ARI Research Note 96-14

The Development of Knowledge Elicitation Methods for Capturing Military Expertise

Gary A. Klein
Klein Associates, Inc.

Fort Leavenworth Research Unit
Stanley M. Halpin, Chief

January 1996

19960418 085

United States Army
Research Institute for the Behavioral and Social Sciences

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

James W. Lussier

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, January	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) August 1986-October 1988

4. TITLE AND SUBTITLE The Development of Knowledge Elicitation Methods for Capturing Military Expertise	5a. CONTRACT OR GRANT NUMBER MDA903-86-C-0170 5b. PROGRAM ELEMENT NUMBER 0605502A
6. AUTHOR(S) Gary A. Klein	5c. PROJECT NUMBER M770 5d. TASK NUMBER 114 5e. WORK UNIT NUMBER S06
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Klein Associates, Inc. 582 E. Dayton-Yellow Springs Road Fairborn, OH 45324-3987	8. PERFORMING ORGANIZATION REPORT NUMBER KATR-863-88-08Z
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-RK 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Research Note 96-14

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

COR: Rex Michel

14. ABSTRACT *(Maximum 200 words)*:

The goal of this SBIR Phase II was to formalize and evaluate a new method of knowledge elicitation, the Critical Decision Method (CDM). A number of studies were conducted using the CDM, and a formal rationale, description, and set of guidelines was developed. Additional work demonstrated the reliability of the method. The CDM was shown to be applicable in Army command-and-control settings. Additional knowledge elicitation methods were developed for team decisionmaking during command-and-control training exercises and for the evaluation of decision support systems. Taken together, the projects performed under this contract exemplify a new discipline of cognitive engineering. They provide a set of methodologies for eliciting and codifying expert domain knowledge to generate systems that improve decisionmaking.

15. SUBJECT TERMS

Knowledge engineering	Expert systems	Decisionmaking
Protocol analysis	Decision support	Command and control

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	22	

ACKNOWLEDGMENTS

I would like to acknowledge Marvin L. Thordsen for his work in organizing, supervising, coordinating, and conducting the various components of this project.

Also, I want to acknowledge the work done by Leslie Whitaker, Roberta Calderwood, Christopher Brezovic, Saul Young, Janet Taynor, Beth Crandall, Timothy Baynes, Sterling Wiggins, and Joseph Galushka on individual studies cited in this report.

Finally, I wish to thank our Contracting Officer's Representative, Rex Michel, along with Dr. Stanley Halpin and Major Edward Sullivan, for all of their enthusiasm, support, and help throughout all phases of this project.

THE DEVELOPMENT OF KNOWLEDGE ELICITATION METHODS FOR CAPTURING MILITARY EXPERTISE

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

THE DEVELOPMENT OF KNOWLEDGE ELICITATION METHODS FOR CAPTURING MILITARY EXPERTISE

Introduction

This SBIR Phase II project (Contract MDA903-86-C-0170) was undertaken in order to develop and evaluate methods applicable to the emerging field of cognitive engineering. Cognitive engineering treats knowledge as a resource. Just as equipment, capital and manpower are all resources, the knowledge acquired by personnel through years of experience in an organization is also a valuable resource. Cognitive engineering seeks to extract, codify and apply this knowledge in ways that benefit an organization and prevent the kind of drain on this resource that can result when experienced people retire or move into other positions.

Three cognitive engineering methods were developed and refined within this project: the Critical Decision Method (CDM), the method of team decision mapping, and a Decision Support Quotient (DSQ) method for evaluating the performance impact of decision support systems.

Cognitive Engineering

The appreciation of expertise has grown during the last 10 years through the development of expert systems for capturing some of the knowledge of experts. In brief, expert systems are a delivery system for coding and applying expertise. Earlier systems had attempted to apply general principles of problem solving; the inadequacies of these systems showed the importance of concrete domain knowledge and helped to usher in the expert systems that focused on domain knowledge. There are other potential delivery systems, so the field of cognitive engineering is not restricted to expert system technology. Norman and Draper (1986) and Mancini, Woods, and Hollnagel (1987) were the first to identify the field of cognitive engineering which was intended to cover methods and strategies for building intelligent decision support systems that use the knowledge of domain experts.

