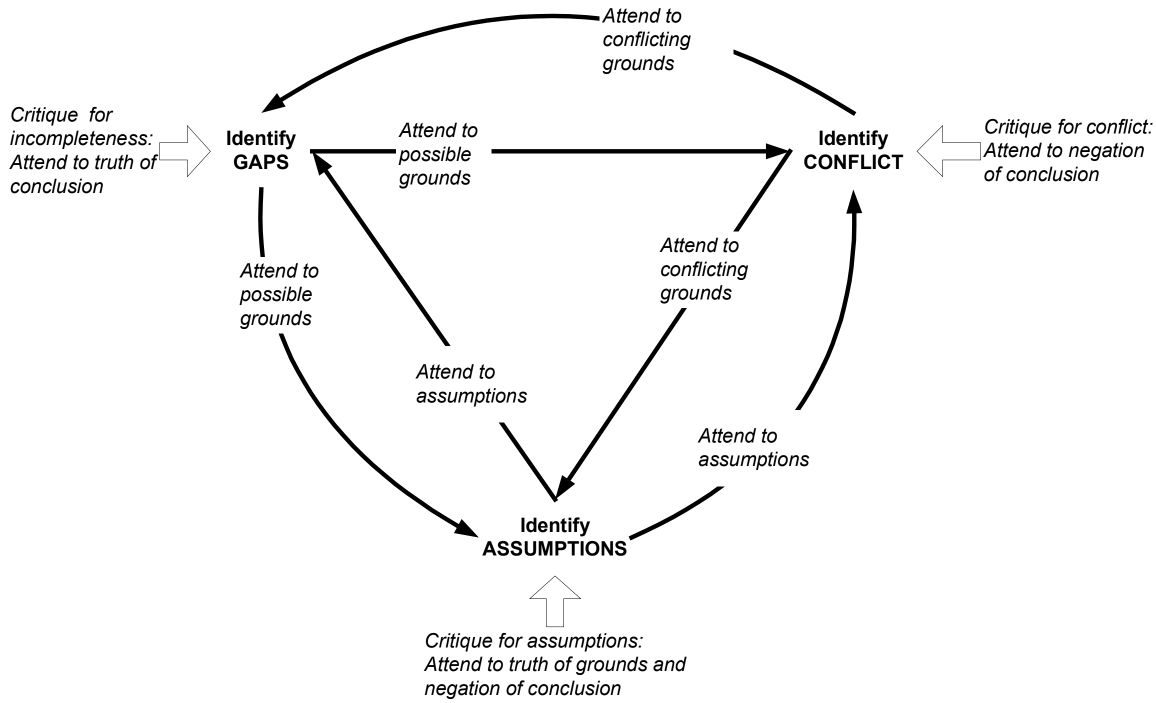


Figure 3. Cycles of critiquing and correcting in the Recognition / Metacognition model. Large arrows represent *critiquing* strategies, which are used to identify different types of uncertainty. Narrow arrows represent correcting steps designed to resolve particular types of uncertainty. In some cases, correction of one type of uncertainty leads to identification of another type of uncertainty.

Critiquing and Correcting Incompleteness

To identify and fill *gaps* in an argument (the case where more than one conclusion is consistent with the current evidence), attention shifts to one of the possible conclusions – in effect, querying its truth. The result is activation of an associated *mental model*, which indicates possible grounds for the conclusion. These grounds are the types of information that have been useful in the past in determining the truth or falsity of the attended conclusion. For example, in order to determine the *intent* of an enemy unit, it is useful to consider the *capabilities* of that unit, as well as its *opportunities*, *goals*, and *actions*.

Attention then shifts to one of the components of the activated mental model for which information is not currently active. For example, the decision maker decides to think about the *capabilities* of the enemy unit whose intent is uncertain. The result may be retrieval of relevant information in long-term memory about that component, or, if relevant information is not retrieved, a decision to initiate external data collection.

A more directive strategy for activating relevant knowledge in long-term memory is to temporarily assume that a conclusion is correct, by persistent attention to that possibility. This and subsequent shifts of attention may activate less immediately accessible information about the likely long-term consequences of an option, or about the less obvious implications of a hypothesis.

Knowledge activated by these attentional strategies may help narrow down the set of plausible conclusions by activating goals or beliefs that further constrain the solution. There are three possible outcomes, as shown in Figure 3: If newly activated knowledge eliminates all but one plausible conclusion, the problem is resolved. If filling gaps turns up constraints that *no* conclusion appears to satisfy, the result is a new problem, *conflict*. Finally, the newly discovered evidence may rest on shaky foundations, e.g., a statistical generalization that does not take into account particularities of the present situation. In this case, the result is another kind of problem, *unreliability of assumptions*.

Critiquing and Correcting Conflict

Correcting incompleteness by filling gaps in evidence is one method for identifying *conflict*. As we have just seen (Figure 3), newly retrieved or collected information may expose hitherto hidden conflict between a conclusion and existing goals or beliefs. Another, more directive strategy for identifying conflict is to temporarily assume (by persistent attention) that a conclusion is *wrong*, in effect tasking the recognition system to activate an account of how that could happen. This tactic heightens the salience of negative information about the conclusion, e.g., possible bad outcomes of an option or reasons why a hypothesis might not be the case. Awareness of this information may have previously been suppressed by stronger positive information.

Conflict among arguments (when there are grounds for both accepting and rejecting a conclusion) can be addressed by shifting attention to the grounds (e.g., sources of information or goals) that are responsible for the conflict. As a result of this shift in attention (and subsequent shifts to which it leads), assumptions underlying the argument may be exposed. It may be learned, for example, that (i) one or more conflicting sources of information are not as credible as previously supposed, (ii) one or more sources of information was misinterpreted in some way, (iii) one or more conflicting goals are not as important as previously supposed, or (iv) one or more options does not in fact conflict with a goal as previously thought. In this case, additional knowledge *removes* constraints on the recognitional conclusion, rather than adding constraints as in the case of filling gaps. Attention shifting reveals that what was previously thought to be a constraint on belief or action (e.g., a report from an information source, or a goal) was based on assumptions (Doyle, 1979; Cohen, 1986).

In the more directive version of this correcting step, the decision maker temporarily assumes (by persistent attention) that a specific source is not credible, or a specific goal is not important, etc., tasking the recognition system to account for how this could be. Such directive techniques can increase the chance that hitherto inactive knowledge in long-term memory about the relevant sources or goals will be retrieved.

There are three possible results of these correcting steps. First, the conflict is resolved if newly activated knowledge convincingly undermines the argument for one of the competing conclusions. For example, newly activated knowledge may establish that one of the conflicting information sources is not credible or that one of the conflicting goals is not important. Second, these correcting steps might undermine the reasons for both conclusions, thus leading back to the problem of gaps in arguments. Third, these correcting steps may lead to the identification of *unreliable assumptions*, if the decision maker must choose between current assumptions and new assumptions that would resolve the conflict (e.g., choose to regard a previously trusted source as untrustworthy).

Critiquing and Correcting Unreliable assumptions

To address *unreliability*, a decision maker must first *identify* key assumptions underlying possible conclusions and then *evaluate* them. Identification of hidden assumptions is not trivial. We have just seen that conflict in evidence can, and should, be used as an indicator of an incorrect assumption in one or both of the conflicting arguments. Yet a decision maker may have a high degree of confidence in the initial recognitional response to a situation, and may be unaware of any opposing considerations, and yet that conclusion may turn out to depend on questionable assumptions (for example, that the present situation resembles previously experienced ones in crucial respects). In addition to conflict, instability of conclusions over time, or variability in the conclusions of different decision makers at the same time, are also symptoms that unreliable assumptions could be playing a role. However, (a) variability per se does not indicate *what* the problematic assumptions are, and (b) variability like conflict is not always available as an indicator.

In a group context, a strategy for identifying assumptions is for decision makers to articulate *reasons* for their divergent conclusions and then to compare these justifications. Openness to such a dialogue is, of course, a natural part of a healthy group decision making process (e.g., Helmreich & Foushee, 1993). When variability does not exist, because there is a single convincing conclusion, disagreement can be induced more artificially, by assigning some individuals the task of "red-teaming" the preferred conclusion or playing the role of devil's advocate. Each potential problem discovered in this way represents an assumption implicit in the favored solution, to the effect that the relevant problem will not materialize.

Skilled decision makers use attention-shifting strategies to simulate these group processes. No matter how confident they are in a particular conclusion, one powerful approach is to assume that the premises of the argument are correct, but that the conclusion itself is *incorrect*, in effect querying the recognition system for an explanation of a failure of a rule. If decision makers are persistent enough, an explanation for the falsity of the prediction or the failure of the plan will be generated. Decision makers may then imagine that this is not the correct explanation for the failure, and force the recognition system to activate another explanation, and so on. Each explanatory possibility activated in this way corresponds to an assumption underlying the original argument from premises to conclusion. If the decision maker wishes to accept the conclusion, the decision maker must be comfortable assuming that each possibility of failure is false.

Assumptions can sometimes be evaluated one by one as they are identified, by shifting attention in order to activate knowledge that bears on their plausibility. However, because of limitations on time, only a small number of assumptions can be dealt with directly in this way. Therefore, the mere fact that a conclusion depends on untested assumptions is not sufficient cause to reject it. In the novel situations where critical thinking is appropriate, some crucial information will inevitably not be available, and no conclusion will fit all the observations or goals perfectly. If gaps and conflicts are to be resolved at all in these cases, it will have to be by means of assumptions.

In fact, real-world decision makers often use an *assumption-based correcting strategy*. They attempt to fill gaps and resolve conflicts in a recognitional conclusion, by retrieving or collecting information if possible but by making assumptions where necessary, until they have a *complete* and *coherent* story. In effect, they ask themselves, "What is the best story I can tell to justify this inference or plan?" They then step back, take a look at the story they have created, and try to evaluate its plausibility *as a whole*. In particular, they ask, "How many truly different assumptions did I have to make to build this story? Are the assumptions I had to make credible in this situation?" If the assumptions are troubling, the decision maker may temporarily drop them, and start again with the gaps and/or conflict that the assumptions were intended to handle (Figure 3). The result may be a new story, supporting a different conclusion. The choice between competing hypotheses or actions is often made based on evaluation of the plausibility of the assumptions underlying competing stories (Pennington & Hastie, 1993).

As Figure 3 and the preceding discussion make clear, meta-recognitional processing is a highly iterative, open-ended, and flexible process. The solution to one type of problem (e.g., filling a gap) can lead to another type of problem (e. g., conflict), which prompts new correcting steps, leading to new problems (e.g., unreliable assumptions), and so on. In the course of this process, recognitional conclusions are improved and/or modified bit by bit through local decisions about what to do next, and an understanding of the strengths and weaknesses of alternative conclusions is developed at the same time. These improvements are accomplished across cycles of shifting attention that either activate long-term memory contents that lay beyond the reach of a single recognitional cycle or lead to external information collection. When further benefits are likely to be outweighed by the costs of additional delay, critical thinking stops, and the decision maker can act immediately on the current best solution to the problem.

Other Views of Decision Making

In most of these respects, meta-recognitional processing contrasts with formal analytical approaches to decision making. Typically, formal methods require a problem structuring stage which specifies in advance the inputs that will be used to model the problem (e.g., Watson & Buede, 1987). The required inputs are not related in any direct way to recognitional responding and the knowledge that it taps, yet decision makers must somehow make precise numerical assessments of variables such as the strength of evidence and importance of goals. Similarly, the steps required to generate outputs from the inputs are determined in advance by the choice of an analytical model. Although some iteration may take place, "thinking" is largely over (and a solution is available) as soon as, but not a moment before, the model is finished according to the prespecified blueprint. Finally, the output is typically an unrealizable statistical abstraction (e.g., "there is a 70% chance of enemy attack"; "the expected utility of option A is equal to 40"), rather than a coherent picture of the situation that can be visualized and planned for. Table 4 compares the view of thinking offered by the R / M model and by analytical and recognition-based models, respectively.

Table 4. Comparison of three paradigms for understanding decision making.

	Analytical Models	**Recognitional Models**	**Recognition / Metacognition Model**
Inputs	Identify all inputs in advance (exhaustive specification of hypotheses, cues, outcomes, goals)	Limited to previously experienced situations and associated responses	Activate knowledge about new hypotheses, options, cues, or goals as current ones are found wanting
Processing	Assign fixed, precise meanings to cues & mathematically aggregate by a set of predetermined steps	Rapid, intuitive, not easily explained or justified	Try to create complete, consistent, and reliable situation picture by dynamically modifying interpretation of cues & goals
Outputs	Unrealizable statistical aggregation	Concrete situation picture, but little insight into its strengths & weaknesses	A single concrete situation picture, with an understanding of its strengths and remaining weaknesses

The Recognition / Metacognition model is a *problem-solving* model. Unlike most problem solving approaches, however, the R / M model identifies *strategies that are explicitly framed in terms of uncertainty*, and specifies how search takes place in a problem space defined by different types and amounts of uncertainty. Each processing step may be determined by global selection of a strategy, or may be determined locally by the results of earlier steps. Both kinds of choice may be affected by persisting epistemic attitudes or individual cognitive styles.

The R / M model explains how experienced decision makers are able to *exploit their experience-based intuition* in a domain (as explained by pattern matching) and *at the same time handle uncertainty and novelty* without resorting to artificial and time-consuming "analytical" methods. Uncertainty is handled not by abstracting from concrete reality, e.g, to estimate probabilities, but by reflecting on recognitions. Metacognitive strategies in effect "annotate" the internal situation model or plan to highlight points of incompleteness, conflict, and unreliability, and then respond to these problems to improve the current model or help the recognitional system to find a better one. To quote Dreyfus (1997, p.28) again, metarecognition is "observation of one's intuitive practice-based behavior with an eye to challenging and perhaps improving intuition without replacing it…"

CHAPTER 2:
A STUDY OF AIRLINE PILOT DECISION MAKING

INTRODUCTION

The Recognition/Metacognition (or R/M) model describes proficient decision making as an series of interactions between rapid recognition-based responding and processes that monitor recognition, inhibit recognition-based action when problems are found, and steer recognitional processes in new directions (Cohen, Freeman, & Thompson, 1998; Cohen, Freeman, & Wolf, 1996). In particular, according to this model, proficient decision makers will delay taking irreversible action when (1) there is uncertainty whose reduction could change the decision, (2) a change in the decision could significantly affect outcomes, and (3) the cost of delay is acceptable (Cohen, Parasuraman, and Freeman, 1998). Skilled decision makers draw on complementary metacognitive skills to handle uncertainty: *critiquing*, in which they actively search for qualitatively different kinds of uncertainty (incompleteness, conflict, or unreliable assumptions) in their current situation understanding and plans, and *correcting*, in which they attempt to resolve or transform problems by recalling or collecting new information and/or by revising assumptions.

Previous study. In a previous study (Freeman, Cohen, & Thompson, 1998; Freeman and Cohen, 1996), we used active-duty commercial airline pilots in a low fidelity simulation to examine the effect of variable (3), the cost of delay. That study tested the prediction that pilots with more flight experience would be more likely than less experienced pilots to adapt their decision making behavior to the available decision time. The time available to make a diversion decision was varied by manipulating the amount of fuel in the aircraft. The prediction was confirmed: Experienced pilots took more time to decide than less experienced pilots when more time was available; conversely, experienced pilots were quicker than less experienced pilots to reach a decision when less time was available. In addition, more experienced pilots made better use of the available time: They were quicker to notice potential problems at the destination airport and quicker to request information regarding alternates.

Hypotheses of new study. This study leverages and extends the previous work. It uses active-duty commercial airline pilots in a similar low-fidelity simulation, in order to test variables (1) and (2) of the R/M hypothesis. In particular, we vary (1) the *degree of uncertainty* about factors that affect the diversion decision, and (2) the *stakes*, or potential swing in outcomes, due to an incorrect decision to divert. Uncertainty is varied by manipulating the reliability and consistency of EFCs (time of expected future clearances) from different sources (ATC, company dispatch, and company station ops). Stakes (i.e., the cost of diversion vs delay) is varied by manipulating the passenger handling facilities available at the diversion airport and the amount of duty time remaining for the pilot in order to conduct a diversion.

The R/M model predicts that with experience, pilots will acquire an increased sensitivity to both the presence and the relevance of uncertainty and high takes, and will make more use of them in regulating their decision making behavior. We thus predicted a more adaptive response to both of the manipulations by more experienced pilots than by less experienced pilots. In particular, we expected that more experienced pilots would take more time to resolve conflicting or unreliable evidence in high uncertainty or high stakes situations than less experienced pilots; on the other hand, they would be expected to act more quickly than less experienced pilots when uncertainty and high stakes are not present. In addition, we anticipate a variety of differences in the way more and less experienced pilots actively seek out and examine evidence and revise their beliefs in the face of evidential conflict.

METHOD

Design.

Two between-subjects variables crossed: *stakes* (high vs low) with *uncertainty* (high vs low). The amount of time (i.e., fuel) available for decision making was held constant. Another factor — years of flight *experience* —varied between participants in the experiment.

Participants.

We recruited participants at the pilots' lounge of a major commercial airline at Dulles Airport. They were paid a nominal fee ($10) to participate in the brief (45-minute) experiment. A total of 64 pilots were used. The participants were assigned to experimental conditions according to a randomization plan that aimed at roughly equal numbers of high and low experience pilots in each condition. High experience was defined by reference to the median years of commercial flying experience (i.e., excluding general aviation and military flying experience, but including major commercial airline and commuter airlines). At the conclusion of the data collection, the median years of commercial flying experience of the sample was 13.

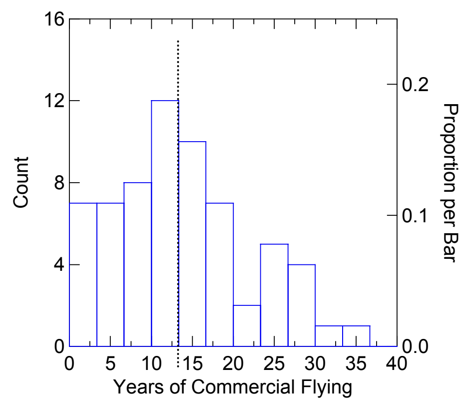


Figure 4. Distribution of participants by years of commercial flying experience. Dotted line shows the median (13 years). The mean was 13.6 years.

The distribution of participants by condition is as follows:

	High Experience		Low Experience	
	High Uncertainty	Low Uncertainty	High Uncertainty	Low Uncertainty
High Stakes	10	7	7	9
Low Stakes	7	6	9	8

Materials & Procedures.

The flight scenario used in this study was a modification of one used in a previous study. Pilots had to decide whether and when to divert from their destination, IAD (Dulles), in the face of a runway accident and adverse weather conditions there.

The task was administered to pilots individually. Participants played the role of Captain and non-flying pilot on a commercial aircraft during the last part of a flight from San Francisco (SFO) to Dulles (IAD) airport in poor, winter weather. Pilots received briefing materials including a flight plan, flight charts, and a status report that specifies the plane's current location, recent weather at the destination and alternates, current fuel load, required fuel load at final approach fix, and a constant speed and burn rate. In sum, pilots received sufficient information to compute the aircraft's location, fuel status, and the time available for decision making until a diversion decision must be made.

