AL/CF-TR-1993-0093

AD-A275 187

ARMSTRONG LABORATORY

DESIGNING FOR PERFORMANCE: A COGNITIVE SYSTEMS ENGINEERING APPROACH TO MODIFYING AN AWACS HUMAN COMPUTER INTERFACE (U)

David W. Klinger
Stephen J. Andriole
Laura G. Militello
Leonard Adelman
Gary Klein

KLEIN ASSOCIATES INC.
582 EAST DAYTON-YELLOW SPRINGS ROAD
FAIRBORN, OH 45324-3987

Marie E. Gomes

CREW SYSTEMS DIRECTORATE
HUMAN ENGINEERING DIVISION

MARCH 1993

FINAL REPORT FOR THE PERIOD JAN 91 - FEB 93

Approved for public release; distribution is unlimited

AIR FORCE MATERIEL COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from the Armstrong Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

DISCLAIMER

This Technical Report is published as received and has not been edited by the Technical Editing Staff of the Armstrong Laboratory.

TECHNICAL REVIEW AND APPROVAL

AL/CF-TR-1993-0093

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

KENNETH R. BOFF, Chief
Human Engineering Division
Armstrong Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 1993

3. REPORT TYPE AND DATES COVERED
Final Technical Report Jan 91 - Feb 93

4. TITLE AND SUBTITLE
Designing for Performance: A Cognitive Systems Engineering Approach to Modifying an AWACS Human-Computer Interface

5. FUNDING NUMBERS
F33615-90-C-0533

PR - 7184
TA - 27
WU - 12

6. AUTHOR(S)
David W. Klinger, Stephen J. Andriole, Laura G. Militello
Leonard Adelman, Gary Klein, and Marie E. Gomes

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Klein Associates Inc.
582 East Dayton-Yellow Springs Road
Fairborn OH 45324-3987

8. PERFORMING ORGANIZATION REPORT NUMBER
AL/CF-TR-1993-0093

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Armstrong Laboratory, Crew Systems Directorate
HQ HSC, AFMC
Wright-Patterson AFB OH 45433-7022

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Cognitive Systems Engineering (CSE) is primarily a blend of technological opportunities, findings from cognitive research, and Cognitive Task Analysis. Using CSE, we were able to produce an efficient and effective redesign of the AWACS Weapons Director (WD) station. The revised WD station was tested using 17 WDs. These WDs performed Defensive Counter Air Missions with both the current interface and the revised interface. The training of the participants on the revised interface was quite brief (4.5 hours). As a result, the WDs did not achieve the same degree of familiarity or automation with the revised interface that they have with the current interface. Yet, when WDs were using the revised system, their performance improved. This is indicated by an increase in performance for a number of process and outcome measures. Also, a skilled WD provided blind ratings of WD performance. These global ratings were significantly higher for the revised interface. The effectiveness of the revised interface suggests that it is possible to pinpoint cognitive task requirements and to make these the driving factors in a design effort. The use of CSE may be a feasible aspect of the design process, enabling system developers to achieve a much stronger effectiveness at relatively low cost.

14. SUBJECT TERMS
Cognitive Systems Engineering, Naturalistic Decision Making, Human-Computer Interface, Cognitive Task Analysis, Retrofitting, Airborne Warning and Control Systems (AWACS),

15. NUMBER OF PAGES
180

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

KLEIN ASSOCIATES INC.

582 E. Dayton-Yellow Springs Rd.
Fairborn, Ohio 45324-3987
Phone: 513/873-8166
FAX: 513/873-8258
E-MAIL: 76360.3035@COMPUSERVE.COM

EXECUTIVE SUMMARY

Cognitive Systems Engineering (CSE) is primarily a blend of technological opportunities, findings from cognitive research, and Cognitive Task Analysis. Using CSE, we were able to produce an efficient and effective redesign of the AWACS Weapons Director (WD) station. The design effort was completed in a relatively short period of time, approximately ten months.

The redesigned WD station was tested at the Aircrew Evaluation Sustained Operations Performance (AESOP) facility at Brooks AFB, TX, using 17 WDs whose performance was tested on scenarios with the current interface and with the redesigned interface. We were only able to provide the WDs with a brief opportunity to learn how to operate the redesigned interface (4.5 hours). In contrast, they averaged 1180 hours using the current interface after having qualified as a WD and completed a training program that itself involved extensive practice on the current interface. As a result, the WDs did not achieve a high degree of familiarity or automatization with the redesigned interface and their subjective workload went up. They often complained that with only a few more hours of practice on the new interface they would have become much smoother.

Nevertheless, their performance showed a marked superiority using the redesigned interface. A number of process and outcome measures were collected and analyzed. A skilled WD provided blind ratings of the WD performance on sessions with the current and the redesigned interface, and the global ratings were significantly higher for the redesigned interface, reflecting an improvement of more than 25%.

The outcome measures echoed this finding. There were significant improvements in how far enemy aircraft were allowed to approach friendly assets, number of enemy aircraft shot down, and number of missiles fired that missed their targets. There were also clear trends favoring the redesigned interface for number of hostile strikes completed, number of WDs who prevented any hostile strikes being completed, and number of aircraft refueled in air versus those returned to base.

The process measures showed the same improvement for the redesigned interface. Reaction time to visual screen alerts was shorter for the redesigned interface, suggesting that actual workload was lower. (WDs

perceived the workload to be higher because of their unfamiliarity with the new system.) Their reaction time to aural messages was minimally slower, but accuracy was sharply increased using the redesigned interface.

The effectiveness of the redesigned interface suggests that it is possible to pinpoint cognitive task requirements and to make these the driving factors in a design effort. Moreover, these Cognitive Systems Engineering activities do not consume a great deal of time or effort. The use of CSE may be a feasible aspect of the design process, enabling system developers to achieve a much stronger effectiveness at relatively low cost. Additionally, the use of CSE could enable the Air Force to realize higher rates of performance along with reduction in training resources.