The process of building this discipline is analogous to petroleum engineering. Petroleum is a vital resource, but 150 years ago there no good uses for it and it was not valued. Now we are very concerned with identifying sources of petroleum, extracting it economically, processing it for different applications, and making those applications. Likewise, before the development of expert systems, the extraction of expert knowledge was of little importance. Now cognitive engineering is a burgeoning field with many new methods and applications.

There are four parallel aspects of cognitive engineering. First is to identify the sources and types of knowledge. Second is to economically elicit the knowledge. Third is to codify the knowledge.

We can say that cognitive engineering is the discipline of developing systems to organize and use the experts' content knowledge for doing work. Expert systems were the first focus of cognitive engineering. They required developers to identify the type of knowledge needed, to elicit that knowledge, to codify it, and represent it in a knowledge base so that the expert system could be used to guide decision making. Earlier attempts to build decision aids were oriented around processes (e.g., Decision Analysis, Bayesian statistics, Multi-Attribute Utility Theory) rather than content knowledge. Expert systems were the first systems built around domain content knowledge.

Expert systems have their limitations. They generally require an explicit track of all rules down to fundamental assumptions, yet many work settings need only to capture the difference between experts and competent performers in order to upgrade the performance of the competent workers. Expert systems often reflect a single expert because it is so hard to reconcile inconsistencies between experts, yet most applications are concerned with general expertise and not with capturing any one individual's knowledge. Another problem is that highly expert performance often depends on tacit knowledge (Polanyi, 1966) which is most difficult to capture in an expert system.

There are other delivery systems for cognitive engineering besides expert systems. One is the Case-Based Reasoning approach (Kolodner, Simpson, & Sycara-Cyranski, 1985) which is oriented around specific concrete cases and analogues as a basis of understanding a domain and using the precedents to make decisions (Ashley & Rissland, 1987; Klein, Whitaker, & King, 1988). A second is the synthesis between expert systems and decision support systems, in which knowledge bases are merged with algorithms and organized to enable the system user to make the decisions. It is inevitable that other concepts for delivery systems will arise now that we have seen what expert systems can and cannot do. The common theme of the delivery systems will be to support the decision maker by making prior experiences and knowledge available for guidance.

There are many potential applications for cognitive engineering. The obvious first one is to build expert systems and decision support systems. A second is to assess a domain for feasibility of building an expert system or decision support system. A third application is to select an efficient knowledge elicitation strategy in order to reduce the cost of developing the system. (This includes the method of eliciting the knowledge and the choice of domain areas on which to focus attention.) A fourth is to evaluate the adequacy of the expert system or decision support system performance. A fifth is to provide feedback for training of decision making at an individual or a team level. A sixth is to build a corporate memory for an organization.

Within this cognitive engineering framework, we can identify a set of needed techniques. Following the discussion presented above, these fall into four categories:

(a) Techniques for identifying the expertise, including identification of the domain experts, and for identifying the type of knowledge needed (e.g., analytical rules, perceptual discriminations, analogues, methods for adjusting analogues).

(b) Techniques for eliciting the knowledge. Hoffman (1987) has reviewed a set of existing methods, such as an analysis of familiar tasks the expert performs, structured and unstructured interviews, limited information tasks, constrained processing tasks, along with the method of tough cases. Other methods include the use of the Kelly Rep Test (Boose, 1984). In 1987, the International Journal of Man-Machine Studies ran a series of special issues on knowledge acquisition for knowledge-based systems in which manual and automated strategies were discussed.

(c) Techniques for codifying the knowledge. The prime example is the set of methods for building knowledge bases for conventional expert systems. Recent developments include Abrett and Burstein's (1988) Knowledge Representation Editing and Modeling Environment (KREME) which is being designed for large scale knowledge-based management.

(d) Techniques for applying the knowledge. Here, the prime example is the technology of expert systems, knowledge-based systems and decision support systems.

Summary of SBIR Phase II Contract

Our work fell into category (b) above, the development and evaluation of techniques for eliciting expert knowledge.

First, we wanted to extend and formalize the Critical Decision Method (CDM) for knowledge elicitation. We had developed the CDM in Phase I to study the expert decision making of fire ground commanders (Klein, Calderwood, & Clinton-Cirocco, 1988). Only recently have behavioral scientists become interested in content-oriented knowledge elicitation methods, which is why there are so few available, and why the formalization of one as promising as the CDM is so important.

