ARI Research Note 96-43

Investigations of Naturalistic Decision Making and the Recognition-Primed Decision Model

Gary A. Klein and Roberta Calderwood
Klein Associates Inc.

Research and Advanced Concepts Office
Michael Drillings, Acting Director

March 1996

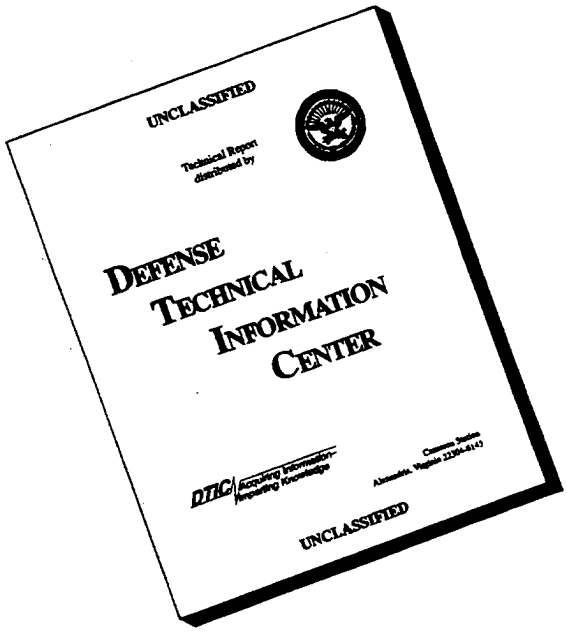
19960621 025

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

DTIC QUALITY INSPECTED 1

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

Klein Associates Inc.

Technical review by

Joseph Psotka

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED (from. . . to)
1996, March	Final	July 1985-July 1988

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Investigations of Naturalistic Decision Making and the Recognition-Primed Decision Model	MDA903-85-C-0327
	5b. PROGRAM ELEMENT NUMBER 0601102A
6. AUTHOR(S) Gary A. Klein and Roberta Calderwood (Klein Associates Inc.)	5c. PROJECT NUMBER B74F
	5d. TASK NUMBER 8516
	5e. WORK UNIT NUMBER C71
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Klein Associates Inc. 800 Livermore Street Yellow Springs, OH 45387	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-BR 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Note 96-43

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

COR: George Lawton

14. ABSTRACT *(Maximum 200 words)*:

This monograph reviews 3 years of research concerned with how experienced personnel make decisions in operational settings characterized by real-time information processing, shifting goals, and high-risk consequences. The study method combined field studies with experiments designed to test specific hypotheses. Study domains were selected so that findings would have high potential for generalizing to military command-and-control decision making. Critical decision interviews were carried out with experienced personnel, including urban fire ground commanders, wildland fire incident commanders, and U.S. Army tank platoon leaders. Interviews were designed to elicit information on the cues, goals, and option evaluation strategies used by these personnel. Based on these interviews, the relationships among such factors as time pressure, experience level, and group interactions were explored.

15. SUBJECT TERMS

Decision making Natural reasoning Time pressure Decision strategies
Expertise Think-aloud protocols Predecsion processes

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	124	

KLEIN ASSOCIATES INC.

800 Livermore Street
P.O. Box 264
Yellow Springs, Ohio
45387

EXECUTIVE SUMMARY

This monograph reviews three years of research concerned with how experienced personnel make decisions in operational settings characterized by real-time information processing, shifting goals, and high-risk consequences.

The study method combined field studies with experiments designed to test specific hypotheses. Study domains were selected so that findings would have high potential for generalizing to military command-and-control decision making. Critical Decision interviews were carried out with experienced personnel including urban fire ground commanders, wildland fire incident commanders and U.S. Army tank platoon leaders. Interviews were designed to elicit information about the cues, goals, and option evaluation strategies used by these personnel. Based on these interviews, the relationships among such factors as time pressure, experience level, and group interactions were explored.

The results of these studies have been used to develop a Recognition-Primed Decision (RPD) model of decision making. This model contrasts with current normative and prescriptive models of decision making, and the implications of this alternative framework are explored. Among the findings that we consider most important are:

* Experienced decision makers come to rely more on situation assessment, while novices rely more on option evaluation strategies.

* Situation assessment seems to involve schematic or prototypical knowledge of cues, goals, and expectancies that apply to a given class of events. Current cognitive research paradigms have not addressed how complex decision events are classified.

* Whereas experts and novices notice the same cues in a situation, novices draw fewer inferences based on these cues. Novices tend to miss the tactical implications of situational cues.

* At least in the domains studied here, decisions are most likely to be made without any conscious deliberation between option alternatives.

* When deliberation does occur, decision makers are more likely to use serial evaluation strategies than concurrent evaluation of options. Serial strategies appear to offer a means of minimizing the calculational burden as well as maximizing the speed with which a decision may be implemented.

* Serial evaluation is associated with satisficing rather than optimizing strategies, and is preferred under time-pressured conditions.

Executive Summary continued:

* Options are frequently evaluated through the use of images or a "mental model" that operates as a simulation for judging whether an option will be successful in a specific case.

* Expert decision makers rely on a process of "progressive deepening" or reasoning into the future.

* Analogical reasoning is infrequently reported, suggesting that the processes involved in selecting and using analogues are relatively automatic and unconscious.

* When analogues are used (often to address non-routine aspects of a problem), they are critical to option selection. Thus, inappropriate analogues are a primary cause of errors.

* Time pressure does not affect the quality of decisions made by experts as much as novices, due to experts' greater reliance on rapid recognitional processes.

Table of Contents

ACKNOWLEDGEMENTS

INTRODUCTION 1

Need for Descriptive Models 2

RECOGNITION-PRIMED DECISIONS 6

Situation Assessment. 8

Serial Evaluation 12

Progressive Deepening 14

IMPLICATIONS 16

Should Proficient Decision Makers Try to Rely on Analytical Methods, Including Concurrent Evaluation of Options, Wherever Possible?. . 20

Do Human Operators Have Too Many Judgment Biases to be Entrusted with Decisions? 21

Summary. 23

REVIEW OF STUDIES 27

Overview of Critical Decision Method 27

Core Procedures. 28

Study 1. 32

Study 2. 34

Study 3. 36

Study 4. 39

Study 5. 41

Study 6. 43

Study 7. 44

REFERENCES. 47

APPENDIX I - Protocol Analysis of Expert/Novice Command Decision Making During Simulated Fire Ground Incidents

APPENDIX A - Probe Points

ACKNOWLEDGEMENTS

This is the final report of a three-year program of research entitled "Analogical Decision Making" sponsored by the U. S. Army Research Institute for the Behavioral and Social Sciences (Contract MDA903-85-C-0327). In addition to the authors, several other investigators made major substantive contributions to the research: Beth W. Crandall, Marvin L. Thordsen, Janet Taynor, and Christopher Brezovic.

We wish to thank our contract monitor on this project, Dr. Judith Orasanu, who has provided us with a most supportive and cooperative atmosphere for pursuing our ideas.

A special thanks must go to the many individuals who participated in our studies. Theirs was in many ways the primary contribution to this effort; what we report tries to maintain the self-knowledge and insights that they offered. They include firefighters from Dayton, Columbus, Cincinnati, Cleveland, and Huber Heights, Ohio and Indianapolis, Indiana; members of the Boise Interagency Fire Control Center; and students and instructors at Ft. Knox's Armor Platoon Leader Training course.

I. INTRODUCTION

This is the final report[1] for the program of research entitled "Analogical Decision Making" sponsored by the U. S. Army Research Institute for the Behavioral and Social Sciences which began in July, 1985 and ended in July, 1988 (MDA903-85-C-0327). The goal of the project was to investigate decision making in operational settings and to develop theory relevant to Army needs. The research approach has relied primarily on obtaining and analyzing verbal protocols of decision events from experienced personnel. We have refined and modified both our data gathering and our analytic methods throughout the course of the project, so that the methodological development has formed a significant part of the research effort. Throughout the project we have attempted to blend features of naturalistic field investigation with the theory and hypothesis testing of behavioral decision making and cognitive psychological research.

Seven separate projects were carried out under this program. Field studies of command-and-control decision making included one on fire ground command decision making, one on wildland fire incident command, and one on U.S. Army tank platoon battle management. An experiment designed to test the effect of time pressure on decision quality was carried out using chess as the decision task. One project reviewed protocols from all of the field studies for evidence of analogical reasoning, and proposed a preliminary taxonomy of the functions of analogues in decision making. One project involved an extensive literature review relevant to the categorization of natural decision events. The review formed the basis for a doctoral dissertation proposed to

---

[1]Portions of this report appear in Klein, G. A. (In press). Recognition-primed decision. In W. Rouse (Ed.), Advances in Man-Machine Systems Research, 5. Greenwich, CT: JAI Press, Inc.

research the effect of context on event classification. The most recent study obtained think-aloud protocols of fire ground command decisions during simulated incidents. This effort allowed experimental verification of hypotheses generated on the basis of previously obtained interview data.

All but one of these projects have been reported previously in technical reports and published articles. Rather than providing an extensive review of each of these projects, it seems more appropriate to provide an overview of those findings that have had the most impact on theory development or applications. Complete citations for each of the referenced studies are provided, but the interested reader may also wish to refer to the titles and summaries of these reports that are provided in Section IV. The single previously unpublished study is included as Appendix I.

The remainder of this report is organized by topics and will not attempt to explain experimental procedures and results in detail or retain the chronological development of the ideas.

The Need for Descriptive Models

The idea that provided the major impetus for the research in this project was that models and research methods in behavioral decision research have been too focused on the analytic processes involved in comparing and selecting from among a predefined set of options. Many laboratory decision tasks are based on some variation of a gamble between two clearly defined bets or alternatives. Even in tasks where multiple options are available and multiple attributes are considered in the evaluation, both the options and the evaluation dimensions are well-defined and remain constant over the course of the decision problem.

Clearly, "real life" decisions rarely come so neatly packaged. Before alternatives can be evaluated or even identified, the decision must be framed or structured in some way that enables relevant goals to be identified and appropriate options to be generated. Moreover, the structure of the decision may change over time as events change and/or a new understanding of a decision problem is achieved. Although the limitations of standard decision research paradigms have long been recognized, there remains a lack of systematic research relevant to dynamic and ill-defined tasks (e.g. Brehmer & Allard, 1986; Edwards, 1962; Gettys, 1983; Rapoport & Wallsten, 1972).

Our primary interest has been on task domains that share the essential characteristics of command-and-control decision making: involving high-level integration of near real-time information for the purpose of deciding how best to utilize force application in a "battle" environment under varying degrees of uncertainty and time pressure (cf. Wohl, Entin, Kleinman, & Pattipati, 1984). Consider these examples of decision events:

> An incident commander is charged with controlling a fire raging through 18,000 acres of forest and range land. This fire has been burning for several days when a fortunate break in weather conditions allows a chance to renew attack efforts which have had to be largely defensive for the past two days. In consultation with his five experienced staff officers, he begins a session that will determine where to place a fire control line. Reports from air and ground surveillance are used as a basis for considering various placements of the direct attack, weighing such factors as the line length required, equipment and manpower proximity, whether or not bulldozers could be used or hand crews would be required, danger if efforts fail, the political ramifications of letting the fire burn a particular area, the effects of renewed bad weather.... A decision must be reached in this 3 a.m. session before the day's front-line attack crews are given orders at 6 a.m.
>
> An urban fire ground commander is called to the scene of a two-story wood frame building fire in a residential area. Reports from citizens at the scene indicate that all of

> the residents are out of the building so he is able to focus attention on setting fire control strategy. The fire is already well-involved, and judging from the location of the visible flames, color of the smoke, and smoke concentrations, looks to have started in a downstairs room, spreading very rapidly into one of the upstairs bedrooms with very little horizontal spread. The rate of spread could indicate a combustible fuel source making the inside attack more dangerous and the prognosis poorer for saving the structure. Standard operating procedure is to order roof ventilation and an outside attack until exposures are protected. However, the vertical spread is so rapid that it looks as if the roof may self-ventilate and an aggressive inside attack may save the bulk of the structure and contents. He hates to see these people lose everything, but he's seen these things go real sour.... He must decide in less than one minute whether to order all crews inside on hose lines or wait for the truck crew to ventilate.

These decision makers must work quickly to clarify the nature of the situation based on their own experience and training and they must decide which of several conflicting goals should be given priority. The situation may change as a result of their own actions or other events, so events must be constantly monitored and reassessed.

Many of the issues of decision research paradigms are simply not directly relevant to these types of decisions. We have found that the very language of decision models is difficult to translate into operational settings. In one of our earlier studies of urban fire ground commanders (FGCs) (Klein, Calderwood, & Clinton-Cirocco, 1985) we were surprised to find the commanders rejecting the notions that they were "making choices," "considering alternatives," or "assessing probabilities." They saw themselves as acting and reacting on the basis of prior experience, and generating, monitoring, and modifying plans to meet the needs of the situations. Because we found no evidence for extensive option generation there was little chance to observe tradeoffs between the utilities of outcomes. Nor could we see any way to

apply the concept of optimal choice. It appeared that a search for an optimal choice could stall the FGCs long enough to lose control of the firefighting operations. The FGCs were more concerned with identifying actions that were "workable," "timely," and "cost-effective."

We originally proposed that understanding of these types of decision events could be increased by focusing on the natural reasoning strategies being used. An early hypothesis was that analogical reasoning was a primary basis for making decisions, based on previous work in inference and predictions (Klein & Weitzenfeld, 1982; Weitzenfeld, 1984). In making predictions, an individual frequently establishes a comparison case based on the similarity of the case to a target case. Similarity is not based on featural matching (e.g. Tversky, 1977), but on an overall judgment about whether the comparison cases contains the relevant causal factors. Adjustments are then made on the basis of differences between the target and comparison cases in order to make a prediction or inference.

Although we later determined that analogical reasoning was too narrow to account for the decision processes we were describing, the theoretical framework that we have developed retains the idea that decision making starts with an understanding of a situation based on previous experiences and knowledge.

We further proposed that laboratory methods using simplified tasks and inexperienced decision makers were an inherently inadequate basis for building models that would have applications to natural decision tasks. We wished to find methods that allowed the contextual constraints that are normally operating to be apparent. This seems to us the best way of generating potentially important hypotheses and for increasing the fit between theory and

practice. We settled on a quasi-naturalistic approach that has generated a rich source of data for generating hypotheses and suggesting fruitful avenues of research.

The next sections describe the theoretical framework, the methods that were developed within this program, and some of the major implications of the model for future research and applications.

II. RECOGNITION-PRIMED DECISIONS

Although the FGCs we studied denied making decisions in the traditional sense of "selecting an option," they were clearly making choices and judgments that affected the course of events. However the FGCs insisted that they rarely deliberated about the advantages and disadvantages of the different options. Instead the FGCs relied on their abilities to recognize and appropriately classify a situation. Once they knew it was "that" type of case, they usually also knew the typical way of reacting to it. They would use the available time to evaluate an option's feasibility before implementing it. Imagery might be used to "watch" the option being implemented, to discover if anything important might go wrong. If problems were foreseen then the option might be modified or rejected altogether.

For this task environment, this recognitional strategy appears to be very efficient. The proficient FGCs we studied could use their experience to generate a workable option as the <u>first</u> to consider. If they had tried to generate a large set of options and then systematically evaluate them, it is likely that the fires would have gotten out of control before they could make any decisions.

The Recognition-Primed Decision (RPD) model is presented in Figure 1. It shows the proficient decision maker becoming aware of events that have

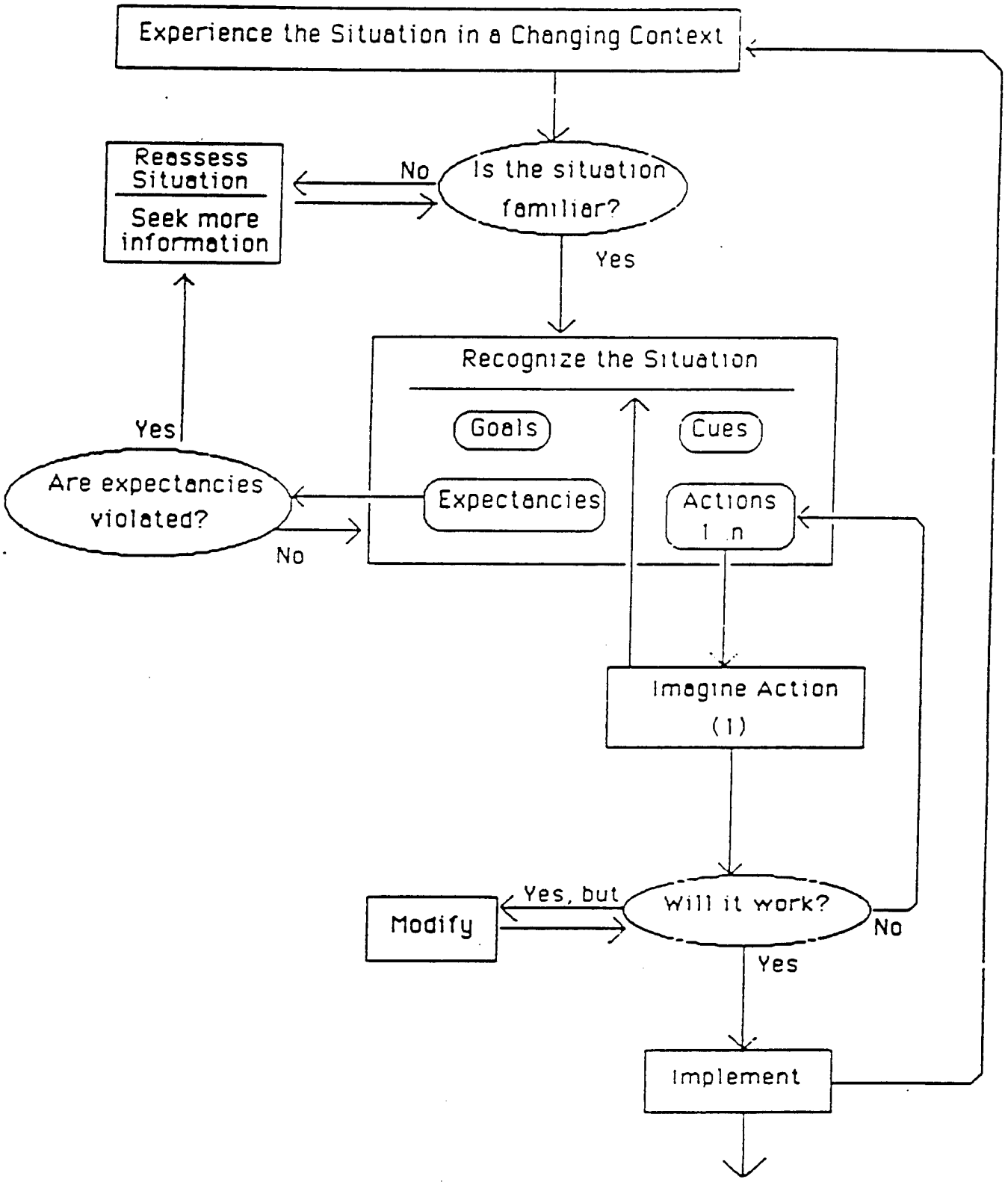


Figure 1: Recognition-Primed Decision (RPD) Model

occurred, and relying on experience to recognize these events as largely typical. The simplest case is one in which the situation is recognized and the standard reaction is implemented. A somewhat more complex case is one in which the decision maker performs some conscious evaluation of the reaction, often using imagery to uncover problems. The most complex case is one in which the evaluation reveals flaws requiring modifications, or the option is judged inadequate and is rejected in favor of another typical reaction.

The evaluation function can operate at several levels. The verification/nonverification of expectancies serves to alert the decision maker that the situational understanding is wrong, and that it is time to rethink it and gather more information. In addition, individual actions are evaluated through progressive deepening to see how they will turn out.

This model clearly includes aspects of problem solving and judgment along with decision making. In naturalistic settings it is rare to find one without the others (e.g., see Wohl, 1981). There are three features of the model that will be discussed: (a) situation assessment, (b) serial evaluation, and (c) progressive deepening.

Situation Assessment

The experts we studied are able to quickly determine if a case poses any new challenges. Their experience with a wide variety of cases assures that most problems they encounter will have many features similar to what they have seen before. In other words, they form a situation assessment based on judgments of prototypicality (Rosch & Mervis, 1975) that activates appropriate "schemas" or "scripts" (Bartlett, 1932; Kintsch & van Dijk, 1975). If the judgment is that the situation is typical, then typical options or standard operating procedures will be generated. But the judgment is that the case is

somewhat atypical, the decision maker will attempt to understand the situation by analyzing its features for a "best fit" to competing prototypes or schema.