ACKNOWLEDGEMENTS

This project reflects the efforts of numerous individuals whose contributions benefitted the final product in many ways. The authors would like to acknowledge:

- The Systems Research Laboratories (SRL) staff of the AESOP facility, Brooks AFB, especially Matt Dalrymple, Dr. David Strome, Neal Takamoto, and Phil Tessier;
- Dr. Sam Schifflet and Capt. Cindy Dominguez of Armstrong Laboratory, Brooks AFB, for conceptual contributions at several critical points in the project;
- Capt. Marcus R. Lesseur, Tinker AFB, for orchestrating the scheduling of the WDs for the evaluation and for verifying the incident accounts;
- Lt. Col. Don Smoot, Wright-Patterson AFB, for providing expert ratings of WD performance;
- Marvin Thordsen, Klein Associates, for his efforts in the early knowledge elicitation sessions;
- Beth Crandall, Klein Associates, for her contributions to the final report.

Finally, we wish to especially thank the AWACS Weapons Directors who participated in the project. The Weapons Directors were from the 552nd, 963rd, 964th, 965th, and 966th squadrons stationed at Tinker AFB, Oklahoma City, Oklahoma. Their insights, feedback, and cooperation were key to the success of the project and we gratefully acknowledge their contributions.

ACRONYM LIST

ADF	Air Defense Fighter
AESOP	Aircrew Evaluation Sustained Operations Performance
ANOVA	Analysis of Variance
AWACS	Airborne Warning and Control Systems
CDM	Critical Decision method
CSE	Cognitive Systems Engineering
CTA	Cognitive Task Analysis
DCA	Defensive Counter Air
FRAG	Fragmented Air Tasking Order
HCI	Human-Computer Interface
IWD	Weapons Director Instructor
NASA-TLX	National Aerospace Administration-Task Load Index
NDM	Naturalistic Decision Making
QAN	Quasi-Automated Nomination
ROE	Rules of Engagement
RPD	Recognition-Primed Decision
RTB	Return-To-Base
SA	Situational Awareness
SAM	Surface-to-Air Missile
SAR	Search-and-Rescue
SD	Senior Director
SME	Subject Matter Expert
SOP	Standard Operating Procedures
SWORD	Subjective Workload Dominance
TD	Tabular Display
WD	Weapons Director

TABLE OF CONTENTS

List of Figures

List of Tables

SECTION 1: INTRODUCTION

The technological advances in computer hardware and software of the 1960s and '70s produced a wide variety of sophisticated, effective systems. Vast amounts of research and development dollars were expended developing, building, testing, and fielding these systems, and many are still in use. But what was once cutting-edge is now standard issue. Many of the computer systems built and fielded over the last two decades are outdated, yet funds required to replace them would be enormous. What are needed are effective, efficient procedures for retrofitting existing systems, to take advantage of advances in computer technology and recent developments in the fields of human factors and cognitive science.

The idea of upgrading existing systems is not new. But many previous efforts have taken a scattershot approach, altering whatever seems most obvious, or adding whatever the latest technological novelty happens to be. The project reported here demonstrates a set of procedures for retrofitting existing systems that begins with identification of key elements of the task, and designs system alterations to support those critical behaviors.

The goal of this project is to demonstrate a successful application of a new method for retrofitting existing systems. This method, Cognitive Systems Engineering, takes what we know about human cognition and develops human-computer interfaces (HCI) to support the cognitive processes of the user. This new method fuses the advances made in the field of computer technology with those made in cognitive science.

Specifically, in this project we targeted the system which supports the Weapons Director (WD) position aboard the Airborne Warning and Control Systems (AWACS) aircraft. These aircraft have been in use since the early '70s, and the most important update in the last 20 years involved a change from the classical round, monochrome radar scope to one which utilizes color. We will show how, given the current AWACS technology, the displays can be further modified to support the decision processes of the users and thereby significantly improve performance. The entire project, which included identifying problem areas, generating system modifications to address these areas, programming the modifications into a simulation facility, and experimentally evaluating the resulting system, was completed relatively inexpensively and with high payoff.[1]

Cognitive Systems Engineering

Cognitive Systems Engineering (CSE) is the application of cognitive science in the design/modification of computer-based information so that the cognitive strengths of the human operator are supported (i.e., decision making and

[1]This project went from domain selection to working system in 10 months, with at least a 20% improvement in performance, for about $300,000.

inference). This perspective provides a framework for the designer to create a system in which human thought processes are treated explicitly and are an integral part of the final product.

Several factors play essential roles in the application of CSE:

- An awareness of the technology available, not only of the target system, but also of the options available to develop the proposed system. The area of computer technology is changing daily, and some understanding of these advances is necessary.

- An understanding of human cognitive processes. What we know about cognition is vastly different now than even 10 years ago. We know more about how people make decisions, and how other cognitive components affect operator performance.

- The application of a Cognitive Task Analysis (CTA). The use of CTA, particularly for interface design, is of central importance. This provides an overall understanding of the user's cognitive processes and where the proposed system can better support those processes.

How does CSE differ from standard human factors approaches? Standard human factors approaches offer little if any leverage in tasks and settings where "performance" is largely cognitive, and operators' actions basically serve to implement the outcomes of their judgments, assessments, and decision making. CSE seeks to take the next step of identifying and documenting the cognitive processes (problem diagnosis and framing, situation assessment, judgment and decision making, inference, problem solving) that direct human behavior in complex tasks, and uses that information to develop decision support systems and HCIs that complement and enhance those cognitive processes.

Because CSE is theoretically grounded in current research on cognitive processes such as attention, memory capacity, situation assessment, and decision making, we can potentially evaluate the strengths and weaknesses of HCI solutions in terms of these processes. We can examine whether an interface reduces memory and attentional requirements, and diminishes workload. We can determine whether the additional capabilities offered to the operator may inadvertently interfere with situational awareness. We can identify a conceptual phenomenon such as workload, target it as something the HCI is designed to reduce, and measure whether this occurred. Providing that we can derive performance measures that reflect these cognitive processes, we should be able to gain a better sense of how the HCI is working than if we only study outcome measures.

CSE is an emerging field. There are few handbooks with which to guide the design team, yet the basic structure is fairly well determined. That is, a behavioral and cognitive task analysis must be performed in order to determine the user needs; these findings need to be utilized in either the design or redesign of a system; and the evaluation must maintain this cognitive focus in order to determine if the proposed modifications are achieving the desired results. Yet, within this structure there is great flexibility. Currently theories regarding how to conduct a cognitive task analysis abound. There is also great disagreement as to the optimum method for determining display features and the evaluation of them.