Second, we wanted to evaluate the Critical Decision Method. We wanted to determine its reliability and validity.

Third, we wanted to test its application to military domains, specifically those involving Army command-and-control battlefield decision making. We also wanted to be alert to the possibility of developing additional knowledge elicitation methods.

Fourth, we wanted to extend the knowledge elicitation methods to cover specific applications to training and decision support systems.

Formalization of the Critical Decision Method (CDM)

The CDM is a retrospective interview strategy that applies a set of cognitive probes to actual non-routine incidents that require expert judgment or decision making. Once the incident is selected, the interviewer asks for a brief description. Then a semi-structured format is used to probe different aspects of the decision making process. Specific procedures have also been developed for analyzing the data.

Although the CDM shares many features with other interview methods, especially those related to Flanagan's (1954) Critical Incident technique, taken as a whole it offers some specific features that distinguish it from these and other knowledge elicitation strategies. (a) It focuses on non-routine cases because these are usually the richest source of data about the capabilities of highly skilled personnel. Therefore, it results in more efficient data collection sessions. (b) It focuses on concrete cases, specifically recalled incidents, rather than on general procedures. These procedures can later be inferred and validated but it is important to maintain the context of the episode to be sure of capturing nuances. (c) It relies on a set of cognitive probes about cues, inferences, strategies, and options that were selected as well as those rejected. (d) It uses semi-structured probing, to avoid the inefficiencies of unstructured interviews while ensuring that the interviews move in a useful direction. The CDM is a protocol analysis method. Concurrent protocol analysis is difficult to gather from actual experts working on difficult cases, especially if there is time pressure and life and death responsibility. The CDM avoids these limitations by probing about previous events. The CDM reduces problems with memory adequacy by concentrating on concrete, vividly experienced events.

We were interested in the CDM for several reasons. One is its effectiveness. By focusing on nonroutine events it goes directly to the special knowledge that experts have. This knowledge may not come out in a study of routine events where expertise is not needed and competent performance is sufficient. A second advantage is its efficiency, since the CDM directs interview time to those cases that require the most expertise and spends little time on mundane issues. Hoffman (1987) has demonstrated the relative efficiency of using such tough cases for eliciting rules for expert systems. A third advantage is that the CDM seems effective for getting at tacit knowledge. Much of the information obtained deals with cues, perceptual discriminations, and ways of assessing situations.

There is a wide variety of cognitive probes that we have used in the CDM, and these are listed in Table 1. The selection of probes for a given study depends on the goals of the study; it would be too cumbersome to use all of these probes in the same session.

The CDM was initially described by Klein, Calderwood, and Clinton-Cirocco (1988), as part of a description of the Phase I SBIR activities. Since the Phase I, we have employed the CDM in a number of other studies (Brezovic, Klein, & Thordsen, 1987; Calderwood, Crandall, & Klein, 1987; Taynor, Klein, & Thordsen, 1987) and have gained this additional experience with it. We have used it to trace the decision strategies of proficient personnel working under time stress (Klein, 1987), and to examine the expertise of computer programmers (Crandall & Klein, 1987).

This experience was synthesized in two papers that attempted to formalize the CDM. The first of these was a Technical Report for the sponsoring organization of the Phase II work, the Fort Leavenworth Field Unit of the Army Research Institute (MacGregor & Klein, 1988). The second was an article in the IEEE Special Issue on Knowledge Engineering (Klein, Calderwood, & MacGregor, 1989).

In these two papers we synthesized the various CDM methods we had used, presented the rationale for different aspects of the strategy, described the cognitive probes and their information value, and presented guidelines for the use of the method.

Evaluation of the CDM

It is very important to evaluate the knowledge elicitation methods being used in order to gauge the reliability and validity of the information gathered. The literature reveals little work done on evaluation issues. We therefore felt it was imperative to assess the CDM.

Table 1

CDM Probes

CDM Probes	Forms of Knowledge				
	Structure	Perceptual	Conceptual	Analogues	Prototypes
Decision Point Options		X			
Cues		X	X		
Causal Factors		X	X		
Goal Shifts					X
Analogues				X	
Errors	X	X	X		X
Hypotheticals	X		X		
Missing Data			X		
Imagery		X			
Task Analysis	X				

We performed two studies of reliability. One as part of the Calderwood, Crandall, and Klein (1987) study and the other as part of the Taynor, Klein, and Thordsen (1987) study. These studies examined the inter-coder reliability of the CDM method by having subsets of the verbatim transcripts (representing one to two-and-a-half hours of interviewing) coded by different researchers working independently. In each case, one coder had participated in conducting and evaluating the original interviews and the other had not been present during the interviews. These data were analyzed and reported in Taynor, Crandall, and Wiggins (1987).