Pilots were able to obtain a variety of different kinds of information on request: They could ask for approach plates. They could contact ATC (Enroute Center and Terminal Approach), company dispatch, airport station operations, and ATIS. These roles were played by an experimenter following a response script presented on a laptop computer. The experimental support software enabled the experimenter to take time-stamped notes concerning participants' actions, requests, and comments.

The main scenario events are shown in detail Table 1 and .Table 2. **Error! Reference source not found.**shows variations associated with the *uncertainty* conditions, while .Table 2.shows variations associated with the *stakes* conditions. The scenario divides into six phases, each marked by the introduction, or availability on request, of some new information. These phases are as follows:

<u>Phase 1. 1300 – 1302:</u> *Time constraints due to fuel.* Status information indicates the amount of fuel remaining, burn rate, and minimum fuel required over final approach. If the pilot does the appropriate calculations, these data imply that the pilot has about 1.5 hours of fuel remaining before final approach. Given the distance from the current location to Norfolk (ORF) and the desirability of an additional fuel safety buffer, this means that a diversion decision to ORF must be made within approximately 30-40 minutes.

Time constraints due to duty time: Information about duty time varies with the stakes condition.

> *High stakes:* Because there was a delay in departure from San Francisco, the pilot's permissible duty time expires in 45 minutes, at 1345. This means that the pilot will be unable to land at Norfolk, refuel, and continue the flight if conditions improve at IAD. (Unless other crews are available at ORF, the passengers will suffer considerable inconvenience.)
>
> *Low stakes:* Duty time expires several hours later, at 17:15. This permits time for the pilot to land at Norfolk, refuel, and continue the flight if conditions improve at IAD. A diversion might introduce less passenger inconvenience than in the high stakes condition.

Initial indication of a problem. Nearing IAD, the pilot receives a call from Center via radio: "Expect holding at DOCS intersection." No explanation is given. ATIS indicates that the destination Dulles (IAD) is closed due to accident and snow, and that the first alternate Baltimore-Washington Airport (BWI) is closed due to snow.

<u>Phase 2. 1302-1310:</u> *Uncertainty versus certainty regarding delay.* Pilot receives a call from Center via radio announcing a late time for expected further clearance (EFC), again without explanation. This EFC is close to the time at which the pilot would have to commit to diverting due to fuel constraints. Calls to airline Station Ops or Dispatch concerning the accident will elicit

the response that "One of your aircraft is lodged in a snow bank." Further information varies by condition:

> *Low uncertainty condition*: Queries to Center concerning the Center's EFC elicit the reply that the EFC is based on recent estimates provided directly by the people doing the work at Dulles, and should be accurate. Queries to company dispatch concerning the Center's EFC elicit a confirmation of its accuracy. Queries to Dulles Station Ops confirm this in more detail, citing past experience with this kind of situation, and past accuracy of estimates of this kind.
>
> *High uncertainty condition*: Queries to center elicit the response that the EFC is merely their best guess at the moment. Queries to dispatch elicit the reply that Dulles Airport Ops is very unsure about the time required for the accident cleanup, and that the actual time could be anywhere in a 30 minute range (from 1315 to 13:45). Queries to Dulles Station Ops confirm this, and suggest more accurate information should be available at 1322.
>
> *Low stakes.* Pilot must call dispatch regarding the hold, to request an alternate. Queries regarding the only available alternative (ORF) will suggest that conditions there are good and expected to remain good. (Moreover, in this condition, the participant crew is not nearing constraints on crew duty time, so a “ground & go” at the alternate is a definite option.)
>
> *High stakes condition:* Queries to dispatch regarding alternates will yield the reply that ORF's support for delivering the passengers to their destination by ground transportation is poor, and there is limited availability of alternate aircrews. (In addition, in this condition, the participant crew will be nearing constraints on crew duty time, making a “gas and go” operation at ORF impossible.)

<u>Phase 3. 1313-1318:</u> *Series of descent orders.* The pilot is presented with an accelerating sequence of requests to descend the stack, and he/she hears orders to preceding aircraft to proceed from holding. There is no change in previous information. The meaning of these descent orders is ambiguous: The planes may be getting clearances to land at IAD, or they may be diverting. If they are diverting, traffic increases at ORF will shorten the time window for diversion available to the pilot.

<u>Phase 4. 1318-1324:</u> *Good news regarding cleanup.* Queries to Station Ops in all conditions reveal that IAD runway cleanup crews are leaving runways now.

<u>Phase 5. 1324-1327:</u> *More good news.* Listening to ATIS reveals that cleanup of IAD runways is complete.

<u>Phase 6. 1327- :</u>*Clearance to land.* Pilot is cleared to land at destination IAD by Center if he/she has not already diverted. Experiment ends

Table 1. Sequence of scenario events, showing information available to all subjects, and to subjects in high and low uncertainty conditions, respectively. Pilots get the italicized information only be specifically requesting it (e.g., ask dispatch about accident, ask Station Ops about snow, etc.).

Elapsed Time in seconds (Clock Time)	**All Subjects**	**High Uncertainty**	**Low Uncertainty**
PHASE 1			
00 (13:00)	Pre-brief: Flight is from SFO to IAD. Alternates are BWI & ORF. Weather forecast for IAD & BWI: Snow, winds gusting to 25 knots (IAD), 20 knots (BWI). Now overhead PUTTZ., FL 240. Expect holding at DOCCS. Fuel: 12,800 lbs. Desired minimum: 5000 lbs. Burn rate: 5000 lbs/hour. *ATIS IAD: Light snow…. IAD closed due to accident on 19L and snow on 19R. Snow removal in progress.* *ATIS BWI: Heavy snow…. At this time, BWI is closed due to snow.*		
PHASE 2			
120 (13:02)	ATC: Hold at DOCCS, FL 190. EFC: 13:31 *ATC: You are #8 in line to land. Some are talking about diversions. Two below you in stack at DOCCS.* *Dispatch: One of our planes slid off runway.* *Dispatch: Weather at ORF looks good.* *Dispatch: None of our flights have diverted from IAD yet, but some are*	*ATC: EFC is our current best guess.* *Dispatch: EFC may be too long or too short, anywhere from 13:15 to 13:45.* *Dispatch: Station Ops isn't sure about accident cleanup.* *Dispatch: We have no details on snow clearance.* *Station Ops: Situation could clear up earlier or later than we originally*	*ATC: We just heard from Dulles. The EFC is based on their time estimate.* *Dispatch: EFC looks accurate.* *Dispatch: Accident clearance is on schedule. Station Ops is pretty definite.* *Dispatch: Snow clearance is on schedule. Airport ops has a good handle on it.*

	asking about alternates. *Station Ops: Not much in the way of delays here until this problem hit.*	*earlier or later than we originally thought. Will know better at 13:22.* *Station Ops: We're not sure how long it will take us to clear accident.* *Station Ops: No telling when airport ops will be done with snow clearance.*	*Station Ops: EFC sounds right. ATC know our current estimates.* *Station Ops: This accident is similar to a previous situation, and we know pretty well how long it will take.* *Station Ops: Airport Ops has been reliable in their runway-clearing estimates all winter.*
600 (13:10)	ATC: Continuing holding at DOCCS.	*Dispatch: Still unsure on EFC.* *Dispatch: No current details on accident or snow cleanup* *Dispatch: No current details on accident or snow cleanup*	*Dispatch: Just talked to Station Ops and EFC at 13:31 still looks right.* *Dispatch Accident cleanup proceeding as expected.* *Dispatch: Snow cleanup proceeding as expected.*
PHASE 3			
780 (13:13)	ATC: Descend to FL 170		
900 (13:15)	ATC: Descend to 150		
PHASE 4			
1080 (13:18)	*Station Ops: Runways should be operational at 13:24.* *Station Ops: Accident cleanup crew is leaving the runway now.* *Station Ops: The snow plows are leaving the runway now.*		
PHASE 5			
1440 (13:24)	ATC: Info LIMA now current. *ATIS IAD (LIMA): Runway 19L plowed and sanded full length. 19R plowed and sanded full length.*		

1560 (13:26:42)	ATC: Descend to 130
PHASE 6	
1620 (13:27:44)	ATC: Clearance to IAD

.Table 2. Information for high and low stakes conditions, respectively. Subjects will receive italicized information only after specific request.

Elapsed Time in seconds (Clock Time)	High Stakes	Low Stakes
00	Pre-brief : Delayed departure from SFO means duty time is up at 13:45.	Pre-brief: Duty time is up at 17:50.
120 (13:02)	*Dispatch: Pax handling is bad at ORF. Expect 2 hour delays for ground transport. No spare crews. Can you do a gas and go if you have to divert?*	*[No problems indicated.]*

Simulation Tool and Measures

A software system was developed to provide a set of cues to a trained commercial pilot as a scenario unfolds. Participating pilots are asked to take the command of captain, and to delegate the role of flying the aircraft to an assumed (and un-modeled) co-pilot. In order to maximize the number of available personnel, we chose to identify a scenario that does not rely on a specific aircraft. This provided us with the largest possible pool of qualified commercial aviation pilots.

The simulation itself provides a low-resolution environment. The experimenter plays the role of various communications, including ATC, Tower, Weather (ATIS), and Dispatch. The scenario begins by presenting an initial context for the experiment and background information regarding the flight and the current circumstances of the flight. In the particular scenarios, we have choose to focus on an approach segment of a flight so as to take advantage of the heightened cognitive workload.

As the scenario unfolds, the pilot is presented with various communications, primarily from the tower. These communications are given verbally to the pilot by the experimenter in response to a timed flow of events that is managed by the software system. As the pilot plays out the scenario, the experimenter has the opportunity to record data on the pilot's behavior. Queries posed to the various information sources (Tower, Dispatch, etc.) are recorded, along with the elapsed time (within the scenario) at which the pilot initiated that communication.

The experimental design provides for the manipulation of certain variables intended to influence stakes and uncertainty. These manipulations were translated into variations in the timing of key events within the scenario. These variations on a common scenario were encoded into distinct files, one per experimental condition, that provide the specific set of events and communications for that experimental condition.

Two programs were provided to the experimenter. One provided a timed flow of events, where those events were drawn from the experimental condition identified for a specific participant. The other provided a means for the experimenter to note the pilot's various information requests, to obtain current communications (such as the ATIS report for IAD at that moment in the scenario), and to record other notes concerning the pilot's decision making process.

Using this tool, the experimenter was able to encode the following pilot decisions and information requests in real time:

Pilot Action	Description
Diversion	The pilot makes a decision to divert.
Use of Calculator	The pilot requests the use of a calculator.
IAD plates	Pilot requests the IAD (Dulles International) approach plates.
BWI plates	Pilot requests the BWI (Baltimore-Washington International) approach plates.
OFS plates	Pilot requests the ORF (Norfolk) approach plates.
ATIS: IADwx	ATIS weather for IAD.
ATIS: BWIwx	ATIS weather for BWI.
ATC: EFC	ATC gives pilot an EFC (Expect Further Clearance).
ATC: Traffic	ATC gives pilot an update on traffic patterns (e.g., stacking and diverting of planes at IAD).
ACARS/Dispatch: EFC	Pilot requests information from ACARS/Dispatch concerning an EFC.
ACARS/Dispatch: Alternates	Pilot requests information from ACARS/Dispatch concerning alternate destinations.
ACARS/Dispatch: Accident	Pilot requests information from ACARS/Dispatch concerning an accident.
ACARS/Dispatch: Snow removal	Pilot requests information from ACARS/Dispatch concerning snow removal.
ACARS/Dispatch: Traffic	Pilot requests information from ACARS/Dispatch concerning traffic patterns.
Station Ops: EFC	Pilot requests information from station operations concerning an EFC.
Station Ops: Accident	Pilot requests information from station operations concerning an accident.
Station Ops: Snow removal	Pilot requests information from station operations concerning snow removal.
Station Ops: Traffic	Pilot requests information from station operations concerning traffic patterns.

The frequency and timing of these events, under different experimental conditions, constitute the dependent measures for this study. A more detailed description of the software aspects of the simulation tool can be found in Appendix A.

RESULTS

We are interested in the effects of three variables – uncertainty, stakes, and experience – on the probability and timing diversion decisions, and on the probability, timing, sources, and topics of information requests. All analyses reported below included all three of these factors. However, to reduce the discussion to manageable proportions, we will address the role of experience and *uncertainty* first, followed by the role of experience and *stakes* (and any interactions with uncertainty).Within each of these discussions, we will first address *diversion decisions* and their timing, then *information requests* and their timing.[8]

Uncertainty and Experience

Diversion Decisions

In this section, we will examine the effects of uncertainty and experience on (i) the probability of diversion, and (ii) the time at which diversion took place for pilots who chose to divert.

Diversion probability. We performed an analysis of variance on the occurrence of a diversion decision as a function of uncertainty, stakes, and experience. Higher uncertainty tended to increase the proportion of pilots who diverted from 26% to 50% ($F(1,60) = 3.563$; $p = .064$). The suggestion in Figure 5[9] that more experienced pilots were slightly more likely to divert in both high and low uncertainty conditions was not significant.

---

[8] Following a suggestion of Abelson (1995, based on Tukey, 1991), we will use the terms "trend" or "tendency" for significance levels that are between p=.05 and p=.15. Tukey suggests a still weaker term, "hint," for significance levels between p=.15 and p=.25. We consistently report trends (.05<p<.15) and hints (only for 15<p<.20) on the assumption that discerning readers may wish to be aware that the data are "leaning" one way rather than the other in the absence of firm conclusions. (Note that this procedure can work against us as well as for us, e.g. the data might not only fail to support our hypothesis, but might suggest the opposite.)

[9] All figures presented in connection with an analysis show least squares means.

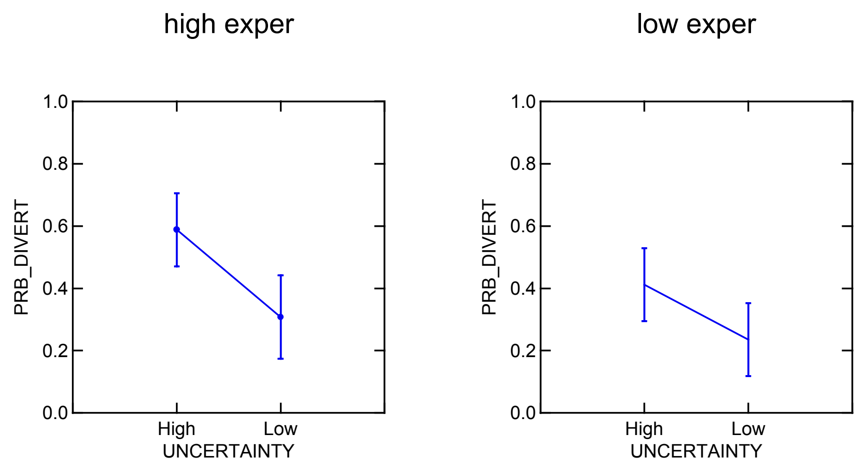


Figure 5. The effect of uncertainty on probability of diversion, broken down by experience level.

Diversion time. We performed an analysis of variance on the time of diversion, including only those pilots who chose to divert, as a function of uncertainty, stakes, and experience. Significant results and trends are as follows:

Stakes	F(1,17) = 3.320; p = 0.086
Experience	F(1,17) = 4.655; p = 0.046 *
Stakes x Experience	F(1,17) = 6.233; p = 0.023 *

As Figure 6 shows, the more experienced pilots, if they did choose to divert, did so *earlier* than less experienced pilots. (We postpone discussion of the effects and interactions involving stakes until a later section.) Figure 7 shows that the experienced pilots diverted earlier than the less experienced pilots in both uncertainty conditions. However, Figure 7 shows something else as well. Although the interaction between experience and uncertainty is not significant, it is worth noting that more experienced pilots, on average, appear to have responded to uncertainty by delaying diversion, while less experienced pilots did not.[10]

[10] Within the high experience group, the contrast between high and low uncertainty represented a trend (F(1,10) = 2.526; p = 0.143

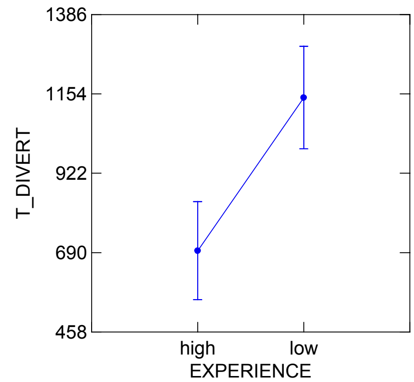


Figure 6. Diversion times as a function of experience. Only pilots who did divert are included in this analysis.

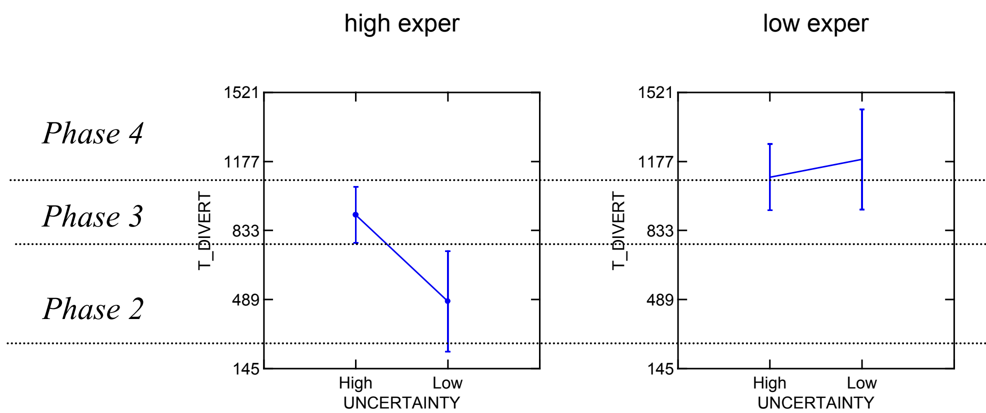


Figure 7. Diversion times in relation to uncertainty and experience for the 25 pilots who chose to divert. Lines indicate phases in which new information becomes available in the scenario.

An interesting way to look at the diversion times in Figure 7 is to regard them not as a function of *time*, but as a function of *information*. From this point of view we can ask, what information was sufficient to trigger diversion under different conditions of uncertainty and experience? Recall that Table 1 divided the scenario into six phases, corresponding to changes in the information that was available to the pilots, on request, at different points in the scenario.

Figure 6 shows the mean diversion times in relation to the boundaries of these information phases.

In Phase 2, pilots in the low uncertainty condition have access to reliable estimates of the time required to clear the runways. With this reliable information, highly experienced pilots on average made the decision to divert in Phase 2, without waiting to resolve uncertainty further. By contrast, pilots in the high uncertainty condition received less reliable estimates of the time required to clear the runways during Phase 2. Experienced pilots in this condition were on average likely to wait to divert until Phase 3, but not longer. Less experienced pilots did not wait longer in high uncertainty conditions than low uncertainty conditions. In addition, the less experienced pilots required at least one additional phase of information in both uncertainty conditions before they reached a decision to divert.