The determining factors for conducting a CTA and the optimum method for determining display features, which make up the overall structure of CSE, are very domain specific. In this instance we were operating within a domain that was rather complex. Split-second decisions are made by operators who work under extreme time pressure and whose actions can spell the difference between success and disaster. Also, when selecting the test domain, we were fortunate enough to locate an excellent high-fidelity simulation facility at Brooks AFB at which to test our display hypotheses.

The AWACS Weapons Director Position

The WD position can be likened to that of an Air Traffic Controller in the sky, with some important differences: commercial aircraft seldom shoot at one another, the Air Traffic Controller never needs to monitor an airborne track in order to determine intent; Air Traffic Controllers are seldom in danger of being shot down (they are not flying in the sky with the aircraft they are controlling); and they do not need to worry about rules of engagement. A WD must contend with all of these and more. Often, WDs work 15-18 hours under high stress, in cold, crowded airplanes. They must direct aircraft under their control, not to a stationary landing strip, but to intercept fast flying enemy aircraft that could shoot down the friendly aircraft if the geometry of the intercept is only slightly miscalculated. During periods of low to moderate air traffic, a WD can comfortably manage all of these tasks. During periods of heavier air traffic, however, a WD is likely to reach overload conditions.

There are typically four WDs aboard each AWACS aircraft. Three of them are actual WDs directing fighters, the fourth, a Senior Director (SD), is a more experienced WD who is essentially the leader/coordinator of the team. The picture of the world that the AWACS radar provides these WDs is divided into three "lanes" or areas. Each WD is assigned a lane. S/he is responsible for all aircraft, friend or foe, in that lane. This means that any enemy aircraft that appear in a WD's lane must be monitored for intent so that actions can be taken if the aircraft becomes a threat. The action taken is usually an intercept by a friendly fighter. Typically, the friendly fighter is one of many that are patrolling the sky providing protection for both airborne and ground assets.

In order to be ready for any emergency, there must be at least one friendly aircraft that can intercept a hostile aircraft. To accomplish this the WD must maintain good "fighter flow." That is, s/he must orchestrate an aircraft flow pattern in which some fighters are engaged in air-refueling, some are returning to their base, some are taking off to replace those that are leaving, and some are patrolling the sky. If the WD gets behind, loses situational awareness (SA), or does not maintain good fighter flow, a hostile aircraft may evade the friendly aircraft and penetrate the line of defense.

The communication aboard an AWACS aircraft appears chaotic at times. Each WD must monitor four outside radio channels and two inside channels. The inside channels are reserved for the SD and other members aboard the aircraft. The four outside channels are for the various pilots in the assigned airspace to communicate with the WD. Typically, a WD has one assigned channel for his/her lane. It is not uncommon, though, for there to be several fighters broadcasting to one WD on more than one channel. That is, fighter pilot A may be using channel 1, while fighter pilot B is using channel 2, yet both are trying to talk to the same WD.

Each AWACS aircraft is a member of a data link. It is the responsibility of each WD to maintain his/her portion of the link. Each WD's lane is combined with all other WDs' lanes into one picture that is viewed by the individuals in a command and control center on the ground. With this network, all AWACS aircraft have access to what other AWACS aircraft are doing and the commander and control center can monitor the overall situation. For instance, if a WD is sending a friendly aircraft to intercept an enemy aircraft, s/he must notify the system so that it can relay this information to the ground. One important function of the WD is to monitor the aircraft's track identifiers, called symbology, and make sure they stay on the correct radar dot. Any breakdown produces a ripple effect so that those down the line do not have an accurate picture of the situation. This makes it difficult, if not impossible, to maintain an adequate line of defense.

To maintain an accurate representation of his/her lane, WDs must continually update the system on board their particular aircraft, which in turn updates the overall network. To accomplish this the WD must execute "switch actions." These switch actions are inputs to the system as to what the WD intends to do. For instance, if a friendly aircraft is going to intercept an enemy aircraft, the WD notifies the system that s/he has "committed" a friendly track against an enemy target by selecting the "commit" switch action and inputting to the system the two tracks of interest. Other switch actions, or system inputs, can be as simple as telling the system to put a track symbology back on a particular radar dot, or as complex as initiating a downed-aircraft point. There are nearly 100 switch actions that can be activated, but only about 20 are used regularly.

The scope, actually a computer monitor, displays the radar's picture of the airspace and the aircraft in it to the WD. Aside from the switch actions, there are two other methods for the WD to communicate with the system. One, an alpha-numeric keypad, is used to type in messages, track numbers, etc., that may be associated with a particular switch action. Two, a trackball controls an on-screen cursor that is used to tell the system which track to perform a certain function on or where to place the center of the screen.

So, the WD monitors the radios, communicates with pilots and attempts to execute switch actions which tell the system what to do, all while trying to maintain the "big picture" of what is going on in his/her lane of defense.

The project reported here was carried out within this rather complex domain to produce a revised set of displays for use by WDs. The displays were designed based on the findings from a cognitive task analysis, and were evaluated in a high-fidelity simulation facility at Brooks AFB. The following sections present our final recommendations for the WD displays, along with the process of development that brought the recommendations to life. We discuss how we "got inside the heads of the users," using cognitive task analytic methods to examine how WDs do their jobs, what cues are important to them, and what information they need to make good decisions. We show how this information directed us to problem areas within the current system, and how we addressed those problems. We describe the usability tests, and show how, from a cognitive perspective, we modified the displays just prior to the formal evaluation in order to achieve a greater impact. We present findings from the evaluation phase of the project. Finally, we discuss implications for future efforts of this type and how they may benefit existing systems.