We have identified four different types of information that are relevant to formulating a situation assessment: (1) plausible goals, (2) critical cues and causal factors, (3) expectancies, and (4) typical actions.

Plausible goals. Part of what it means to understand a situation is to understand what will be possible to accomplish. Goals here do not refer to the context-free goals of decision-theoretic models (generally the maximizing of some value), but to specific outcomes that a decision maker tries to achieve (i.e. the context-bound goals of Keeney & Raiffa, 1976).

Because the concept of "goals" can be vague, we suggest that a useful way to conceptualize goals is through contrast sets (Olson, 1970), since the selection of a goal also has implications about goals that were not selected. For example, an FGC could claim that the goal was to "do my job," but this is scarcely enlightening since there is no reasonable alternative. The more specific goal of "performing an interior attack" is meaningful because the alternatives are to "perform an exterior attack," "perform search-and-rescue," "call in a second alarm," or "abandon the effort and take precautions that the fire does not spread." In other words, the pragmatic meaning of a goal is communicated by contrasting it to alternative goals.

Critical cues and causal factors. In field settings, there are usually a great many events to attend to, and novices often feel overwhelmed with stimuli. Proficient decision makers do not feel this overload. The RPD model hypothesizes that the formulation of a situation assessment includes prioritizing critical cues, helping insure that attention is not diverted to less important cues or events.

For example, FGCs have learned to quickly scrutinize the color of flames. (This cue has meaning in terms of fire temperature and, by inference, the types of materials that are burning. Similarly, patterns of smoke convey information about the intensity of the fire by the pressure with which smoke is being pushed out of the building. Perceptual cues usually convey causal implications. Additional types of information also have causal implications. The nature of roof construction may imply vulnerability to damage, affecting the safety of sending firefighters onto the roof. However, there are times when the nature of roof construction is less important (e.g., when the danger is from smoke, not flames). The situation assessment includes attention to relevant cues and types of information.

Critical cues may also determine the timing of actions. In some of the fire ground decisions we studied, the decision maker's expertise was in recognizing when to act ("I held off ventilating the roof until I could see that the fire was beginning to spread to the attic").

Expectancies. Expectancies function to prepare decision makers for action and to provide clues for testing whether the situation is correctly understood. The situation assessment includes expectancies of what is likely to happen, and when. These expectancies can include a sequence of events, or a time course for events.

For an inexperienced decision maker, expectancies are poorly formed, vague, and hard to test. In contrast, an experienced decision maker holds clear expectancies. If events do not fit these expectancies, the resulting discrepancy raises questions about whether the situation assessment was correct. Thus, in one case we studied, an experienced FGC directed a stream of water at the area believed to be the seat of a fire. If correct, he

expected the pattern and color of the smoke to change within 20-30 seconds. When he saw no changes after about 45 seconds he suspected the seat of the fire was elsewhere.

The recognition of feasible goals includes expectancies about having sufficient time for accomplishing the goals. Goals are often linked to a timetable, which may be revealed in messages indicating whether plans are on schedule, or ahead of schedule, or "behind the power curve." If time expectations are violated it will create a feeling of urgency, and possibly a shift in situation assessment. For example, the FGC in the case cited above recognized that there was no longer enough time to extinguish the fire in the apartment building before it became a danger to the occupants. Smoke was already showing on the fourth floor, indicated that fire had spread to that area. There was no longer opportunity to direct water down at the fire.

Typical actions. A familiar situation evokes a familiar set of actions. It may even be useful to postulate an "action queue" of typical reactions to a situation at hand.[2] Therefore, the concept of situation assessment includes the identification of typical responses.

It has been asserted that proficient decision makers generate options on the basis of typicality--what is usually done in such a situation. Other possibilities are that an option is generated on the basis of recency (what was done the last time this came up), availability, or other factors. The generation of options by proficient decision makers is not fool-proof, and certainly does not invariably lead to optimality. The advantage is that it

---

[2]Robert Holt (personal communication) has helpfully suggested the concept of an action queue in this context.

generally leads to acceptable actions with little expenditure of time and effort.

The recognition of a situation includes recognizing actions as well as recognizing goals and cues. One thing we learned from interviewing proficient decision makers is that much of their attention was on their own reaction, primarily the set of orders to give to their troops. In other words, a fire might feel like a "search-and-rescue" situation more than a "single-family 2-story home with brick exterior." The emphasis is on a functional understanding of what to do, as well as a structural understanding of cues and relationships.

Serial Evaluation

Serial evaluation refers to the assessment of options one at a time until a satisfactory one is found. Serial evaluation is different from concurrent evaluation of options, whereby a set of options is generated and evaluated comparatively.

This incident, taken from our interview files, illustrates serial evaluation.

> The commander of an emergency rescue squad arrived at the scene where a young woman, either drunk or on drugs, had either jumped or fallen from a highway overpass. She probably was attempting suicide by trying to fall to the highway below. But she missed, falling instead on a support strut for a highway sign. She was lying face-down on this strut, semi-conscious. A hook and ladder truck pulled up, and was directed to the highway below, to block traffic. Two members of the squad climbed out, and one pinned her legs to the strut while the other pinned her arms. The decision was how to raise her to safety without endangering the crew. The head of the squad told us that he first considered using a Kingsley harness which is the standard rescue equipment, but in imagining its use he could see that it would not work. Since it is attached from the front, he couldn't see how to push the woman to a sitting position without risk to all three people. He imagined attaching the Kingsley harness from the back, but

> saw it would put excessive strain on the woman's back and so rejected that option. Next, he thought of using a Howd strap, which ties onto a victim (in contrast to a Kingsley harness which snaps on), but saw that it ran into the same problems so he rejected it. The next option he considered was using a ladder belt--a heavy belt that firefighters wear when climbing up several stories on a ladder, so that they have a snap to attach to the top rung of the ladder in case they lose their balance. He imagined lifting her up a few inches, slipping the ladder belt under her waist, buckling it closed (only 1 buckle is involved) and tying a rope to the snap attachment. This is the option that he selected, and the rescue was made.

In this example, the decision maker considered three options (five if you count attaching the Kingsley harness and Howd strap from the back), but at no time contrasted the strengths and weaknesses of one option versus another. Instead, each option was examined in turn until a workable one was identified. The decision maker reported that the whole decision took less than a minute.

The work of Simon (1955) is probably the best known discussion of a serial generation and evaluation strategy in the behavioral science literature. Simon described the use of satisficing as a means of quickly and efficiently finding an effective option. Satisficing is a process by which choices are evaluated one at a time until a satisfactory one is found, and then implemented. It differs from an optimization strategy in which virtually all options have to be generated and evaluated in order to determine which is best. Simon claimed that efficient business executives relied on satisficing. Cyert and March (1963) have also noted the importance of searching for the first option that works instead of trying to find the best option.

The RPD model extends the concept of satisficing in several ways. It asserts that the first option selected from the "action queue" is the most typical option, and therefore has a high likelihood of being effective. Therefore the proficient decision maker begins with a promising option, making

satisficing a more powerful strategy than if options were being generated randomly. The RPD model also asserts that options are selected roughly in order of their typicality, although other influences are undoubtedly present, such as recency and availability. Typicality itself is certainly linked to the concept of representativeness (Tversky & Kahneman, 1974), which under some conditions can mislead subjects. There is a trade-off here. Proficient decision makers have gained the ability to be generally successful and to react quickly. They accept the risk of being sub-optimal, and of occasionally having to spend time evaluating an option that is unacceptable.

A serial evaluation strategy as posited by the RPD model continuously makes available to a decision maker a preferred course of action. If time pressure forces a response, decision makers are prepared. In contrast, a concurrent evaluation model would leave a decision maker unprepared for action during the time course of the analysis. Only when all the analyses were completed would it become clear which course of action to select.

Progressive Deepening

Progressive deepening is the process of imagining how an option will be carried out within a specific situational context. It is the attempt to anticipate each important step, to notice the most likely reaction(s) to that step, to find the best way(s) to handle that reaction. It is an important component of recognitional decision making. Progressive deepening enables a decision maker to forecast the adequacy of a course of action. Within behavioral decision theory, options are evaluated by comparing them to each other with regard to how well they satisfy a set of criteria. In contrast, the RPD model asserts that one action is evaluated at a time. This is done by

imagining how the action would be implemented within the specific setting.[3] It is like running an "instant pre-play" to see if anything might go wrong.

An example might be useful. The head of a rescue unit arrived at the scene of a car crash. The victim had smashed into a concrete post supporting an overpass, and was trapped unconscious inside his car. In inspecting the car to see if any doors would open (none would), the decision maker noted that all of the roof posts were severed. He wondered what would happen if his crew slid the roof off and lifted the victim out, rather than having to waste time prying open a door. He reported to us that he imagined the rescue. He imagined how the victim would be supported, lifted, and turned. He imagined how the victim's neck and back would be protected. He said that he ran this imagining through at least twice before ordering the rescue, which was successful.

One of the first descriptions of this strategy in the psychological literature was in the work of de Groot (1965/1978). He coined the term "progressive deepening" to describe how chess grandmasters follow out a line of play and make sure it does not lead to any blunders. De Groot studied chess players trying to pick the best move in a difficult position. In the 40 protocols he presents, the chess players considered anywhere from 2 to 11 options but almost never compared one option to another.

We have expanded on de Groot's (1965/1978) work by hypothesizing that within the context of recognitional decision making, progressive deepening can help the decision maker in a number of ways: find weaknesses in an option; find ways to repair these weaknesses and thereby improve the option; discover

---

[3]Appreciation is expressed to Alexander Levis for discussions about the importance of imagery for decision making.

new opportunities that arise through implementing an option; alert the decision maker to previously ignored dynamics of the situation, thereby helping to modify the situation assessment.

In some ways, the topic of progressive deepening will overlap with the concept of contingency planning. However, contingency planning sometimes refers to systematic examination of plans. If the contingency planner checks for possible errors and oversights by examining as many assumptions as possible within the time available, this can be a very tedious process that could bog down in an exponential explosion of different factors and possibilities. In contrast, contingency planning by progressive deepening enables a skilled performer to be alert to important flaws in a plan without having to examine everything, and without having to decide what to examine and what to ignore (which entails first examining everything).

Progressive deepening also affects situational understanding. As actions are imagined, new features of the situation may be found, so Figure 1 shows an arrow leading back from the box "Imagine Action" to "Recognize the Situation."

III. IMPLICATIONS

The RPD model developed under this contract offers several important contrasts to normative and prescriptive models. What does the concept of recognitional decision making have to say about prescriptive decision models? First, let us examine such a model.

A strong position on prescriptive decision making has been taken by Janis and Mann (1977) who recommended that decision makers should be generating and contrasting options whenever possible. They claimed that decision making is stressful, that people avoid it when possible, and that many times where concurrent evaluation between options is appropriate and necessary, it does

not happen. Janis and Mann did not intend their advice for situations where there was extreme time pressure, but it is instructive to examine their ideas nonetheless.

For Janis and Mann (1977), there are seven criteria to be used in judging whether decision-making procedures are of high quality. They define the "ideal" decision maker (p. 11) as one who should:

-- thoroughly canvass a wide range of alternative courses of action;

-- survey the full range of objectives to be fulfilled and the values implicated by the choice;

-- carefully weigh whatever he knows about the costs and risks of negative consequences, as well as the positive consequences, that could flow from each alternative;

-- intensively search for new information relevant to further evaluation of the alternatives;

-- correctly assimilate and take account of any new information to which he is exposed, even when the information or judgment does not support the course of action he initially prefers;

-- reexamine the positive and negative consequences of all known alternatives, including those originally regarded as unacceptable, before making a final choice;

-- make detailed provisions for implementing or executing the chosen course of action, with special attention to contingency plans that might be required if various known risks were to materialize.

Janis and Mann (1977) assert that "failure to meet any of these seven criteria when a person is making a fundamental decision (one with major consequences for attaining or failing to attain important values) constitutes

a defect in the decision-making process" (p. 11). They are well aware of the problems of making decisions under moderate time stress, and cite research showing that time pressure leads to a failure to make effective use of relevant and available information. They label the condition of "hypervigilance" as one where high conflict exists, along with a belief that a satisfactory solution exists, but with an apparent lack of time to search and deliberate. Here, the decision maker is hypothesized to display indiscriminate "openness" to all information. Janis and Mann (1977) repeatedly complain that people terminate information searches before all relevant data are examined.

The framework Janis and Mann (1977) are using is fairly typical of decision research that has attempted to formulate techniques to improve decision quality. For most of this work, the perspective taken is one described above as concurrent evaluation, and as analytical decision making. The decision maker is viewed as "faced with alternatives", which can be specified as branches emanating from a single point in a search tree. It is also natural to speak of the decision maker "considering the consequences" of each alternative in terms of an analysis of future states (odds/ probabilities) weighed against alternative goals (preferences/utilities). Techniques such as Decision Analysis and Multi-Attribute Utility Analysis (MAUA) have been derived to help the decision maker work out the various consequences of options.

This perspective leads to the conclusion that humans are limited and biased decision makers (e.g. Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). It has been a logical step, therefore, to focus decision support on methods of debiasing judgments (Fischhoff, 1982; Kahneman & Tversky, 1979),

calibrating probability estimates (Kadane & Lichtenstein, 1982), instructing in optimal combinatorial methods (Zakay & Wooler, 1984), and the like.

The potential significance of this work is enormous. If it is possible to develop general methods to improve decision making, then these methods could be trained and they could be embedded within decision aids to provide a large improvement in decision quality.

Unfortunately, the payoffs have yet to be seen. Decision aids, to support the use of Decision Analysis and MAUA, do not seem to have been well accepted in operational settings. With a few exceptions, decision training has not been shown to be very effective, and under time pressure such training has not shown any benefit (Rouse, 1978; Howell, 1984; Zakay & Wooler, 1984).

One way to understand the prescriptive implications of a recognitional decision model is to examine some of the standard recommendations for improving decision making.

Should Proficient Decision Makers Generate as Many Options as Possible?

From the perspective of recognitional decision making, the answer is "No."

This recommendation is heard from both decision researchers (e.g., Gettys, 1983) and practitioners writing popular books and articles (e.g., Janis & Mann, 1977). In the time-pressured environments we studied, there simply was not enough time to follow such advice. It takes time to generate and evaluate options, and delays may be intolerable. In addition, the situation may shift during the analyses so that the whole process has to start over again. Even in the absence of time pressure we rarely observed proficient decision makers trying to generate large sets of options. In our

research it is the novices who seem intent on generating options. Experts can recognize what to do right away and do not need to search further. Therefore, advice to generate large option sets is telling people to act like novices. The time to develop large option sets is when a situation is encountered that is unfamiliar, or when there are disputes.

Should Proficient Decision Makers Try to Rely on Analytical Methods, Including Concurrent Evaluation of Options, Wherever Possible?

Again, the answer is "No."

A recognitional decision strategy is a valid use of expertise in decision tasks and is a strong alternative to a generation/comparison strategy. In a recognitional strategy the expertise of the decision maker comes out in the identification of the appropriate option to consider, rather than in the way evaluation dimensions are selected or weighted or options are ranked.

The danger is not just that, by requiring proficient decision makers to perform analytical decision making and concurrent evaluation, they will then be forced into performing sub-tasks that are time consuming and inefficient. The greater concern is that they will be unable to make effective use of their own expertise. The Decision Analysis and MAUA approaches may not leave much room for the recognitional skills of experienced personnel. Therefore the risk of using these approaches is that decision performance will become worse, not better. In addition, trainees may not have a chance to develop expertise if they learn to rely on the analytical methods rather than developing their own recognitional capabilities. Remember that novices lack the recognitional skills needed to effectively perform recognitional decision making. It should not be difficult to convince novices to rely on analytical decision aids,

thereby limiting their opportunity to ever develop the experience base necessary for recognitional decision making.

Do Human Operators Have Too Many Judgment Biases to be Entrusted with Decisions?

The answer to this question is also negative.

Research on such biases as availability, representativeness, poor use of probabilistic data, and so on, has created an impression that people are inherently flawed decision makers. The implication is that we should develop training programs to reduce biases, rely on decision aids, and train special decision consultants.

These biases have been demonstrated in settings where context has been carefully limited, tasks are well-defined, and experience level is usually low. In other words, the opportunity for effective recognitional decision making has been limited, and attempts to apply recognitional decision making lead to errors. How well do these data generalize to actual decision tasks? Christensen-Szalanski and Beach (1984) has argued that these classical decision biases are artifacts of laboratory methodology and of the analytical perspective; they showed that studies of novice decision makers usually found evidence for biases whereas studies of experts usually documented their strengths. Christensen-Szalanski (1986) has also shown that judgment biases, even if they exist outside the laboratory are of little importance in decision making since the proportion of actual decisions they affect will be quite small. That is, the biases primarily operate on very low frequency events (e.g., a clear bias in the way physicians diagnose pneumonia would lead to an average of only one missed diagnosis per year). And if the frequency of such events increases, so will experience with them, thereby diminishing the bias.

Furthermore, Christensen-Szalanski cites a series of studies showing that "biased" decision processes do not lead to much reduction in the quality of the decisions. For example, miscalibration bias is clear cut but has almost no effect on accuracy of forecasting.

Furthermore, some tendencies that show up as biases for well-defined laboratory tasks may be of value in the field. "Biases" such as availability and representativeness reveal the fact that proficient decision makers have learned to rely on episodic memory. They can store earlier experiences as potential analogues to guide future performance. Surely the skilled use of episodic memory would be a strength for proficient cognitive performance in general, rather than a weakness for handling abstract story problems about female bank tellers (e.g., Tversky & Kahneman, 1983).

The optimization model has gained prominence as a normative model in psychology, despite its bases in statistical analysis and economic theory. Decision Analysis and Multi-Attribute Utility Theory with their emphasis on generating many options and then evaluating them systematically, has been likened to a process of evolution (Cooper, 1987). Success is achieved by starting with diversity and then applying stringent criteria for continuation. In studies with naive subjects, it has made good sense to encourage them to be creative and generate many options. Furthermore, if we have more faith in our conscious abilities to analyze and evaluate options than in the non-conscious abilities to generate the options, then we have a procedure where the important part is done in a way that can be controlled, observed, and improved. However proficient decision makers can rely on recognition to generate an option that is usually workable. We may not know how they do

this, and we may never be able to bring this skill under conscious control, but we should not ignore or deny it. And we should try not to develop prescriptions that interfere with recognitional decision making.

Summary

How can we characterize the major points of difference between recognitional and analytic decision making? There are seven important differences.

One difference is the focus on situation assessment rather than on the selection of an option. A recognitional model presupposes that the task has been recognized as being familiar in some important ways, so that the decision maker understands the plausible goals, cues and variables to monitor, expectancies, and the typical reactions. Descriptive models of analytical decision making have generally not addressed situation assessment. Prescriptive models like Decision Analysis regard situation assessment as a construction of states of the world, but have more to say about the consequences of selecting individual options.

A second difference is the mechanism for generating options. For analytical decision making, the assumption is of a fairly random process requiring careful evaluation. In contrast the RPD model describes how proficient decision makers can generate promising options as the first ones considered.

A third difference is that the RPD model concentrates on satisficing whereas analytical models have emphasized optimizing.

A fourth difference is about the nature of the evaluation. Analytical models deal with strong techniques for performing concurrent evaluation. The

RPD model describes decision makers as relying on progressive deepening to perform serial evaluation. The decision maker's task is seen as anticipating the outcome of implementing the typical reaction.

A fifth difference is in the treatment of options. The RPD model views options as being elaborated during the progressive deepening process. Limitations are found but the decision maker often tries to find ways of overcoming them, thereby strengthening the option. In contrast analytical models treat options as completed; attempts to modify and improve options would disrupt the evaluation process.

Sixth is the use of imagination. The evaluation process relies on the decision maker's ability to imagine how the option will be carried out, using world knowledge to anticipate pitfalls.

Seventh is response availability. Decision makers almost always will have an option that is ready to implement if time runs short; they only have to curtail their evaluation. In contrast an analytical strategy prevents a decision maker from knowing which option is favored until all the computations are completed.

What type of model is the RPD model? It is a descriptive model, derived from observations made in field studies. We primarily examined proficient decision makers, often driven by time pressures, but some of the studies examined less proficient decision makers and incidents that were not so time pressured. We have studied non-routine incidents, and in some studies we have probed only decision points that were non-routine, since we expected that such decisions would be most likely to require analysis. Even so, we found that the proficient decision makers were relying on recognitional strategies.