In studying the reliability of data in the Calderwood, Crandall, and Klein (1987) study, rater agreement on identifying decision points ranged from 81% to 100%. There were 29 decision points originally identified and the new coder was able to identify all of these along with some additional ones that had been treated as standard responses.

The next question dealt with how reliably the decisions could be classified in terms of identified strategies. The same 29 decision points were independently coded by the same coders based on the interview transcripts. Exact agreement using a five-category system was 66.5% based on a weighted average, which was significantly higher than chance ($p<.001$). When a looser criterion for agreement was used (i.e., matching codes that were at least adjacent in the five-category system), agreement went up to 87.8%. These data show very good reliability. They also suggest that we should not use a fine-grained analysis since the distinctions are too subtle for reliable coding.

Reliability of decision coding was also studied in the wildland project (Taynor, Crandall, & Wiggins, 1987). The materials were 18 probed decision points which were coded by the original experimenter as well as by a new coder. Essential agreement using the adjacent category method described above was 88.9%, which replicates the earlier finding. Data from this study also let us examine agreement about the coding of the same decision points retested at five months after the incident. The average rate of agreement was 82.5%. We concluded, therefore, that the CDM methods for cognitive probing and for coding produce highly reliable results.

It proved more difficult to study the validity of the CDM data. It is not clear how to validate the information captured to see whether it is accurate. One study was performed that addressed this issue indirectly. Whitaker and Baynes (1988) assumed that if the CDM information was useful, then providing it to decision makers working through a scenario should improve their understanding of the scenario, and it should improve their ability to predict the decision making of the expert who had been interviewed. That is, by providing the situation assessment information from an interview to a different fireground commander, we should be able to improve accuracy of predictions. Seven scenarios were developed, one for practice and six for data gathering. A total of 24 firefighters with many years of experience (but no command promotions) participated. This attempt failed, the only study in this project that was unsuccessful.

However, the failure was highly instructive. We defined a number of important conditions that would affect the impact and utility of situation assessment reports. Specifically, we had stripped the situation assessment reports of any linkage to response options in order to avoid biasing the results. In so doing, we, in effect, made sure that the situation assessment reports had no information value for the task of making predictions about response selection. This may be why the prediction rate was not improved by the situation assessment reports. In fact, subjects reported confusion because they expected to find implications of the situation assessment within the reports and could not. In bending over backwards to be fair, we may have gone too far. Another design problem that we found in retrospect was the decision options were too complex to be easily distinguished. In addition, the nature of the situation assessment report was not optimal. It focused on the cues the expert was noticing rather than on the inferences drawn from these cues. Other studies (e.g., Brezovic, Klein, & Thordsen, 1987; Calderwood, Crandall, & Klein, 1987) have shown that experts and novices differ more regarding the inferences drawn than regarding the cues used to draw them. With hindsight, this paradigm was not suited for demonstrating the impact of the situation assessment reports. In conclusion, we learned a great deal from this study even though we were not able to demonstrate an effect.

A second study was performed to examine the issue of validity (Crandall, 1988). Highly experienced fireground commanders were shown simulated scenarios of nonroutine incidents, and were probed for situation assessment using the CDM method. Another group with comparable experience was shown the same scenarios and was asked to simply "think-aloud". Compared to this undirected data collection, data gathered using the CDM probes contained significantly more information on the commander's situation assessment, including critical cues and goals. CDM also revealed an underlying structure that linked causes with actions. This link was not present in the undirected responses. There were no differences between the groups in the number of action statements. These findings demonstrate the type of information gain realized through the use of CDM interviews. The results also show that think-aloud verbal protocol methods may not reveal key aspects of experts' knowledge.