Figure 8 and Figure 9 depict the profile of diversion frequency by information-phase for the different conditions. The high experience group (Figure 9) presents a sharper and more consistent profile in both high and low uncertainty conditions. Higher uncertainty causes both the mean and modal diversion times to shift from phase 2 to phase 3. The profile of the low experience group (Figure 8) has considerably more spread in both uncertainty conditions, with no well-defined modes in the high uncertainty condition.

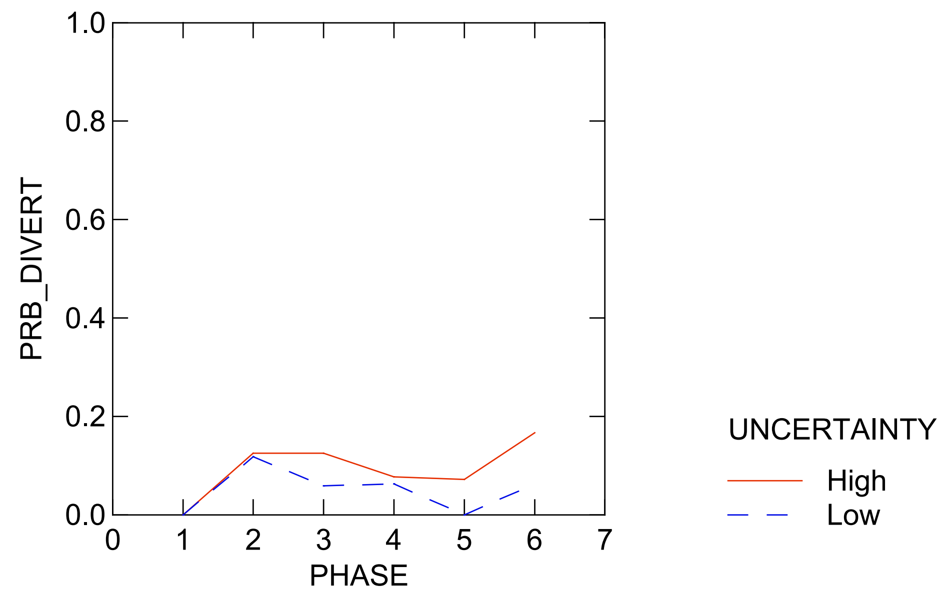


Figure 8. Diversions by phase for low experience pilots.

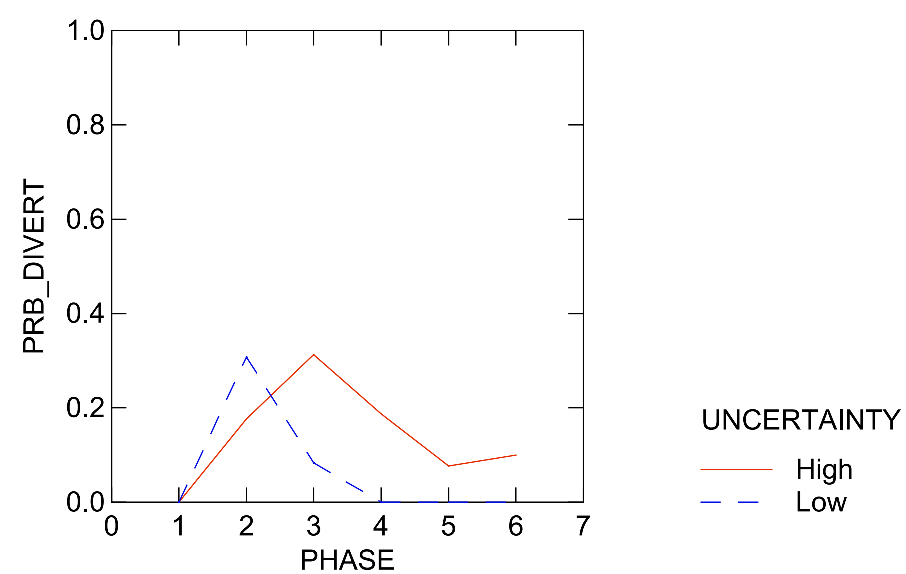


Figure 9. Diversions by phase for high experience pilots.

Requests for information about the causes of delay.

The two uncertainty conditions differ in terms of the information provided to pilots on request from three sources – Air Traffic Control (ATC), company dispatch, and Dulles Station Ops. All three sources indicate either high or low reliability of the EFC (time of expected further clearance), depending on the uncertainty condition. The uncertainty conditions also differ in terms of more specific information from company dispatch and Dulles Station Ops. This information confirmed either the high or the low reliability of the EFC by providing additional, high or low reliability details about the accident and snow clearance.

All pilots in all conditions made repeated requests to ATC for clarification of the EFC. However, the only significant effects on these requests involved stakes, which we will turn to later. Experience and uncertainty, by contrast, had more dramatic effects on the way pilots made use of dispatch and station ops to obtain confirming or disconfirming *details*. We performed a multivariate analysis of variance on the number of information requests made by the pilots, with five orthogonal factors: the source from which information was requested (dispatch or station ops), the topic of the request (EFC, accident, or snow), uncertainty (high or low), experience (high or low), and stakes (high or low).[11] The significant results and trends were as follows:

Source of request	F(1,336) = 54.951; p = 0.000 ***
Topic of request	F(2,336) = 5.887; p = 0.003 **
Uncertainty	F(1,336) = 19.057; p = 0.000 ***
Experience	F(1,36) = 2.712; p = 0.101 (T)
Source x Topic	F(2,336) = 8.698; p = 0.000 ***
Source x Uncertainty	F(1,336) = 3.397; p = 0.066 (T)
Source x Experience	F(1,336) = 4.635; p = 0.032 *
Source x Uncertainty x Experience	F(1,336) =2.324; p = 0.128 (T)
Source x Stakes[12]	F(1,336) = 3.649; p = 0.057 (T)

The effect of source is straightforward: There were far more requests for information from dispatch (a mean of 1.2 requests per pilot) than from station ops (a mean of .3 requests). In regard to topic, the EFC attracted the most queries (a mean of 1.0), with snow removal second (a mean of .7), and the accident third (with a mean of .5). The highly significant interaction of topic and source is shown in Figure 10. Pilots varied in the frequency with which they queried dispatch about the different topics, but were consistently low in the rate at which they queried station ops on any topic.

[11] It was not possible to include ATC in this MANOVA, because ATC could not be queried specifically regarding the snow and accident; hence, topics would not have been fully crossed with sources.

[12] This was the only effect that met our reporting criteria involving stakes. We will discuss the role of stakes later.

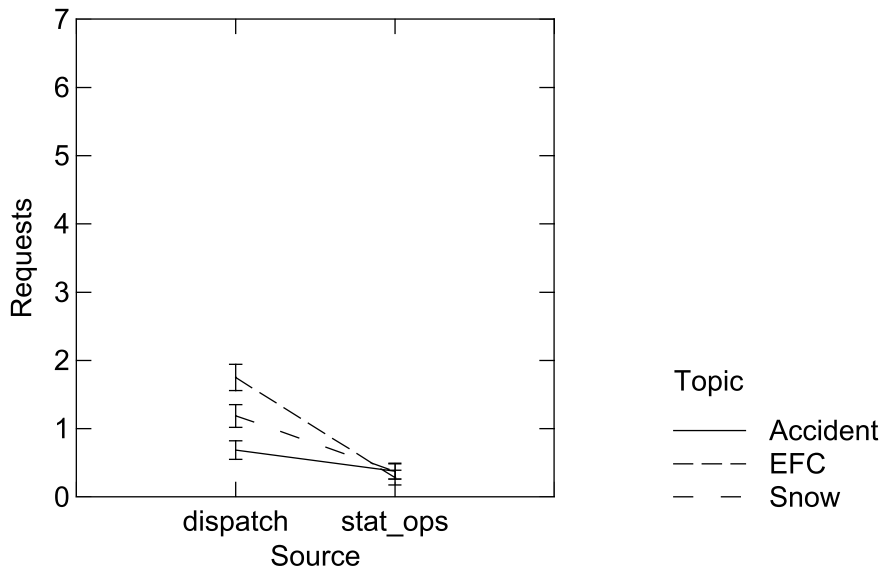


Figure 10. How pilots chose to request information, by source and topic.

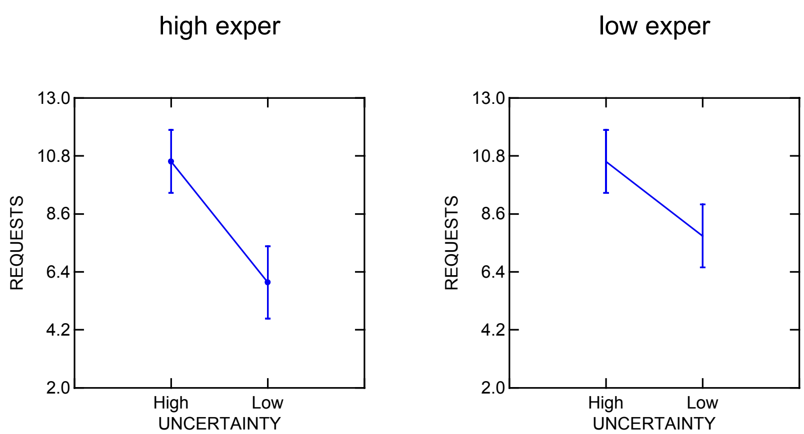


Figure 11. Effect of uncertainty on the total number of information requests, broken down by experience level.

As shown in Figure 11, high uncertainty significantly increased the average number of queries for both low and high experience pilots. Figure 12 shows the tendency for uncertainty to accentuate the preference for dispatch over station ops, even though the total number of requests from both sources increased.

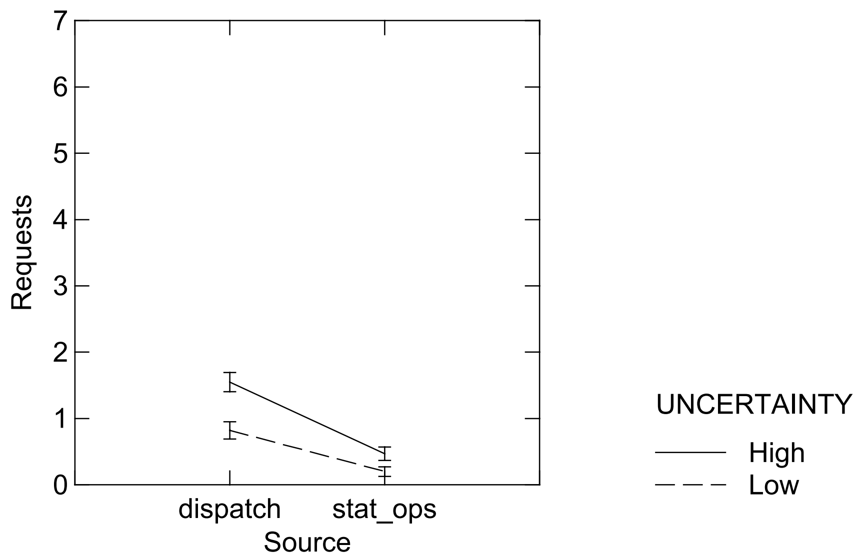


Figure 12. Effect of uncertainty on information requests, broken down by source.

There was a trend for more experienced pilots to make *fewer* requests for information than less experienced pilots. As shown in Figure 13, this was due entirely to the significant interaction between experience and source of information. Experienced pilots requested significantly less information from *station ops* than less experienced pilots ($F(1,60)=5.730$; $p=.020$), but requested approximately the same information from dispatch. In other words, the tendency to rely more on dispatch than station ops increased with experience under both levels of uncertainty. Experienced pilots became more focused and selective.[13]

The focusing effect of experience tended to be stronger under high uncertainty conditions than low uncertainty conditions. Figure 14 shows that the preference for dispatch over station ops appears to increase due to uncertainty for highly experienced pilots, but remained unaffected by uncertainty for less experienced pilots. This figure also suggests that the increased preference for dispatch over station ops under high uncertainty (Figure 12) was due to the more experienced pilots.

---

[13] Of course, this interpretation assumes that station ops information was either less valuable or redundant. If this is not the case, one might conclude that experience made pilots excessively narrow in their information use.

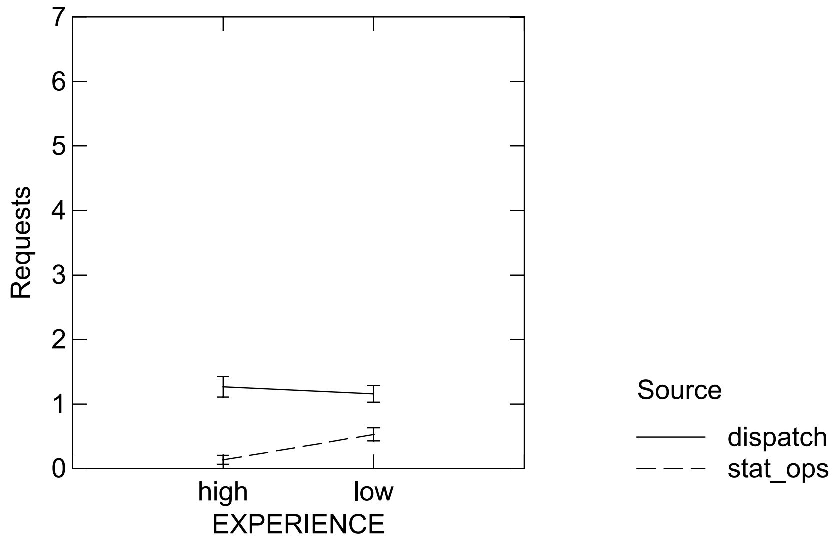


Figure 13. The effect of experience on the frequency and source of information requests.

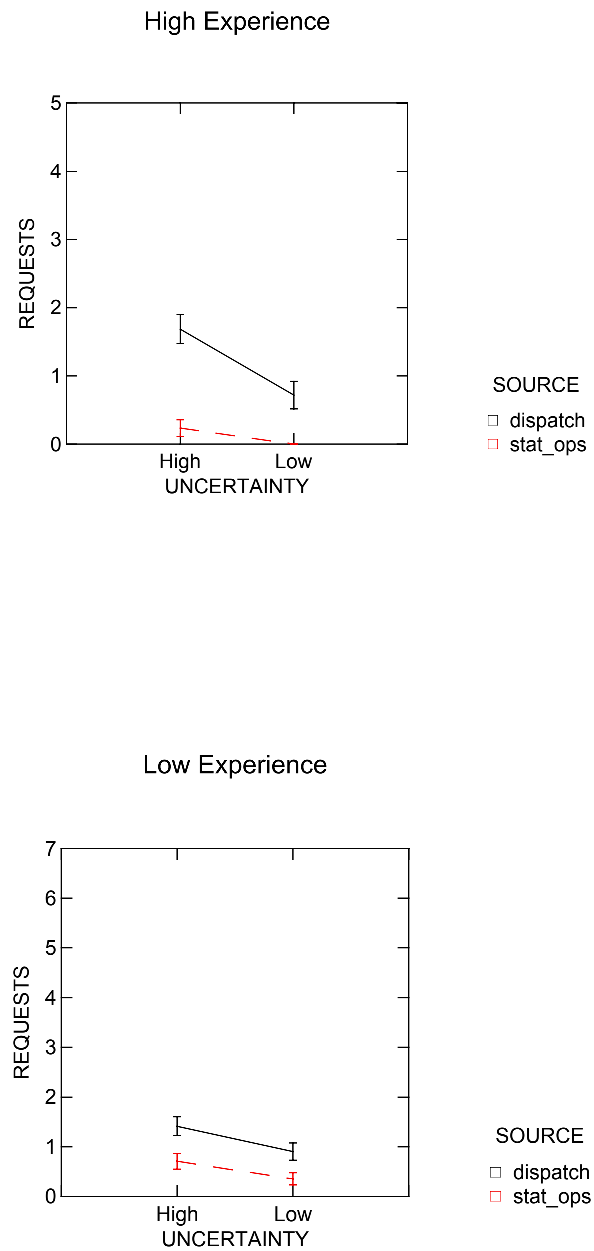


Figure 14. Joint effects of experience, uncertainty, and source of information requests.

Use of information sources

Another way of looking at the issue of information requests is to distinguish (i) the total number of *different sources* that pilots used, from the (ii) average number of information requests they made *per source*. An analysis of variance was performed on both measures, using uncertainty, stakes, and experience as factors. This analysis is broader in several respects that those reported in the previous section. First, we were able to include queries to ATC as well as dispatch and Station Ops. Second, we expanded the analysis to include requests for all kinds of information, such as traffic conditions or alternates, and not just those that pertained directly to uncertainty about the causes of delay (i.e., EFC, snow and accident removal). We will consider the number of information sources first, then turn to the number of queries per source.

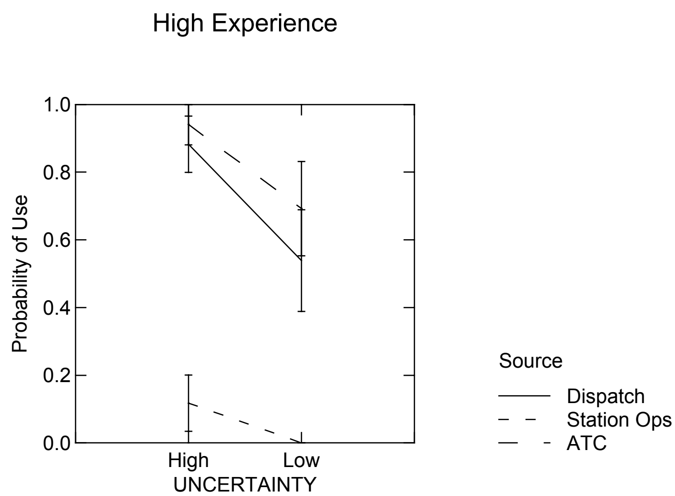
We examined the effects of experience, uncertainty, and stakes on the chance that a pilot would *use* an information source at all, as distinct from the number of requests made to that source. We performed a multivariate analysis of variance, with three all-or-nothing dependent variables corresponding to whether or not a pilot used an information source (ATC, dispatch, and station ops). Significant results and trends are as follows:

Multivariate (ATC, Dispatch, Station Ops)	
Uncertainty	F(3,54)= 3.018; p = .038 *
Experience	F(3,54)= 4.245; p = .009 **
Stakes x Experience	F(3,54)= 3.409; p = .024*
Uncertainty x Stakes x Experience	F(3,54)=1.617; p = .196
ATC	
Uncertainty	F(1,56)= 4.603; p = .036 *
Experience	F(1,56)= 4.2725; p = .043 *
Uncertainty x Experience	F(1,56)= 1.848; p = .180
Stakes	F(1,56)= 2.802; p = .100 (T)
Dispatch	
Uncertainty	F(1,56)= 5.304; p = .025 *
Stakes x Experience	F(1,56)= 5.304; p = .025 *
Uncertainty x Stakes x Experience	F(1,56)= 3.496; p = .067 (T)
Station Ops	
Uncertainty	F(1,56)= 3.027; p = .087 (T)
Experience	F(1,56)= 9.576; p = .003 **

Uncertainty significantly *increased* the overall number of sources used, and experience significantly *reduced* the overall number of sources used (as shown by the multivariate tests). In addition, the univariate tests show that uncertainty and experience had significant, and opposing effects on the likelihood of using *specific* information sources. These effects can be seen in Figure 15. Uncertainty significantly increased the tendency to query ATC about the EFC, while experience slightly but significantly reduced the tendency to query ATC.[14] On the other hand,

[14] Recall that pilots received much information from ATC automatically, without the need for an explicit request. This analysis pertains only to active requests by pilots for further clarification from ATC.

experience had a very large negative impact on the tendency to use Station Ops, while uncertainty had a slight tendency to increase the use of Station Ops.