The RPD model is also a conceptual model, a framework for understanding how people function under operational conditions. In its present form it lacks the set of clear postulates that would allow it to be used to generate testable hypotheses. Thus, it is hard to imagine how the RPD model could be rejected. In the presentation of research findings, data were cited that supported the model but there was no presentation of negative research, because it is not clear what would constitute negative findings. For this reason, it anticipated that the model will have greater value for applied questions than for generating basic research.

Is the RPD model pertinent to decision making? Since the model describes processes where concurrent evaluation of options is avoided, and since it has bee claimed that they are the core of making a decision, the model may more appropriately be considered a description of problem solving than of decision making.

Berkeley and Humphreys (1982) treat decision making as concurrent evaluation. They define decision making as "the moment of choice among alternative immediate acts the decision maker has under consideration." (p. 203) This general type of definition is fairly standard in the field. Under this definition, we have not been studying decision making.

However, this definition may be too restrictive. It defines a phenomenon that is rarely encountered outside of laboratory conditions. The people we studied included fireground commanders with over 20 years of experience, and they claimed that they hardly ever used concurrent evaluation of options. Yet they were handling tasks that called for the allocation of personnel and equipment. And they were able to identify a number of "decision points" where reasonable options existed, options that someone with less experience might

well have chosen. Isenberg (1984) has also reported that business executives could recall few instances in their entire careers where they made decisions using concurrent evaluation about options. If these people are not making decisions, how relevant is the concept of decision making to applied psychology?

Of course, our data consist of verbal reports. It is quite conceivable that the people we studied were actually performing concurrent evaluation, but were doing it unconsciously or had forgotten about it. But the burden of proof shifts to the proponent of unconscious concurrent evaluation, to demonstrate that this phenomenon occurs and to explain how it is done. Until such proof is offered, there is no compelling reason to believe in the phenomenon of unconscious concurrent evaluation.

A more useful definition of decision making may be: identifying a course of action at a point where meaningful options exist. Under this definition, the decision maker does not have to consider more than one option actively. What makes it a decision is that meaningful options do exist and that the decision maker can articulate them if necessary. The focus here is on the task, not inside the head of the decision maker. It would allow us to compare how experts and novices perform the same task, contrasting their strategies. We could also study how changing the task conditions affected experts and novices differently. With the Berkeley and Humphreys (1982) definition it is natural for researchers to study naive subjects to make sure they are not using recognitional capabilities to avoid concurrent evaluation of options.

Lipshitz (1987) takes an even more extreme position than ours. He argues that decisions are fictions, artificially created under laboratory conditions. In naturalistic settings people function in a seamless web of intentions and

activities. He feels that it is a mistake to claim that individual decisions are being made, and that distinct goals are being identified. These concepts may be helpful in communicating, but they are also misleading. Looking at the flow of activities during a critical incident, it is clear that there are an infinite number of potential choices, and countless possibilities for decision points. Similarly, goals are not simply presented or deactivated. The actual situation may include a variety of intentions which will increase or decrease in importance, sometimes gradually and sometimes suddenly; it is misleading to segment out one or two intentions at the point where they were suddenly given prominence and pretend that these were the decision maker's goals.

Lipshitz's (1987) arguments are consistent with the observations we gathered during our interviews. We have adopted a less stringent definition than Berkeley and Humphreys (1982) because we judge their definition as too limited to be valuable for applied research. Lipshitz's position is even less stringent than ours but we will have to see how useful it is for providing direction for improving task performance. This is a key criterion--how helpful is each theoretical position to professionals working on applications? Theoretical hairsplitting can go on forever; guidance is needed today.

IV. REVIEW OF STUDIES

Our primary means of data collection was a Critical Decision method (Klein, Calderwood, & MacGregor, in press) developed under this and related contracts. It will be helpful in interpreting the specific results reviewed in this section to briefly outline key features of the method.

Overview of Critical Decision Method

The Critical Decision method is a retrospective interviewing strategy that shares many features with other methods, especially those related to

Flanagan's (1954) critical incident technique. Specific features of the technique include its focus on non-routine specific cases, and the use of probes to elicit information that may not be offered spontaneously.

The method was designed to strike a balance between a host of research objectives and practical constraints. For example, direct observations of command decisions coupled with an on-going verbal protocol of a commander's thought processes was first considered (see Hoffman's (1987) Method of Tough Cases). However, such an approach was deemed impractical in this case. Not only are challenging incidents relatively rare in any single location and expensive to cover because of the extreme time pressure, but the nature of the task makes any risk of outside interference untenable. We have used on-site observations to develop requisite domain knowledge prior to performing the actual elicitation task, and whenever possible to augment the data gathering.

At another extreme, simply asking fire ground commanders for unstructured accounts of their decisions would have resulted in little more than unrelated "war stories." Our goal was to focus the expert on those elements of an incident that most affected decision making, and to structure responses in a way that could be summarized along a specified set of dimensions while still allowing the details to emerge with the commander's own perspective and emphasis intact.

Core Procedures

The procedures adopted for the Critical Decision interviews represent our solutions to meeting these goals and practical considerations. The basic procedure can be summarized in the following steps:

Step 1: Select incident. Incidents were typically self-selected by the commander with the criterion that the case should represent a "command

challenge;" that is, they should illustrate a situation in which a decision had a significant impact on the outcome (either successfully or unsuccessfully). This selection criterion is common to critical incident methods (e.g. Klemp & McClelland, 1986) as a means of obtaining the most detailed and accurate reconstructions. In some cases incidents were selected by the interviewers or by supervisors during on-going operations.

Step 2: Obtain unstructured incident account. The participant is asked to describe the incident from its onset (e.g., the time he received the alarm) to the time when the incident was judged to be under control. For the most part this account proceeded without interruption by the interviewers, except for minor points of clarification. The procedure accomplished several goals. First, it created a context for understanding on the part of the interviewer. Second, the account served to activate the officer's memory of the event as a context for questioning. In addition, we judged that the procedure helped us achieve a high level of cooperation from the officers by establishing us as listeners rather than interrogators. During on-site observations this step might be very brief or eliminated.

Step 3: Construct incident timeline. After the incident had been related, the interviewer proceeded to reconstruct the account in the form of a timeline that established the sequence and duration of each event. Events included both objectively verifiable occurrences (e.g. "the second alarm equipment arrived two minutes later") and thoughts and perceptions reported by the officer (e.g. "the color of the smoke indicated the presence of a toxic substance," "I thought I might have to call a second alarm at this point"). The timeline served to establish a shared awareness of the "facts of the case." Many times inconsistencies in the account could be detected and

corrected on the basis of the timeline, and missing facts filled in. In addition, questions about the timeline focused the officer's attention on events from more than a single time perspective, an approach having demonstrated utility for obtaining accurate eyewitness testimony (Geiselman, Fisher, MacKinnon & Holland, 1985).

Step 4: Decision point identification. During the timeline construction, specific decisions were identified for further probing. In some cases the verbal cues marking a decision were obvious (e.g. "I had to decide whether it was safe enough to send my crews inside"), but this was not always the case. In other cases, it would be clear that an officer was taking one of several possible courses of action or was making a judgment that affected the outcome, but there was no clear indication that the officer saw himself as "making a decision" at this point. A decision point was probed if the participant confirmed that other reasonable courses of action were possible or that another participant (perhaps one with less or greater expertise) might have chosen differently.

Step 5: Decision point probing. Different studies have used different probes, depending on the objectives of the projects. Interview Guides included in the complete study reports indicate the wording of questions that were systematically asked as part of the interview. Table 1 summarizes the probe types that have been routinely used.

Questions to elicit the details of cue usage were almost always asked first as part of the timeline construction, and represented the current information that was likely to have been heeded at each event time. Prior knowledge was also probed. We had a special interest in eliciting any recall

of prior experiences that influenced the officer's size-up or expectancies about a situation. Such specific remindings were coded as analogues.

Table 1

Critical Decision Interview Probes

Probe Type	Probe Content
CUES	What were you seeing, hearing, smelling...?
KNOWLEDGE	What information did you use in making this decision, and how was it obtained?
ANALOGUES	Were you reminded of any previous experience?
GOALS	What were your specific goals at this time?
OPTIONS	What other courses of action were considered, or were available to you?
BASIS	How was this option selected/other options rejected? What rule was being followed?
EXPERIENCE	What specific training or experience was necessary or helpful in making this decision?
AIDING	If the decision was not the best, what training, knowledge, or information could have helped?
TIME PRESSURE	How much time pressure was involved in making this decision? (Scales varied.)
SITUATION ASSESSMENT	Imagine that you were asked to describe the situation to a relief officer at this point, how would you summarize the situation?
HYPOTHETICALS	If a key feature of the situation had been different, what difference would it have made in your decision?

Study 1: A Critical Decision Study of Expert and Novice Fire Ground Command Decisions

Goals. The goals of this study were to model the decisions made by experienced urban fire ground commanders (FGCs) using the Critical Decision method and to develop methods of analysis that would aid in understanding the role of experience in decision making.

Method. Critical decision interviews were carried out with 12 "Expert" and 12 "Novice" FGCs employed by six professional midwestern urban fire departments. Experts had an average of 19.5 years of fire fighting experience with an average of 11 years as an officer. Novices had an average of 10 years of fire fighting experience and less then two years as an officer.

Coding. Each decision was classified into one of nine types. The types were defined in terms of the intersection of two conceptual dimensions. Dimension #1 reflected serial or concurrent evaluation. This continuum is anchored on one end by choice involving little or no deliberation by the FGC. For these events, the FGC's actions appeared to be based primarily on his previous experience with similar events. When conscious deliberation did occur it frequently involved identifying and clarifying the nature of the situation itself or the specifics of action implementation or timing. These processes are commonly relegated to "predecision" stages or studied as aspects of monitoring or supervisory control, but we found them to be inseparable from decision making in this natural context. At the other end of the continuum were decisions fitting the definition of decision making more closely, in which action choices were deliberated in an attempt to meet multiple and sometimes conflicting goals. Dimension #2 reflected the degree to which the decision primarily involved evaluation of situation assessment (SA) or option (option) components of the decision problem.

decision primarily involved evaluation of situation assessment (SA) or option (option) components of the decision problem.

Additional coding categories assessed whether a specific analogue had entered into the decision process, whether imagery was used, how quickly the decision was made, and whether the decision involved future planning.

Results. Fifty-four percent of the decision points were SA decisions. In these cases, identification and recognition of the situation allowed a choice of action to be generated and implemented without further deliberation. In 14% of the decision points, implementation and timing of a highly preferred or standard option was the most crucial issue. Even in the 32% of the cases that involved evaluation between options, 14% were serially evaluated. Thus, only 18% of the decisions fit the classical definition of decision making as concurrent evaluation between options.

Experts and Novices were roughly equally likely to deliberate about options. However, Experts used an approximately equivalent mix of serial and concurrent strategies whereas Novices appeared to rely more on concurrent deliberation. Experts were also more likely to deliberate about situational aspects of the decision problem, whereas Novices deliberated more about option implementation and timing. Experts also appear to construct novel option solutions much more frequently than Novices, and to report the use of imagery and evaluate potential options more frequently than Novices. Finally, Experts were almost twice as likely as Novices to consider future contingencies in their decision making.

Publication: Calderwood, R., Crandall, B., & Klein, G. A. (1987). Expert and novice fire ground command decisions (KATR-858(D)-87-02F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327) for U.S. Army Research Institute, Alexandria, VA.

Study 2: A Critical Decision Study of Distributed Decision Making in Wildland Firefighting

Goal. The goal of this study was to investigate decision strategies used by highly experienced commanders as they coordinated the efforts of thousands of firefighters during a large wildland fire. In this way we hoped to learn about decision-making strategies employed by command level experts in a high risk, often rapidly changing, distributed decision environment.

Coding. Coding followed the scheme developed in Calderwood, Crandall, and Klein (1987), except that the possibility that multiple decision strategies might be used was recognized by allowing multiple codes for a single decision point.

Method. This was an observational study carried out over eight days. Highly expert, command level, wildland firefighters working within the Incident Command System were observed and interviewed as they managed the suppression of a large forest fire. Seventeen very experienced members of two national Overhead Teams served as participants in this study. Critical Decision interviews were conducted by two on-site observers to determine the nature of the decision making strategies these experts used while performing their command-and-control activities.

Findings. As predicted, these experts relied heavily upon recognitional decision-making strategies. This was more pronounced in areas in which they had the greatest expertise. At many decision points they did not need nor have the luxury to deliberate among options. However, for decisions involving organizational issues and interpersonal negotiations (28% of the incidents identified as critical), we found a predominance of analytical strategies in which several options were evaluated concurrently.

Many of the complications of distributed decision tasks we had anticipated did not occur. There was little problem with information overload. Communication channels were limited but were used effectively. There was open communication about differences in the way situations were perceived and goals were formulated, but these were controlled so as to maintain team cooperation and morale.

Publication: Taynor, J., Klein, G. A., & Thordsen, M. (1987). Distributed decision making in wildland firefighting (KATR-858(A)-04F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Study 3: A Critical Decision Study of Decision Making in Armored Platoon Command

Goals. This study was conducted to investigate the validity of laboratory based decision models for describing how Novices attain experience in armored platoon command. A primary focus was the Novice decision makers' description of contextual cues present at the time of the decisions. Investigators constructed a representational system for the cues and topics related by the Novice decision makers. These were compared to reports of the same decision situations related by experienced instructors who had observed evaluated them.

Method. CDM interviews were conducted with three classes of Armed Officer Basics over days three to six of field training exercises at Fort Knox, KY. Two observer/interviewers identified decision situations and interviewed student platoon leaders within twenty minutes of their completing the exercise scenarios. One observer rode in the platoon leader's tank during the exercises and then interviewed the trainee. The other interviewed the instructor who rode on top of the platoon leader's tank.

Coding. Three types of data were collected from the students in the interviews: the type of decision situation and decision strategy used, the cues and knowledge available to the student during the time of the decision, and self-performance ratings on a) tank and b) platoon actions as a result of the decisions. Instructors also reported cues and knowledge available and rated students' actions on the same performance scale presented to the students.

Results. The contextual cues and areas of knowledge students reported in their decision accounts were very similar to information offered by the

instructors. This suggests that performance difficulties were not generally the result of inattention to appropriate environmental cues but misinterpretation of the cues' importance. Analysis of the students' decision strategies revealed two main methods of resolution: 1) limited option deliberation and 2) recognition-primed decision implementation. These two methods were approximately equal in frequency. The students' high use of the latter type of strategy is consistent with our earlier research on more experienced personnel in other domains and supports the validity of a recognitional model for decision making at lower levels of expertise as well.

The number of analogues reported by the students was fairly stable across the observed training period and demonstrates that novices also use previous experience to guide decision making. Interestingly, analogues were helpful only about half of the time. On the remaining occasions the impact of analogues was mixed, ranging from neutral to disruptive.

One area in which some very interesting results surfaced was in the differential use of "hypotheticals" by the armored officer basic students (AOBs) used as compared to the more experienced track command instructors (TCIs). "Hypotheticals" reflected an evaluation of possible alternative future states. Overall, the AOBs showed a much weaker inclination to consider these hypotheticals. In addition, the more abstract the hypotheticals were, the greater the discrepancy between the number considered by the TCIs versus the AOBs. Terrain and factors concerning one's own tank were considered to be concrete hypotheticals, while platoon control, other friendly support, communications, and enemy unit hypotheticals were considered more abstract.

As mentioned earlier, the frequency of situational assessment statements was roughly equivalent for AOBs and TCIs. However, the primary area where a

discrepancy in SA did occur was in use of hypotheticals. While it does not appear that the AOBs are less attentive to SA information, it does seem that they are not yet able to select the most effective information to use to generate available options.

Publication: Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision making in armored platoon command (KATR-858(B)-87-05F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Study 4: How Do People Use Analogues to Make Decisions

Goal. The purpose of this project was to examine the data we had gathered using Critical Decision method to learn more about analogical reasoning and its role in decision-making.

Method. Data from Critical Decision interviews collected in five studies was reexamined for evidence of analogue use. The studies included two studies of urban fire ground command (Calderwood et al. 1987; Klein, Calderwood, & Clinton-Cirocco, 1986), a study of decision training during tank platoon exercises (Brezovic, Klein, & Thordsen, 1987), a study of decisions made during a wildfire incident (Taynor, Klein, & Thordsen, 1987) and a study of Air Force design engineers (Klein & Brezovic, 1986).

This data base contains over 400 decision points in all, culled from interviews with over 100 decision makers. The data were analyzed and compared in order to gain a broader understanding of the role of analogical reasoning in decision making. A total of 33 analogues were identified in enough detail to analyze the functions served by the analogue.

Results. Three functions of analogical reasoning were identified, 1) understanding situational dynamics, 2) generating options, and 3) evaluating the probable success or failure of implementing an option. Several tentative conclusions were also offered:

* Analogical reasoning is reported relatively infrequently by experts, perhaps because the individual cases have often merged into prototypes.

* When analogical reasoning occurs, it is often critical for success. For experts, it often emerges during non-routine cases.

* Novices appear to rely more heavily than experts on analogical reasoning, but have not learned how to apply the analogues, modify them, or

reject them. Therefore almost half the analogue use by novices results in poor choices.

Publication: Klein, G. A., & Calderwood, R. (1988). How do people use analogues to make decisions? Case-Based Reasoning Workshop, sponsored by Defense Advanced Research Projects Agency (DARPA), Clearwater Beach, FL.

Study 5: The Effect of Time Pressure on Expert Decision Making

Goal. The study investigated the effect of time pressure on the decisions made by chess players at two different levels of skill. The hypothesized results were based on assumed differences in the temporal requirements of calculational and recognitional modes of processing. Calculational processes, such as generating move-countermove sequences in order to evaluate outcomes, are relatively time dependent. When time constraints are imposed, calculations must be either truncated or omitted, thereby impairing performance. Recognitional processes, on the other hand, are defined as rapid and holistic. Performance based on recognitional processes should therefore be relatively insensitive to time constraints. Thus, we anticipated an interaction between time pressure and playing skill on move quality in chess. An obtained interaction would provide converging evidence for the claim that highly-skilled decision makers rely more on their recognitional abilities than do less skilled individuals.

Method. The rated quality of moves for very strong (masters) and weaker (class B) players was compared for tournament games played under regulation (at least 50 moves in two hours) and blitz (6 minutes total playing time per player) time rules. Tournaments were arranged as double round-robins wherein each of three players at each skill level played each of the other players four times, twice for regulation and twice for blitz games. This design resulted in 24 games, six in each of the conditions resulting from crossing the player-skill and game-type factors. Moves were rated for quality on a 5-point scale by a chess grandmaster.

Results. Results of the analysis of move quality ratings supported the predicted interaction between skill level and game type. That is, the

decrement in move quality for blitz games compared to regulation games was greater for the class B players than for the masters. The validity of the interaction was supported by the fact that masters were more able to maintain higher quality moves in the blitz condition at the same time that they generated a substantially greater number of moves, and proportionately more complex moves, than the class B players. These results were interpreted as supporting the view that more highly skilled players can rely more extensively on rapid recognitional processes than less-skilled players. Of course, this does not rule out the possibility that given adequate time, more highly skilled players may also calculate more extensively and more profitably than weaker players.

Publication: Calderwood, R., Klein, G. A., & Crandall, B. W. (In press). Time pressure, skill, and move quality in chess. American Journal of Psychology.

Study 6: Classification of Decision Events

Goal. Over 50 published articles were reviewed relevant to event classification, prototypes, schemas, skill development, and "pre-decisional" decision processes. The goal was to synthesize the findings in order to suggest a framework for studying situational assessment processes.

Findings. Several surprising "holes" in our knowledge relevant to these areas were uncovered and summarized.

* There are surprising few links between the decision making literature and the literature on concept formation and categorization. Yet, the RPD framework suggests that a major component of decision making is in how an event is understood and classified.

* Little is known about natural event classification. Classification stimuli have tended to be objects or unidimensional concepts.

* The closest analogue to event classification may be in problem-solving studies which have used psychological scaling techniques to uncover and represent the "dimensions" on which similarity judgments are made. Many of these studies have compared the derived representations of Expert and Novice performers in order to draw inferences about the nature of skill development. However, these investigations have not considered how context may influence the judgments on which the clustering solutions are based. Nor have they considered how context might interest with skill.

Publication: Based on this literature review, a dissertation was proposed by the second author and accepted by the Psychology Department of the University of New Mexico. The study results will be sumbitted for publication.

Study 7: Protocol Analysis of Expert/Novice Fire Ground Decision Making During Simulated Incidents

Goal. The goal of this study was to experimentally demonstrate the suggestive findings obtained in previous studies in this series. Whereas the previous investigations have relied on retrospective interviews to probe for information, the present study obtained think-aloud protocols during simulated incidents. Thus, this approach would offer the first opportunity to judge the content and strategy differences of Expert and Novice decision makers unconfounded by differences in the type of incident and the information available in the situation.