Application of CDM to Individual Decision Training

In the study mentioned above (Whitaker & Baynes, 1988) that attempted to demonstrate the validity of the CDM situation assessment reports for improving prediction accuracy, we also looked at whether there was any improvement over trials, since that would suggest that situation assessment reports and the paradigm in general were of potential value in training. Again, this study was not successful. And, once again, the apparent reasons for failure were very instructive. To keep the sessions short, the experimenter did not provide feedback on the reasons why the non-selected options were not chosen, thereby reducing the opportunity for learning. Also, the scenarios were collected from incidents involving different fireground commanders so there was no chance to learn anyone's individual style. Finally, there were only six trials, probably too few for a training effect in personnel with several years of experience. We had no expectations of finding training effect. It was just a variable to inspect in addition to the main purpose of this study, which was to evaluate validity. The study did reveal some key issues for study if using the CDM in training, so it served a useful function.

Knowledge Elicitation for Team Decision Making

The concept of eliciting knowledge from a team was at first confusing, since we envisioned no "team mind". Yet our observations of command and control decision making surprised us because we did see evidence for a team mind. We then set about developing methods for tracking it.

We sent teams of observers to three sites where command and control decision training exercises were being conducted. These were Fort Riley, Fort Hood, and the Command and General Staff College at Fort Leavenworth. The exercises at Fort Riley and Fort Leavenworth enabled us to gain familiarity with the domain and with the requirements for behavioral observation. The third exercise at Fort Hood involved 25 participants plus another 25 exercise controllers who directed the ARTBASS battalion command and control training system. We stationed an observer in the trailer with the S3 (Operations) function to watch the performance of from four to seven Army planners who spent five hours generating a defensive operations plan for the next day's battle. We also conducted CDM interviews after the training exercise.

We found, however, that the CDM interviews were often redundant. We did not need to ask people which options they considered, because our tape recordings of the exercise revealed the options they were discussing. In fact, during the exercise the planners proceeded to examine options and develop them in ways we had found for individual decision makers. Since the discussions followed a decision framework so systematically, the record of these discussions was the record of a team mind. It was a verbal protocol of a decision making team. In some ways the recording was not as good as individual CDM protocols, since it did not cover everything that everyone in the room was thinking. In other ways it was better since the verbalizations were spontaneous and did not interfer with the task. The verbalizations were, in fact, a part of the task.

Therefore, we were able to draw decision maps of the way the team identified goals, acknowledged information receipt, and identified and evaluated options from a transcription of the tape recording. Figure 1 presents an example of a team decision map. Using this map, we were also able to analyze the five-hour planning session into content units, and to label the theme for each unit and the reasons it ended and another began.

These methods appear relevant for training team decision making. They enable observers to provide feedback to the team members, to show areas of efficiency and inefficiency, and to make team members sensitive to the effect of their actions on subsequent planning events. The methods have less relevance to knowledge engineering. Their potential use in training does make them an important product of this research effort.

Subsequently we observed more advanced command and control training exercises at the Command and General Staff College, Fort Leavenworth, in order to demonstrate that we could use this type of team decision mapping to provide feedback in real time to the trainees (Thordsen & Klein, 1991). We observed four days of a five-day Command Post Exercise and, at the end, we did provide the instructors and trainees with feedback about the process of the decision making (rather than on the content which is the focus of traditional feedback). The process feedback was received very well and we were invited to return for the next class to be conducted. We also hope to build on this decision mapping work in an upcoming SBIR Phase I effort with the Army Research Institute in the area of executive decision training.