SECTION 2: COGNITIVE SCIENCE FRAMEWORK

The goal of CSE is to identify and document the cognitive processes that direct human behavior in complex tasks, and use that information to develop systems that complement and enhance these cognitive processes. The CSE process involves the application of findings and methodologies from cognitive science. Cognitive science encompasses a broad body of academic and applied research concerned with the human mind and the reception, storage, retrieval, transformation, and communication of information. Cognitive scientists seek to understand perceiving, thinking, remembering, and other mental events. The perspective has a number of points of contact with human factors. However, human factors research has emphasized basic sensory, perceptual processes and/or biomechanical aspects of performance. This approach has been extremely helpful in identifying physical, biomechanical aspects of human performance issues, and in offering solutions to enhance performance of those tasks. However, CSE goes a step further, by providing a method to examine the cognitive processes of the operator. The CSE approach offers considerable leverage in tasks and settings where "performance" is largely cognitive, where operators' actions serve to implement the outcomes of judgments, assessment, and decision making.

In our own work, we have chosen to focus on issues surrounding judgment, decision making, and problem solving in real-world settings. This approach is known as "naturalistic decision making." Work in the area of naturalistic decision making (NDM) provides a theoretical perspective that allows us to examine cognitive performance issues and subsequently to apply our knowledge of decision processes to support quality decision making and reduce operator errors through the design of better systems.

To learn how decision makers handle the complexities and confusion of operational environments, NDM researchers have moved their research out of controlled and predictable laboratory settings and into the field to study domains that are complex and challenging. Experienced operators of complex systems are primarily the subjects of study.

It is beyond the scope of this section to provide a thorough review of the NDM literature. However, we will discuss a few important concepts that have direct application to the WD task. Several other sources provide thorough discussions of NDM including Klein (in press) and Klein, Orasanu, Calderwood, and Zsambok (in press).

NDM researchers have studied a wide variety of domains. Investigators report findings from domains as diverse as firefighting, anti-air warfare command and control, and power plant control. These domains share a number of characteristics that affect how decision makers make decisions. Thus, much of what is learned in one domain may be applicable to another. The question concerns what findings from other domains we can apply to the decisions that WDs make.

An Overview of Naturalistic Decision Making

NDM is a recent approach. Decision researchers such as Payne (1976) and Beach and Mitchell (1978) had pointed out that the heavily analytical strategies prescribed by classical decision researchers are not practical for many tasks, and that under conditions such as time pressure and uncertainty, people are more likely to invoke simpler strategies. Similar to the classic decision models, however, these contingency models still focused on how people select the best course of action from comparison among a set of several alternatives. Several years later, Rasmussen (1985) and Wohl (1981) formulated more detailed descriptions of NDM and linked the two functions of diagnosing a situation and selecting a course of action. Neither Rasmussen nor Wohl are academic researchers. They were working to resolve design problems in real-world domains: nuclear power plant displays and Navy command and control. Thus, it may have been easier for them to perceive the relationship between diagnosing a situation and selecting an appropriate course of action for that situation. The importance of considering situation diagnosis will become clearer as we describe features of NDM later.

During this same period, a few researchers had begun to investigate naturalistic settings. Hammond, Hamm, Grassia, and Pearson (1987) showed that highway engineers made effective use of analytical decision strategies for tasks such as estimating traffic load. But for other tasks, such as estimating accident rates, the engineers did better using intuitive strategies. Shanteau and Phelps (1977) found that expert judges were able to make reliable and accurate decisions without following analytical procedures. Their work stands in sharp contrast to the earlier decision research that emphasized strategies for selecting one course of action from many.

The critical events for NDM occurred in the late 1980s. There had been a growing realization that decision making involved more than picking a course of action, that decision strategies had to work in operational contexts, that intuitive or nonanalytical processes must be important, and that situation assessment had to be taken into account. A number of researchers presented models showing how decision makers could use their experience to handle operational contexts. Klein and his colleagues (Klein, 1989; Klein, Calderwood, & Clinton-Cirocco, 1986) reported on fireground commanders, tank platoon leaders, and design engineers. Noble, Boehm-Davis, and Grosz (1986) reported on Naval command-and-control personnel. Pennington and Hastie (1981) reported on jurors. Beach (1990; Beach & Mitchell, 1978) studied business decisions. Lipshitz (1989) reported work with Army officers. This research went beyond pointing out the limitations of the classical models of decision making; it presented models of how people make decisions in real-world, operational settings. With the emergence of these models, NDM research achieved coherence as an approach for studying basic and applied issues.

Characteristics of natural environments. What makes natural environments particularly challenging for decision makers? What characteristics of natural environments cause the classical decision literature to lose its relevance for decision makers in the real world? Research has identified the essential characteristics of naturalistic decision environments (Klein, in press; Orasanu & Connolly, in press). Table 1 presents nine features that are particularly interesting. Not every domain includes these variables, and some naturalistic reasoning strategies apply even when most of these features are missing. Nevertheless, the features in Table 1 cover the most challenging aspects of operational settings. To help people think clearly under pressure, we must understand how people make decisions under the conditions listed.

Previous models of decision making avoided the features listed in Table 1. The classical theories of decision making (Baron, 1988; von Winterfeldt & Edwards, 1986) grew out of mathematics and game theory (Keeney & Raiffa, 1976). These models showed how decision makers should use their estimates and judgments to make optimal choices. The models were formulated for straightforward, well-defined tasks. The models were not intended for cases where time was limited, goals were vague and shifting, and data were questionable. Therefore, the classical models are not useful describing how people work in dynamic, time-compressed settings.

Features of naturalistic decision models. The most important finding that has emerged from NDM research is that in operational settings people rarely compare options to select a course of action. That is, they do not decide what to do by comparing the relative benefits and liabilities of various alternative courses of action. For example, Klein et al. (1986) investigated how fireground commanders make decisions about deploying their crew members during difficult urban fires. The commanders insisted that they never tried to determine whether one option was better than another. Quite often, they implemented successfully the first course of action that came to mind. For researchers trained to expect that decision making necessarily involved comparison between options, this was totally unexpected. How can skilled decision makers select effective courses of action without comparing options?

NDM research has produced extensive evidence indicating that decision makers can use their experience to size up the situation, recognize it as typical in some ways, and identify the typical way of responding. Therefore, skilled decision makers may never have to consider more than one option when making decisions. The different strategies for contrasting options rarely come into play. Of course, there are times when it is important to contrast optional courses of action, particularly for individuals who do not have sufficient experience. But for most cases, including very difficult incidents, the critical step for the experienced decision maker is to assess the situation. Once the decision maker understands the situation, an appropriate course of action is easily identified.