The study was designed to address several inter-related issues of relevance to RPD model development:

* Does this alternate method provide convergent evidence for the serial evaluation strategy described by the RPD model?

* Does the method provide a technique for examining progressive deepening and imagery as a means of option evaluation?

* What aspects of situation assessment are spontaneously reported -- what cues, inferences, and goals are associated with command decisions?

* To what extent are these factors associated with domain expertise?

Method. Three simulations of fire ground incidents were developed. Key events of actual incidents were recreated using an audio-visual format to present the details of the incident from the commander's perspective. The simulation presents relevant radio communication and a series of graphic slides of an incident from the time of the initial alarm to a point where the incident has been brought under control. All events are depicted from the point of view of the FGC. A narrator supplies needed background information

that would be known to the commander or would become available in other ways during an actual incident. Key events are portrayed in near real-time. In their final form, each simulation contains multiple decision points that span the duration of the incident. The tape is stopped at these points allowing the participant, assuming the commander's role, to think-aloud about any decisions he would make at this point.

Twenty-two professional firefighters participated, 11 Experts and 11 Novices. Expert/Novice ranking was made on the basis of overall command experience and an ability rating made by the Chief of Suppression Officer of a major urban fire department from which participants were drawn.

Coding. A coding procedure was developed and tested for inter-coder reliability. The method classified each protocol remark into 12 independent categories related to cues, knowledge, actions, and goals. In addition, evidence for both RPD and concurrent decision strategies, imagery and analogue use, progressive deepening, and possible errors were noted when they occurred.

Results. Analysis of the frequencies of the remark categories substantiated the hypothesized differences in decision "focus" for Experts and Novices. Experts appeared to pay more attention to assessing the situation (noticing cues and making inferences based on the cues), whereas Novices pay relatively more attention to generating and evaluating options.

A content analysis based on a conceptual node graph of the remark categories was performed. This graphing method proved to be a powerful interpretive tool for abstracting within group commonalities and highlighting between group differences. The node graphs supported the idea that both situation assessment and action schemas were richer and more elaborated in the

Expert group. It also revealed underlying "decision points" that were difficult to detect in individual protocols.

Publication: The full report (Protocols Analysis of Expert/Novice Command Decision Making During Simulated Fire Ground Incidents by R. Calderwood, B. W. Crandall, & T. H. Baynes) of the study is included as Appendix I of this final report.

REFERENCES

Barlett, F. C. (1932). Remembering. Cambridge, England: Cambridge University Press.

Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the 'bias heuristic'. Acta Psychologica, 50, 201-252.

Brehmer, B., & Allard, R. (1986). Dynamic decision making: A paradigm and some experimental results. Sweden: Uppsala University.

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision making in armored platoon command (KATR-858(B)-87-05F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Calderwood, R., Crandall, B., & Klein, G. A. (1987). Expert and novice fire ground command decisions (KATR-858(D)-87-02F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327) for U.S. Army Research Institute, Alexandria, VA.

Christensen-Szalanski, J. J. J. (1986). Improving the practical utility of judgment research. In B. Brehmer, H. Jungerman, P. Lourens, and G. Sevon (Eds.), New directions in research on decision making. North Holland: Elsevier.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. American Psychologist, 39, 75-78.

Cooper, W. S. (1987). Decision theory as a branch of evolutionary theory: A biological derivation of the savage axions. Psychological Review, 94, 395-411.

Cyert, R. M. & March, J. G. (1963). A behavioral theory of the firm. Englewood Cliffs, NJ: Prentice-Hall.

de Groot, A. D. (1965/1978). Thought and choice in chess (2nd ed.). NY: Mouton.

Edwards, W. (1962). Dynamic decision theory and probabilistic information processing. Human Factors, 4, 59-74.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. A. Slovic, & A. Tversky (Eds.), Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P. & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. Journal of Applied Psychology, 70, 2, 401-412.

Gettys, C. F. (1983). Research and theory on predecision processes (TR 11-30-83). Norman, OK: University of Oklahoma, Decision Processes Laboratory.

Hoffman, R. R. (1987). The problem of extracting the knowledge of experts from the perspective of experimental psychology. AI Magazine, 8, 53-67.

Howell, G. E. (1984). Task influence in the analytic intuitive approach to decision making final report (ONR Contract N00014-82 C-001 Work Unit (NR197-074)). Houston, TX: Rice University.

Isenberg, D. J. (1984). How senior managers think. Harvard Business Review, 6, pp. 80-90.

Janis, I. L., & Mann, L. (1977). Decision making: A psychological analysis of conflict, choice, and commitment. NY: The Free Press.

Kadane, J. B., & Lichtenstein, S. (1982). A subjectivist view of calibration (82-6). Eugene, OR: Decision Research.

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

Kahneman, D., & Tversky, A. (1979). Intuitive predictions: Biases and corrective procedures. TIMS Studies in Management Science, 12, 313-327.

Keeney, R. L., & Raiffa, H. (1976). Decisions with multiple objectives: preferences and value tradeoffs. New York: Wiley.

Kintsch, W., & Van Dijk, T. A. (1975). Comments on summaries of stories. Langages, 40, 98-116 (in French).

Klein, G. A. (In press). Recognition-primed decision. In W. Rouse (Ed.), Advances in Man-Machine Systems Research, 5. Greenwich, CT: JAI Press, Inc.

Klein, G. A., & Brezovic, C. P. (1986). Design engineers and the design process: Decision strategies and human factors literature. Proceedings of the 30th Annual Human Factors Society, 2, 771-775. Dayton, OH: Human Factors Society.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. Proceedings of the Human Factors Society 30th Annual Meeting, 1, 576-580. Dayton, OH: Human Factors Society.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1985). Rapid decision making on the fire ground (KA-TR-84-41-7). Yellow Springs, OH:

Klein Associates Inc. Prepared under contract MDA903-85-G-0099 for the U.S. Army Research Institute, Alexandria, VA.

Klein, G. A., Calderwood, R., & MacGregor, D. (In press). Critical decision method for eliciting knowledge. IEEE Transactions, Man, Systems and Cybernetics.

Klein, G. A., & Weitzenfeld, J. (1982). The use of analogues in comparibility analysis. Applied Ergonomics, 13, 99-104.

Klemp, G. O., & McClelland, D. C. (1986). Executive competence: What characterizes intelligent functioning among senior managers? In. R. J. Sternberg & R. K. Wagner (Eds.), Practical intelligence: Nature and origins of competence in the everyday world. MA: Cambridge University Press.

Lipshitz, R. (1987). Decision making in the real world: Developing descriptions and prescriptions from decision Maker's retrospective accounts. Unpublished manuscript. Boston University, Center for Applied Sciences.

Nisbett, R., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. In B. Brehmer, H. Jungermann, P. Lourens, and G. Sevon (Eds.), New Directions in Research on Decision Making. The Netherlands: Elsevier Science Publishers B.V.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. Psychological Review, 77, pp. 257-273.

Rapoport, A., & Wallsten, T. S. (1972). Individual decision behavior. In P. H. Mussen and M. R. Rosenzweig (Eds.), Annual review of psychology. Palo Alto, CA: Annual Reviews Inc.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the

internal structure of categories. Journal of Cognitive Psychology, 7, 573-605.

Rouse, W. B. (1978). Human problem solving performance in fault diagnosis task. IEEE Transactions on Systems, Man, and Cybernetics, 4, SMC-8, 258-271.

Simon, H. A. (1955). A behavioral model of rational choice. Quarterly Journal of Economics, 69, 99-118.

Taynor, J., Klein, G. A., & Thordsen, M. (1987). Distributed decision making in wildland firefighting (KATR-858(A)-04F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA.

Tversky, A. (1977). Features of similarity. Psychology Review, 84, 327-352.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90(4), 293-315.

Weitzenfeld, J. (1984). Valid reasoning by analogy. Philosophy of Science, 51, 137-149.

Wohl, J. G. (1981). Force management decision requirements for Air Force tactical command and control, IEEE Transactions on Systems, Man, and Cybernetics, 618-639.

Wohl, J. G., Entin, E. E., Kleinman, D. L., & Pattipati, K. (1984). Human decision processes in military command and control. In W. B. Rouse (Ed.),

Advances in Man-Machine Systems Research, Vol. 1, (pp. 261-307). Greenwich, CT: JAI.

Zakay, D., & Wooler, S. (1984). Time pressure, training, and decision effectiveness. Ergonomics, 27, 273-284.

APPENDIX I

Protocol Analysis of Expert/Novice Command Decision Making During Simulated Fire Ground Incidents

Roberta Calderwood, Beth W. Crandall, and Timothy H. Baynes

Klein Associates Inc.
Yellow Springs, OH 45387

July, 1988

Portions of this document can be found in Klein, G. A. (in press). Recognition-primed decision. In W. Rouse (Ed.), Advances in Man-Machine Systems Research, 5. Greenwich, CT: JAI Press, Inc.

Table of Contents

List of Figures ii

List of Tables iii

Acknowledgements iv

Introduction I-1

Recognition-Primed Decision Making I-2

Serial Evaluation I-4

Progressive Deepening I-6

Situation Assessment I-7

Development of FGC-Simulation I-11

Method I-14

Study Participants I-14

Procedure I-15

Protocol Analysis I-17

Coding Procedure I-17

Coding Reliability I-20

Results I-25

Analysis of Category Frequencies I-25

Content-Node Analysis of Remark Categories I-33

Evidence of Decision Strategies I-45

Summary and Conclusions I-49

References I-53

List of Figures

Figure		Page
1	Recognition-Primed Decision (RPD) Model	I-5
2	Percent of Cue-deliberation and Action-deliberation	I-31
3	Cue and Knowledge Probe Point 1.1	I-37
4	Action and Goal Nodes for Probe Point 1.1	I-39
5	Cue and Knowledge Nodes for Probe Point 2.1	I-43
6	Action and Goal Nodes for Probe Point 2.1	I-44

List of Tables

Table		Page
1	Sample Coded Protocol	I-21
2	Percent Agreement for Remark Segments for Coders	I-23
3	Reliability of Remark Category Coding: Kappa Coefficient. . .	I-24
4	Average Number of Remarks for Experts and Novices.	I-26
5	Instances of Situation Assessment Deliberations.	I-47
6	Instances of Action Deliberations.	I-48

Acknowledgments

This study has been supported by the U.S. Army Research Institute for the Behavioral and Social Sciences Contract No. MDA903-85-C-0327. The authors wish to thank Dr. Judith Orasanu for her encouragement and support. Our deepest appreciation goes to the personnel of the Dayton Fire Department who participated in the study. Thanks also to Jim Nutt, Chief of Miami Township Fire Department for his help in developing the simulations.

Protocol Analysis of Expert/Novice Command Decision Making During Simulated Fire Ground Incidents

Roberta Calderwood, Beth W. Crandall, and Timothy H. Baynes

This research is part of a series of studies investigating how decisions are made in operational settings by trained personnel. Our focus has been on environments in which strategic and tactical decisions must be made under conditions of uncertainty, risk, and time pressure such as urban firefighting (Calderwood, Crandall, & Klein, 1987; Klein, Calderwood, & Clinton-Cirocco, 1986), wildland firefighting (Taynor, Klein, & Thordsen, 1987), and tank platoon battle management (Brezovic, Klein, & Thordsen, 1987).

For the present study, verbal protocols were obtained from professional urban firefighters during simulated fire incidents in which they were asked to assume the role of fire ground commander (FGC). The FGC is responsible for establishing strategy and overseeing tactical maneuvering of personnel and equipment in response to a fire emergency. Decisions include where and how to attack the fire given such factors as risk to crews and civilians, availability of water and other resources, and risk to exposed property. A major goal of the simulation development was that scenarios be realistic and complex enough to ensure a high level of engagement. Thus, the approach afforded an opportunity to examine decision making in a more ecologically valid way than has generally been the case in decision research (Neisser, 1976), and to obtain a rich and distinctive source of data for addressing issues of inference, expectancies, and tacit knowledge that are part of decision making in real-world settings.

Although the participants were all experienced firefighters who had from 7 to 27 years with the department, we were particularly interested in how different levels of experience and skill might influence the nature of the decision processes that would be reported. Therefore, officers were selected who represented a wide range of experience and ability.

Surprisingly little attention has been given to how expertise in decision making develops. In fact, since the landmark articles by Daniel Kahneman and Amos Tversky in the 1970's (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1971, 1973, 1974), decision researchers have tended to stress almost exclusively the ways in which decision makers are biased and suboptimal (see Christensen-Szalanski & Beach, 1984; Hammond, 1987; Lopes, 1987). We think this tendency has been misguided and that skill in decision making develops, as it does in other human endeavors, in terms of the representation of domain-specific knowledge (e.g., Dreyfus & Dreyfus, 1986; Glaser, 1981).

The method of protocol analysis developed for this study, along with the specific study questions addressed can best be understood in relation to the descriptive decision model that we are developing within this program of research. Therefore, this model will be briefly described and some of the findings and limitations of previous studies will then be reviewed.

Recognition-Primed Decision Making

The previously cited studies of command-and-control decision making relied on a variant of Flanagan's (1954) critical incident method that we have called the Critical Decision method (see Klein, Calderwood, & MacGregor, in press). The method uses a timeline reconstruction of a specific incident and focused probes to obtain retrospective protocols describing a decision event. Questions relate to the cues and knowledge that were heeded, the goals that

were operating, any alternatives that were generated, and how the implemented course of action was selected. Although there are obvious limitations to the method related to people's ability to recall and verbalize their own reasoning processes (e.g., Nisbett & Wilson, 1977), it is well recognized that such protocols can provide valuable insights about consciousness that would otherwise be impossible to obtain (Ericsson & Simon, 1984).

In these studies we have interviewed and observed over 100 individuals and probed almost 400 decision points. This extensive data base has provided a rich source of information about areas in which current models of decision making may be inadequate or misleading and has pointed toward potentially fruitful avenues for further research.

One of the most striking and consistent findings across these studies is how little evidence was found for evaluation strategies that rely on a direct comparison of the strengths and weaknesses of a set of generated options. Instead, experienced decision makers are most frequently found to rely on their abilities to quickly classify a situation on the basis of their prior experiences with similar cases. Once classified, options are automatically suggested, based either on standard operating procedures or on analogues that have been successfully employed previously. Only in cases where the initial recognition-based option is judged to be unworkable is a second option generated and examined for feasibility. This process continues in a serial fashion until a workable option is found.

We have described this process as a Recognition-Primed Decision (RPD) strategy (Klein, in press) depicted in Figure 1. Three aspects of the model are of primary importance: serial versus concurrent evaluation, progressive

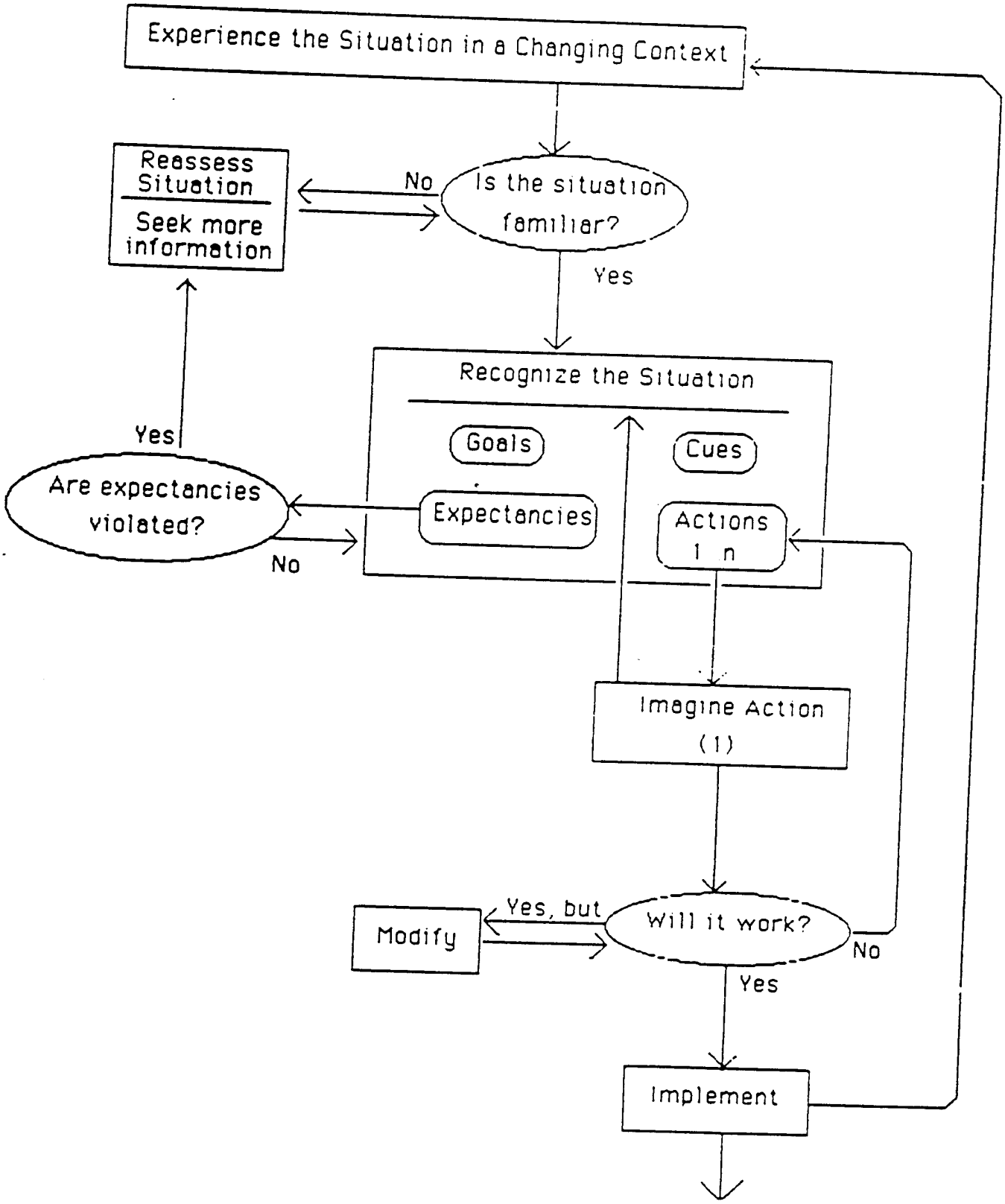
deepening, and the recognition of situations which -- we refer to as situation assessment.

Serial Evaluation

The serial evaluation of options described in the RPD model was the dominant strategy found in both of the urban fireground studies (Calderwood et al., 1987; Klein et al., 1986). We wish to claim that the RPD strategy will be more prevalent for experienced decision makers, relying as it does on memories for previously encountered similar events. Some support for the claim is suggested by the fact that in the tank platoon study (Brezovic et al., 1987), where the platoon leaders had relatively less experience than any of the other decision makers studied, less than half (42%) of the decisions were classified as RPD. Further, in the wildland study (Taynor et al., 1987) operational decisions, with which the commanders were most experienced, were more frequently found to be RPDs than were decisions involving organizational and management problems. However, the Calderwood et al. (1987) study of urban fire ground commanders found no difference in the frequency of RPDs between the most-experienced (expert) and least-experienced (novice) commanders once differences in the number of decisions made by individuals in these groups were taken into account.

No direct comparisons of decision strategies along the expert-novice dimension were possible in these studies. In each case, the whole situation as well as the decision maker differed. Because we have held the decision scenario constant in the present study, we can examine these issues under more controlled conditions. In the present study we sought support for the contention that concurrent option evaluation is the hallmark of a novice and not the end towards which decision makers should aspire.

Figure 1 -- Recognition-Primed Decision (RPD) Model



Progressive Deepening

Standard approaches to behavioral decision theory assume that option comparisons are carried out through some form of cost-benefit analysis. That is, dimensions are applied to all options under consideration using some process to determine which option is "best" on the aggregate of these dimensions (e.g., multi-attribute utility theory). The serial evaluation process of the RPD conceptualization leaves no basis for such comparative calculations. Instead, the decision maker must somehow evaluate the "goodness" of an option in isolation.

We believe that the primary means of evaluating an option involves a process of imagination in which the decision maker runs through a mental simulation of the outcomes of implementing an option. We have adopted the term "progressive deepening" to refer to this evaluation, a term coined by de Groot (1965/1978) to describe how chess grandmasters follow out a line of play to make sure it does not lead to blunders. The protocols we have collected contain some vivid examples of these mental simulations. One example of the progressive deepening strategy is an incident involving the rescue of an unconscious woman who was suspended from a highway overpass. For each type of rescue harness the commander considered, he imagined how it would be put on and how the woman's back would be supported once she was free of the structure. Several options were rejected when the image revealed a moment when the risk would be too great either to the woman or to his crew.