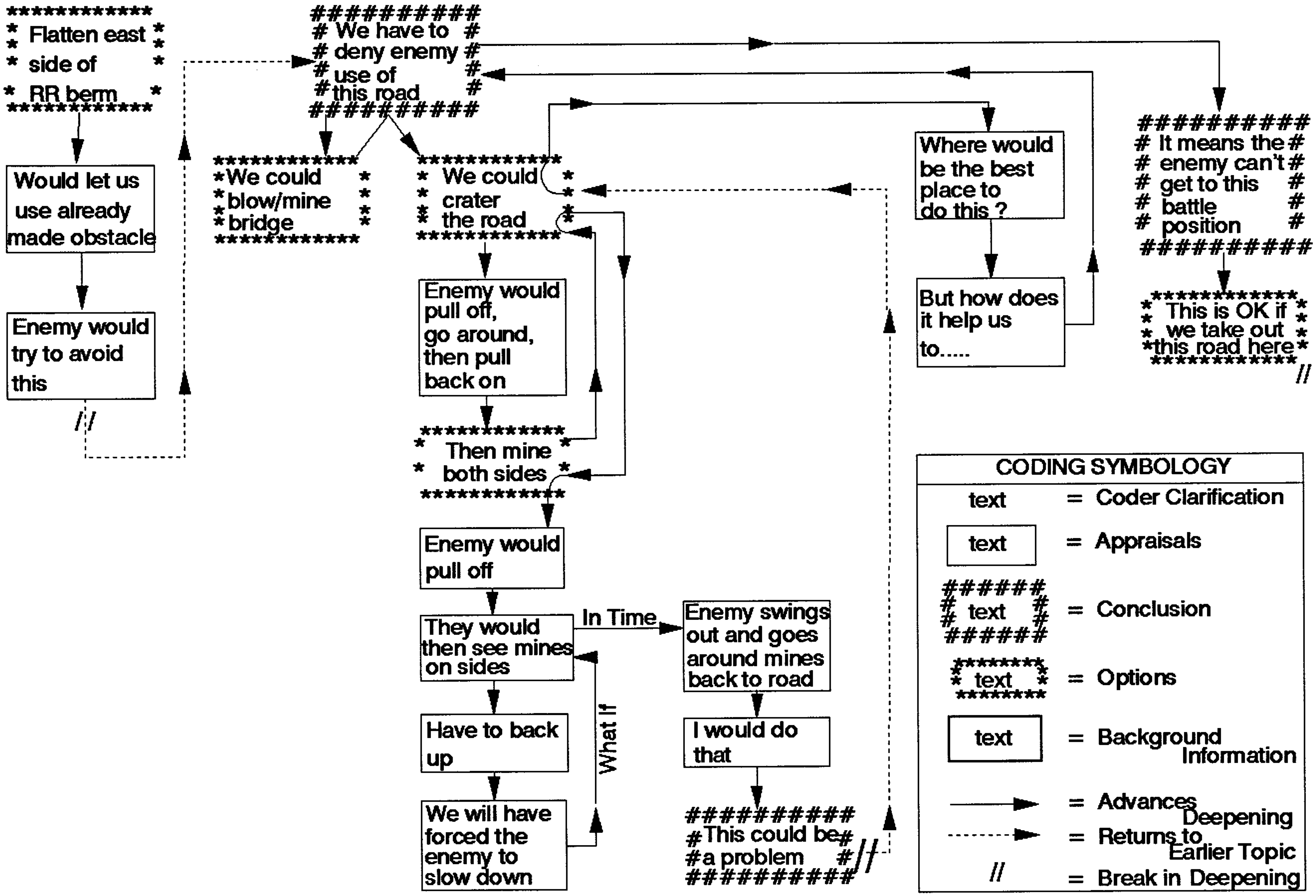

Figure 1. Team Decision Map

Evaluation of Decision Support Systems

If the goal of cognitive engineering is to build decision support systems that capture and apply domain knowledge, then we will need tools for evaluating the impact of these systems on task performance.

Therefore, one of the final tasks of this SBIR Phase II project was to develop such an evaluation tool. The evaluation would have to contrast task performance of a decision maker with and without the decision support system. It would also need to contrast the performance of the organizational unit with and without the decision support system, and also to address the adequacy of the user-computer interface.

We had recently developed a parallel tool for evaluating knowledge-based systems (Klein, 1989; Klein & Brezovic, 1989; Klein & King, 1988). We call this instrument an AIQ^{sm} test. "AIQ" stands for Artificial Intelligence Quotient. In a standard IQ test there is little meaning in an overall score and the same is true for the AIQ measure. What matters is the profile of subscale scores showing strengths and weaknesses. We adopted the AIQ approach, and modified the AIQ instrument by deleting sections dealing with system performance alone (since a decision support system would be used only in conjunction with an operator).

This evaluation approach, a Decision Support Quotient (DSQ) was applied to the Brigade Planner System recently developed at White Sands to support command and control decision making at brigade level. The application of DSQ to Brigade Planner was successful (Thordsen, Brezovic, & Klein, 1988), adding an additional cognitive engineering technique to the methods developed under this SBIR project. Table 2 shows the profile of strengths and weaknesses for Brigade Planner as evaluated by the DSQ.