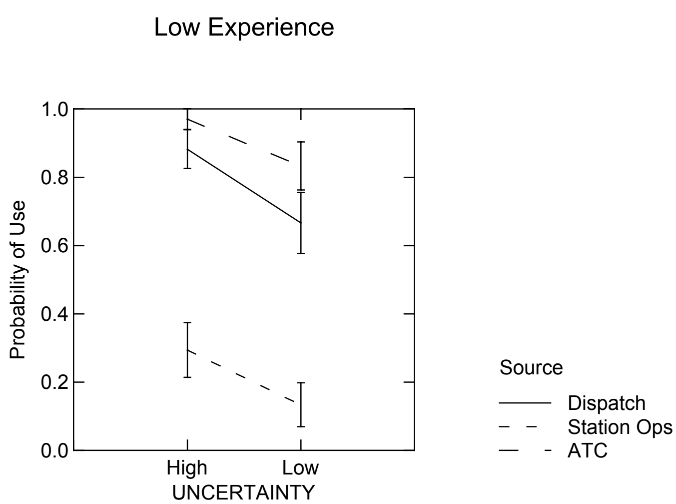


Figure 15. Effects of experience and uncertainty on the probability that a pilot will use a given information source.

A subsidiary analysis looked at the *average number of different information sources* used by pilots. The significant results and trends from that analysis are as follows:

Uncertainty	F(1,56)= 9.938; p = 0.003 **
Experience	F(1,56)= 10.418; p = 0.002 **

The average number of information sources that pilots used *increased* under conditions of uncertainty, for both more and less experienced pilots, but experience led pilots to use *fewer* sources in both uncertainty conditions (see Figure 16). In particular, highly experienced pilots used an average of approximately *one* of the three information sources when uncertainty was low. Figure 15 suggests that under low uncertainty an experienced pilot might have used *either* dispatch *or* ATC, but was not likely to use both. In other words, they were treated as if they were *equivalent, or substitutable, sources of information*. Station Ops was never used by the experienced pilots in low uncertainty. When uncertainty increased, experienced pilots added only one additional source on average. Figure 15 suggests that experienced pilots then used *both* dispatch *and* ATC, as *complementary, or mutually verifying sources*, with only a small chance of looking at Station Ops as well.

Less experienced pilots, on the other hand, used an average of two sources even in low uncertainty. Figure 15 suggests that this was primarily a combination of dispatch and ATC, but also included a small chance of using Station Ops. In high uncertainty, the less experienced pilots used both dispatch and ATC, with a good chance of using all three sources.

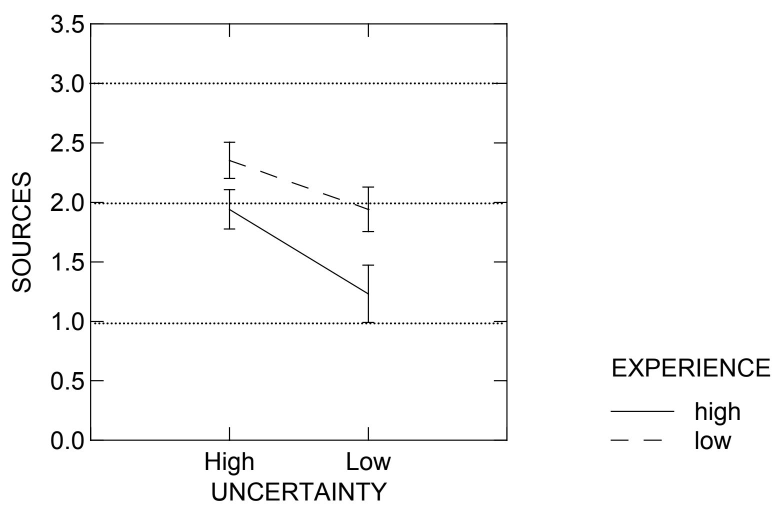


Figure 16. Effects of uncertainty and experience on the average number of information sources used (ATC, dispatch, station ops).

We now look briefly at the other side of the coin: the average number of queries *per information source that was used*. An analysis of this variable led to the following results:

Experience	F(1,51)= 4.515; p = 0.038 *
Stakes	F(1,51)= 5.745; p = 0.020 *
Uncertainty x Stakes x Experience	F(1,51)= 13.854; p = 0.055 (T)

As shown in Figure 17, experience *increases* the average *number of queries* that pilots pose to the sources that they do use, at both levels of uncertainty. Uncertainty, on the other hand, has no effect on queries per source, but only increases the number of sources that are used.

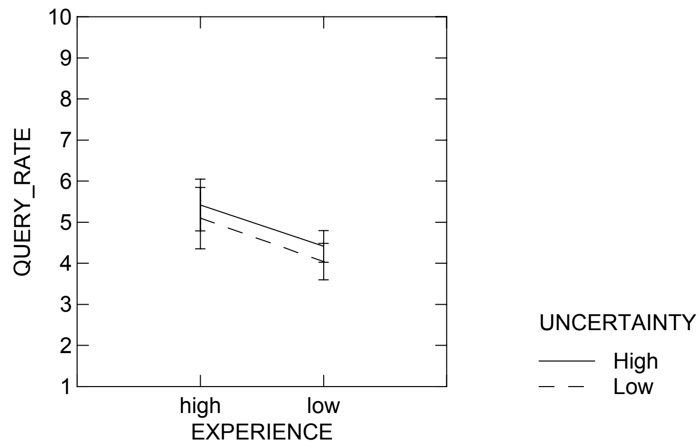


Figure 17. Effect of experience on average number of queries per information source, broken down by uncertainty level.

In sum, as suggested earlier, the reduction in information requests due to experience (Figure 11) can be attributed entirely to the reduced *number of information sources* that experienced pilots consider (Figure 15 and Figure 16). This reduction, in turn, is due primarily to two factors:

- a lower likelihood of using station ops (Figure 13 and Figure 15), and
- a lower likelihood of using any one source (e.g., dispatch) to verify another source (e.g., ATC) unless uncertainty warrants it (Figure 16).

Finally, more experienced pilots make more intensive use of the information sources that they do consult than less experienced pilots (Figure 17). This might be part of a reasonable strategy to monitor events more effectively for *changes*. Such *high-frequency situation monitoring* trades off against *uncertainty resolution*, i.e., consulting multiple sources to verify a *single* reported event. When the latter is unnecessary (because of low uncertainty), experienced pilots appear to put more emphasis on the former.

<u>Timing of information requests.</u>

In this section we examine the timing of pilots' first requests for information on different topics. We perform a set of analyses on the impact of uncertainty, stakes, experience, and *specific topics* on the timing of information requests. It was not possible to combine these analyses into a single MANOVA, because such an analysis could only use pilots who actually made information requests in each category, and the resulting data matrices are too sparse.

The analyses focused on five topics about which pilots could request information: the EFC, the accident, snow removal, traffic, and alternates. The significant results and trends were as follows:

EFC	[no effects]
Accident	
Uncertainty x Experience	$F(1,18) = 3.476$ $p = 0.079$ (T)
Uncertainty x Stakes x Experience	$F(1,18) = 2.200$; $p = 0.155$
Snow	
Uncertainty	$F(1,26) = 4.415$; $p = 0.045$ *
Uncertainty x Experience	$F(1,26) = 7.605$; $p = 0.011$ **
Uncertainty x Stakes x Experience	$F(1,26) = 4.938$; $p = 0.035$ *
Traffic	
Uncertainty	$F(1,32) = 17.282$ 0.000 ***
Experience	$F(1,32) = 4.325$; $p = 0.046$ *
Stakes x Experience	$F(1,32) = 3.221$; $p = 0.082$ (T)
Alternates	
Uncertainty x Stakes x Experience	$F(1,56) = 2.361$; $p = 0.130$ (T)

Uncertainty and experience interacted in their effects on inquires about snow removal and the accident. Figure 18 and Figure 19 reveal a consistent pattern of timing differences for queries about these two topics. In high uncertainty, experienced pilots asked about snow and the accident *at the same time as or earlier than* less experienced pilots; but when uncertainty was low, more experienced pilots inquired about these topics *later*. This is consistent with more efficient use of information sources by highly experienced pilots to resolve uncertainty when and only when it exists. [15]

[15] There were no significant results or trends regarding the timing of requests for information about the EFC.

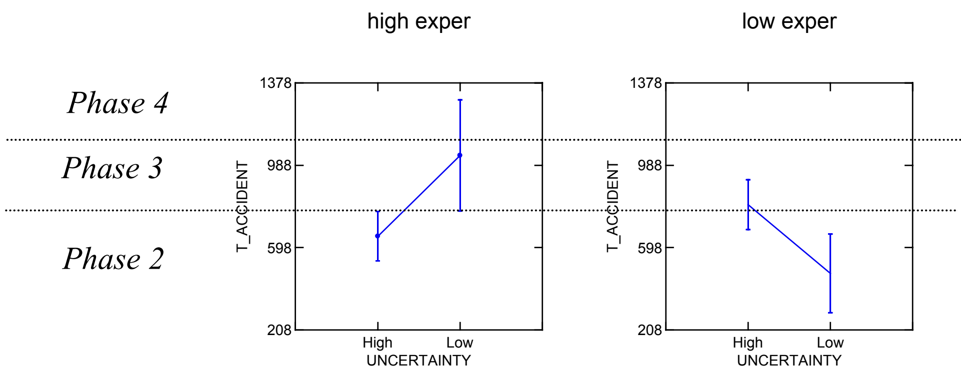


Figure 18. Interactive effects of experience and uncertainty on the time of the first inquiry about the accident.

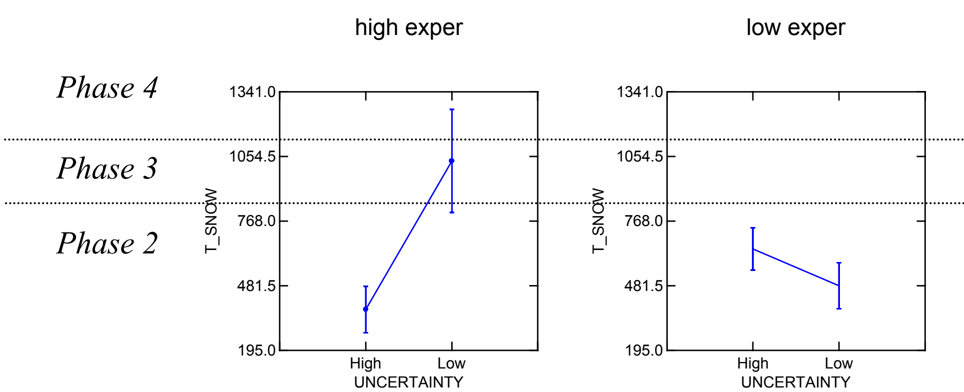


Figure 19. Interactive effects of experience and uncertainty on the time of the first inquiry about snow removal.

In addition, both uncertainty and experience had a strong effect on the timing of inquiries about traffic. Such inquires are relevant as pilots move from situation assessment to planning a possible diversion. As Figure 20 shows, the more uncertain the diversion was, the later such requests came, for both experienced and less experienced pilots. However, experienced pilots made these inquiries earlier than less experienced pilots in both uncertainty conditions. This is consistent with the shift in diversion time for highly experienced pilots from Phase 2 in the low uncertainty condition to Phase 3 in the high uncertainty condition (Figure 9). It is also consistent with the hint in Figure 6 that experienced pilots who diverted did so earlier than less experienced pilots. Earlier diversions and earlier requests for information about traffic may both be due to the speedier resolution of uncertainty by experienced pilots. This in turn may be related to the more efficient use of information sources.

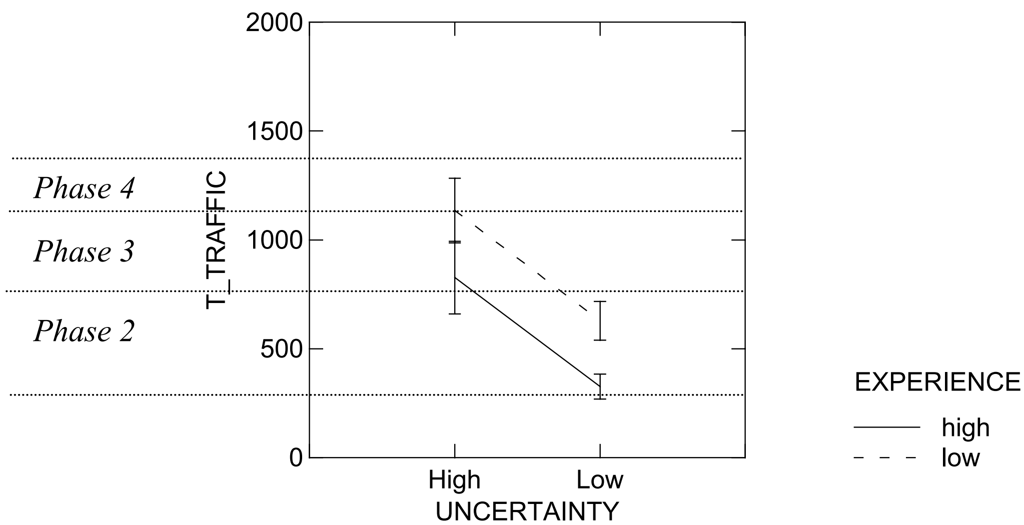


Figure 20. Effects of uncertainty and experience on the time of initial queries about traffic. Compare this to Figure 3, effects of uncertainty and experience on the time of diversion.

<u>Stakes and Experience</u>

In addition to varying uncertainty about the need for diversion, we also varied the stakes involved in the diversion decision. Recall that stakes were varied in two interrelated ways. First, pilots in the high stakes condition were near the end of their duty time. That meant that they would be unable to divert to the alternate (Norfolk), refuel, and then resume the flight to the destination (a “gas and go”). Secondly, passenger handling facilities at the alternate were poor in the high stakes condition. This meant that no spare crews were available to fly the passenger to the destination, and that ground transportation involved significant delays. In the low stakes condition, on the other hand, pilots had plenty of duty time left, and passenger handling conditions at Norfolk were good.

The intent of the high stakes condition was to make diversion more costly. We expected that this would reduce the probability of diversion and/or cause more experienced pilots to delay diversion, in comparison to the low stakes condition, where diversion involved less sacrifice of passenger convenience.

We turn now to the impact of stakes on diversion decisions and on requests for information. We shall also consider any interactions of stakes with experience and with uncertainty.

Diversion Decision

The manipulation of stakes had no effect on the probability of diversion. Stakes also failed to interact with experience or with uncertainty in its effects on diversion probability. Figure 21 shows that the differences that were in fact observed involved a (non-significantly) *higher* frequency of diversions in the high stakes as compared to the low stakes conditions, regardless of level of experience.

When diversion did occur, moreover, high stakes tended to move the decision *earlier* rather than later. This trend is best understood in the light of a significant interaction between stakes and experience. As shown in Figure 25, it was the *less* experienced pilots who responded to high stakes by diverting earlier than in the low stakes condition. The more experienced pilots were not affected by stakes at all.

These results suggest rather strongly that the stakes manipulation did not work as intended, at least for the less experienced pilots. The shortened duty time may have influenced the less experienced pilots to make a more rapid decision, to avoid infringement of the duty time rule. In effect, it shortened their decision window, even more strongly than the constraints imposed by fuel. For example, if pilots took more time to make a diversion decision, and then encountered further delays in landing at the alternate, they might have exceeded their allowable flight time. It is plausible that this combination of legal and safety factors overrode any implications for increased passenger inconvenience after a diversion.

We can speculate regarding the reason for the lack of effect of stakes on the more experienced pilots: (i) They may have judged the duty-time constraint to be less pressing than did the less experienced pilots. (ii) They were already making very rapid diversion decisions in all conditions (Figure 6), and may have felt no need, or perhaps ability, speed up any more. (iii) The added passenger inconvenience resulting from diversion in the high stakes condition may have received more weight in their decision making than for the less experienced pilots. The latter was our original interest, but any or all of these considerations would be reasonable.

In this light, it is worth highlighting the exceptionally *long* latency for diversion decisions by less experienced pilots when the stakes were low. Many of these pilots were diverting at a time (Phase 4) when information from station ops had become available (if requested) indicating that the runways at Dulles would soon be cleared. The low stakes situation may have influenced less experienced pilots to stretch out the decision making process more than necessary, and to be less diligent in monitoring the situation. In this respect, then, the *lack* of influence of stakes on the more experienced pilots was beneficial; they were better off not slowing down in the low stakes situation.

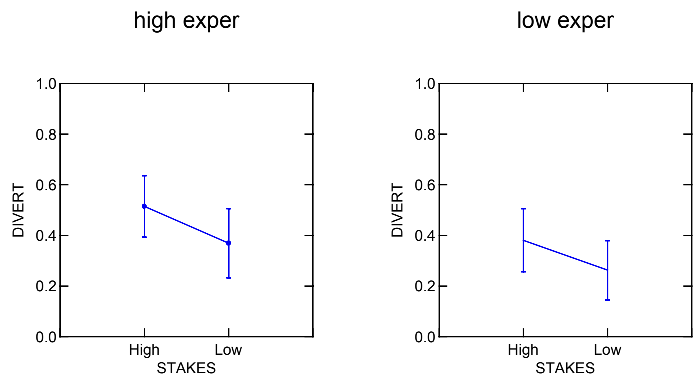


Figure 21. Probability of diversion as a function of stakes, broken down by experience.

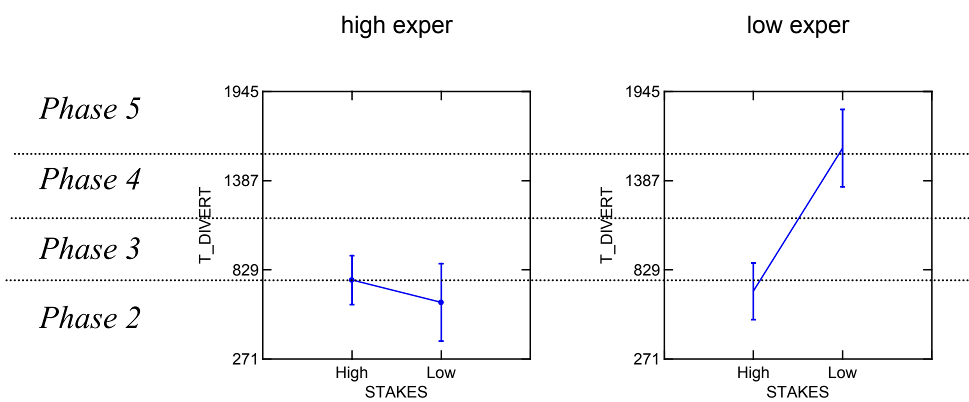


Figure 22. Time of diversion under high and low stakes conditions, broken down by level of experience.