We suspect that this ability to imagine or project a scene into the future is an important component of skilled decision making. In the expert-novice study of urban firefighting (Calderwood et al., 1987), the Expert FGCs reported using imagery in over twice as many decisions as did the Novice FGCs

(20% vs. 8%). A related code classified each decision point as to whether it involved a reference to a possible future "state-of-the-world." In this too, twice as many Expert decision points as Novice decision points (48% vs. 24%) were future oriented.

In the study of tank platoon leaders (Brezovic et al., 1987), a major difference between the protocols of the student Armored Officer Basics (AOBs) and the more experienced Tank Crew Instructors (TCIs) was in the reported cases of what were termed "hypotheticals." These were statements that reflected consideration of future actions by either platoon or enemy troops. In each of 16 content categories, the TCIs had a higher percentage of hypotheticals.

Again, the present study provides an opportunity to produce converging evidence for these findings in a case in which the situations being viewed are held constant across the decision makers and the sources of biases present in our interview method are absent.

Situation Assessment

Behavioral decision theory has generally defined decisions in terms of what Berkeley and Humphreys (1982) call "the moment of choice." That is, decision models describe how an option is selected once the relevant options and evaluation dimensions have been generated. It does not generally try to account for the "pre-decision" processes (Gettys, 1983) involved in detecting and structuring the decision problem, defining relevant goals, and generating plausible courses of actions.

From the RPD perspective, however, processes prior to option selection are seen as critical. Options are evaluated in terms of the individual's "situation assessment" -- the understanding of a situation based on a sense of

familiarity. We have proposed that situation assessment entails at least four conceptually distinct dimensions of recognition: (1) critical cues and causal factors; (2) expectancies; (3) typical actions; and (4) plausible goals. We have found it essential to track each of these dimensions as they change over the course of an incident in order to understand the decision maker's choice of action.

Our conceptualization of situation assessment is related to the concepts of "schema" (Bartlett, 1932; Schank, 1986) and "mental model" (Gentner & Stevens, 1983) that are part of many current cognitive theories. The general notion is that incoming information is categorized, selected, edited, and organized on the basis of a person's general knowledge about a domain.

We expected an analysis of the content of protocol data obtained in this study to shed light on the nature of situation assessment processes. Specifically, we hoped to examine which cues were being heeded at each decision point, what inferences were drawn based on the cues, and which goals were most important in determining a selected course of action. In previous studies in this program of research, conclusions about these factors have been limited by the fact that there was no way to know what cues had actually been present in the situation or what other inferences might have been drawn. Nor has there been any way to compare how experts and novices might differ in their situational understanding, as each incident was unique to the individual reporting it to us.

The most systematic attempt to derive situation assessment categories in this research program was undertaken in the study of tank platoon leaders (Brezovic et al., 1987). This study had the advantage of on-site interviewing of decision makers during force-on-force field maneuvers over a three-day

period. Interviews were carried out both with the AOBs serving as platoon leaders and with the TCIs who had several more years of experience and who were responsible for evaluating the AOBs during the exercises. Thus, within the limits of physical proximity, the AOBs, TCIs, and the interviewers had the same information available.

Protocols were analyzed for 16 situation assessment categories relating to friendly and enemy control, actions, training, support, and time dimensions. The contextual cues and areas of knowledge the AOBs reported were very similar to the TCIs, suggesting that differences were not generally the result of inattention to appropriate environmental cues.

Rather, the primary differences seemed to reside in use of "hypotheticals" -- i.e., statements that reflected consideration of future actions by both platoon and enemy troops. In most of the categories considered, the TCIs had more remarks coded as hypotheticals than did the AOBs. Also of interest was the fact that for the TCIs, there were about the same number of hypotheticals for platoon and enemy categories, whereas the AOBs were much more focused on their own platoon's movements. The AOBs seemed less able to imagine how an enemy would react than to anticipate behaviors of their own platoon.

Results obtained in the study of the Expert-Novice fire ground decision making (Calderwood et al., 1987) are also relevant to this issue. This study was an initial attempt to use protocol analysis to capture the two logically distinct processes of situation assessment and option selection. Each deliberated decision was classified as having primarily involved deliberation about situation assessment (SA-decisions) or options (Option-decisions). These dimensions correspond to the operational distinction between the

questions "What is my situation?" and "What am I going to do about it?" We hypothesized that the expert decision makers would make relatively more SA-decisions than novices, in keeping with findings in related fields. For example, Larkin (1981) found that expert physics problem solvers expend more effort in constructing some kind of analogue to a physical representation before starting to solve the problem, whereas novices are more likely to proceed almost immediately to setting up equations. Similarly, Sternberg (1986) found that more intelligent problem solvers tended to put more of their time in the encoding of an analogy problem and less time operating on these encodings.

The results of this study can only be viewed as suggestive given the previously described difficulty of directly comparing decisions made by the expert and novice participants. Nonetheless, when the percents of SA-decisions and Option-decisions for each incident protocol were computed, experts had an equivalent percent of each type (30% and 30%) whereas novices had a higher percent of Option decisions (29%) than of SA decisions (18%).[1] Thus, the pattern of these relative percentages were in the predicted direction.

Again, the present study enables the examination of situation assessment by examining the content of decision protocols generated by more- and less-experienced commanders in response to the same events.

Study Goals

This study represents a significant departure from previous studies in this series. The previous investigations relied on retrospective interviews

---

[1]Note that the remaining percentages reflect the frequency of "Automatic" or non-deliberated decisions for which the SA versus Option distinction was not made.

to probe for information. The present study obtained think-aloud protocols during an incident by stopping the action at pre-selected "probe points." This necessitated creating a simulated task environment in which we could control the information being received by the decision maker. Considerable effort was expended in this project developing the simulation materials and protocol analysis methods.

The study was designed to address several inter-related issues of relevance to RPD model development:

*Does this alternate method provide convergent evidence for the serial evaluation strategy described by the RPD model?

* Does the method provide a technique for examining progressive deepening and imagery as a means of option evaluation?

* What aspects of situation assessment are spontaneously reported -- what cues, inferences, and goals are associated with command decisions?

* To what extent are these factors associated with domain expertise?

Development of Urban Fire Ground Command Simulations (FGC-Simulation)

In designing a simulation format, we faced an initial dilemma: how to realistically engage a commander in the decision making process while at the same time preserving control over the input features of an incident so as to allow comparisons across commanders. This problem exists because decisions in this environment occur in the context of action sequences, where the outcome of any particular decision affects the subsequent course of events. One approach to this dilemma has been interactive videodisc and computer presentation technologies that allow action sequences to be played out along some pre-specified number of alternative "branches." Although this approach allows repetition and comparison of scenarios within the limits of the

specified branches, the supporting technology tends to be very expensive both to develop and to run. More importantly, it assumes that the selected branches accurately represent and effectively exhaust the natural response categories of the decision maker.

Our solution was to develop scenarios that were organized around predefined decision points (as opposed to options) that represented key events within the overall incident. The scenario is interrupted at each of these decision points and verbal protocol data obtained. Once the participant has responded, the scenario is restarted and the narrator provides information about the actions that were actually taken at that point by the "real" FGC who was in charge of the incident. This device serves to reorient each participant to a common set of circumstances before allowing the incident to develop further.

Scenarios were based on interviews with FGCs obtained in a previous study (Calderwood et al., 1987). Events were recreated using an audio-visual format to present the details of the incident from the commander's perspective. The simulation presents relevant radio communication and a series of graphic slides of an incident from the time of the initial alarm to a point where the incident has been brought under control. All events are depicted from the point of view of the FGC. A narrator supplies needed background information that would be known to the commander or would become available in other ways during an actual incident. Key events are portrayed in near real-time.

We were fortunate to have a pool of incident accounts on which to base the scenario scripts. Nonetheless, the process of selection and development presented several challenges, for we wanted to depict the look and sound of the fireground as completely and accurately as possible. In choosing

incidents that would be suitable for simulation development, we adopted the following criteria:

Complexity. We wanted to retain the dynamic, complex nature of decision making on the fireground. This argued for using incidents that involved a series of decisions made in response to shifts in situational elements, and a variety of cues present in the situation.

Cue availability. Given our interest in situation assessment processes, presence of critical cues that were immediately available to the commander was another salient dimension. Our emphasis here was on cues that could be depicted without our prestructuring or interpreting them for the participant. This meant that the majority of critical cues in an incident had to be visual or auditory, because of the difficulty of representing olefactory or tactile cues in a simulation.

Incident type. We wanted the incidents to represent a range of issues and of types of tactical and strategic decisions. At the same time, we had been cautioned by trainers and others in the fire service, that fireground simulations are sometimes devised that are so "far out" that they are not believable. If the simulations appeared full of tricks or too much like a game, we feared that the participants' engagement would suffer. Thus, the incidents had to be believable and to represent the normal range of decision making for officers in this geographic area.

We began by developing script outlines and storyboards of the accompanying graphics for seven incidents. Working with an experienced officer/trainer, we selected two scenarios for full development and a third that could function as a practice. The incidents were judged to present different types of command challenges. The two study simulations were both

structure fires, one at an occupied apartment building and one at an unoccupied restaurant in a large historical building. The practice scenario involved an overturned tanker truck on a highway.

Development of the final scenarios was an iterative process. Graphic artists were retained to draw the structures and depict visual cues of the fire and surrounding area at several key points. A sound engineer oversaw development of the audio portion of the simulation. Radio communication was supplied by recording voices speaking through two-way radios. A professional actress was retained to narrate the incidents. Background noise and sound effects appropriate to the fireground were added. At each step in this process, we sought feedback about the realism of the representations from the FGC consultant.

We piloted the simulated incidents with four firefighters, two of whom are highly experienced FGCs. Final revisions and corrections were made on the basis of their comments and the graphics were then converted to slides.

Method

Study Participants

Participants were all professional firefighters employed by the City of Dayton Fire Department, Dayton, OH. We worked with the Department's Chief Suppression Officer to recruit volunteers who represented a range of command experience. Because the scenarios required the participant to take charge of a major incident, it was not feasible to use new firefighters or firefighters with virtually no command experience. The 22 FGCs who comprised the final sample are seasoned firefighters who were judged to vary considerably in command experience and skill by their chief. Expert FGCs had an average of 18.9 total years (range = 10-27 years) and Novices had an average of 13.5

total years of firefighting experience (range = 7-23 years). Experts had an average of 11.0 years in command positions (range = 6-20 years), as opposed to an average of 4.4 years of command experience in the Novice group (range = 1-12 years).

Procedure

In the final form each FGC-simulation scenario contained multiple decision points that span the duration of the incident. The scenario involving the overturned tanker truck was used in the present study for practice. It contains three decision points and two graphic depictions. The apartment-fire scenario contains five decision points, three graphic depictions of the scene, and two overhead maps that show apparatus and hydrant placement. The Inn-fire scenario contains six decision points, three graphics, and one overhead map.

The simulated incidents were presented using a tabletop audio/slide projector. Sessions were conducted individually by one of two trained examiners who were blind to rank or expertise classification.

After a brief introduction, the simulation format was explained. Participants were asked to imagine themselves in the command role in each incident, and told that at certain points the tape/slide presentation would be interrupted. They were instructed that when this occurred, they should:

> ". . . say aloud all the things you are thinking to yourself, and even to describe any images or memories that come to mind. Although we are interested in your plans and any actions you might take, we are also interested in what you are hoping to accomplish with your actions, what you are noticing, and any other options you are

considering. It is safe to say that we are interested in anything that pops into your head."

It was explained that once they had told us what they were thinking, the tape would be restarted and would tell them what the FGC actually did in this incident. We noted that some of the actions/decision depicted might be considered controversial, and that they might not always agree with what the FGC did. In that event, we asked that they save criticism until the end of each incident, when they would have a chance to comment. It was thought that providing an opportunity to critique the incident at the end would foster additional engagement, and keep participants from becoming sidetracked in disagreements about how the incident was being handled.

The participants were told that the initial scenario was for practice, to give them an idea of the format and answer any questions. After the practice, they were reminded that they were the commander throughout the incidents, and that they should say "whatever comes into your head." After answering any questions, we presented the remaining two scenarios. Verbal protocol data in response to all three scenarios was recorded on a separate audiotape, once permission to tape record had been obtained. No participant declined to be recorded.

At each predefined decision point, the audio/projector was stopped and the examiner asked: "What are you thinking about at this point?" At the end of the participant's verbal response, the examiner asked: "Anything else?" Participants were not prompted in any other way. When the participant declined further comment, the examiner restarted the tape. Participants occasionally raised questions about aspects of the simulation or the incident itself. If the request was for information contained in the taped narration,

or available on the screen, examiners would respond with the requested information. If, however, the request was for additional information not contained in the scenario, the examiners politely declined. This was done to ensure comparability across participants.

Biographical information (e.g., years of firefighting experience, years of command) was obtained at the end of each session when data gathering was complete. Sessions took from 45 minutes to 2 hours, depending on the amount of talk.

Protocol Analysis

Coding Procedures

All of the utterances produced during the probe points for the three incidents were transcribed. All of the speech recorded for each subject for a particular probe point within an incident constitutes a protocol. Each protocol was segmented by the transcriber into paragraphs reflecting naturally occurring pauses in the speech. Protocol transcriptions were keyed to a subject number but did not indicate the experience level or name of the participant.

A sampling of protocols from each probe point was read prior to beginning formal coding procedures and three probe points were selected as a sample of the larger set of 14 probe points: the first and fifth probe points in the apartment fire incident (Probe points 1.1 and 1.5) and the first probe point of the Inn incident (Probe point 2.1). The first probe points in each incident contained more words (and presumably more remarks) than later probes and would establish a context for interpreting later probe points. The single later probe point (1.5) was chosen arbitrarily.

Remark and category definition. The first step of the protocol analysis was to define the basic units of analyses. The goal was to segment the connected discourse into "chunks" at a meaningful level for addressing relevant research questions. This might have been anything from single words to short phrases to higher-level topics abstracted from groups of sentences. We chose as our unit of analysis the remark, which we defined as a word, phrase, sentence, or group of sentences that have a common topic or referent. The list of remark types constituting the coding categories are described below. Thus, remarks are somewhat circularly defined in relation to the content categories we used to address the research questions of interest.

Every remark contained in a protocol was classified as belonging to one and only one of the following categories:

1. CUE - (type). Remarks that express an awareness of the information provided by the scenario. They are the facts of the present case as viewed by the subject (Example: "I can see the fires in the basement). The cue "type" is a summary of the information content -- what was noticed.

2. CUE-Deliberation - Remarks that express uncertainty about the meaning of a cue or set of cues (e.g., "This could be a ...") indicating a need to deliberate or come to a decision about the current state-of-the-world.

3. CUE-Anticipation - Remarks that involve a prediction about an anticipated future state-of-the-world based on present cues or inferences. (Example: "From the looks of it, fire's going to run that wall right into the attic.")

4. KNOWLEDGE - Remarks that express domain-relevant knowledge of fire ground factors. These may occur in assessing the meaning of cues in the present case or to evaluate or generate an action plan. General knowledge may be about equipment, resources, crew functions, building structure, or fire dynamics,

etc. (Example: "Balloon construction means that a fire is likely to spread vertically very rapidly;" "Blackish smoke indicates the presence of a hazardous material.") This is booklearning that is applied, modified, and interpreted in assessing cues in the specific incident.

Knowledge expressed as standard operating procedures (SOP) were considered as a special case of knowledge. These remarks reflect standard strategy and tactics on the fireground (example: "Life is the most important"), or about procedures that are considered standard by this department in particular. (Example: "Our second-in engine functions as the supply," "trucks have priority for the front of the building").

5. ACTION - (type). Remarks that express the current actions or plans that the FGC will implement.

6. ACTION Deliberation - Remarks that express uncertainty about the action (Example: "I could/might do ...") indicating the need for deliberation or further evaluation before the action is implemented.

7. CONTINGENCY - Action remarks that indicate that a plan would only be put into effect <u>when</u> or <u>if</u> a future condition is met. (Example: "If it gets to the attic and mushrooms, I'll pull the crews out and go to master streams.")

8. GOAL - (type) - explicit statements about the purpose or reason for taking an action. These specify what the FGC hopes to accomplish. (Example: "We need to get that fire vented so we can clear the smoke and heat out of there.")

9. REMINDING - Remarks that refer to prior experiences that the simulation has brought to mind. (Example: "We've got a structure a lot like this one right up here on Main Street.")

10. APPRAISAL - Remarks that reflect an evaluation of "How am I (as FGC) doing?" or "How will this all turn out?" (Example: "This is a tough fire, I think we'll lose the building.")

11. META-COGNITIONS - Remarks that express how the FGC would go about thinking, making decisions, etc. (Example: "I'll be keeping in mind that the crews will wear out fast in this heat -- gotta keep an eye on that.")

12. CRITIQUES/QUESTIONS - comments or questions about the simulation itself or the simulation FGC's handling of the incident.

13. MISCELLANEOUS - incomplete or indiscernible remarks or remarks that do not fit any of the study categories.

A sample coded protocol is presented in Table 1.

Coding was done for all the protocols in a given probe point before going on to the next. We found that it greatly facilitated coding to have the context of a probe point clearly in mind, especially in designating the cue and action types being expressed. It also meant that the data for a particular probe point might be discarded and the protocols re-coded if reliability for that point did not reach acceptable levels. Inter-coder reliability was assessed periodically throughout the coding.

Three coders (the authors) were randomly assigned protocols from each probe point, although not all coders coded the same number of protocols.

Coding Reliability

After developing the coding definitions and procedures, several protocols were coded by each of the three coders (the authors) in order to test and refine our understanding of the coding categories. Formal assessment of inter-coder reliability was carried out concurrently with the coding of the three decision points coded for the present study. This was done in the

TABLE 1

Sample - Coded Protocol

[1] Cue [2] Cue [3] Cue [4] Cue [5] Cue [6] Knowledge:SOP	[1]Okay, I can see that we have a fairly large structure,] [2]fire on the second floor;] [3]doesn't appear that the first floor is involved at this point.] [4]It appears that engine 1 laid out coming in, they laid out their supply coming in,] [5]I don't see where there is an attack line down yet.] But, [6]initial companies, that would be their first response, to go ahead and lay the initial attack line, make entry into the building, make the stairwell to the second floor, and start checking it for the fire.]
[7] Cue [8] Action [9] Action [10] Knowledge:Fireground factors	[7]Since this is a relatively old building,] [8]my first response would be to go ahead and stage a second alarm,] [9]I would need DP&L, it is more than likely they are a large natural gas consumer.]
[11] Action [12] Appraisal [13] Cue [14] Cue	[11]I would go ahead and stage a second medic,] [12]this is going to be a tough fire] and [13]we may start running into heat exhaustion problems;] [14]it is 70 degrees now and it is going to get hotter.]
[15] Action [16] Action [17] Action-Elaboration [18] Goal	[15]The first truck, I would go ahead and have them open the roof up] and [16]the second truck I would go ahead send them on inside] and [17]have them start ventilating, start knocking the windows out and working with the initial engine crew, false ceilings and get the walls opened up.] [18]Get to the source of the fire, get it knocked down.]
[19] Cue [20] Action [21] Anticipation [22] Action [23] Cue-Deliberation [24] Anticipation [25] Metacognition [26] Contingency [27] Anticipation [28] Contingency [29] Action-Deliberation [30] Action	[19]I am assuming at this point I do not have any other engines on the scene.] [20]As soon as I can, go ahead and order the second engine to hook up to the supply and pump to engine 1,] [21]I am assuming engine 2 will probably be there in a second] and [22]have them pump to engine 1, supply them.] [23]I don't know how long the supply line lay is,] [24]but it appears we are probably going to need more water than one supply line is going to give us.] [25]So I would keep in mind,] [26]unless we can check the fire fairly rapidly,] [27]that we are going to have to have some more water,] [28]so start thinking of other water sources.] [29]Consider laying another supply line to engine 1] and [30]stand back and watch.]

interest of efficiency and also as a check on any drift in our agreement over time.

Two separate aspects of reliability were assessed -- inter-coder agreement in segmenting the protocols into remark units, and inter-coder agreement in classifying the remark segments into coding categories. The first task, assessing the reliability of remark segmenting, is difficult because differences between a match and a mismatch are usually a matter of degree of agreement. That is, a difference in any given segment can carry over to several subsequent segments. The second task, assessing the reliability of remark classifications, is interdependent with the first task. It is hard to classify a "thing" the same way if there is no agreement about which "thing" is being classified. In related research, these problems are sometimes bypassed by having a single criterion coder responsible for segmenting protocols, or by only computing classification reliability on remarks that were segmented with good agreement. Both of these procedures would appear to inflate the degree of agreement among coders starting from an unstructured verbal protocol.