Table 2

DSQ Ratings for Brigade Planner System

Evaluation Parameter	Rating (1=low, 5=high)
A. Operator + System Performance	
1. Content Coverage	4.66
(a) # cases	5
(b) completeness	4
(b) % time	5
2. Power	4.00
(a) speed	4
(b) success rate	4
(c) quality	4
3. HOE	4.00
4. Flexibility	1.50
(a) handles incomplete and missing data	2
(b) distinguish between certain/uncertain data	1
5. Expandability	1.00
(a) external maintenance	1
(b) internal data base updating	1
6. Cognitive Skill Requirements and Demands	3.25
(a) domain experience requirements	2
(b) time pressure on user	4
(c) mental effort on user	4
(d) potential frustration effects	3
B. User-Computer Interface Adequacy	
1. Explanatory	1.00
2. Error Handling	2.50
3. Tutoring	1.00
4. Audit Trail of Lines-of-Reasoning	1.00
5. Structure of Work Session	2.50
C. System Impact on the Organization	
1. Time to Solution	4.00
2. Quality of Solution	4.00
3. Impact of Constant Use on Expertise	N/A
4. Logistic Demands	3.00

Conclusions

This SBIR project has been completely successful in accomplishing its objectives -- the development, formulation, and evaluation of tools for cognitive engineering.

Three specific methods were developed. The first was the Critical Decision Method (CDM), which was derived for Phase I and evaluated and applied to military command and control tasks in Phase II. The CDM has proven to be an effective direct technique for eliciting higher levels of expertise, including the perceptual and conceptual bases of expertise. We have used it in many settings for a variety of applied functions. There are no reports of any other knowledge elicitation method more carefully worked out or evaluated than the CDM.

Two other cognitive engineering tools were developed -- the suite of team decision mapping methods for tracking ongoing command and control exercises (Thordsen, Galushka, Young, & Klein, 1990; Thordsen & Klein, 1991) and the Decision Support Quotient for evaluating the task performance of decision support systems (Thordsen, Brezovic, & Klein, 1988).

During the two years of the Phase II contract, six studies were performed and additional anlyses conducted to generate nine articles and papers: Thordsen et al. (1990) analyzing the decision strategies of Army battle commanders during command and control training exercises at Fort Leavenworth, Fort Riley, and Fort Hood; Thordsen & Klein (1991) analyzing the decision strategies of command and control trainees at Fort Leavenworth; Thordsen, Brezovic, and Klein (1988) evaluating the performance capabilities of a decision support system, the Brigade Planner System; Klein, Calderwood and MacGregor (1989) describing the CDM; MacGregor & Klein (1988) describing the CDM; Whitaker & Klein (1988a) examining situation assessment; Taynor, Crandall, and Wiggins (1987) assessing the reliability of the CDM; Whitaker and Baynes (1988) applying the CDM to a training task; Whitaker & Klein (1988b) participation in a workshop on command and control knowledge elicitation techniques.

Research and analysis for an additional five papers were partially supported by this contract: Klein (1989) presenting a recognitional model of decision making; Brezovic, Klein, & Thordsen (1987) analyzing the decision strategies of tank platoon leaders; Taynor, Klein, & Thordsen (1987) analyzing the decision strategies of wildland incident commanders; Klein & Peio (1989) developing a prediction paradigm for assessing the decision making of experts and novices; Klein (1987) describing a method for evaluating expert systems.

Another accomplishment of this effort is the further elaboration of a framework for the cognitive engineering domain. We have gained in our appreciation of the need and requirements for cognitive engineering. We have a deeper understanding of the importance of the tools that have been developed, and a clearer concept of the tools that are still needed.

Now that we have developed effective and efficient methods for knowledge elicitation, the next step is to work on the delivery systems. We need to have a wider range of systems for applying domain knowledge for training and for operational decision support. The popularity of expert systems is testimony of the need for such delivery systems. We now realize that expert systems do not have universal applicability. We must work on strategies to augment expert systems and to enable easier capture, representation, and application of expertise.

REFERENCES

Abrett, G., & Burstein, M. H. (1988). The KREME knowledge editing environment. *Proceedings of the Knowledge Acquisition for Knowledge-Based Systems Workshop*, pp. 1-0 - 1-19.

Ashley, K., & Rissland, E. (1987). Compare and contrast, a test of expertise. *Proceedings of AAAI-87.*

Boose, J. H. (1984). Personal construct theory and the transfer of human expertise. *Proceedings of the National Conference on Artificial Intelligence*, 27-33.