Information requests

As noted in previous sections, the stakes variable was associated with a number of effects and trends in regard to information requests. The following is a summary:

Number of requests for information about causes of delay	Source x Stakes	F(1,226) = 3.649; p = 0.057 (T)
Use of ATC	Stakes	F(1,56) = 2.802; p = 0.100 (T)
Use of Dispatch	Stakes x Experience	F(1,56) = 5.304; p = 0.025 *
	Uncertainty x Stakes x Experience	F(1,56) = 3.496; p = 0.067 (T)
Information requests per source	Stakes	F(1,51) = 4.515; p = 0.036 *
	Uncertainty x Stakes x Experience	F(1,51) = 3.854; p = 0.055 (T)
Time of first request re accident	Uncertainty x Stakes x Experience	F(1,18) = 2.200; p = 0.155 (T)
Time of first request re snow	Uncertainty x Stakes x Experience	F(1,26) = 4.938; p = 0.155 (T)
Time of first request re traffic	Stakes x Experience	F(1,32) = 3.221; p = 0.082 (T)
Time of first request re alternates	Uncertainty x Stakes x Experience	F(1,56) = 2.361; p = 0.130 (T)

The role of high stakes in causing earlier diversions would naturally cut short the information collection process. Therefore, it is not surprising that the two main effects of stakes – on use of ATC and on information requests per source –were each negative (Figure 23).

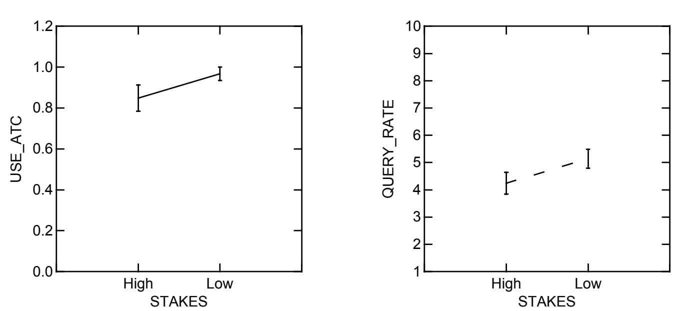


Figure 23.High stakes reduced the probability of using ATC as an information source, and reduced the number information requests per source.

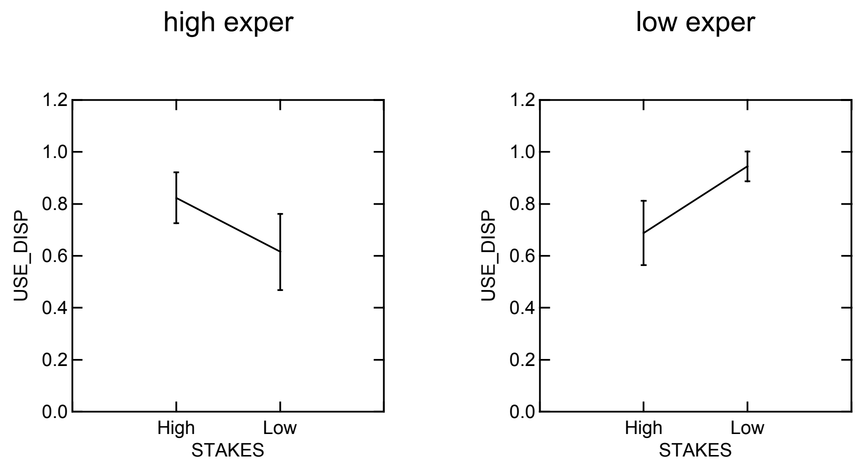


Figure 24. Interactive effect of stakes and experience on the probability of using dispatch as an information source.

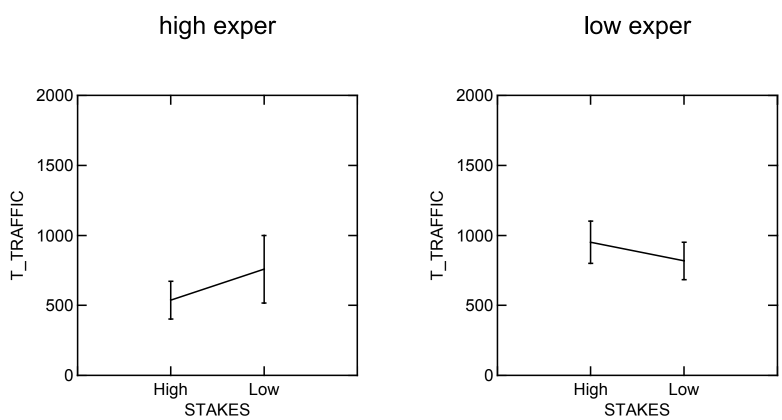


Figure 25. Interactive effects of stakes and experience on the time of first requests regarding traffic.

More interestingly, as seen in Figure 24, experience reversed this effect of stress in at least one case. More experienced pilots were more likely to use dispatch in high stakes conditions rather than less likely. In addition, experienced pilots were earlier in requests regarding traffic when stakes were high (Figure 25). Both of these findings suggest a more active response of experienced pilots to high stakes.

Providing support of this conclusion is an interesting series of three-way interactions involving uncertainty, stakes, and experience. Though no one of these interactions is more than marginally significant, they fall into a strikingly consistent pattern. We will consider them in turn:

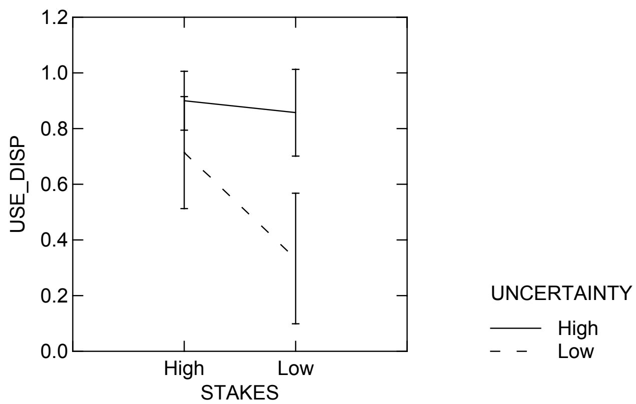
Use of dispatch: We saw earlier (Figure 14) that more experienced pilots tended to use fewer information sources than less experienced pilots, especially in low uncertainty conditions when information was less needed (Figure 15). In particular, the probability of using dispatch as a source was high for less experienced pilots regardless of uncertainty level, but approached 1.0 for more experienced pilots *only* when there was high uncertainty. Figure 26 adds to this picture, by showing that *high stakes* also motivates more experienced pilots to increase their use of dispatch. The experienced pilots appear to be following a rule, that dispatch is a necessary information source if *either* uncertainty is high *or* stakes are high. There is no such clear pattern for less experienced pilots. If anything, they maintained a high use of dispatch in all conditions *except* high stakes.

Recall that diversion times for experienced pilots were not affected by stakes (Figure 22). Therefore, the *increased* use of dispatch by experienced pilots under high stakes cannot be accounted for indirectly by change in diversion time due to stakes. On the other hand, the *reduced* use of dispatch by less experienced pilots under high stakes might in fact be a passive consequence of their early diversion under high stakes. There was a very slight, non-significant increase in diversion probability for both experienced and inexperienced pilots under high stakes (Figure 21). However, since this is the same for both groups, it would not account for their very different information collection strategies under high stakes.

Time of first queries about accident and about snow. Along similar lines, we saw earlier that more experienced pilots were *quicker* to inquire both about the accident (Figure 17) and about snow (Figure 18) under high uncertainty than under low uncertainty conditions. This tended to be reversed for low experience pilots. Figure 27 and Figure 28 add to this picture. Figure 27 shows that *high stakes* also motivate early inquiry about the accident, even under low uncertainty, for more experienced pilots. Similarly, Figure 28 shows that more experienced pilots inquire earlier about snow removal when stakes are high, even when uncertainty was low. The experienced pilots seem to be following an information collection strategy that prompts *early queries* about the causes of delay (the accident and snow) under conditions *either* of high uncertainty *or* high stakes.

Again, the less experienced pilots show no clear pattern. Once again also, there is no plausible indirect explanation of the behavior of the experienced pilots, since their diversion times were not affected by stakes.

High Experience



Low Experience

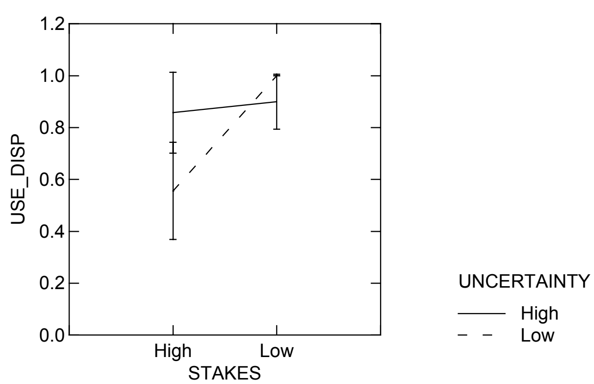


Figure 26. Interactive effects of stakes, uncertainty, and experience on probability of using dispatch as an information source.

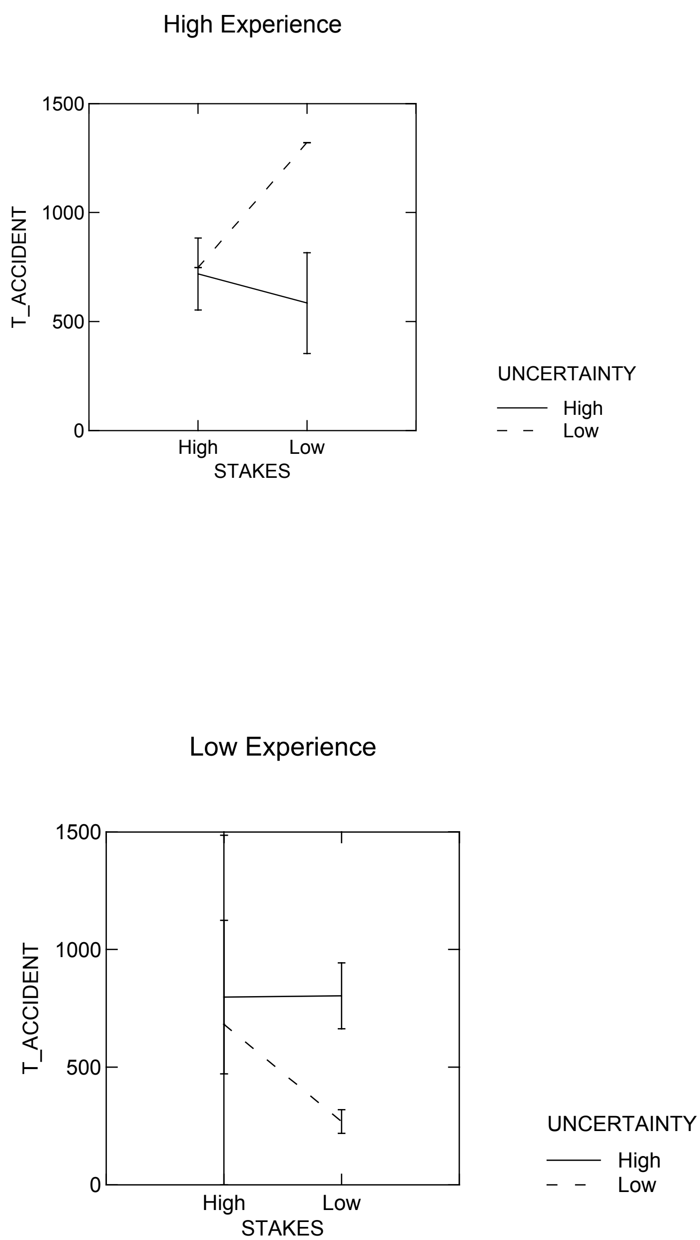
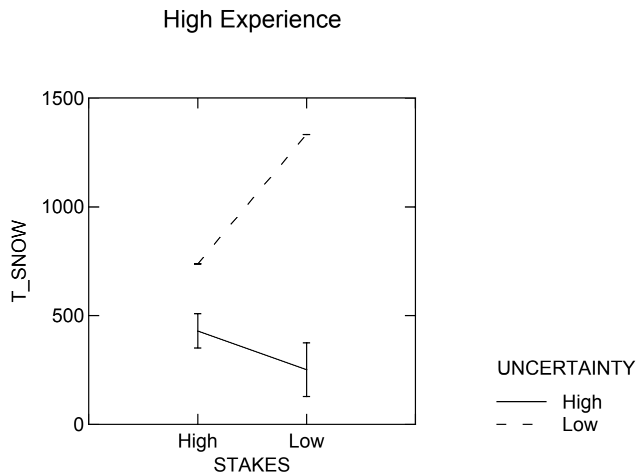


Figure 27. Interactive effects of stakes, uncertainty, and experience on the time of the first information request regarding the accident.



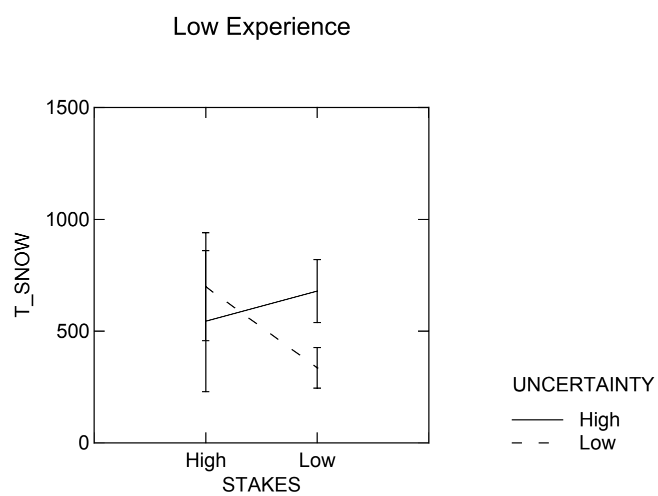


Figure 28. Interactive effects of stakes, uncertainty, and experience on the time of the first information request regarding snow removal.

Number of queries per information source. We found earlier that experience led to a significant increase in the number of queries per information source (Figure 16). We have also seen that high stakes slightly reduces the number of information requests per source (Figure 24). Figure 29 qualifies this picture. Stakes influence query rate *only* for experienced pilots in low uncertainty. In particular, high stakes reduce the intensity with which *experienced* pilots query their sources when uncertainty is *low*. This may again involve a more efficient information collection strategy. When stakes are high, there is a greater premium on the use of time than when stakes are low. Unless uncertainty is high, this apparently means that experienced pilots tend to spend less time monitoring the situation. Again, there is no clear pattern in the responses of less experienced pilots.

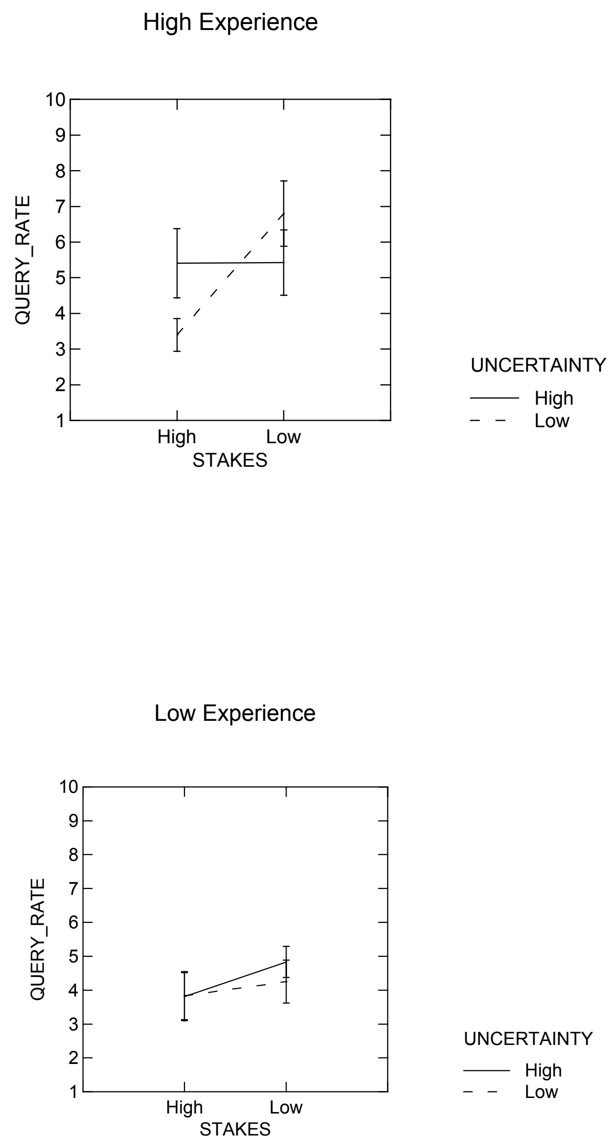


Figure 29. Interactive effects of stakes, uncertainty, and experience on average number of information requests per information source.

Conclusions: Experience and Critical Thinking Skills

More experienced pilots differed from less experienced pilots in a variety of ways that suggest the relevance of the critical thinking skills that we discussed in Chapter 1. In terms of the R/M model (Figure 1), these skills include:

- The Quick Test: Determining when to continue critical thinking and when to commit to a decision, based on a balance of time, stakes, and uncertainty.

- Critiquing: Identifying uncertainty in information, including gaps, conflict, and unreliable assumptions.
- Correcting: Adopting appropriate strategies for resolving uncertainty, such as collecting appropriate information to fill gaps, or verifying sources of information to resolve unreliable assumptions.

We now discuss how the findings in the above study bear on each of these categories of skill.

Faster Decisions and the Quick Test

There was no effect of experience on the probability of diverting. but when more experienced pilots diverted, they diverted *earlier* than less experienced pilots (Figure 6). In addition, unlike less experienced pilots, they did not divert any later in low stakes conditions than in high stakes conditions (Figure 22). More experienced pilots were also more consistent among themselves in the timing of diversions than less experienced pilots (Figure 8 and Figure 9).

This pattern of diversion decisions suggests that experienced pilots reached similar conclusions about diversion, but did so consistently faster than less experienced pilots. Our previous findings (Freeman, Cohen, & Thompson, 1998; Freeman and Cohen, 1996) showed that experienced pilots vary the amount of time to make decisions according to the time constraints, and the available decision window, while less experienced pilots are much less likely to do so. According to the Recognition / Metacognition model, a key skill in handling uncertainty is balancing the costs of time against the potential costs of wrong decisions (the Quick Test). Although it was not the main focus of this study, a judgment of this sort may have been operating in the present scenario.

Nevertheless, this does not explain *how* experienced pilots were able to make effective decisions in less time than the less experienced pilots.