Our solution was to adopt a sampling strategy that eliminated the interdependence of these two aspects of coding. Because we used a word-processor to print the columns of transcribed text, the text was broken arbitrarily at the end of a line. We chose the first word of every third line as an anchor on which to compare coders. A subset of protocols from each probe point were chosen at random and coded by at least two coders. Six protocols for probe point 1.1 were coded by all three coders. To increase efficiency, only partial overlap between coders was carried out for probe points 1.5 and 2.1. Coder C coded only two protocols on which to assess

reliability from each of the two remaining probe points. Coders A and B were responsible for coding the bulk of the protocols and they overlapped on five protocols for probe point 1.5 and seven protocols for 2.1.

Reliability of remark segmenting was assessed by comparing segments containing each anchor word on a three point scale. High match indicated almost perfect agreement (ignoring prepositions or articles) in designating both the beginning and ending of a remark segment. Medium match indicated a good degree of overlap in the segments -- these segments either started or ended in the same place on the transcript. Low match indicated discrepant segmenting of remarks. That is, coders' remark segments neither began nor ended similarly. Percent of remarks at each level of match, for each coder pair, are presented in Table 2. These data indicate a high level of agreement among the three coders in remark segmenting.

Inter-coder agreement on classification of remarks into content categories was assessed using the kappa coefficient (Cohen, 1960). Kappa is a chance-corrected measure of nominal scale agreement, first developed to assess

Table 2

Percent Agreement for Remark Segments for Coders A, B, and C

Level of Match	A with B[a]	B with C[b]	C with A[c]
High	69.8	64.6	65.5
Medium	28.6	34.1	33.7
Low	1.5	1.3	0.9

[a]Based on 167 remarks.
[b]Based on 204 remarks.
[c]Based on 119 remarks.

reliability of patient assignment to medical diagnostic categories. It provides a more conservative measure of inter-coder reliability than do simple percent of agreement measures which often provide inflated indices of coding reliability (Fleiss, 1981). Kappa coefficients for major coding categories for coder pairs AB and BC are presented in Table 3.[2] Several coding categories occurred so infrequently that their reliability could not be assessed. In these cases, data were either combined into a higher-level coding category (e.g. "Anticipation" was collapsed into the "Cue" category) or were dropped altogether.

Table 3

Reliability of Remark Category Coding: Kappa Coefficients

Category	A with B[a]	B with C[b]
Cue	.52	.56
Knowledge: Fireground factors	.26	.51
Knowledge (SOP)	.60	.55
Action	.68	.69
Contingency	.58	.23
Goals	.58	.73
All Categories	.54	.50

[a]Based on 174 remarks
[b]Based on 111 remarks

[2]In the interests of efficiency, all three coders did not code all protocols. The number of protocols that coders A and C coded in common was judged too small to allow good measurement of their coding reliability.

In general, kappa values that exceed .75 are considered excellent agreement beyond chance, kappas between .40 and .75 indicate fair to good agreement beyond chance and those below .40 are considered poor. Based upon these criteria, levels of inter-coder agreement in the present study are generally quite good. As a check on variation in reliability across probes, overall kappa coefficients were computed separately for probes 1.1 and 2.1. These values were very close: .61 for probe point 1.1 and .57 for probe point 2.1.

Results

One concern in a study of this type is how well participants respond to the simulation format and whether they are able to verbalize their thinking. Participants generally reported that they found the task interesting and they seemed to have little trouble talking during the probe point pauses. The number of words spoken during the selected probe points ranged from 39 to 723. Averages for the three selected probe points were 337 words for probe point 1.1, 166 words for probe point 1.5 and 289 words for probe point 2.1. Example protocols from two Experts and two Novices from probe point 1.1 are included as Appendix A.

Analysis of Category Frequencies

Remarks contained approximately 13 words on the average. As expected from the word count, the first probe points (1.1 and 2.1) contained more remarks on the average (24.95 and 22.85, respectively) than did probe point 1.5 (13.75). The average number of remarks for the Experts and Novices for each of the selected probe points is shown in Table 4. There do not appear to be large or systematic differences in the number of remarks for these

Table 4

Average Number of Remarks for Experts and Novices for Three Probe Points

Probe Point	Experts	Novices
1.1	26.27	22.18
1.5	14.09	13.45
2.1	22.36	23.27

groups; nonetheless, the protocol frequencies are expressed as a percent of the remarks in the protocol in order to equate the conditions.

The average percentages of 11 remark categories for the Experts and Novices are shown in Table 5. Several of the coding categories will not enter into this interpretative analysis, although they were retained in the data pool for purposes of obtaining overall remark percentages. For example, across probe points, repetitions of previous remarks comprised 7.1% of the coded remarks. We felt that including these in the analysis was not informative enough to justify the added complexity. The Critique/Question category comprised another 3.4% of the remarks. These are discounted because they are irrelevant to the decision-making issues being addressed. The Appraisals, Remindings, and Meta-Cognition categories comprised less than 2% of remarks in each probe point. These categories do not enter into the present discussion, although we did examine each of the examples of these categories for clues they might contain about decision-making processes (these exasmples are discussed in a later section). Finally,

Table 5

Average Percent[a] of Decision Remark Categories
for Experts and Novices for Three Probe Points (N=22)

Category	Probe 1.1		Probe 1.5		Probe 2.1		Differences
	E	N	E	N	E	N	
Cue	18.5	16.5	22.6	19.0	15.3	15.4	
Elaboration	1.3	1.1	0.3	3.0	0.6	0.0	
Anticipation	3.6	1.9	2.7	2.7	2.4	3.5	
Deliberation	5.1	7.4	4.6	2.0	11.3	4.2*	E > N*
Knowledge: Fireground factors	7.9	4.3	4.0	1.6	6.7	3.5	E > N#
Knowledge (SOP)	4.8	5.9	7.8	0.8	6.0	3.5	
SITUATION ASSESSMENT	41.2	37.1	42.0	29.1	42.3	30.1	E > N#
Action	20.2	25.9	29.2	29.7	22.1	39.2	
Elaboration	7.1	2.7*	1.4	1.3	4.5	4.1	E > N*
Contingency	2.4	4.4	.2	7.7*	2.0	2.9	E < N#*
Deliberation	1.5	4.8*	4.9	4.6	2.2	1.2	E < N*
ACTION ASSESSMENT	31.2	37.8	35.7	43.3	30.8	47.4	E < N#
Goals	8.3	6.5	4.5	3.3	7.1	3.6	E > N#

*Large difference support
#Replication support

[a]Percentages are proportion of all remarks contained in a protocol. Tabled findings do not sum to 100%.

remarks coded Miscellaneous comprised only 4.6% of the coded remarks, and these were also dropped to simplify the analysis.

We know of no generally accepted methods for establishing statistical significance for data such as these. We adopted the following criteria for determining which of the observed differences between Experts and Novices should be considered "meaningful:"

1) The absolute difference between the category means is greater than the pooled estimates of the standard deviation of the protocol percentages for the category within that probe point (Large Difference support).

2) The direction of the differences between Experts and Novices for a category is consistent across the three probe points (Replication support). These criteria seemed to provide a reasonable balance between the costs of Type I and Type II errors for this exploratory analysis.

Table 5 is organized into groupings corresponding to the production-rule form suggested by the RPD model: if CONDITION, then do ACTION. In the RPD model, the CONDITION determination is termed a situation assessment. The situation assessment is based on incoming information and case-relevant knowledge. If there is uncertainty with regards to the CONDITION, then deliberation must occur to determine what condition is most appropriate. Deliberation may involve seeking more information, thinking about alternative interpretations of the cues, or anticipating future developments by projecting a scene forward in time. Situation assessment should be reflected in the Cue and Knowledge categories.

It became clear in coding the protocols that the Cue and Knowledge categories really represented a continuum of inferential interpretation. For example, remarks in probe point 2.1 frequently referred to the fact that the

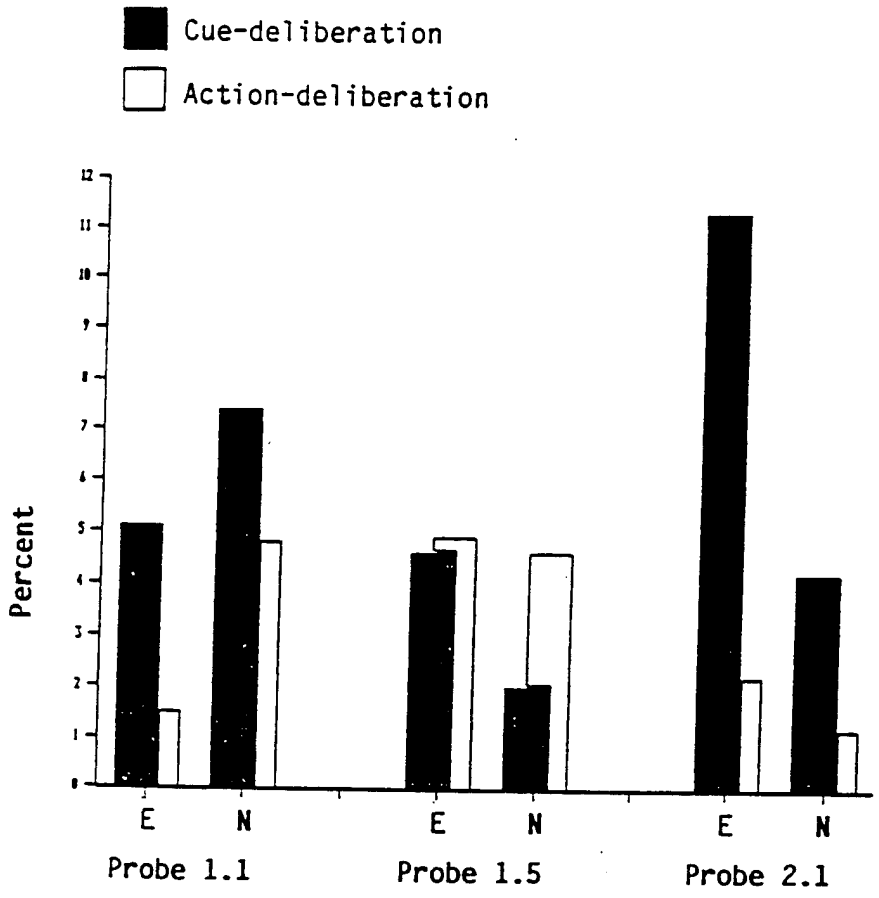
building was old. This remark was coded as CUE - (type = structure is old). This information had been provided as part of the background of the case, it required no inference. However, in a real incident a building's age may have to be inferred based on an interpretation of other cues. Other remarks concerned which crews or equipment were available, such as "I have an extra engine available." Such remarks were frequently so matter-of-fact that one coder tended to classify them as Cues. Another coder tended to classify these remarks as Knowledge, because a judgment of "availability" required knowledge of the relationship between resources and needs. This ambiguity accounts for the low inter-coder agreement for the Knowledge category (kappa = .26) for coders A and B. When these categories are combined, agreement is raised into the "good" range (kappa = .57).

A higher overall percent of Experts' remarks were classified as Situation Assessment remarks than were Novices' (Expert Situation Assessment = 41.8%, Novice Situation Assessment = 31.1%) and the direction of this difference is supported by the Replication criterion. However, these differences are not large and inspection of Table 5 reveals that they are primarily related to the Knowledge and Cue-deliberation remarks. The fact that these categories are designed to reflect inferential and reasoning processes, provides support for the view of expertise on which the RPD model is based. That is, to the extent that remark categories reflect relative "amounts" of processing, the Experts appear to deliberate more frequently about the nature of the situation than do Novices. Said differently, the consistently higher percentage of Situation Assessment remarks support the notion that Experts expend more effort in building an accurate "mental model" of a situation on which to base decisions about what actions to take. The difference found in the Cue-deliberation

category for probe point 2.1 will be discussed in the next section in which the content of the category remarks is examined.

Examination of the Action Assessment categories reveals that Novices have a higher overall percentage of remarks in these categories than Experts (Novice Action Assessments = 42.8%, Expert Action Assessments = 32.6%). The direction of this difference is supported by the Replication criterion. This is further support for the hypothesized differences in the deliberation strategies of Experts and Novices. Of particular interest is the Deliberation category in probe point 1.1 which meets the Large Difference criterion. This category provides the most direct support for the hypothesis, generated on the basis of retrospective interviews with FGCs (Calderwood et al., 1987), that Novices deliberate about options more than do Experts. The difference found in the Contingency category of probe point 1.5 also favors the Novices. Contingency planning is conceptually related to the Deliberation code. In one case the deliberation is about what to do <u>now</u>, in the other it is about what to do <u>in the future</u>.

The fact that Experts have a higher percentage of Cue-deliberation remarks than Novices, while the opposite is true for Action-deliberation remarks should <u>not</u> be interpreted to mean that Novices make absolutely more Action-deliberation remarks than Cue-deliberation remarks. In probe points 1.1 and 2.1, <u>both</u> Experts and Novices deliberate more about Situational cues than about action alternatives. These relationships are portrayed in Figure 2. This is an interesting finding, given that most decision models are mute on the nature of problem structuring and classification.



E = Expert
N = Novice

Figure 2: Percent of Cue-deliberation and Action-deliberation for Experts and Novices in Selected Probe Points

Over the three probe points, there were slightly more Goal remarks for the Experts than for Novices (Expert Goals = 6.6%, Novice Goals = 5.5%), and the relationship meets the Replication criteria. This is not a surprising finding given the general assumption that intelligent performance is distinguished by being "goal-driven" (e.g., Larkin, 1981; Holding, 1985; Anderson, 1981). What did surprise us was how uninformative the goal statements usually were for illuminating the basis for action. Indeed, Action remarks were frequently hard to discriminate from Goal remarks. Take, for example, the remark "I would order a line inside to locate and attack the fire." This could be parsed as Action = take line inside; Goal = locate fire and Goal = attack fire. Another remark, "I would order the engine crew to attack the fire." The goal "to attack the fire" is stated as an action but really means the same thing as the previous remark. Some support for the confuseability of actions and goals comes from the reliability assessments of these categories. Considered separately, the reliability of the categories for both sets of coders was in the "good" range (kappa <.75). Collapsing the categories together raises the kappa into the excellent range (kappa >.75) for both sets of coders.

The fact that some of the differences are present in some probe points and not others should not surprise us. Each situation would be expected to create its own unique context that will highlight specific components of a decision process. We are far from being able to specify the conditions under which the observed results will occur. In the next section we describe our analysis of the content of these protocol categories as a beginning in understanding these conditions.

Content-Node Analysis of Remark Categories

We assume that expertise represents a kind of operative knowledge that allows experts (in this case decision makers) to perform under a wide range of conditions in their domain. It is a capacity to achieve problem solutions, and not a property of behavior (see for example Johnson, Zualkerman, & Garber, 1987). Thus, the "correctness" of a decision is not a sufficient index of proficiency. Not only can right decisions be made for the wrong reasons and vice versa, but in real-world tasks it may be impossible to define what a "correct" decision is (Edwards, Kiss, Majone, & Toda, 1984). The goal of understanding how experts structure and represent knowledge is impetus for the rise in studies based on content analysis of protocols in recent years (e.g., Graesser, 1981; Hammond, Hamm, Grassia, & Pearson, 1984).

For this study, we have developed our own method of representing the content of the protocol remarks. The method was designed to illuminate commonalities in a way that would aid interpretation of Expert/Novice differences in relation to the RPD model. The method is based on node units which describe the content or topics of the remark categories. A node is defined as being present whenever remarks in at least two protocols are judged have referred to the same topic. The assumption is that for two individuals to comment on the same cues, make the same inference, or require the same action, the remark reflects some component of the underlying schemas that have been activated by the simulation task. Nodes are indicated by a word or phrase summarizing the remark topic.

The first step was to examine the content of the coded remarks and to organize them into conceptually meaningful sets. This was done by listing the content of all remarks in a given category for a set of protocols. Each protocol was then analyzed for whether it contained an instance of each remark

item. If two or more protocols had the same item, it became a node to be represented on the node graph for that probe point.

A decision had to be made about how best to display nodes graphically. One possibility was to organize the nodes into a temporal ordering that might reflect causal connections between cue nodes and action nodes. However, the temporal ordering of categories varied considerably and no mechanism was found for abstracting the temporal relationships. In the absence of theory that would generate a logical or natural ordering of the nodes, we decided to retain a simple list structure that would indicate the degree of overlap in the sample of protocols for the nodes. In other words, we wished to distinguish nodes which were mentioned in all or many of the protocols from those for which there was minimal overlap. This index may reflect the typicality or centrality of the node for the underlying knowledge.

Four classes of nodes were defined:

1) Cue Nodes = Cues and Cue-elaborations

2) Knowledge Nodes = General Knowledge, SOPs, Cue-deliberation, and Anticipation

3) Action Nodes = Actions, Action-elaborations, and Action-deliberation

4) Goal Nodes = Goals and Goal-elaborations

The content-node analysis was carried out on the two initial probe points (probe point 1.1 and probe point 2.1). The other probe point (probe point 1.5) represented a very novel situation for the FGCs we studied (another fire alarm was received while they were still engaged in fighting the apartment fire) and it did not seem to have sufficient overlap in the action items to be meaningful. The overlap in the situation assessment categories was also quite

slim, perhaps because it was a "later" decision and some important information was not repeated. Additional work is needed before we will be able to say how general the proposed method is.

In the node graphs (Figures 3-6), symbols are used to distinguish each of the node categories. The open symbols indicate that the node type was present in fewer than 5 of the protocols within the group (low density); hatched symbols indicate overlap on five to seven of the protocols (medium density); and filled symbols indicate overlap on eight to eleven of the protocols (high density). The side-by-side presentation of the Expert and Novice graphs facilitates noticing which nodes are absent for either group (indicated by a "?") and differences in the node densities. In the Cue graph, some of the nodes seem to fall into natural groupings that are also indicated. The label for the grouping is bracketed to indicate that it is based on our own grouping strategy rather than being tied directly to any features of the protocols themselves.

Many of the differences between Experts and Novices in the node graphs are based on very small samples, so care must be taken not to overemphasize any particular difference. What we are seeking are general patterns of contrast that can be used to generate hypotheses for future research or that can be meaningfully related to specific contextual variables.

Occasionally the node analysis revealed a "branch" in the nodes representing a two-choice alternative, usually of the form "X or not-X," so these nodes are shown linked together on the graph. Linked nodes are particularly interesting because they seem to reflect probabilistic inferences or option-selection processes that would not have been apparent in a single protocol. That is, a remark topic may indicate only the outcome of a decision

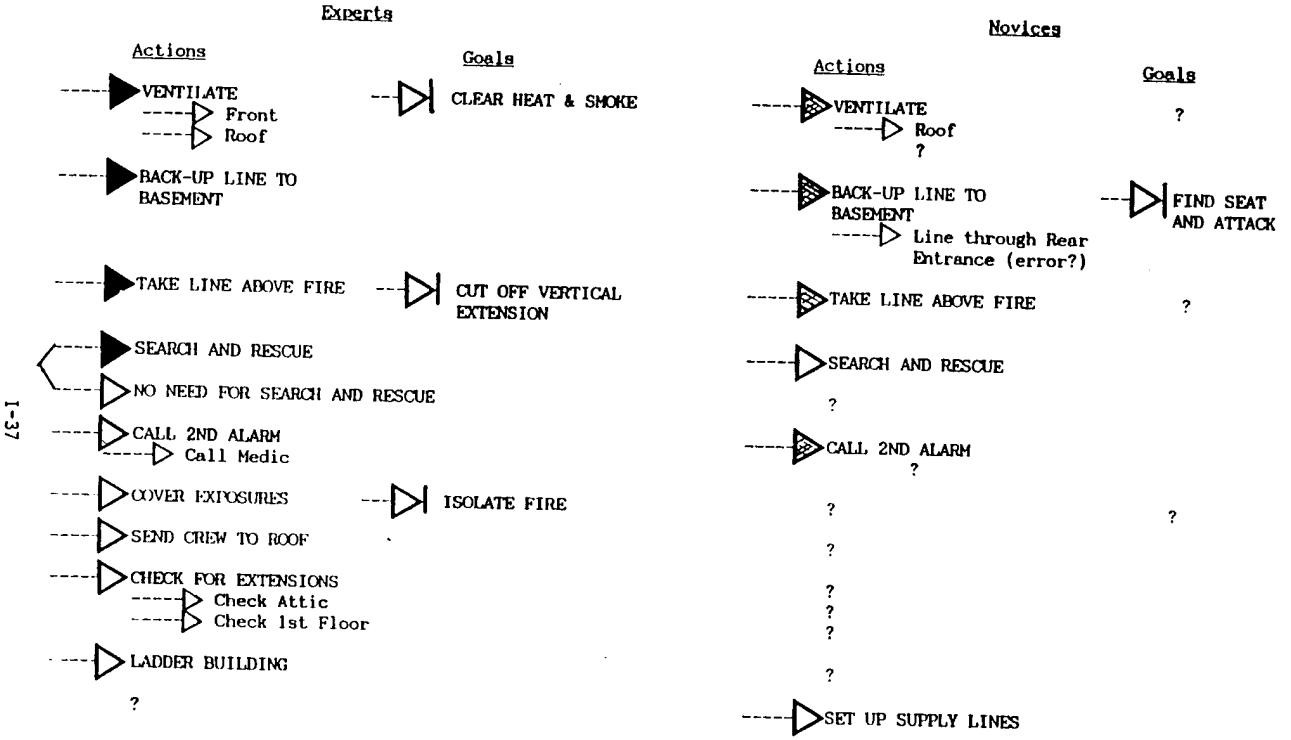
process in a single protocol. However, by seeing the complementary outcome that is reached by other individuals, one can make inferences about the underlying decision processes.