Table 1

Characteristics of Naturalistic Domains

Characteristic	Description
Time pressure	Decision makers have limited time in which to make decisions and implement responses.
Dynamic settings	Situations are not static; they evolve over time.
High risk	The consequences of errors are high for either the decision maker or others in the situation.
Shifting goals	Dynamic conditions change what is important. As situations evolve the decision maker must be able to modify goals.
Feedback loops	Actions taken will alter the situation, and thus may dramatically affect the subsequent goals and actions.
Ambiguous, missing, and questionable data	Available data rarely paint a clear picture. Pieces of information may conflict with each other, be missing altogether, or be of unknown quality.
Cue learning	Experienced decision makers associate meaning with constellations of environmental cues and with changes in cue clusters. These meanings are not available to novices.
Experienced decision makers	Most decision makers have some level of task experience, ranging from journeyman to expert. Decision makers in real-world settings are rarely novices.
Teams	Decision makers often work together in groups as teams.

Adapted from Orasanu and Connolly (in press). *The reinvention of decision making*. In G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok (Eds.), Decision making in action: Models and methods. Norwood, NJ: Ablex Publishing Corporation.

NDM research has produced extensive evidence indicating that decision makers can use their experience to size up the situation, recognize it as typical in some ways, and identify the typical way of responding. Therefore, skilled decision makers may never have to consider more than one option when making decisions. The different strategies for contrasting options rarely come into play. Of course, there are times when it is important to contrast optional courses of action, particularly for individuals who do not have sufficient experience. But for most cases, including very difficult incidents, the critical step for the experienced decision maker is to assess the situation. Once the decision maker understands the situation, an appropriate course of action is easily identified.

An incident reported by Kaempf, Wolf, Thordsen, and Klein (1992) il'ustrates this point. The commander of an AEGIS Navy cruiser was faced with a decision about whether to shoot down a pair of F-4s that threatened the cruiser. On the surface, the decision was about two different courses of action: to engage or not to engage. On a deeper level, the decision was about assessing the situation, determining the intent of the fighter pilots.

On this particular day, the cruiser was escorting an unarmed ship through the Persian Gulf when two Iranian F-4s took off and began to circle near the end of a nearby runway. Each successive orbit brought the fighters closer to the two ships. Both aircraft turned on their search radars; the lead pilot turned on his fire control radar and acquired the ships. By the rules of engagement in effect, this was a hostile act, and the AEGIS commander would have been justified in firing on the aircraft. However, his mission was to reduce hostilities, not increase them. The AEGIS commander decided that the two aircraft were not going to attack. How did he form this assessment?

The Captain tried to imagine that the F-4s were hostile. However, he could not imagine that a pilot preparing to attack would be so conspicuous, flying around in plain view. The pilots further announced their presence by activating their radars, even using their radars unnecessarily by keeping them on when their circles carried them away from the ships. The Captain just could not imagine how pilots planning to attack would behave in that way.

In contrast, the Captain could imagine how the pilots were trying to harass him. All of their actions appeared consistent with this hypothesis, whereas the hostile intent hypothesis had some major flaws. Therefore, the Captain inferred that the F-4 pilots were simply playing games. Once the Captain reached this decision about the situation, then determining a course of action was simple. He would take action to prepare his ship, but would not engage the aircraft. This incident illustrates several insights derived from NDM research.

First, most often people try to satisfice and find the first workable solution rather than optimize by finding the best solution. Simon (1955) was the first to make this distinction. In operational settings, it is very difficult to determine what the best course of action is, even with hindsight. Decision strategies that try to calculate the optimal course of action only work when time is plentiful and the goals are clearly defined. For example, no one can say that the AEGIS commander was right or wrong in not firing at the F-4s as soon as they illuminated their fire control radar. In this case it worked out, because he avoided an incident by increasing his level of risk while retaining the ability to defend his ship.

Second, situation assessment decisions are distinguishable from course of action decisions. Sometimes, decision makers need to diagnose what is happening, and select one diagnosis from among several. At other times, the decision maker must determine which action to take. In the F-4 example above, the commander was faced with a diagnosis decision.

Third, in operational settings, people use their experience to arrive at situation assessments. In addition, they can use context to help them draw inferences.

Fourth, in most cases, the situation assessment makes the appropriate course of action obvious. Many operational domains have extensive standard operating procedures (SOP) and preplanned responses. The purpose of these procedures is to aid the decision maker by identifying the appropriate courses of action. Planners spend considerable effort anticipating possible contingencies and identifying the appropriate response for each. This removes from the decision maker the burden of generating courses of action, but increases the burden of correctly assessing the situation.

This situation is also true for a WD on the AWACS. SOP and rules of engagement (ROE) dictate what actions the WD should take in most situations. However, the WD has the responsibility of diagnosing the situation. Once this is accomplished the SOP and ROE dictate which actions to take. Thus, the tough task for the WD is to make the diagnosis decision.

Finally, decision makers frequently must act with incomplete and often conflicting information. Often, decision makers do not receive the information that would make a decision easy. This may be due to a variety of causes: poor communications, inadequate sensors, malfunctioning equipment, mistakes by others, poor environmental conditions. These factors may also lead to conflicts among the information received from different sources. Experienced decision makers expect these problems and learn methods for handling situations in which they receive inadequate information.

These insights from NDM research portray decision makers as capable of using experience to handle difficult situations without having to evaluate different options. This stands in direct contrast to many decision training programs based on classical models of decision making: generate many different options, carefully list the strengths and weaknesses of each, calculate the best, and then act. Anything less is seen as deficient. According to the NDM framework, this advice may be useful for novices who lack experience diagnosing situations. But the advice is incompatible with the way that proficient operators make decisions. The available data (Isenberg, 1984; Klein, 1989; Soelberg, 1967) clearly show that decision makers do not follow the classical advice of contrasting options, yet they are quite successful. Experts make good decisions frequently without comparing options. Furthermore, departing from the classical advice is what experts are able to do. Thus, it is a model to emulate, not correct. Clearly there are times when option comparison is called for particularly for the less experienced. But in highly procedural jobs, these times are relatively rare. Typically, NDM models:

- explain how people can use experience to make decisions
- describe how decision makers can use situation assessment to identify a course of action
- describe how decision makers can settle on a single, feasible course of action without having to consider many possibilities
- describe how decision makers can be poised to act, rather than having to wait to complete their comparisons and analyses.