More Selective Use of Information: Critiquing and Correcting

A natural explanation for the speed of decision making is that experienced pilots were more selective about the information sources they used. They used significantly fewer different information sources overall than less experienced pilots. In particular they favored dispatch over station ops more than the less experienced pilots (Figure 13; Figure 15), using station ops not at all in the low uncertainty condition. They relied on average on one information source when the situation was clear, and consulted a second source only when the situation was uncertain (Figure 16). Less experienced pilots typically used two sources even when uncertainty was low. More experienced pilots made more information requests per source from the sources that they did use than the less experienced pilots (Figure 17). Similarly, when uncertainty about the delay was high, more experienced pilots were quick to make queries about the causes of delay, such as snow removal (Figure 18) and the accident clean-up (Figure 19). But they were much slower to ask these questions when the uncertainty was low. Less experienced pilots either did not vary their information request latencies with uncertainty, or varied in the opposite way.

These findings are consistent with the acquisition of better skills for handling uncertainty by means experience. According to the Recognition / Metacognition model, as noted, such skills can be analyzed in terms of critiquing (finding different kinds of uncertainty) and correcting (collecting or recalling information, and revising or evaluating assumptions to resolve the uncertainty). The experienced pilots in this study showed both types of skill. First, they regarded station ops as a *less reliable source* of information (critiquing), and chose not to use it (correcting). Second, more experienced pilots used a second source to *verify or flesh out* information only when the information had low reliability. This represents a strategy of critiquing (identifying low reliability information) and correcting (seeking out a second source to verify the first).

The influence of stakes on information requests illustrates critiquing for another kind of uncertainty. Experienced decision makers not only identify *unreliable* information. They are also able to identify and prioritize *gaps* in information. Highs stakes increased the importance of information about the alternates (in particular, passenger handling capabilities) and about traffic (which would influence the time to land at an alternate in relation to duty time). As a result, more experienced decision makers were more likely to use dispatch as an information source (Figure 24), and quicker to ask questions about traffic (Figure 25), in high stakes conditions than in low stakes conditions. Less experienced pilots did not adapt to stakes in a differentiated way. They were less likely to request information from *any* sources on *any* topics in the high stakes condition.

The more experienced pilots also adapted their requests for clarification about the accident clean-up (Figure 27) and about snow removal (Figure 28). They made more requests for this information when stakes were high, and it was important to fill gaps in knowledge about conditions at the destination, even if uncertainty was low.

In sum, the more experienced pilots seemed to adopt a reasonable *correcting* strategy for handling uncertainty. They increased their use of specific information sources under conditions of *either* (1) high uncertainty, to resolve unreliability in information about the delay, *or* (2) high stakes, to fill gaps in information that had a strong bearing on outcomes in the present situation (Figure 26).

Mental Models of Time Orientation

It is important to understand the role of the kinds of time-oriented *mental models* discussed briefly in Chapter 1. There we differentiated (i) *proactive* mental models, which depict actions adopted in order to influence events, (ii) *predictive* models which adopt actions to take advantage of events already expected to occur, and (iii) *reactive* models, which respond to events that are already in progress. In the present study there was little evidence of the proactive time orientation. Pilots might have been proactive, for example, if they had tried to influence dispatch or ATC decisions in some way. This was not within the scope of the simulation that we provided, and probably not realistic in a similar, real-world situation either. On the other hand, there was ample evidence of a predictive time orientation. In order to regulate the amount of time they spent before deciding to divert, pilots had to anticipate events involving (a) fuel, (b) duty time, (c) accident clean-up, and (d) snow removal. (c) and (c), of course, introduced considerable uncertainty in the high uncertainty conditions.

The critiquing and correcting activities that we observed among experienced pilots were enlisted in the creation and verification of predictive models (involving the accident and snow removal). There is a highly reciprocal relationship between such models and the cognitive strategies that we described above. Pilots were motivated to create predictive models by considerations of time, stakes, and uncertainty. And without the requisite skills in critiquing and correcting, they would have been unable to identify and fill relevant gaps in the models, or to verify and evaluate their reliability.

Mental Models of Purpose

In addition to time orientation, another crucial element of a decision maker's mental model is an understanding of *purpose*. Typically, more experienced decision makers are able to adopt a longer-range, or broader, view of the objectives of the organization of which they are a part. In the case of commercial airline pilots, for example, this may involve handling a rapidly evolving, unfamiliar situation by taking into account both safety goals and, if safety goals are satisfied, more specific company goals. For example, in a previous study of commercial airline pilots (Cohen, M.S. 1993.), we found that more experienced pilots, but not less experienced ones, were *influenced* by company dispatch advice about diversion in a highly uncertainty weather situation. However, critiquing and correcting skills were a necessary prerequisite for this influence. Experienced pilots did not simply follow dispatch advice. They verified it against the situation in the light of safety goals. If the advice could be followed without compromising safety, they did so. If not, they did not. By contrast, less experienced pilots simply *disregarded* dispatch advice altogether.

These behaviors exemplify the kinds of critical thinking skills that can be addressed by training.

CHAPTER 3:
IMPLICATIONS FOR TRAINING PILOT CRITICAL THINKING SKILLS

In this chapter, we outline a strategy for training the critical thinking skills of commercial airline pilots, based on the results reported in Chapter 2 and the theory described in Chapter 1.. According to Salas & Cannon-Bowers (1977), a training strategy orchestrates (1) *tools* (such as feedback and simulation) within (2) *methods* (such as instruction, demonstration, and practice), in order to convey (3) a *content*.

In developing a training strategy, attention must be paid to the underlying theoretical conception of decision making. Different theoretical conceptions are associated with differences in content, methods, and tools – in short, along each of the dimensions that characterize a training strategy. We will briefly examine the implications of different models of decision making for the content, tools, and methods of training. We then move on to a more detailed look at a training strategy based on the extended R / M model.

ROLE OF THEORY IN TRAINING STRATEGY

Table 3 outlines the most salient differences in content, tools, and method among training strategies based on (i) formal models of decision making, (ii) recognition-based models, and (iii) the Recognition / Metcognition model, respectively.

From the point of view of formal models of decision making, the *content* of training is a set of general-purpose techniques (Baron & Brown, 1991). The principle *tool* for defining this content is logic or decision theory, regarded as normative models of thinking (e.g., Watson & Buede, 1987). The primary *method* of presentation is explicit classroom instruction, ranging from focus on formal algorithms (e.g., Laskey & Campbell, 1991), to focus on more qualitative issues such as problem structuring (e.g., Mann, Harmonio, & Power, 1991). Examples of decision problems are not emphasized as content, but are used as *tools* for a variety of purposes: i.e., to motivate the formal techniques during instruction (Adams & Deehrer, 1991), to demonstrate their generality across domains (Mann et al., 1991), and for paper and pencil practice in the component procedures. Problems are selected to illustrate the algorithm or technique that is currently being taught. Often, the problems are artificially prestructured rather than presented naturalistically; i.e., the available options and the probabilities and utilities of their outcomes are explicitly stated. There is typically little emphasis on the ability to match the appropriate method to problems of different types (Beyth-Marom, et al., 1991) or on time-stressed conditions, in which the full analytical method may be infeasible.

Table 3. Differences in training strategies typically associated with different views or models of decision making.

	Models of decision making		
	Logical / Probabilistic Reasoning	**Rapid Recognition**	**Recognition / Metacognition**
Content of training	General purpose formal modeling and reasoning techniques.	Specific situation - response associations.	Mental model types and critical thinking strategies.
Tools used in training	Normative model of decision processes. A small number of paper & pencil examples.	Compilation of cues and responses used by proficient decision makers. Realistic simulation of a large number of representative scenarios.	Cognitive model of proficient real-world knowledge structures & decision processes. Realistic simulation of a moderate number of challenging scenarios, mixed with more routine situations.
Methods of training	Explicit instruction. Practice with procedural feedback.	Little instruction. Practice with immediate feedback re correct response.	Explicit instruction. Practice with delayed or self-administered process feedback.

At the opposite extreme, decision training based on the recognitional point of view attempts to convey examples of decision problems and their solutions as the *content* of training, not general-purpose techniques. Rapid and direct retrieval of the appropriate response to a wide range of situations is the training objective, not choice of the optimal response from a set of alternatives. The primary *method* in recognitional training is practice with a large set of representative problems. Little or no attention is given to explicit instruction, and trainees are usually not encouraged to verbalize the reasons for their decisions during practice. Immediate feedback regarding the correctness of the trainee's response ensures that the situation and the response to be associated with it are represented simultaneously in working memory (Reiser, Kimberg, Lovett, & Ranney, 1992). Two additional features of practice may be used to develop rapid, automatic responding: "Overlearning" – produced by exposure to a large number of trials with consistent stimulus-response mappings (Shiffrin & Schneider, 1977), and practice under time-constraints (Schneider, 1985). *Tools* like high-fidelity simulation may be used to increase the similarity of training conditions to real-world task environments (Means, Salas, Crandall, & Jacobs, 1993).

The R/M model yields an approach to training that is distinct from both formal and pattern recognition models. The *content* of critical thinking training is neither a small set of general-purpose methods nor a vast quantity of specialized patterns and responses. The focus is on a moderately sized set of mental model types (such as purpose, intent, team member reliability, and time orientation) and critical thinking strategies that critique and correct those mental models when direct recognitional retrieval is inadequate. Unlike specialized patterns, both the mental models and the thinking strategies are generalizable in many respects across domains that are characterized by (a) time constraints and (b) uncertainty about human action either within or outside the decision maker's own organization. Unlike general-purpose methods, they are most effectively taught by building on pre-existing familiarity with a particular domain (Kuhn, et al., 1988).

Methods for training for critical thinking include both explicit instruction and practice. Prior instruction on concepts and processing strategies has been found to facilitate learning during subsequent practice (Nickerson, Perkins, & Smith, 1985). In particular, such instruction can provide trainees a new conceptual framework for understanding the skills being trained. For example, the notion that problems can and should be solved by a mechanical application of decision rules must be replaced by a more flexible, iterative, and constructive approach to selecting an action (Brown & Palincsar, 1989). Making principles explicit also helps students transfer what they have learned to varied settings (Collins, Brown, & Newman, 1989).

Practice in critical thinking involves realistic, but non-routine situations, even if they are relatively improbable (Lesgold, Lajoie, Bunzo, & Eggan, 1992). As a result, trainees are exposed to more challenging situations than they would be likely to experience in a representative sampling of the domain. During practice, the explicit articulation of problem-solving strategies is encouraged, to foster reflective self-awareness (Shoenfeld, 1987; Scardamalia & Bereiter, 1985). Problem conditions may be varied – e.g., more and less time-stressed, more and less routine, more or less high stakes – so that trainees learn to decide when to rely on direct recognition and when to use critical thinking strategies.

Feedback focuses on appropriate processes rather than on correct responses. Indeed, the notion that there is a single "correct" answer may often be counterproductive in the kinds of ill-structured or novel problems for which critical thinking is appropriate (King & Kitchener, 1994). Immediate feedback may also be counterproductive. First, it short circuits students' efforts to understand the problem in depth. Delayed feedback, on the other hand, allows for discovery learning through free exploration of the problem (Bennett, 1992). Second, immediate feedback short circuits students' efforts to evaluate their own performance. Instead, trainees can be asked to provide, or at least control, their own feedback, to foster self-reflective skills. For example, trainees may participate in a group discussion after practice, in which they critique the performance of others and respond to feedback regarding their own performance (Shoenfeld, 1987).

A important *tool* for providing feedback is expert modeling of the thinking processes to be trained (Collins, Brown, & Newman, 1989; Druckman & Bjork, 1991). This, too, may be turned into a constructive exercise by asking trainees themselves to compare their own performance with the performance of the expert model (Bloom & Broder, 1951).

A CRITICAL THINKING TRAINING STRATEGY

Table 4 outlines the essential features of a critical thinking training strategy based on the above guidelines. It shows tools, methods, and content associated with the R / M model. We will discuss critical thinking training tools in the remainder of this section, before turning to a more detailed overview of the training content in Chapter 7.

Table 4. Tools, methods, and content of the R / M critical thinking training strategy.

Tools	Methods	Content

• Cognitive task analysis (e.g., critical incident interviews)
• Theory-based definition of critical thinking skills
• Survey of training needs
• Interactive, graphical user interface
• Challenging practice scenarios
• Performance measures (process & outcome)
	• *Information-based:*
	  • Frame decision making as flexible & iterative
	  • Prepare students to use specific concepts & strategies during practice
	  • Demonstrate decision processes
	• *Practice-based:*
	  • Realistic, challenging
	  • Mix with routine
	  • Encourage verbalizing thought processes
	  • Regard feedback as a skill to be trained
	  • Guided practice with feedback and modeling of target behavior
		• Focusing on purpose
		• Critical thinking about purpose
		• Orienting to the enemy in time
		• Critical thinking about time orientation
		• Using initiative

THEORY-BASED DEFINITION OF CRITICAL THINKING SKILLS

Based on the findings of Chapter 2 and the theoretical model described in Chapter 1, the following skills appear to characterize experienced commercial airline pilots. Proficient decision makers in this domain:

1 Develop and use appropriate *mental models*

1.1 *Purpose*: Develop and use models of higher-order or longer-term purposes. Frame decisions in a larger context. This includes an awareness of issues pertaining to both safety (e.g., weather, duty time) and company business (e.g., fuel and passenger convenience).

1.2 *Time orientation*: Develop models of the relationship of own actions to other events, and use these models to develop proactive, predictive, and reactive plans. Take initiative where appropriate.

2 Adopt appropriate *critical thinking strategies* with respect to these mental models

2.1 Identify and seek to fill critical information *gaps* in models . For example, make expectations explicit and monitor events for consistency with expectations.

2.2 Identify and seek to resolve *conflicts* between situation understanding and observations, or between plans and goals. For example, mentally simulate plans to see if they achieve all goals; generate contingency plans, or branches, to compensate for risk

2.3 Identify and evaluate *assumptions* underlying situation models or plans. For example, construct a story that you must believe in order to accept a situation model or plan, and evaluate the story; if the story is implausible, try to develop an alternative mental model, and evaluate that.

2. 4 Determine when and if to commit to action based on available time, stakes, and uncertainty. *Regulate* critical thinking process by balancing costs and benefits.

APPENDIX A:
SIMULATION TOOL

To run the experiment, the experimenter must execute the **nasa4a.prl** and complete some initial questions concerning the participant # and the experimental condition. Once those questions have been answered, the experimenter must execute the second program, **nasa4b.prl**. These two programs (**nasa4a.prl** and **nasa4b.prl**) will synchronize on a semaphore (**flag**) and neither will proceed until both are running. Once both programs are executing, the experimental scenario will begin to unfold. The experimenter will be directed to present events to the participant as they occur (by **nasa4a.prl**), and the participant will be able to query information sources as mediated by the experimenter's interface (**nasa4a.prl**). Data on the experimental session will be recorded into a file for later analysis.

The remainder of this section will describe the specific implementation in some detail. Further documentation on specific functions is in available in the main program files themselves, "nasa4a.prl" and "nasa4b.prl". These program were written in the scripting language "Perl". Implementations of this language are freely available for almost any computer platform. Perl version 5.0 or better is recommended for running the software developed in this project. The reader is referred to www.perl.org and www.activestate.org to locate a version of Perl suitable for their hardware platform and operating system.

Name	Type	Description
nasa4a.prl	Perl program	Main program - provides the menu for the experimenter's interaction with the scenario. The events are generated by a separate program (nasa4b.prl). A *scenario response table* (e.g., RSHISHIU.prl) is used to dynamically determine the contents of the experimenter's menu as the scenario evolves over time.
nasa4b.prl	Perl program	Main program - generates events as the scenario unfolds over time. A *scenario events table* (EVALL.prl) is used to generate these events.
flush.pl	Perl library	Contains an I/O routine used when synchronizing data streams.
jewels.prl	Perl library	Contains an assemblage of generally useful routines, some of which were used to implement the main programs.
flag	Semaphore	Used to synchronize the activities of the two main programs, nasa4a.prl and nasa4b.prl.
new	Directory	Contains the scenario specific files used in the experiment.

new/EVALL.prl	Scenario events table	Used in all experimental conditions. (This is actually a Perl script that is included by nasa4b.prl).
new/RSHISHIU.prl	Scenario response table	Stakes=HI, Uncertainty=HI.
new/RSHISLOU.prl	Scenario response table	Stakes=HI, Uncertainty=LO
new/RSLOSHIU.prl	Scenario response table	Stakes=LO, Uncertainty=HI
new/RSLOSLOU.prl	Scenario response table	Stakes=LO, Uncertainty=LO
data	Directory	All data generated during experimental runs is written into this directory. The files in this directory are named "nasa_#", where "#" is the participant number assigned by the experimenter.

Scenario events table: EVALL.PRL

Events are generated as defined by a *scenario events table*. In this experiment, only a single skeleton of events was used. Those events are encoded within the EVALL.PRL file. The experimental design manipulations were accomplished by varying the responses and or timing of those responses, as defined by the *scenario response files*.

A scenario events table is a Perl script that is programmatically included within the **nasa4b.prl** program. The script is responsible for defining an associative array named **EVENTS** that relates a moment in time within the scenario to an event that must be delivered to the experimental participant. The general format for an event is:

$EVENTS{"13:00:00"} = "event-label\npart-1\rpart-2…\rpart-n";

The time of the event is expressed using a military time format. The event label is separated from the rest of the communications using a newline character ("\n"). The distinct components of the communications are delimited using a carriage return character ("\r"). Lines that begin with a hash mark ('#') are ignored and serve as comments. Blank lines are also ignored. The last non-blank line in the file must be "1;" - this indicates to Perl that the file has been successfully processed.

The following event occurs at 13:00 (the start of the scenario). It has a label that is intended to be meaningful to the experimenter: "ATC; Descent order 1." That label is followed by a dialog pattern in which the ATC communication alternates with responses by the pilot. E.g., ATC will announce "United 222, This is Dulles Approach." and the pilot is expected to respond, e.g., "United 222", after which ATC will continue the communication "United 222, Descend to 17,000".

$EVENTS{"13:13:00"} = "ATC; Descent order 1.\nATC - \"United 222, This is Dulles Approach.\".\r(PILOT - response).\rATC - \"United 222 descend to 17,000.\".";

Scenario response tables: {RSHISHIU.PRL, RSHISLOU.PRL, RSLOSHIU.PRL, RSLOSLOU}

The Subject communicates with a variety of information sources, and the scripted contents on those information sources evolves with the scenario over time. There can be more than one response from the same information source concerning the same topic (e.g., EFC, Snow Removal, Traffic, etc.). To facilitate this, we read a table that not only defines what those responses are, but defines the information sources as well and links them into menu options within the experimenter's menu.

A scenario response table is a plain text file that is parsed by the **nasa4a.prl** program. The format of the file is Seven whitespace delimited fields (the last field may contain embedded whitespace). Lines that begin with a hash mark ('#') are ignored and serve as comments. Blank lines are also ignored.