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 the U.S. Army Research Institute for the Behavioral and Social Sciences, 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 the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Crandall, B. (1988). *A comparative study of think-aloud and critical decision knowledge elicitation methods.* Yellow Springs, OH: Klein Associates Inc.

Crandall, B., & Klein, G. A. (1987). *Key components of MIS performance.* Yellow Springs, OH: Klein Associates Inc. Report prepared for Standard Register, Dayton, OH.

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

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

Klein, G. A. (1989). Recognition-primed decisionmaking. In W. Rouse (Ed.), *Advances in Man-Machine System Research, 5.* Greenwich, CT: JAI Press, Inc.

Klein, G. A. (1987). AIQ: A test for the performance of expert systems. *Proceedings of the AOG/AAAIC 87 Joint Conference.* Dayton, OH.

Klein, G. A., & Brezovic, C. P. (1989). Evaluation of expert systems. In S. J. Andriole & G. W. Hopple (Eds.) *Defense Applications of Artificial Intelligence.* Boston, MA: D. C. Heath & Co.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1988) *Rapid Decision Making on the Fire Ground* (ARI Technical Report 796). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Klein, G. A., Calderwood, R. & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, pp. 462-472.

Klein, G. A., & King, J. A. (1988). Evaluating the performance of knowledge-based systems. *Proceedings of the AAAI-88 Workshop, Validation and Testing of Knowledge-Based Systems,* Austin, TX.

Klein, G. A., & Peio, K. J. (1989). The use of a prediction paradigm to evaluate proficient decision making. *American Journal of Psychology,* 102, pp. 321-331.

Klein, G. A., Whitaker, L. A., & King, J. A. (1988, May). Using analogues to predict and plan. *1988 Case-Based Reasoning Workshop*. Sponsored by the Defense Advanced Research Projects Agency (DARPA), Clearwater Beach, FL.

Kolodner, J., Simpson, R., & Sycara-Cyranski, K. (1985). A process model of case-based reasoning in problem solving. In M. Kaufmann (Ed.), *Proceedings of IJCAI-85* (pp. 284-290).

MacGregor, D. & Klein, G. A. (1988). *Knowledge elicitation of recognition-primed decision making* (ARI Technical Report 799). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mancini, G., Woods, D. D., & Hollnagel, E. (1987). *International Journal of Man-Machine Studies: Cognitive Engineering in Dynamic Worlds*. NY: Academic Press.

Norman, D. A., & Draper, S. W. (Eds.) (1986). *User centered system design*. Hillsdale, NJ: Erlbaum.

Polanyi, M. (1966). *The tacit dimension*. Garden City, NY: Doubleday & Co., Inc.

Taynor, J., Crandall, B., & Wiggins, S. (1987). *The reliability of the critical decision method* (KATR-863[B]-87-07F). Yellow Springs, OH: Klein Associates, Inc. Prepared under contract MDA903-86-C-0170 for the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

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 the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Thordsen, M., Galushka, J., Young, S., & Klein, G. A. (1990). *A Knowledge elicitation study of military planning* (ARI Technical Report 876). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Thordsen, M., Brezovic, C. P., & Klein, G. A. (1988). *A decision support quotient for evaluating the performance of decision support systems* (KATR-863-88-06Z). Yellow Springs, OH: Klein Associates, Inc. Prepared under contract MDA903-86-C-0170 for the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Thordsen, M., & Klein, G. A. (1991). *Methods for providing direct feedback about decision processes for command and control classroom exercises* (ARI Research Note 91-20). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A231 760)

Whitaker, L. A., & Baynes, T. (1988). *Situation assessment fireground commander training study* (KATR-863-88-06Z). Yellow Springs, OH: Klein Associates, Inc. Prepared under contract MDA903-86-C-0170 for the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Whitaker, L. A. & Klein, G. A. (1988a). Situation awareness in the virtual world: Situation assessment report. In *Proceedings of the 11th Sympoium of Psychology in the Department of Defense*, April.

Whitaker, L. A., & Klein, G. A. (1988b). *Knowledge elicitation for developing more usable expert systems*. Paper presented at the Second Annual Space Operations Automation Land Robotics Workshop (SOAR), Dayton, OH, July.