NDM highlights the relative importance of diagnostic decisions for people to succeed in dynamic, time-compressed situations. Previous literature focused on the phenomena of choosing a course of action. NDM research demonstrates that these course of action decisions play a relatively minor role in operational settings. Success in dynamic, time-compressed settings requires that people make accurate diagnostic decisions.

SECTION 3: REQUIREMENTS ANALYSIS

It is the cognitive component of the requirements analysis that is often overlooked in the design and modification stages of system development. It is with this naturalistic framework that we began the requirements analysis. Many times designers rely solely on traditional methods such as behavioral task analyses and IDEF charts. While these methods yield valuable information they do not address important cognitive elements such as decision making and situational awareness (SA). In this project we used these traditional methods, but also included a cognitive task analysis based on in-depth interviews with users. This allowed us access to the contextual information that previous work in NDM identifies as critical to cognitive processing in dynamic settings such as the AWACS, as well as information about the specific decision-making processes used by WDs.

In this section, we describe our activities during the requirements analysis phase of the project. We begin with an explanation of how we became familiar with the behavioral aspects of the WD task. This is followed by a description of the cognitive task analysis that we conducted, including an in-depth description of two interview techniques that we have found to be effective in eliciting cognitive elements of the task. We then discuss the analysis of the interview data and describe how this led to modifications of the WD station aboard the AWACS aircraft.

Analysis of the WD Task

To build our knowledge of the domain, we began investigating the available written materials. These included numerous WD handbooks and student guides. These provided us with a basic knowledge of the standard operations of the position. In addition, we had access to IDEF charts, which provide detailed outlines of the steps necessary for completing a particular task. The IDEF chart for the WD position identified the tasks involved in committing a friendly aircraft track against another airborne track.

The cognitive task analysis consisted of three concept-mapping sessions, followed by 13 Critical Decision method interviews, and an analysis of all the interview data. These are discussed below.

Concept Mapping. A concept map is a schematic device for representing meaningful relationships among concepts. Originally devised as an instructional and evaluation tool for use in academic settings (e.g., Gowin & Novak, 1984), concept mapping has been used more recently as a knowledge elicitation method by Air Force operations researchers for analyzing user needs and developing work station designs (McFarren, 1987; McNeese, Zaff, Peio, Snyder, Duncan, & McFarren, 1990). In conducting a concept-mapping interview, the interviewer asks the participant a very general question concerning the participant's area of expertise. The conversation is then recorded by the interviewer in the form of a concept map. Each concept is enclosed in a circle, with arrows connecting the concepts. The arrows are labeled with the relationships among the concepts. The result is a map depicting the organization of important concepts and their relationships from that individual expert's point of view.

Because concept maps provide an overt, explicit representation of individual concepts and the linkages among them, they allow knowledge interviewers and SMEs to exchange views and correct misunderstandings as the map is being developed. Concept maps obtained from different SMEs can be used to examine the commonalities and idiosyncracies that exist in a knowledge base and to generate a comprehensive knowledge representation of domain expertise.

In this case, concept-mapping sessions were conducted at the onset of the project to provide a broad overview of the domain from the perspective of several WDs. In individual interviews, each WD was asked "How do you organize the WD task in your mind?" The concept map was recorded by the interviewer as the discussion unfolded.

Completed concept maps can be difficult to understand. They often appear disjointed and fractured. Concept mapping sessions tend to be fast-paced sessions in which an expert describes concepts and links them with other concepts that could be anywhere in the map. As the expert explains these linkages, they seem logical to the interviewer who begins to see the domain in much the same way. For those not present for the interview, the context is missing. For them, the reasons for particular groupings and linkages are not always apparent or may appear arbitrary. Therefore, it is sometimes useful to reorganize the maps in order to fully utilize their information content. Figure 1 is a concept map as it was constructed during an interview. The links and concepts are obvious, yet it is often difficult to glean much information from the structure. Figure 2 offers a reorganization of the map. This version closely resembles a flow chart, with the more general concepts in boxes, and the supporting concepts below. For some this structure is easy to understand, for others the loss of flexibility from the original map makes it seem less coherent. Typically, those who are familiar both with the domain and with concept mapping methods prefer the original map. For others the flowchart diagram may be preferred.

Typically the concept maps help focus additional knowledge elicitation. For this project, we conducted three concept-mapping sessions to familiarize ourselves with the domain. This allowed us to understand the basic concepts of weapons directing, and to become familiar with the language. We were able to obtain a glimpse of the WDs' mental models or ways of organizing information pertaining to the WD task. These maps also sparked our curiosity in various areas. For instance, after examining the concept maps, we became interested in the transition from tactical control to close control. We wanted to know more about the big picture. How is it conveyed to the pilot and others in the network? Is there a standard method for conveying all the information? Which cues are critical and when are they most critical? These questions made their way into the Critical Decision method (CDM) interview sessions that followed.