Field Name	Description
TIME	Time is an integer field that expresses the elapsed time within the experiment in MM:SS, that is, two digits for minutes and two digits for seconds.
FOOTNOTE	This field was an historical artifact. It referred to a footnote within a spreadsheet in which all responses were coded before being transferred to the individual *scenario response tables*.
MENU	Menu is an integer keyed to the numbers on the menu in program that uses these data. Therefore, if you need to create a new menu item, do so in that program, and then use that item number in the first field of the appropriate record(s) here.
SOURCE	The information source, e.g., Dispatch, ATC, Station Operations, ATIS, etc.
TOPIC	The topic of the message, e.g., *acc*=accident, *snw*=snow, *EFC*=Expect Further Clearance, etc.
GROUP	The group of subjects within the experimental design that will receive this response.
MSG	The last field is the full text of the message and may have arbitrary textual content EXCEPT that it may not include a tilde ('~') character.

The only fields that are currently processed are the Menu, and Time (elapsed time in seconds). All other fields are printed to screen at some point, but are essentially free text, that is, you can write whatever you'd like in those fields.

The following is an example record from a scenario response table. This response would be given to a participant if they queried dispatch regarding the accident, beginning two minutes into the scenario and until another response table record is encountered that changes the response text. "29" refers to a menu item within the experimenter's menu. "LOUNCERT" indicates that this response will only be generated in the low uncertainty condition.

02:00 29 33 Dspch acc LOUNCERT "Cleanup is proceeding as expected. Station Ops has a good handle on how long it will take."

Data records: All experimental data is recorded within files in the **data** directory. These files have a record format consisting of tab-delimited fields. These fields are {Subject#, ClockTime, ElapsedTime, MenuItem#, Text, Note}. The Note field is optional and is therefore not always present.

Field Name	Description
Subject #	The subject number assigned to the pilot by the experimenter.
Clock Time	Measured in Hours : Minutes : Seconds.
Elapsed Time	In seconds since the start of the scenario.
Menu Item #	Identifies a menu item in the experimenter's menu.
Text	The textual label from the *scenario response table*, e.g., "ATC : EFC"
Note	A note as recorded by the experimenter during a session, e.g., "Prefers 19Right because has lower minimum. Let flying guy brief the approach. Only suggestion is that he keep autopilot on until runway in sight with medium auto brakes. If goes below 1800 will have to head to Norfolk. Tell flight attendants."

Experimenter's menu. The experimenter's menu is generated by the **nasa4a.prl** program. It provides a range of methods to record specific information requests and actions taken by the pilot, as well as methods that query the simulation environment to identify the current responses for an information request by the pilot. (Pending events are generated by **nasa4b.prl** program and are displayed in a separate window on the computer desktop.)

This table summarizes the specific menu options that were available to the experimenter. Other options may be readily added by changing the appropriate data definition within the **nasa4a.prl** program. The menu codes are used to label records in the data file generated for each experimental session.

Selecting a menu option typically results in appending a new record to the data file for that experimental session. The format of such records is described above.

Code	Label	Description
1	Misc. notes	Used for taking notes on questions and behaviors of the pilot that are not readily categorized by the other menu options.
2	Divert	Used to record when the pilot make a decision to divert.
3	Calculator	Used to record when the pilot requests the use of a calculator.
4	IAD plates	Pilot requests the IAD (Dulles International) approach plates.

5	BWI plates	Pilot requests the BWI (Baltimore-Washington International) approach plates.
6	OFS plates	Pilot requests the ORF (Norfolk) approach plates.
7	Exp specified	Experimenter specified. This menu item is reserved for specific use as determined by the experimenter. The experimenter is expected to document the meaning of any such use so that the data may remain readily interpretable.
8	Exp specified	Experimenter specified - *ibid.*
9	Exp specified	Experimenter specified - *ibid.*
11	ATIS: IADwx	ATIS weather for IAD.
12	ATIS: BWIwx	ATIS weather for BWI.
21	ATC: EFC	ATC gives pilot an EFC (Expect Further Clearance).
22	ATC: Traffic	ATC gives pilot an update on traffic patterns (e.g., stacking and diverting of planes at IAD).
31	ACARS/Dispatch: EFC	Pilot requests information from ACARS/Dispatch concerning an EFC.
32	ACARS/Dispatch: Alternates	Pilot requests information from ACARS/Dispatch concerning alternate destinations.
33	ACARS/Dispatch: Accident	Pilot requests information from ACARS/Dispatch concerning an accident.
34	ACARS/Dispatch: Snow removal	Pilot requests information from ACARS/Dispatch concerning snow removal.
35	ACARS/Dispatch: Traffic	Pilot requests information from ACARS/Dispatch concerning traffic patterns.
41	StationOps: EFC	Pilot requests information from station operations concerning an EFC.
42	StationOps: Accident	Pilot requests information from station operations concerning an accident.
43	StationOps: Snow removal	Pilot requests information from station operations concerning snow removal.
44	StationOps: Traffic	Pilot requests information from station operations concerning traffic patterns.
t	Display current time	Display the current scenario time for the experimenter.
x	Exit	The experimenter must confirm this selection in order to end the experimental session.

APPENDIX B: ORDER, EXPLAINABILITY, AND CONFLICTING EVIDENCE

INTRODUCTION

This appendix describes a proposed simulation study designed to examine cognitive factors underlying inference by commercial airline crews under uncertainty and time pressure. In particular, it focuses on the use of base rate information in a context in which such information conflicts both with other base rate information and real-time cues. The proposed design has the following novel features:

1. *Information relevant to base rates is presented in a realistic form, as conditioning variables and surrogate experience.* Laboratory research suggests that people tend to ignore base rate information when "individuating" evidence regarding the hypotheses is available (Kahneman, Slovic, & Tversky, 1982). In these studies, base rate information was presented explicitly in the form of stated numerical frequencies. (For example, in one well-known study by Kahneman and Tversky, subjects were told that 85% of the cabs in a city are blue, and 15% are green [base rates]; a witness who is 80% accurate testified that the cab responsible for a hit-and-run accident was green [individuating evidence]. What is the probability that the cab responsible for the accident was blue?). By contrast, other studies show that when base rates are learned from actual experience of events, they appear to influence decisions appropriately (Nisbett, Borgida, Crandall, & Reed, 1976). In real-world domains, however, base rates may be learned in a variety of other ways. Most base rate learning is not the experience of a simple numerical frequency of an event. Rather, it is the association of varying numerical frequencies with different conditions, or the learning of causal models that predict different likely outcomes under different conditions. The relevant base rate of an event on a particular occasion thus depends on the prevailing conditions (e.g., mountainous terrain and winds make mountain waves more likely). In the present study, such conditions are explicitly stated in the flight departure package, where they must be noticed and remembered by pilots. Another important source of base rate information is vicarious experience, reflected in the reports of others. In the present study, we examine whether pilots can act effectively on base rates reflected in the number of reports of mountain wave activity and reports of pitot heater failure by other pilots (pireps, flight crew alert bulletin).

2. *The study presents base rates for independent causal processes (mountain wave and pitot heater failure).* These two base rates represent separate bodies of knowledge that would not normally be expected to interact or to be stored in an integrated fashion (cf., Pearl, 1988, p. 184). The only link between these causal processes is that each provides a possible explanation for a subsequent event: a rise in indicated airspeed during initial ascent. In traditional base rate studies alternative hypotheses are mutually exclusive values of a single variable (e.g., blue cab versus green cab, engineer versus lawyer), and base rates are therefore non-independent. In this study, successful decision makers must retrieve causal knowledge from two separate domains, and construct and compare two causal stories.

3. *The study varies the reliability or explainability of base rate information.* For example, pireps may be a highly reliable predictor of mountain wave activity during a flight if the reports are very recent and from aircraft very close to the intended flight path. The same number of pireps may be a far less reliable predictor (and thus more easily explained away) if they are less recent and/or from aircraft at different altitudes and locations. (Notice that it is because of point 2 above that we can independently vary the reliability of the base rate information for the two

competing causes.) In the absence of conflict, decision makers may assume that base rate information is correct; but when the base rate information conflicts with individuating evidence or with base rates for competing potential causes, they may probe for potential unreliability in order to arrive at a single coherent explanation. Traditional studies sometimes vary the frequencies represented by base rates, but do not (intentionally) vary their reliability. Nevertheless, variations in reliability may explain some of the paradigmatic findings in this area. For example, it has been claimed (Kahneman et al., 1982) that base rates are not neglected when they are causally linked to the hypothesis in question (e.g., if subjects are told that the percentage of accidents caused by blue cabs is 85%, they can infer that blue cab drivers are less competent, hence, more likely to cause accidents) rather than simply statistically linked (the percentage of blue cabs in the city is 85%). This effect, however, may not reflect the preference for causal reasoning so much as it reflects explaining away base rate information that (1) conflicts with individuating evidence (witness's testimony) and (2) is unreliable (because it neglects important causal factors such as the relative competence and training of drivers for the two cab companies). In other words, non-causal base rates tend to be less reliable predictors, and are thus easier to explain away when they conflict with individuating evidence. The notion of reliability explains the additional finding that non-causal base rates are in fact used when there is no conflict with individuating evidence (hence, no need to explain away the non-causal base rates). In the present study, we vary the reliability of base rate information (pireps regarding mountain wave activity), without affecting the causal nature of the underlying reasoning.

4. *The study varies the order in which base rate information for competing hypotheses is presented.* A popular paradigm for studying how decision makers handle conflicting evidence is to vary the presentation order for evidence confirming and disconfirming a hypothesis. Such studies have themselves produced mixed results, yielding primacy effects in some cases (greater influence for the same evidence when it appears first than when it appears second) and recency effects in others (greater influence when the same evidence appears last) (e.g., Einhorn & Hogarth, 1987; Adelman, Bresnick, Black, Marvin, & Sak, 1995). A variety of explanations for the different results have been suggested: e.g., explanations of primacy in terms of attention decrement, expected redundancy, overgeneralization from an initial small sample, locking in of an early conclusion, correlation between order and importance, anchoring and insufficient adjustment, or reinterpreting later cues; and explanations of recency in terms of anchoring and overadjustment, or contrast effects. In this study, we vary the presentation order of independent base rate information (for mountain wave activity and pitot heater failure, respectively), rather than the presentation order of individuating evidence. A crucial difference is that the two sets of base rate information are not in conflict with one another when they are actually presented, but only later in the scenario after the event to be explained has occurred (the unusual rise in indicated airspeed). This ambiguous event, when it first occurs, may be explained as a consequent of either cause, and thus induces conflict retrospectively. Resolving the conflict may trigger a process of retrieving the initial base rate information, probing its reliability (perhaps for the first time), and selecting the most plausible causal story (involving a reinterpretation of the significance of one of the base rates). Order effects (whether primacy or recency) in this context are not explicable in terms of processes that would have to occur during the actual presentation of the base rate information (such as expected redundancy of later information, overgeneralization from early information, anchoring and adjustment, or contrast effects) since such inference processes are not likely to take place during the presentation of the base rate information (i.e., before the occurrence of the ambiguous consequent). A careful examination of

the pattern of order effects may shed light on the processes of retrieval and retrospective reinterpretation involved in resolving the subsequent conflict. For example, a possible pattern would involve an effect of presentation order on judgments (e.g., the assessed probability of a mountain wave) after occurrence of the ambiguous event but not before, and an interaction of presentation order with the reliability/explainability of the base rate information regarding mountain waves.

5. *There is a natural continuous dependent measure of the pilot's current belief regarding the most likely cause.* If the rise in indicated airspeed is caused by a mountain wave, the appropriate response by the pilot (or autopilot) is to reduce power and raise the nose of the aircraft until actual airspeed decreases to the desired level. On the other hand, if the rise in indicated airspeed is caused by frozen pitot tubes, the above response will lead to a reduction in airspeed and, if continued, to stall and crash; hence, the pilot should not adjust the throttle and should prevent the autopilot from doing so. A key dependent measure, therefore, will be the initial response and response latency by the pilot to the rise in indicated airspeed, and the latency of any change in that response. For this reason, it is unnecessary to interrupt the scenario to ask pilots to assess probabilities regarding various possible causes of the rise in indicated airspeed. (Other dependent measures are also available, however, such as information requests. Moreover, we will ask pilots for explicit judgments of the probability of a mountain wave before the flight actually begins.)

METHOD

Subjects. We will need 60 or more captain/first officer crews, with 747-400 experience. If possible, variations in crew experience would be useful, i.e., some crews with significantly more commercial flight experience than others.

Design. Two between-crew independent variables will be crossed: mountain wave base rates (3 levels) X base rate presentation order (2 levels), for a total of six experimental groups. Both variables involve differences in the departure package given to pilots before the flight. The three levels of mountain wave base rates are as follows:

Condition	Departure Package contains:
High & reliable base rate	Conditions consistent with mountain waves (mountainous terrain, high velocity winds at intermediate altitudes, strong jet stream, low altimeter); numerous recent pireps in vicinity report mountain waves
High but unreliable (explainable) base rate	Conditions consistent with mountain waves (mountainous terrain, high velocity winds at intermediate altitudes, strong jet stream, low altimeter); numerous pireps report mountain waves, but they are not recent and pertain to different regions from scheduled flight
Low base rate	Conditions do not favor mountain waves (mountainous terrain, but moderate velocity winds at intermediate altitudes, high altimeter); no pireps re mountain waves

The two levels of base rate presentation order involve: (i) Flight crew alert bulletin regarding pitot heater problems placed on top of departure package, hence, before information regarding mountain waves, versus (ii) flight crew alert bulletin regarding pitot heater problems placed on bottom of departure package, hence, after information regarding mountain waves.

Procedure. Crews will receive standard flight packages, including (i) a flight plan, (ii) set of aeronautical charts, and (iii) a flight crew alert bulletin regarding problems with pitot heater malfunctions. After examining this material, the pilots will be asked to assess the probability that they will encounter a mountain wave during the flight. They will then fly the scenario described below. The first officer in each crew will be assigned as pilot flying. This is intended to maximize workload for the captain, who must both diagnose the problem and watch the first officer. (Conversely, if the first officer were the pilot non-flying, the first officer would probably not feel compelled to watch the captain, but would devote full attention to diagnosing the problem.) In addition, this may deprive the pilot flying of a clear view of the standby airspeed indicator (which is on the captain's side).

Some time may be required to orient pilots to the simulator. Participants should be asked to bring their own airline's operating manuals.

Scenario. Following a routine takeoff by a 747-400 on a flight from Seattle (SEA) to Kennedy airport (JFK), the aircraft is held at 9000 feet for about ten minutes. During this time, heating elements on three of the four pitot sensors fail, and the combination of moisture in the atmosphere and freezing temperatures at the aircraft's current altitude cause the pitot tubes to freeze. (The pitot attached to the standby airspeed indicator does not freeze). As the plane ascends, the increasing differential in pressure between pitot and static sensors will cause indicated air speed to rise inappropriately.

How the pilot reacts to the increase in indicated airspeed will depend on his or her understanding of its cause. There are at least two candidate causal stories (or mental models) in which this same event may occur. Figure 30 portrays both causal stories: An (incorrect) mountain wave explanation unfolds from left to right at the top of the figure and is shown with solid causal arrows, and the (correct) pitot heater failure story unfolds from left to right along the bottom of the figure and is shown by dotted causal arrows. Notice that the two stories share the ambiguous event, rise in indicated airspeed. In this figure, the response of the pilot (or autopilot) is assumed to fit the mountain wave mental model. If the rise in indicated airspeed really were caused by a mountain wave, this response (reducing power, raising the nose) would restore the desired speed and result in stabilized flight. However, since the correct story involves pitot tube failure, the result is a positive feedback loop: The pilot's actions result in accelerated climb, hence, further increases in differential pressure between static and frozen pitot sensors, and further reductions in indicated airspeed. If the pilot continues to try to correct this by reducing power, the cycle continues until the plane stalls and crashes.

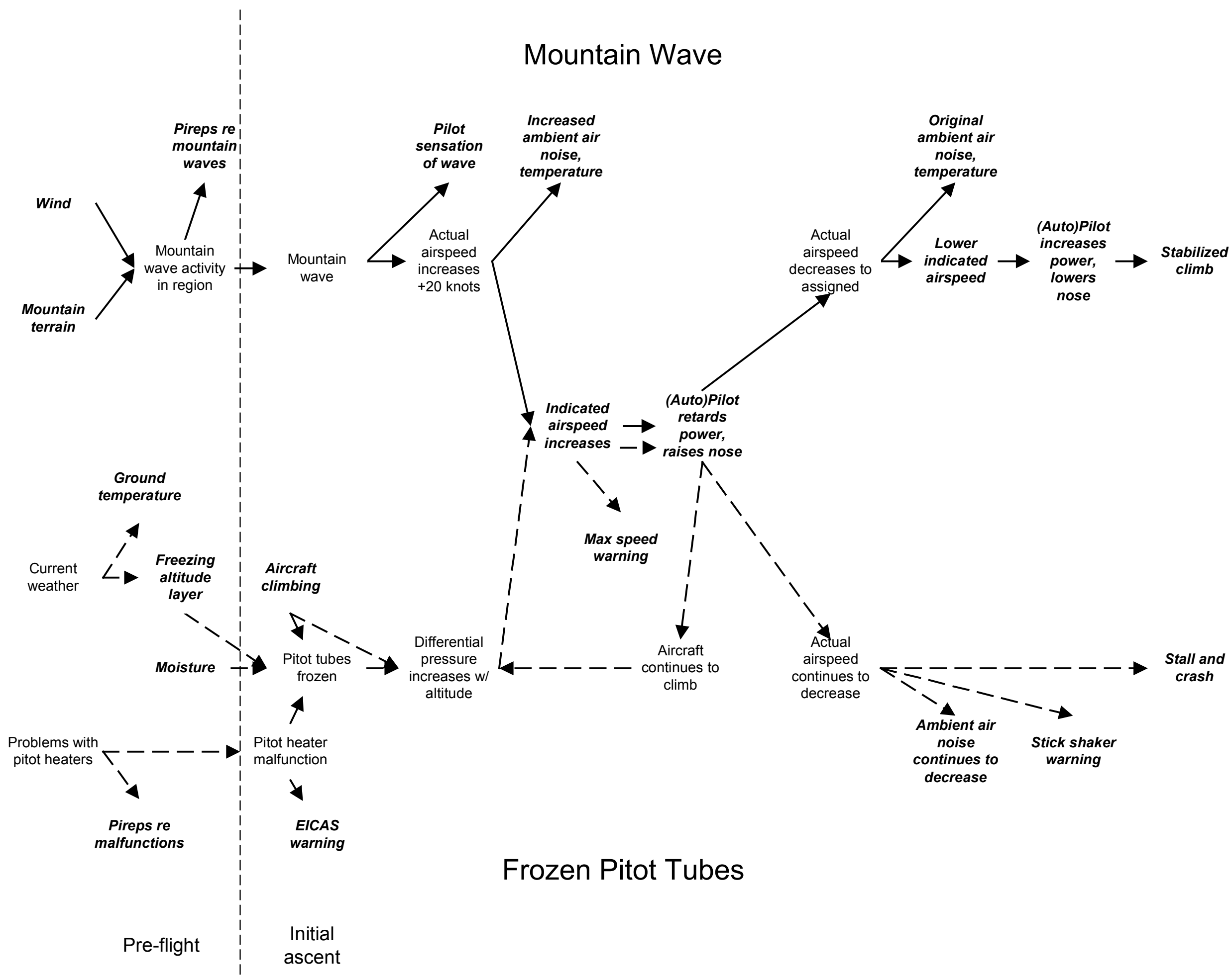


Figure 30. Flow of events as they might be interpreted (solid lines) and as they actually happen (dotted lines).