Probe Point 1.1

Interpretation of the nodes for probe point 1.1, reqiores knowing the context for the decision event. In this scenario the participant FGC hears an alarm to an apartment fire. After receiving details about the time of day, weather conditions, and responding units, he learns several facts about the structure that were known to the FGC of the incident -- that this is a poorly constructed building, that it has punk-board flooring and a second roof added to the original. He also hears a description of the entrances and number of apartments in the building. As he "travels" to the scene, he hears a size-up being dispatched from the first-arriving officer on the scene. He also hears the order for the engine crew to take an inch-and-a-half line into the basement. Upon "arrival" the participant is shown a slide depicting the scene as it would appear. The scene shows details of the apartment structure, heavy smoke is shown covering the building front and flames are shown escaping from a basement apartment. The first-arriving crews are shown exiting the basement. The participant FGC then hears a report from the first-arriving officer indicating that the intense smoke and heat are preventing crews from reaching the fire.

The Cue and Knowledge node graph for this probe point is shown in Figure 3. This graph represents a simplification of 49 separate Cue remark topics and 27 separate Knowledge topics.

The Cue nodes fall into categories of fire dynamics, the structural features of the building, the building's occupancy, personnel and equipment resources, and current tactics. Although these node categories are present



NOTE: The -----▷ and --▷| symbols represent action and goal nodes, respectively. Node density is indicated as follows: open symbols represent low-density nodes, hatched symbols represent medium-density nodes, filled symbols represent high-density nodes. The "?" indicates missing nodes that were found for the contrasting group.

Figure 3: Action and Goal Nodes for Probe Point 1.1

for both the Experts and Novices, there are some specific differences in the nodes that are worth considering. The Expert node labeled "look of the fire" refers to the intensity and scope of the fire suggested in the graphic illustration. There is no corresponding node for the Novices, who may be less able to notice the subtleties of the perceptual cues. The unspecified content of "look" is consistent with the idea that perceptual cues are among the most difficult to articulate. The Critical Decision method used in other studies in this series was designed to probe for clarification of the nature of such cues.

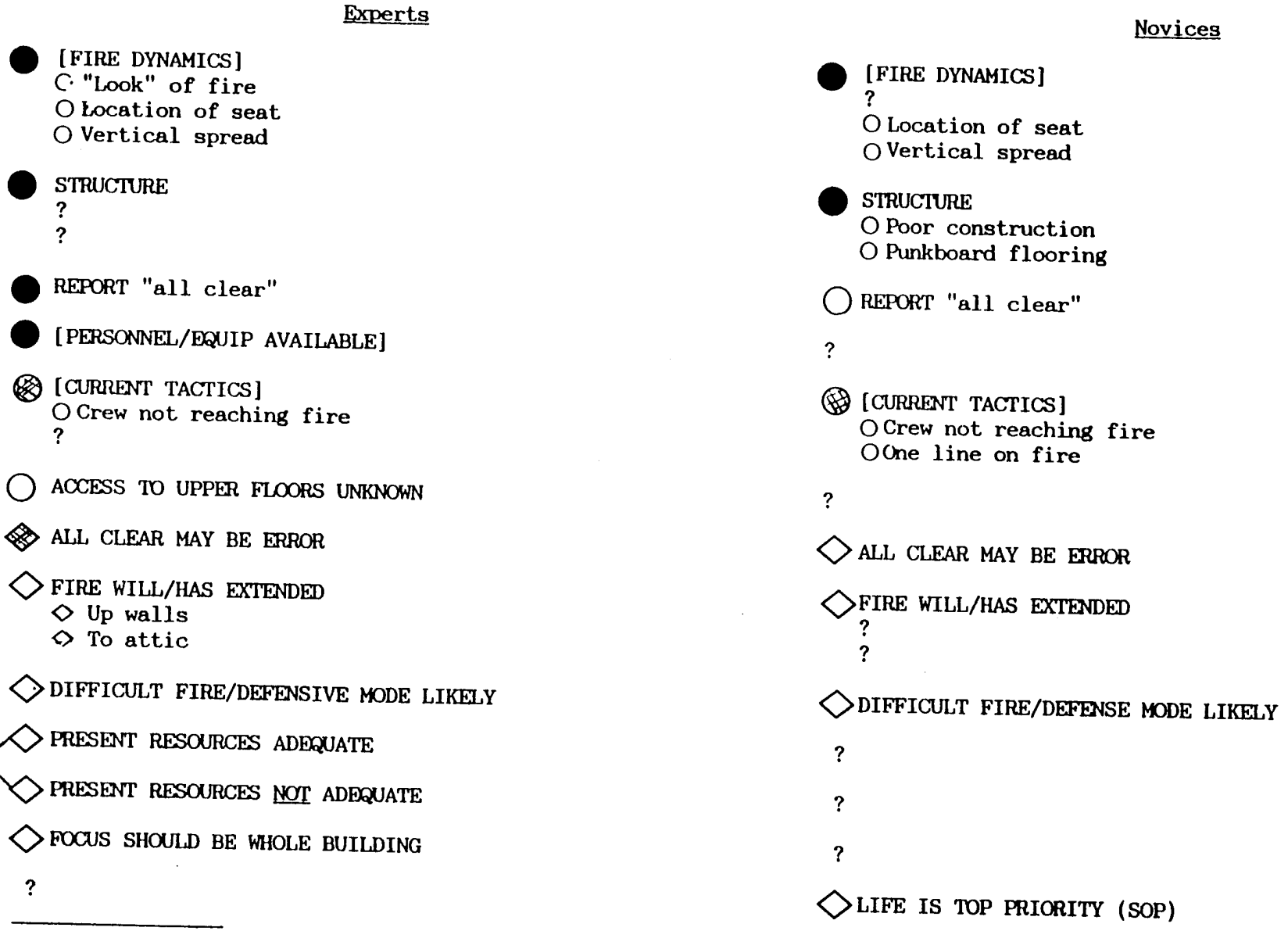
The absence of the structure nodes ("poor construction" and "punkboard flooring") in the Expert graph is interesting. These cues were reported in the audio portion of the scenario, but are not directly present in the graphic illustration of the scene. The analysis cannot tell us whether the presence of the nodes means the cue is more important to the Novices or whether it is simply unstated by the Experts for some reason. The Novices may need to remind themselves of previously stated cues more than Experts.

The Experts made more remarks pertinent to the availability of personnel and equipment, in line with the idea that they more easily adopt a "command" perspective that involves managing these resources. At a more global level, the Experts' graph has more high-density Cue nodes than the Novices, possibly indicating a more coherent schema underlying these topic statements. Alternatively, density differences could indicate more or less overall remark topics, but this does not seem to be the case here. Experts and Novices had roughly equal numbers of topics represented in the protocols (35 and 31, respectively).

Differences in the Knowledge nodes are among the most interesting because these are closest to the inferential and reasoning processes that might distinguish different decision outcomes. The nodes were extracted from a total of 27 topics in the Knowledge categories with an identical number of unique topics (20) for the Expert and Novice protocols. The most striking difference in these Knowledge nodes is the fact that a branch present in the Expert graph is missing in the Novice graph. This branch represents a dichotomy between a judgment that the presently available resources <u>are</u> or <u>are not</u> adequate. Recall that the high density Cue node indicating attention to resource availability was also absent in the Novice graph. It seems safe to assume that issues of resource allocation were more salient to the Experts.

Another node present in the Expert but not the Novice graph concerns the topic of "focus." Some of the Expert protocols contained a reminder to pay attention to the whole situation and not just the involved apartment. This node has the flavor of a maxim or general rule. The idea of learning to expand one's focus beyond the most salient or immediate problem has been made frequently in our interviews with firefighters. The single Knowledge node that is present in the Novice graph but not in the Expert graph, "life is top priority," also has this maxim or rule-like quality.

Turning to the Action and Goal node graph (Figure 4) one can again find several differences between the Experts and Novices. The Action nodes represent the overlap from a total of 35 distinct remark topics, 32 for the Experts and 22 for the Novices. For the first three nodes, representing the most overlap in the remark topics, the same nodes are present in both graphs, but there was somewhat more overlap among the Expert protocols. Although both graphs contain the "back-up line to basement" node, there is a



NOTE: The "○" and "◇" symbols represent Cue and Knowledge nodes, respectively. Node density is indicated as follows: open symbols represent low-density nodes, hatched symbols represent medium-density nodes, filled symbols represent high-density nodes. The "?" indicates missing nodes that were found for the constrasting group.

Figure 4: Cue and Knowledge Nodes for

tag on the Novices' node indicating the possibility of taking the line in the rear entrance. It was clearly stated in the scenario that there was no rear entrance, so this action indicates an error.

The Expert Knowledge graph contains a branch for whether to implement a search and rescue. The issue here is whether to accept the "all clear" cue given in the scenario. Although more of the Experts indicated that the search and rescue would be needed (see the high-density node), it was clearly more of an issue for these officers then for the Novices. Only half of this branch is represented in the Novice graph, and only by a low-density node. Four of the low-density nodes present in the Expert graph are absent in the Novice graph. All of these represent potentially important tactical considerations, whereas the single node which is absent from the Expert graph represents a tactic that has already been accomplished in the scenario.

There are three Goal nodes in the Expert graph and only one in the Novice, but these are all low-density nodes representing straightforward outcomes of the specified actions. These are general goals that, like standard operating procedures, represent fire ground tactics that would apply to almost any structural fire of this kind. Situation-specific goals are curiously absent.

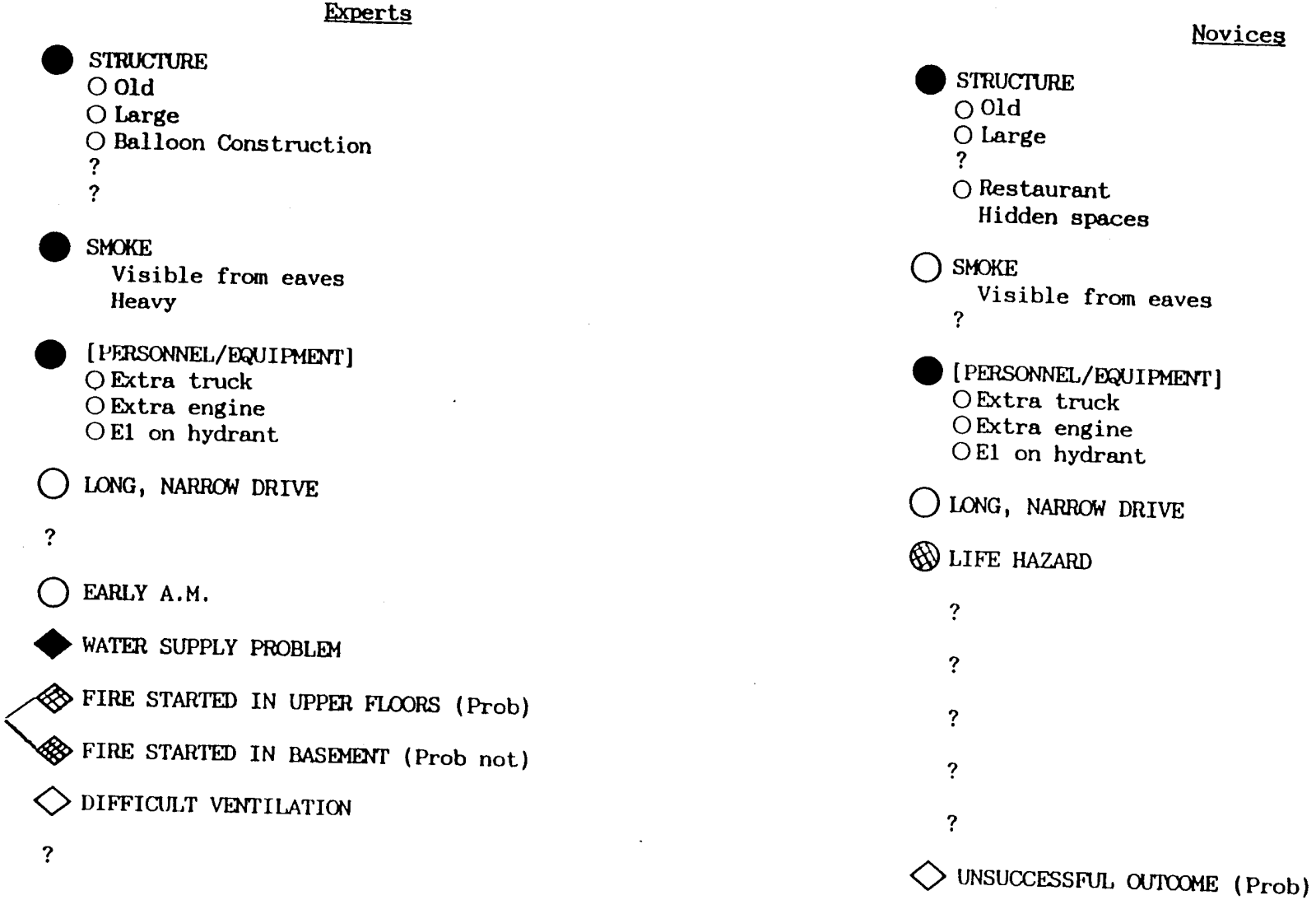
Probe Point 2.1

In this scenario, a report is received of a fire at a restaurant housed in a well-known historical building. The narrator supplies information about the time of day, the weather, and the building construction. The building is said to be brick and heavy timber, to have been renovated and enlarged with the additions being balloon construction. The responding units are indicated and the size-up is heard over the radio during the participant-FGCs transit to

the scene. The participant-FGC is shown a slide indicating the details of the scene as it would appear on his arrival and indicating the tactical maneuvers that were taken by the first-arriving officer.

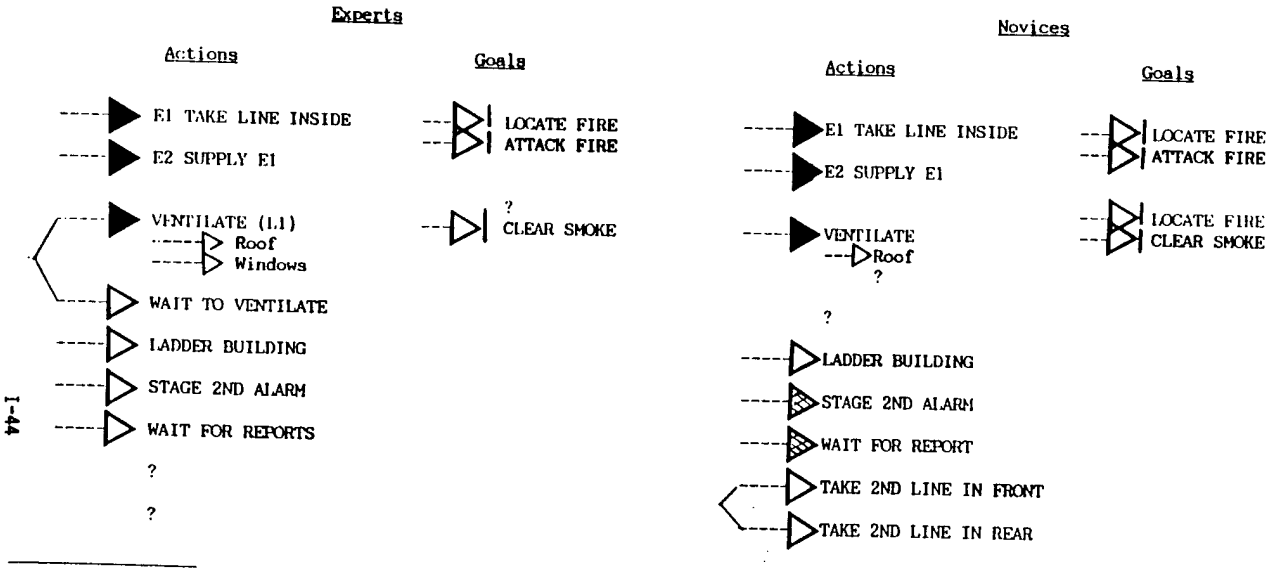
The Cue and Knowledge nodes for this probe point are shown in Figure 5. Roughly equivalent Cue categories are present in the Expert and Novice graphs, although there are several minor differences in the specific node topics. More striking are the differences in the Knowledge nodes. The high-density "water supply problem" node in the Expert graph is not even present in the Novice graph, showing a lack of attention to resource issues that was also seen in probe point 1.1. A decision branch related to the probability that the fire started in the upper floors versus the basement shows up in the two medium-density nodes indicated. Neither of these nodes is present in the Novice graph, nor is the node indicating that ventilation may be difficult. The only Knowledge node present in the Novice graph but not in the Expert graph was an appraisal indicating that the outcome of the incident would probably not be successful.

The Action and Goal node graphs for the Experts and Novices (Figure 6) are not as strikingly dissimilar as they were for probe point 1.1. Rather than the richer and more elaborated actions for the Experts for that probe point, here the numbers of distinct action topics for the two groups were almost identical (24 and 23 for Experts and Novices, respectively). The most notable difference is perhaps the presence of a branch in the Expert graph between the "ventilate" and "wait to ventilate" nodes. Knowing when to ventilate is one of the most frequently mentioned characteristics of expertise in this domain. The low-density branch in the Novice graph for the nodes "take second line in front" and "take second line in rear" that is



NOTE: The "○" and "◇" symbols represent cue and knowledge nodes, respectively. Node density is indicated as follows: open symbols represent low-density nodes, hatched symbols represent medium-density nodes, filled symbols represent high-density nodes. The "?" indicates missing nodes that were found for the contrasting group.

Figure 5: Cue and Knowledge Nodes for Probe Point 2.1



NOTE: The ----▷ and ---▷| symbols represent action and goal nodes, respectively. Node density is indicated as follows: Open symbols represent low-density nodes, hatched symbols represent medium density nodes, filled symbols represent high-density nodes. The "?" indicates missing nodes that were found for the contrasting group.

Figure 6: Action and Goal Nodes for Probe Point 2.1

not present in the Expert graph is hard to interpret, given that the second engine has not yet arrived. Experts may simply have been better at limiting their remarks to actions that would be taken at the time of the probe. This possibility can be examined when probe point 2.2 is analyzed in the near future.

No remarkable differences are evident in the Goal nodes. Again, the goals tend to be the general goals of good fire ground tactics and do not seem to illuminate the present situation or factors associated with expertise.

Evidence for Decision Strategies

The analysis of the remark topics presented thus far does not directly address the evaluation strategies described by the RPD model. Such strategies might only be discerned by considering the meanings of a series of remarks taken together, and in context. Possible clues to these strategies were noted on each protocol and examined separately from the remark frequency and content analyses. In addition, we tried to be sensitive to instances of analogue use, prototypes, and errors of judgment or interpretation.

The data relevant to these processes were disappointingly sparse. In the 66 protocols examined and coded for this study, we found only 11 instances of deliberated decisions, and a handful of instances of progressive deepening, specific analogues, prototypes and imagery.

Nonetheless, the examples that were found offer a tantalizing look at the cognitive processes and strategies that underlie command decision making. Of the 11 cases of deliberated decision making identified in the protocols, six were cases of serial decision strategies -- two involving cue-deliberation and four involving action-deliberation. In the remaining five cases of concurrent decision strategies, four involved Cue-deliberation and one involved Action-

deliberation. Excerpts from the protocols containing these deliberated decisions are presented in Tables 6 and 7.