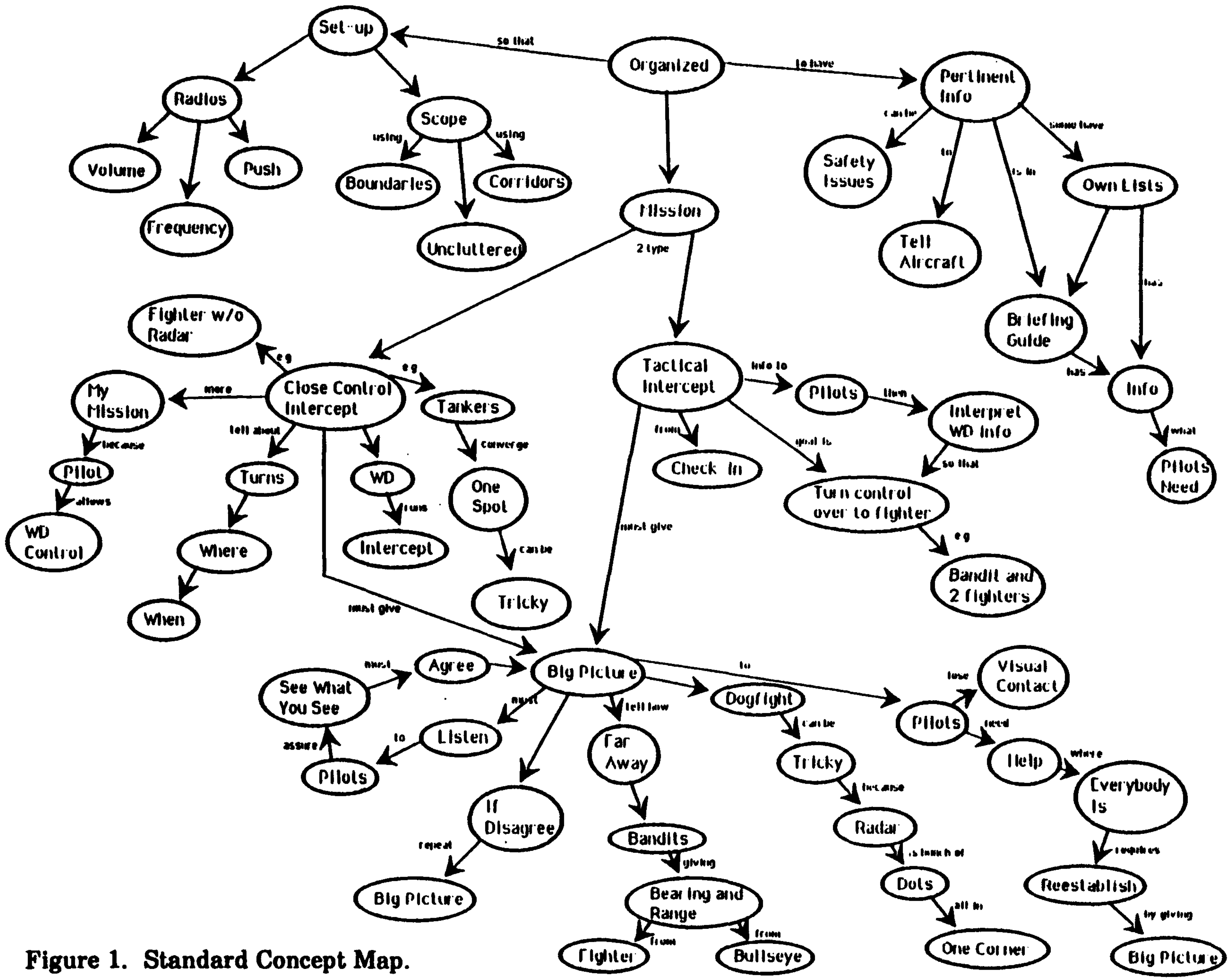


Figure 1. Standard Concept Map.

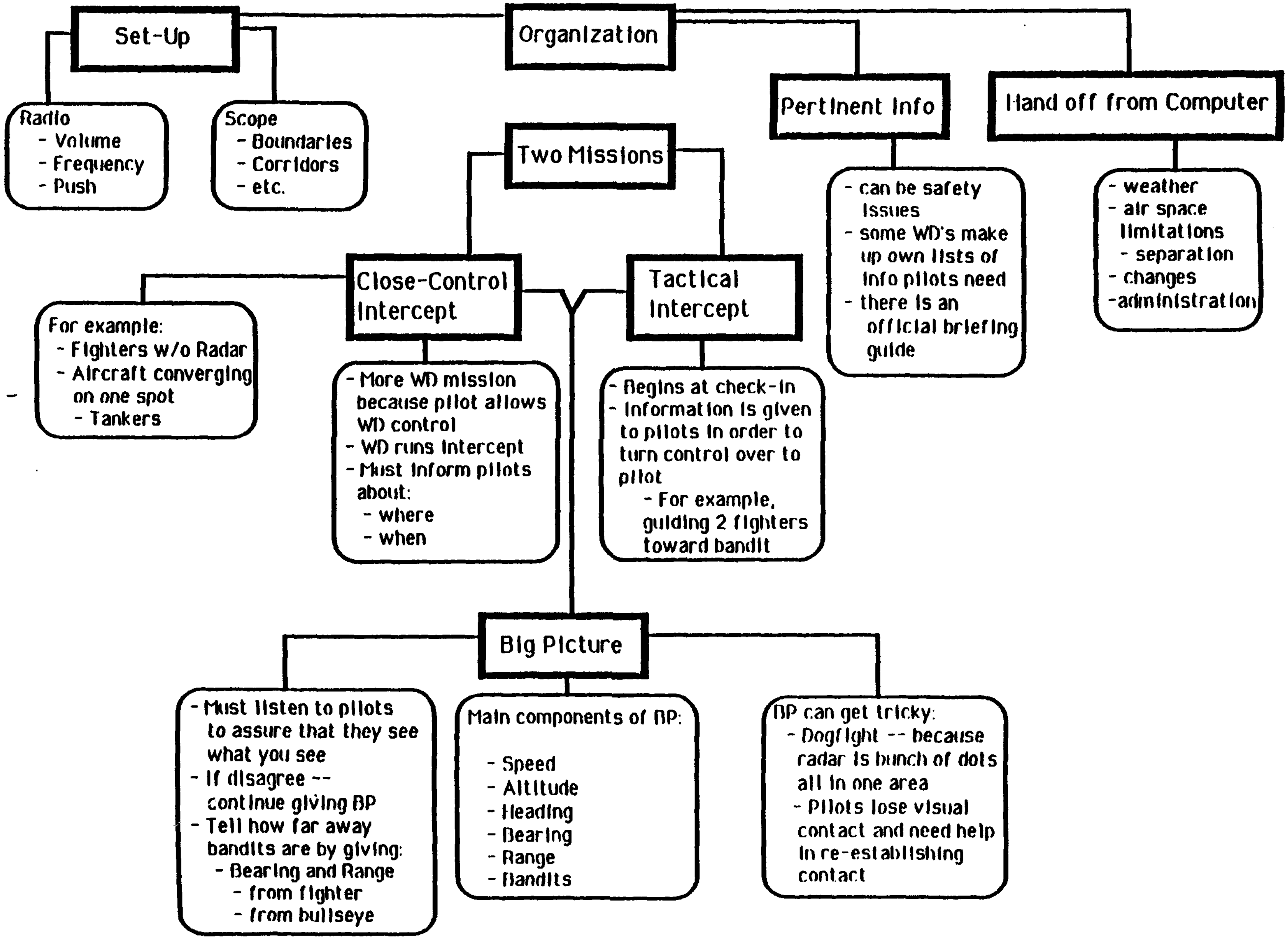


Figure 2. Flow chart representation of a concept map.