Pilots can avoid this outcome by correctly diagnosing that the increase in indicated airspeed is erroneous. They may do this by attending to a variety of real-time cues: an EICAS warning regarding pitot tube heater failure (which, however, does not spell out the implications for indicated airspeed); the continued increase in indicated airspeed until a maximum speed warning occurs; the continued reduction in ambient air noise (indicating a reduction in actual airspeed); and the conflict between the main airspeed indicators and the standby airspeed indicator (which does not fail). Any of these cues may also cause them to reconsider the reliability of base rate information supporting the competing mountain wave explanation.

The following table summarizes the key inferentially relevant information available to pilots at different stages of the experimental scenario, and dependent measures applicable to each stage:

	Pre-Flight	**Holding at 9k**	**Continued ascent**
Evidence for (against) mountain wave	Base rate info re mountain waves: pireps, winds, altimeter, terrain		(No increase in ambient air noise) (No mountain wave sensations)
Critical event			Indicated airspeed increases
Evidence for pitot heater failure	Base rate info re pitot heater failures: flight crew alert, temperature	EICAS warning re pitot heater failure	Decrease in ambient air noise, no decrease in IAS, max speed warning, stick shaker
Dependent measures	Estimated probability of a mountain wave	Information seeking & crew discussion (e.g., referring to departure package for information re mountain waves or pitot heater)	Initial response & latency after IAS increase; subsequent response after continued IAS increase

Materials. The following is an outline of the briefing materials and scenario timetable to implement the above scenario.

Flight Departure Package

Departure Package Item	Condition A: High, reliable mountain wave base rate	Condition B: High, unreliable mountain wave base rate	Condition C: Low mountain wave base rate
Aircraft:	747-400	Same	Same
Origin:	Seattle, 23:59 scheduled departure	Same	Same
Destination:	Kennedy, standard arrival time	Same	Same
Alternates:	Standard alternates	Same	Same
Seattle rwy:	16L	Same	Same
Altimeter:	29.78	Same	29.92
Flight plan:	Standard instrument departure (SID)	Same	Same
Weight:	½ normal passenger load; ½ normal freight load	Same	Same
Fuel:	Standard fuel for destination + diversion to first alternate + 45 minutes + 30 minutes of holding	Same	Same
Seattle wx:	500 overcast; 1.5 mile visibility; rain & fog; temperature 10°C (50°F); wind 140 at 8kts; high velocity winds at intermediate altitudes, strong jet stream and westerly[16] flow.	Same	500 overcast; 1.5 mile visibility; rain & fog; temperature 10°C (50°F); wind 140 at 8kts.
Enroute terminal forecasts:	Kennedy: 400 overcast; 1 mile visibility; rain & fog; temperature 42°F; winds SE at 7kts.	Same	Same

[16] We have assumed that westerly winds are most likely to generate mountain waves in the vicinity of Mt. Rainier. If easterly winds are more likely to do so, please substitute easterly winds.

Pireps (to be appended to wx report):	Four pireps reporting moderate mountain wave activity in the vicinity of planned flight path from 20:00 to 22:45. No control problems reported. Tops of overcast reported at flight level 19K feet by C-500 at 22:50. Light to moderate rime ice reported by 727 between 8-13K feet along J-90. Log does not mention system problems or unusual maintenance.	Four pireps reporting moderate mountain wave activity, none more recent than 18:00 and located significant distance from planned flight path.. Remaining items same	(No mountain wave activity indicated on pirep.) Remaining items same

In addition, the flight departure package will include the following:

Flight Crew Alert Bulletin

All B-747 Flight Crews

We have recently experienced a number of pitot heat failures on our -100, -200, and -400 series aircraft. Boeing Commercial Airplane Co., the FAA, and our airline are working closely in order to develop a fix. In the meantime, exercise extreme caution when operating in areas of known or forecast icing conditions if you receive indications of a possible pitot heat failure.

Captain Bob Welch
VP Flight Operations

This alert will be place on the top of the departure package for half the crews, and on the bottom for the other half.

The instructions that we are preparing for subjects will include the following text. (The text will be identical for all groups.) We would appreciate your comments:

- Local time is: 23:30. You are preparing for takeoff from Seattle. Please follow standard procedures: enter the flight plan into the FMC, obtain ATIS information, complete the pre-engine start checklist, push back from the gate, start engines and taxi out to runway 16L for takeoff.
- Please assign the first officer as flying pilot, as this is his leg of the trip.
- Please use the operating procedures of the carrier for which you fly.

- If you have any questions, please ask them now.

Timetable of scenario

The following timetable indicates the events that should occur during the flight portion of the scenario and the point at which each event should occur. They are the same for each of the experimental conditions.

We have not attempted to estimate the rate at which indicated airspeed should rise in this scenario, given the frozen pitot tubes. Hopefully, the simulator has a formula for computing this from the growing differential in air pressure between the frozen pitot tubes and the unfrozen static pitot tubes as the aircraft ascends. Please assume that the pressure in the pitots is slightly higher than that found at the altitude at which the pitots freeze (9K feet) because the air in the pitot is compressed a little by the ice plug.

Do not inhibit any effects or events that occur naturally in the simulator. We want these pilots to have the full experience the simulator has to offer, including all of the system alarms (e.g., EICAS heater failure messages, overspeed clacker, underspeed stick shaker), messages, or other events that would be expected under the conditions in this scenario. We are particularly interested in simulating the natural ambient air hiss over the cockpit (which decreases as airspeed declines) and external temperature changes.

Cue to start events	All Groups
Scenario start	Pre-engine start, startup, taxi and takeoff: No indicators of problems
AC reaches 4K feet.	ATC issues 9000 foot altitude restriction due to traffic. ATC notifies pilot he should expect to climb to filed altitude within 10 minutes. (This delay is intended to give pitot tubes time to freeze).
AC climbing from 4K to 9K feet	ATC issues radar vector to 030, to clear traffic to permit climbing. Show target on TCAS if possible. ATC issues radar vector from 030 to 130, to clear traffic to permit climbing. Show target on TCAS if possible.
AC at 9K feet (AC will be approaching or over the Cascade Mts. on radar vectors to airways)	The following events occur in a rapid stream: Total air temperature varies between -1°C and 0°C. EICAS issue standard messages for failure of heaters in Capt's pitot, FO's pitot, and right aux pitot 2. Do not trip circuit breakers for pitot heaters. Do not issue warnings concerning left aux pitot 1 or any statics. Heaters for these do not fail. Indicated airspeed now changes erroneously with any change in altitude. ATC issues radar vector from 130 to 090, to intercept airway.
2-3 minutes after EICAS warnings (sufficient time to freeze pitot tubes)	ATC clears ac to ascend to 37K feet.
Remainder of ascent	Pilots attempt to detect, diagnose and solve the problem of the erroneous indicated airspeed.

REFERENCES

Adams, M. J. & Feehrer, C. E. (1991). Thinking and Decision Making. In J. Baron & R. V. Brown (Eds.), Teaching Decision Making to Adolescents (pp. 79-94). Hillsdale, NJ: Erlbaum.

Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(3), 484-495.

Anzai, Y. (1991). Learning and use of representations for physics expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise* (pp. 64-92). Cambridge: Cambridge University Press.

Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D.L. Forrest-Pressley, et al. (Eds.), *Metacognition, cognition, and human performance. Vol I: Theoretical perspectives*. NY: Academic Press, Inc.

Baron, J. & Brown, R. V. (Eds.). (1991). *Teaching decision making to adolescents*. Hillsdale, NJ: Erlbaum.

Beach, L.R., & Mitchell, T.R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review*, *3*, 439-449.

Bennett, K. (1992). The use of on-line guidance, representation aiding, and discovery learning to improve the effectiveness of simulation training. In J.W. Regian & V.J. Shute (Eds.), *Cognitive approaches to automated instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates, 217-244.

Berry, D.C., & Broadbent, D.E. (1987) Interactive tasks and the implicit-explicit learning process. *Psychological Research, 49*, 7-15.

Beyth-Marom, R., Fischhoff, B., Quadrel, M.J., & Furby, L. (1991). Teaching decision-making to adolescents: A critical review. In J. Baron & R.V. Brown (Eds.), *Teaching decision-making to adolescents.* Hillsdale, NJ: Erlbaum.

Blau, P.M. (1970). A formal theory of differentiation in organizations. *American Sociological Review, 35*, 201-218.

Brown, A.L., & Palincsar, A.S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates, 394-452.

Burns, T., & Stalker, G. M. (1962). The management of innovation. London: Tavistock.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 221-246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R.A. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 333-380). San Francisco: Jossey-Bass Publishers.

Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). NY: Academic Press.

Castellan, N.J., Jr. (1993). Paradoxes in individual and group decision making: A plea for models. In N. J. Castellan (Ed.), *Individual and group decision making* (pp. 125-134). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.S. Steinberg (Ed.), *Advances in the psychology of human intelligence* (Vol. I).Hillsdale, NJ: Lawrence Erlbaum Associates.

Christensen-Szalanski, J.J.J. (1993). A comment on applying experimental findings of cognitive biases to naturalistic environments. In G.A. Klein, J. Orasanu, J., R. Calderwood, & C.E. Zsambok, (Eds.), *Decision making in action: Models and methods.* Norwood, NJ: Ablex.

Cohen, L.J. (1981). Can human irrationality be experimentally demonstrated? *The Brain and Behavioral Sciences, 4*(3), 510-517.

Cohen, M.S. (1986). An expert system framework for non-monotonic reasoning about probabilistic assumptions. In J. F. Lemmer & L. N. Kanal (Eds.), *Uncertainty in artificial intelligence* (pp. 279-293). Amsterdam: North-Holland Publishing Co.

Cohen, M.S. Taking risks and taking advice: The role of experience in airline pilot diversion decisions. Proceedings of the Seventh International Symposium on Aviation Psychology, Columbus, OH, 1993.

Cohen, M.S. (1993). The naturalistic basis of decision biases. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 51-99). Norwood, NJ: Ablex.

Cohen, M.S., Freeman, J.T., & Wolf, Steve. Meta-Recognition in Time Stressed Decision Making: Recognizing, Critiquing, and Correcting. Human Factors (38,2), pp. 206-219. 1996.

Cohen, M.S., Parasuraman, R., & Freeman, J.T. (1998). Trust in Decision Aids: What is it and how can it be improved? *Proceedings of the 1998 Command and Control Research & Technology Symposium*, Monterey, CA.

Cohen, M.S., & Freeman, J.T. (1996). Thinking naturally about uncertainty. In *Proceedings of the Human Factors & Ergonomics Society, 40 Annual Meeting*. Santa Monica, CA: Human Factors Society.

Cohen, M.S., Freeman, J.T., & Thompson, B.B. (1998). Critical thinking skills in tactical decision making: A model and a training method. In J. Canon-Bowers, & E. Salas (Eds.), *Decision-making under stress: Implications for training & simulation.* Washington, DC: American Psychological Association Publications.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Donaldson, L. (1995). *American anti-management theories of organization: A critique of paradigm proliferation*. New York: The Cambridge University Press.

Dreyfus, H.L. (1997). Intuitive, deliberative, and calculative models of expert performance. In

C.E. Zsambok & G. Klein, (Eds.), *Naturalistic decision making* (pp. 17-28). Mahwah, NJ: Erlbaum.

Druckman, D., & Bjork, R.A. (Eds.). *In the mind's eye: Enhancing human performance*. Washington, DC: National Academy Press, 1991.

Ericsson, K.A., & Smith, J. (1991). Prospects And limits of the empirical study of expertise: an introduction. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise*. Cambridge: New York: University Press.

Fischhoff, B., & Johnson, S. (1990). The possibility of distributed decision making. In National Research Council. (Ed.), D*istributed decision making: Report of a workshop*. Washington DC: National Academy Press.

Forrest-Pressley, D.L., MacKinnon, G.E., & Waller, T.G. (Eds.) (1985). *Metacognition, cognition, and human performance*. NY: Academic Press.

Gardenfors, P., & Sahlin, N. (1982). Unreliable probabilities, risk taking and decision making. *Synthese*, *53*.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K.A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hammond, K.R. (1993). Naturalistic decision making from a Brunswikian viewpoint: Its past, present, future. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods*. (pp. 205-227). Norwood, NJ: Ablex Publishing Corporation.

Holyoak, K. J. (1991). Symbolic connectionism: toward third-generation theories of expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise*. Cambridge: New York: University Press.

Janis, I.L. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.

Kahneman, D., Slovic, & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. NY: Cambridge University Press.

Kerr, N.L. MacCoun, R.J. & Kramer, G. P. (1996). "When are N heads better (or worse) than one?": Biased judgment in individuals versus groups. In E. Witte & J.H. Davis, *Understanding group behavior: Consensual action by small groups (Vol. 1)*. Mahwah, NJ: Lawrence Erlbaum Associates.

King, P.M. & Kitchener, K.S. (1994*). Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass Publishers.

Klein, G. A. (1993). A Recognition-Primed Decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex.

Kleinman, D.L., & Serfaty, D. (1989). Team performance assessment in distributed decision making. In R. Gilson, J.P. Kincaid, & B. Goldiez (Eds.), *Proceedings of the Interactive Networked Simulation for Training Conference*. Orlando, FL: Naval Training Systems Center.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego: Academic Press, Inc.

Larkin, J.H. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 311-344). Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J., McDermott, J., Simon, D. P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science, 20*(208), 1335-1342.

Laskey, K., & Campbell, V. (1991). Evaluation of an intermediate level decision analysis course. In J. Baron & R. Brown (Eds.), *Teaching decision making to adolescents.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Leddo, J., Chinnis, J.O., Jr., Cohen, M.S., & Marvin, F.F. (1987). *The influence of uncertainty and time stress on decision making* (Technical Report 87-1). Falls Church, VA: Decision Science Consortium, Inc.

Lesgold, A., Lajoie, S., Brunzo, M., & Eggan, G. (1992). SHERLOCK: A coached practice environment for an electronics troubleshooting job. In J.H. Larkin, & R.W. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Levi, I. (1986). *Hard choices: Decision making under unresolved conflict*. NY: Cambridge University Press.

March, J.G. (1996). Exploration and exploitation in organizational learning. In M. D. Cohen & L. S. Sproull (Eds.), *Organizational learning* (pp. 101-123). Thousand Oaks, CA: Sage Publications.

McIntyre, R.M., & Salas, E. (1995). Measuring and managing for team performance: Lessons from complex environments. In R.A. Guzzo, E. Salas, & Associates (Eds.), *Team effectiveness and decision making in organizations* (pp. 9-45). San Francisco: Jossey-Bass Publishers.

McKeithen, K.B., Reitman, J.S., Rueter, H.H., & Hirtle, S.C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology, 13.*

Means, B., Salas, E., Crandall, B., & Jacobs, T.O. (1993). Training decision makers for the real world. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), <u>Decision making in action: Models and methods.</u> Norwood, NJ: Ablex Publishing Corporation, 306-326.

Metcalfe, J., & Shimamura, A.P. (1994). *Metacognition*. Cambridge, MA: The MIT Press.

Nelson, T.O. (1992). *Metacognition: Core readings*. Boston: Allyn and Bacon.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nickerson, R.S., Perkins, D.N., & Smith, E.E. (1985). *The teaching of thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Orasanu, J.M. (1993). Decision making in the cockpit. In E.L. Wiener, R.L. Helmreich, & B.G.

Kanki (Eds.), *Cockpit resource management* (pp. 137-172). NY: Academic Press, Inc.

Orasanu, J.M. & Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 327-345). Norwood, NJ: Ablex.

Patel, V.L., & Groen, G.J. (1991). The general and specific nature of medical expertise: A critical look. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 93-125). Cambridge: Cambridge University Press.

Payne, J.W., Bettman, J.R., & Johnson, E.J. (1993). *The adaptive decision maker*. NY: Cambridge University Press.

Reiser, B.J., Kimberg, D.Y., Lovett, M.C., & Ranney, M. (1992). Knowledge representation and explanation in GIL, an intelligent tutor for programming. In J.H. Larkin & R.W. Chabay (Eds.), *Computer-assisted instruction and intelligent tutoring systems: Shared goals and complementary approaches*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K.A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise*. Cambridge: Cambridge University Press.

Scardamalia, M., & Bereiter, C. (1985). Fostering the development of self-regulation in children's knowledge processing. In S.F. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Research and open questions*. Nillsdale, NJ: Lawrence Erlbaum Associates.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors*, *27*(3), 285-300.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127-190.

Schoenfeld, A.H. (1987). What's all the fuss about metacognition? In A.H. Schoenfeld (Ed.), *Cognitive science and mathematics education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Scott, W. R. (1998). *Organizations: Rational, natural, and open systems.* Upper Saddle River, NJ: Prentice Hall.

Serfaty, D., MacMillan, J., Entin, E.E, & Entin, E.B. (1997). The decision making expertise of battle commanders. In C.E. Zsambok & G. Klein. *Naturalistic decision making.* Mahwah, NJ: Erlbaum.

Shanteau, J. (1992). The Psychology of experts: An alternative view. In G. Wright & F. Bolger (Eds.), *Expertise and decision support.* New York: Plenum Press.

Smithson, M. (1989). *Ignorance and uncertainty: Emerging paradigms.* New York: Springer-Verlag.

Sniezek, J.A., & Henry, R.A. (1990). Revision, weighting, and commitment in consensus group judgment. *Organizational Behavior and Human Decision Processes, 45*(1), 66-84.

Thompson, B.B., Cohen, M.S., & Shastri, L. (1997). *A hybrid architecture for metacognitive learning*. Arlington, VA: Cognitive Technologies, Inc.

Thompson, J.D. (1967). *Organizations in action.* New York: McGraw-Hill.

Van Creveld, M. (1985). *Command in war*. Cambridge: Harvard University Press.

Voss, J. Perkins, D.N. & Segal, J.W. (1991). *Informal reasoning and education*. Hillsdale: Erlbaum.

Voss, J.F., Wolfe, C.R., Lawrence, J. A., & Engle, R.A. (1991). From representation to decision: An analysis of problem solving in international relations. In R.J. Sternberg & P.A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 119-158). Hillsdale, NJ: Lawrence Erlbaum Associates.

Watson, S.R., & Buede, D.M. (1987). *Decision synthesis: The principles and practice of decision analysis.* New York: Cambridge University Press.

Freeman, J.T. and Cohen, M.S. (1996). Meta-recognitional Skills in Decision-Making by Commercial Airline Pilots. NASA-Ames Research Center, Moffett Field, CA. Contract #NAS2-14075.

Freeman, J.T., Cohen, M.S., and Thompson, B.T. (1998). Time-Stressed Decision-Making in the Cockpit. Proceedings of the Association for Information Systems 1998 Americas Conference, Baltimore, MD.