Based on these few instances, we offer several highly tentative observations:

First, the instances of deliberated decision making do not appear to be accounted for by the simulated incidents themselves. They do not consistently occur at any particular point in the incident or in response to particular aspects of the situation represented.

Second, there does not appear to be commonality in what the FGCs are deliberating about (e.g., apparatus placement, where the fire is located, resource availability). There is virtually no overlap in the content of these deliberated decisions.

Third, examination of the serially deliberated decisions suggest that this strategy is activated when the FGC notes new or previously unnoticed cue information. Even when the serial strategy clearly involves action deliberation, it does not appear to occur because of some previously unrealized action possibility.

Recognizing that the absolute frequencies of occurrence are very low, we would note nonetheless that comparison of Experts' and Novices' data reveals directional differences in line with other Expert/Novice findings obtained in this program of research. Of the 11 deliberated decisions, 64% were obtained from Novice FGCs' protocols. Moreover, Novice FGCs' deliberated decisions more often involved concurrent than serial strategies (57% vs. 43%). This was not the case for the Expert FGCs, whose deliberated strategies more often involved serial than concurrent strategies

TABLE 6

Instances of SA Deliberation

DECISION STRATEGY	FGC EXPERTISE	ISSUE	PROTOCOL EXCERPT
Serial Deliberation	Novice	What resources are immediately available?	I've got engine 3, medic and rescue 2 does not have any hose off it. Wait a minute rescue 2 does too. Engine 3, the medic, and chief 2, whatever he was, I can send them down to investigate and see what is going on down at the nursing home. At the same time, I can have the call go in for mutual aid for the nursing home.
Serial Deliberation	Novice	How to ensure access for arriving apparatus?	I would have had the second incoming engine set up for supply and supply this engine. The aerial truck, gosh, depending on the location of the hydrant, this is going to block the possible entrance for other apparatus. I think this would have been a concern to make certain that the access was available for the other incoming apparatus. Once that was taken care of I would have waited for the report from my engine crew as far as if they found the fire.
Concurrent Deliberation	Novice	Is the structure safe for entry?	Obviously there has probably not been anyone in that building since one or two o'clock in the morning. So if a fire was going in there, it had a good chance to get started, good chance it started in the kitchen area and it has worked up through that balloon frame construction to the roof area. Heavy timber construction, obviously more stable than some of the others. The building is old, but it is a lot safer than things like structural steel. Pending right now, I am looking at the possibility of an interior attack.
Concurrent Deliberation	Novice	Do interior conditions mean back draft is likely?	You know, if that line, and another line and the line up here can't handle it within like 10 or 15 minutes and we know there are no people in there, we are going to have to surround it & drown it. Hope it doesn't back draft on them, but I don't think it is going to, it looks like it is already venting itself; but I don't have any idea what is like up here. That is why I wanted to send a truck crew inside, that is why I wanted to send one to the roof to open up.
Concurrent Deliberation	Novice	Where is life hazard greatest?	We don't have a life safety problem as such in here. We have to dwindle our crews down which is going to make it more dangerous to operate, we don't have the resources here but we are going to have to marshall as full a response as possible to that nursing home, because we have a life safety hazard which supercedes this building. In short, we are at the point now where we are more willing to write this building off and attend the problem of a life safety problem down the street.
Concurrent Deliberation	Expert	Where is the seat of the fire?	From the view of the building, it would appear that the fire has probably started somewhere on the second floor, it is not necessarily iron clad, it could have started somewhere in the back and has proceeded up. But there is heavy smoke and contamination, which would indicate that the fire is already in the attic and is banking down and coming out the eaves. I imagine this is probably smoke coming out the windows here, so that is probably involving in the second floor and already banked up into the roof.

TABLE 7

Instances of Action Deliberation

DECISION STRATEGY	FGC EXPERTISE	ISSUE	PROTOCOL EXCERPT
Serial Deliberation	Expert	What resources to release to nursing home fire.	I am going to release everyone in staging to respond to that other fire. If it is just down the street, they can get there in just a couple of minutes or so. Meanwhile the dispatchers, or whomever, can get, begin their mutual aid request. I will then consider if there is anything else I can release here. I don't believe that I want to release any of the fire fighters that are committed to the scene just yet. I am going to keep the ones that I have on the scene, the ones I have staged I will release to the nursing home. There is another chief here, I will let him go to take command of that fire. That's it.
Serial Deliberation	Expert	How to put out fire.	At this point in time, of course, you know something that sort of gets me about this scenario is the fact that we only have 750 gals. of water. Because in order for me to put master streams of into operation, I am going to have to have more than 750 gallons of water a minute. If I put foam into that building, chances are the way this fire is going the heat is going to break the foam down. There is no way I am going to be able to do any good with foam. It is very highly doubtful so I think I would go with a deck gun first. But like I say, I don't know how much water is available to me here.
Serial Deliberation	Expert	Need for second alarm.	The fire is in the basement, it is hard to tell, flames coming from just this one window in this one area. Probably would call for extra equipment, get a second alarm there, depending on the--I do not know if I would call a second alarm with just that amount of fire showing. Get another line down in the basement, try to get it ventilated so crews could get back in there.
Serial Deliberation	Novice	Where to attack fire.	It is probably good line placement to try to get to that fire in that situation, we know it has a combustible floor. If you take that line in on the first floor anywhere, you are just going to be shooting in the dark. Sometimes the best way to attack a fire, given that line between your fire and life hazard, you go right to the fire.
Concurrent Deliberation	Novice	Who to appoint interior sector.	I have two truck companies. I want one truck company to ventilate I want the second truck company to go in and become my inside sector commander. I already have a chief on the scene (I'm not used to that). I'll assign that chief as inside sector commander, if he in, or I will use a captain, but I want to know what is going on inside.

(75% vs. 25%). This does offer support for the notion that Novices are more likely than Experts to employ deliberation in decision making.

Summary and Conclusions

The present investigation was carried out as the final study in a series of interrelated studies of command-and-control decision making. As such, it has been guided in conceptualization and in choice of methods by the rich and often provocative data obtained in those previous efforts. The results of our earlier studies -- carried out in a variety of natural settings -- had raised a number of questions about the validity and utility of standard decision models for understanding decision behaviors in the time-pressured, high risk and complex situations represented by command-and-control. These studies served as the basis for development of our RPD model, which has continued to guide our thinking, and which we see as offering an alternative conceptualization to standard decision models that emphasize option generation and evaluation.

As results from these several studies accumulated, and as we developed and refined our Critical Decision method, it became clear that we needed to examined certain key issues under more controlled conditions. For example, the studies offered repeated and compelling suggestions of the importance of situation assessment processes for decision making, especially as it is carried out by highly proficient decision makers. But without the means to know more precisely what information was available in a situation, and to present the same set of situation features and pivotal events within a given incident, we were left at an uncomfortably speculative level in terms of model testing and development.

The simulations developed for this study were designed to reproduce as closely as possible the experience of fire ground command during an actual incident. The intensity of involvement we observed in our participants, as well as the sheer amount of talk they generated in response to the simulations, indicates that we were able to represent the key elements of an actual incident with enough authenticity to engage these FGCs' knowledge and decision processes. The "think-aloud" method also offers evidence of the cognitive content and processes that underlie decision making, without the potential response biases introduced by the guided probes and semi-structured interview methods of CDM.

The remark frequencies are assumed to provide an index of the relative attention given to different aspects of a decision event. These data substantiated the critical role that situation assessment in command decision making. In addition, convergent evidence was provided for the hypothesized relation between these processes and relative degrees of domain skill.

For Expert FGCs, remarks related to situation assessment consistently exceeded those given to action assessment, while the opposite was true for Novices.

The content node analysis of the protocol remarks was undertaken in order to investigate the nature of the schema that are presumed to underlie the commanders' decisions. The method allowed several interesting features of the protocols to be illuminated. The node graphs for the Experts tended to be richer and more elaborated than the Novices', but the qualitative differences in specific nodes were the most interesting. The Experts' graphs revealed different issues being addressed on the basis of a similar set of cues. These

issues involved the very types of causal inferences that one would expect to be associated with more highly developed domain knowledge.

One unexpected finding that emerged from the node-graphing procedure was the ability to detect node branches that were not readily discernible in the individual protocols. In these cases, it is assumed that a split in the "solutions" reached about some ambiguity in the situation reflect an underlying decision process that was only infrequently articulated in the protocols. If this assumption is correct then we have demonstrated that the decision "space" is not the same for individuals at different levels of skill and experience. This is quite different from the standard view which treats differences in decision making in terms of assigning probabilities and values to a set of pre-defined alternatives. We hope to pursue the implications of this idea in future research.

We were also surprised at the relatively low frequencies of goal remarks, by either Expert or Novice FGCs. Moreover, it seemed to us that when goals were discussed, there was often a nebulous, generic quality to them that was quite different from the situation and action assessments being offered. We have been struck by the consistent difficulty we have had across this series of studies in getting people to talk informatively about goals. It has begun to occur to us that the problem may not be one of inadequate research methods. Rather, people may have an extremely difficult time thinking/talking about goals independently of the actions they supposedly guide or the situations they are intended to address. Given the many decision support systems that are organized around goal specification and clarification, we had expected to find evidence of the utility of goals in naturally occurring decision making.

We think the failure to find such evidence in this study or in others in this series is intriguing, and plan explore this aspect of decision making further.

Finally, findings from the present study indicate that evidence on cognitive processes and decision strategies is not often revealed in "think aloud" protocol data. The Critical Decision method was originally developed to elicit such information in the context of retrospective reports of actual events. Pairing the simulation format with CDM would appear to offer a powerful research tool for studying this aspect of decision making under more controlled conditions. An initial study using this approach is presently under way as part of another contract (MDA903-86-C-0170) and the results look promising. Nonetheless, when FGCs' protocols did contain evidence of cognitive process, it was clear-cut and compelling. The protocols provide supportive evidence for the RPD model. They indicate that people do use serial decision strategies, and that Novices rather than Experts are likely to deliberate during decision making.

At a more general level, the protocols have led us to reconsider certain aspects of the RPD model. The supposition that decision making occurs as the outcome of a production rule: if CONDITION, then do ACTION suggests a linearity to decision events. That is, the decision maker assesses a situation, recognizes it as familiar, and proceeds to act -- guided in his or her choice of what to do by that sense of familiarity. We are increasingly less comfortable with the ordered quality of this conceptualization, while continuing to adhere to its recognitional components.

We would acknowledge that division of decision making into separate, independent situation assessment and action components is a useful convention, especially for comparing the RPD model with other models of decision making.

Nonetheless, we are increasingly convinced that the division is an artificial one. We did not find FGCs talking first about the situation and once their concerns about the nature of the situation and been satisfied, only then moving on to action remarks. Rather, their attention seemed to move back and forth between elements of the situation and the actions intended to address them. They are constantly assessing BOTH situational factors and action factors -- recognizing categories of each and matching one to the other as needed.

References

Anderson, J. R. (Ed.). (1981). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Barlett, F. C. (1932). Remembering. Cambridge, England: Cambridge University Press.

Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the 'bias heuristic'. Acta Psychologica, 50, 201-252.

Brezovic, C. P., Klein, G. A., & Thordsen, M. (1987). Decision making in armored platoon command (KATR-858(B)-87-05F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA. AD# A231775

Calderwood, R., Crandall, B., & Klein, G. A. (1987). Expert and novice fire ground command decisions (KATR-858(D)-87-02F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327) for U.S. Army Research Institute, Alexandria, VA.

Christensen-Szalanski, J. J. J., & Beach, L. R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. American Psychologist, 39, 75-78.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

de Groot, A. D. (1965/1978). Thought and choice in chess (2nd ed.). NY: Mouton.

Dreyfus, H. L., & Dreyfus, S. E. (1986). Mind over machine: The power of human intuitive expertise in the era of the computer. NY: Free Press.

Edwards, W., Kiss, I., Majone, G., & Toda, M. (1984). A panel discussion on "What constitutes 'a good decision'?" In K. Borcherding, B. Brehmer, C. Vlek, & W. Wagenaar (Eds.), Research perspectives on decision making under uncertainty, pp. 5-27. (Selected Proceedings of the Ninth Research Conference on Subjective Probability, Utility and Decision Making, Groningen, August-September 1983). NY: Elsevier.

Ericsson, K. A. , & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Flanagan, J. C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Fleiss, J. L. (1981). Statistical methods for rates and proportions. (2nd ed.) NY: Wiley.

Gentner, D., & Stevens, A. L. (Eds.). (1983). Mental models. Hillsdale, NJ: Erlbaum.

Gettys, C. F. (1983). Research and theory on predecision processes (TR 11-30-83). Norman, OK: University of Oklahoma, Decision Processes Laboratory.

Glaser, R. (1981). The future of testing. American Psychologist, 36, 923-936.

Graesser, A. C. (1981). Prose comprehension beyond the word. NY: Springer-Verlag.

Hammond, K. R. (1987). Toward a unified approach to the study of expert judgment. In J. L. Mumpower, L. D. Phillips, O. Renn, and V. R. R. Uppuluri (Eds.), NATO ASI Series F: Computer and Systems Sciences: Vol. 35. Expert Judgment and Expert Systems (pp. 1-16). Berlin: Springer-Verlag.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1984). The relative efficacy of intuitive and analytical cognition. Boulder, CO: Center for Research on Judgment and Policy.

Holding, D. H. (1985). The psychology of chess skill. Hillsdale, NJ: Erlbaum.

Johnson, P. E., Zualkerman, I., & Garber, S. (1987). Specification of expertise. Int. Journal Man-Machine Studies, 26, 161-181.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 127-251.

Klein, G. A. (In press). Recognition-primed decision. In W. Rouse (Ed.), Advances in Man-Machine Systems Research, 5. Greenwich, CT: JAI Press, Inc.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. Proceedings of the Human Factors Society 30th Annual Meeting, 1, 576-580. Dayton, OH: Human Factors Society.

Klein, G. A., Calderwood, R., & MacGregor, D. (In press). Critical decision

method for eliciting knowledge. IEEE Transaction Systems, Man and Cybernetics.

Larkin, J. H. (1981). Enriching formal knowledge: A model for learning to solve textbook physics problems. In J. Anderson (Ed.), Cognitive skills and their acquisition (pp. 311-334). Hillsdale, NJ: Erlbaum.

Lopes, L. L. (1987). The Rhetoric of Irrationality. Paper presented at Colloquium in Mass Communication, Madison, Wisconsin, November. (Currently under revision.)

Neisser, U. (1976). Cognition and reality. San Francisco: Freeman.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Schank, R. (1986). Explanation patterns. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1986). Inside intelligence. American Scientist, 74, 137-143.

Taynor, J., Klein, G. A., & Thordsen, M. (1987). Distributed decision making in wildland firefighting (KATR-858(A)-04F). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-85-C-0327 for U.S. Army Research Institute, Alexandria, VA. AD# A225413

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. Psychological Bulletin, 2, 105-110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 4, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. Science, 185, 1124-1131.

APPENDIX A

PROBE POINTS

EXPERT FGC - PROTOCOL #06

Probe Point 1.1

Okay, they tell me everyone is out. I am not sure if that meant out of that apartment or out of the building. I am going to assume that there could still be someone in the building itself, I am concerned with the whole building, not with just the apartment.

I want to ventilate the building so that those crews can get into that particular apartment. I will probably ladder the front of the building; I will have crews at the rear of the building with ladders; I want a crew to go in and check the rest of the building at this end and search and rescue if needed down in that area.

If the ventilation can help, and they get the fire, that is fine. In the meantime, I am going to call for some additional medics, I only had one sent, in a fire like this I want more on the scene. If I don't need them I can release them quickly enough. I will probably stage some extra apparatus until I am sure we can contain this fire in this apartment. If it doesn't appear that the ventilation of the fire is going successfully, then I can use that equipment, it will be on hand.

I will probably send a crew to the roof, just to stand by; but I don't want them doing anything just yet. I also will have some extra lines, they have an inch and three-quarter, I want another line at least up above the fire and also between the fire and the unexposed portion of the building. That's all.

EXPERT FGC - PROTOCOL #09

Probe Point 1.1

Okay, they said everyone was out, by that I would be assuming that they mean this whole building had been cleared.

The fire is in the basement. It is hard to tell, flames coming from just this one window in this one area. Probably would call for extra equipment, get a second alarm there, depending on the--I don't know if I would call a second alarm with just that amount of fire showing. Get another line down in the basement, try to get it ventilated so crews could get back in there and also try to get crews up on the other floors to conduct a search to be sure everyone is out. The two minutes that the crews have been on the scene, it is unlikely that they would have been able to complete a real search of a building, so second in crew probably to ventilate in that area of the fire. Get another backup line, try and drive, with two lines together, try to get down to the fire area. When other people got there, have a third line taken to the second floor, the floor above, check for fire extension vertically. With the poor construction, if it has been burning long enough to be popping out the window like that, it is quite likely that it has been spreading upward. I would like to have at least two lines in the basement, one line on the second floor and people searching the second and third floor. Old buildings are prone to have open vertical shafts, fire can spread. If that is the case, then wait for initial reports from the other crews searching the building. If that would be the case, I would probably complete the second alarm.

Still, looks basically like an offensive fire, just need to clear it out a little bit to have people make the attack. Looks pretty simple right now. Again, I can't tell from the picture how the stairs to the upper floors would run. Are there separate stairs for each set of apartments there? They might have said, but I missed it. Protect the stairs, make sure they stay clear, make sure we got everybody out.

Ventilation to get people to the fire to put the fire out, eliminate the hazard and also to remove the people from the hazardous area. Pretty much simultaneous operations. We will just assign the crews and wait for further reports.

Probe Point 1.1

Oh, gosh, picturing the whole fire scene, picturing the size of the building which creates a problem. Immediately I thought that with that large of a building, I would have called additional equipment and they said it was occupied. Poor construction is another, being aware of that, I would have another reason for calling additional equipment.

I do not think--I wouldn't have permitted police crews inside the building, for just that reason. They have no equipment. They would be exposed to who knows what. I would have probably been setting up the other engine while they were in route. I would have had them spot certain areas, supplied the actual first apparatus on the scene. The ladder truck, there again like I said, I would have probably called for additional equipment being it was occupied. Would not have permitted the police in there, I would have had my personnel evacuate the building, using their protective equipment and breathing apparatus which the police do not have.

Probably, when I have gotten to the conclusion that our crews say they cannot get to it, I might after the actual evacuation of the building has been completed, set up for some type of ventilation. Your priorities would be just on the scene, would be your life. Concern yourself with the people first, then after that was maintained, or taken care of, then I would attend to the actual fire problem. But that large of a building, with our standards as we do it now, would automatically be to call for extra equipment and probably medic crews also. I can't remember if they said they had medic crews dispatched to this or not.

[INT: One medic]

Okay, I would have probably asked for additional medic crew being that it looks like there might have been as high as twenty people in that building. Then you just deal with the problems as they occur. But my initial size up of the situation would have been as such.

NOVICE FGC - PROTOCOL #12

Probe Point 1.1

Well, I have to assume that because Rescue 1 laid the line down the basement, I have to assume it is like an engine, probably with just additional tools than what a normal engine crew would carry. The heavy type tools, power tools to effect a rescue and that type thing. Other than that, I assume from the way Rescue 1 was used that it is like an engine only with additional equipment.

I have fire coming out of the basement out of the middle of the building. I wish I knew this building a little bit better. Wait a minute, they said there were two apartments in the basement.

Now I got police doing rescue in the rest of the building. I don't see any other doors to the building so I am going to have to assume there is a door at this end and maybe one at the other end. So we have doors up there.

The second crew on the scene, I don't remember who that was, I am going to try to get into the apartment or apartments in the neighborhood of the fire to try to stop extension that way. Third crew I am going to send down behind rescue 1 to back them up and try to gain entry to the fire area. It is a long haul down that basement. I don't know if both the apartments are on one side or if there is one apartment on each side in the basement. Anyhow, those two crews, working together, should be able to get into the area of the fire. The forth crew, well really it is going to be my fifth crew--

The second crew really, go back a little bit, first job is to make sure rescue 1 has water. Okay, then the next available crew will go to the area around over the top of the fire, if at all possible, try to make sure there is no extension. The next crew, which is really my fourth crew, send down the basement to assist rescue 1. Truck crew, I am going to have trying, to look and see if there is any other way they can ventilate that basement, try to get some more smoke out of there. Also, I am going to have maybe one or two guys out of the truck crew start upstairs and check after the police to make sure everybody is out.

Then, let's see, then I've got one medic, okay. I am going to--I've got everybody assigned so I am going to stage maybe two more engines and a truck to have just in case everything turns to dirt on me. I want to have somebody else there that I can assign quickly if I need them. With that many crews and with fire limited the way it is, I really should not have a whole lot of trouble.