Critical Decision method (CDM). CDM is a knowledge-engineering strategy based on Flanagan's critical incident technique (Flanagan, 1954). Using recollection of a specific incident as its starting point, CDM employs a semi-structured interview with specific, focused probes designed to elicit particular types of information from the interviewee. Solicited information includes goals that were considered during the incident, options that were generated, evaluated, and eventually chosen, cue utilization, contextual elements, and situation assessment factors specific to particular decisions. CDM protocols provide detailed records of the information gathering, judgments, interventions, and outcomes that surround problem solving and decision making in a particular task or domain.

Researchers at Klein Associates developed CDM to elicit the decision strategies used by experienced fireground commanders and emergency rescue personnel at the scene of a fire or emergency. We found that many of these decisions relied on subtle perceptual cues and assessments of rapidly changing events that were not easily articulated. Thus, an interview format was developed that allowed experts to focus on and describe aspects of their tasks that are normally difficult for them to articulate. CDM has been demonstrated to yield information richer in variety, specificity, and quantity than is typically available in experts' unstructured verbal reports (Crandall, 1989). The method has been used in over a dozen studies and in domains as varied as fireground command, battle planning, critical care nursing, corporate information management, and commercial and helicopter piloting (e.g., Calderwood, Crandall, & Baynes, 1988; Calderwood, Crandall, & Klein, 1987; Crandall & Calderwood, 1989; Crandall & Gamblian, 1991; Crandall & Klein, 1988; Klein, 1989; Klein, Calderwood, & Clinton-Cirocco, 1986; Klein & Thordsen, 1988; Thordsen & Calderwood, 1989; Thordsen, Klein, Michel, & Sullivan, 1988; Thordsen, Klein, & Wolf, 1990; Wolf, Klein, Thordsen, & Klinger, 1991).

In the current study, individual CDM interviews were conducted with 13 WDs. They were asked to describe an incident in which their skills were challenged. After an initial description of the incident, the interviewer and WD constructed a timeline of the incident in order to better understand the flow of events. The WD was asked to fill in gaps and add anything else that came to mind. Once the course of events had been determined, the interviewer followed up with cognitive probes aimed at obtaining information concerning the WD's cognitive processes: the problem solving and decision making that surrounded the events depicted in the timeline. Of particular interest was the WD's focus of attention and how it changed as the incident progressed, the cues that were attended to during different phases of the mission, the implications those cues had for the WD, and the options that were considered.

Although memories for such events cannot be assumed to be perfectly reliable, the method has been highly successful in eliciting perceptual cues and details of judgment and decision strategies that are generally not captured with traditional reporting methods (Crandall, 1989). Moreover, CDM provides this information from the perspective of the person performing a task, and can be particularly useful in identifying cognitive elements that are central to its proficient performance. Detailed descriptions of CDM and the work surrounding it can be found in Klein (1989) and Klein, Calderwood, and MacGregor (1989).

We have included summaries of four incidents described during the interview sessions in Appendix A. The information obtained in these interviews was analyzed in order to extract specific information regarding cue utilization, contextual elements, focus of attention (or lack of focus), and option generation. This information was then incorporated into storyboards illustrating our design recommendations.

Analysis of the CDM Interview Data

The CDM interviews were transcribed immediately following the interview sessions. We began analyzing them by looking for ways to organize the information obtained in these lengthy, in-depth interviews. In order to gain the most insight from these data, we utilized several different analytic techniques.

Initially, we dissected individual incidents into static "snapshots" depicting various stages of situation assessments. Each snapshot was carefully examined for critical cues and features. We noted the WD's focus of attention at that point in time, cues contributing to that particular assessment of the situation, information that would have been helpful but was not available, and strengths of the system in that specific situation. As this list was compiled, we also generated potential display features and modifications intended to support the WD's strengths during that situation, compensate for human limitations (i.e., memory, attention), and fix any weaknesses noted in the existing system.

A second technique was inspired by one of the CDM interviews. This particular WD had worked as a ground controller on a manual radar system before coming to AWACS and this perspective inspired us to look at our data from a different point of view. He believed that the simplicity of the manual system forces a controller to stay fully engaged in the task, while many of the extra features included in the AWACS technology serve more to distract than aid the controller. For example, on the manual system, the controller tracks aircraft by marking the location, with each consecutive radar sweep, on the radar screen with a grease pencil. The AWACS computer tracks each aircraft and provides the WD with a record of the last six 10-second intervals; the WD needs only monitor the screen, s/he does not actively record the movement of aircraft. This WD claimed that it is too easy to "get lost in your scope" on the AWACS. Controlling aircraft is an art in which a balance is maintained between pilot communication and accurate situation assessment. It is easy for a new WD to get so caught up in the computer technology on the AWACS that this balance is forgotten. This interviewee's preference for the manual system affected not only the way he performed his job as a WD, but also the way in which he trained new WDs.

We began considering the strengths of the manual system and what may have been lost with the introduction of the advanced AWACS technology. As we went over an incident, we would look for instances in which the WD may have been "lost in the scope"–concentrating too much on the computer screen and not enough on communicating with pilots. We also looked for instances in which the AWACS showed advantages over a manual system. We again began to generate display ideas, this time features that would compensate for the tendency of the AWACS computer to act as a distractor without eliminating the added strengths of the AWACS technology.

Finally, a third technique involved examining specific aspects of the WD task. The WD task was broken down into critical functions such as performing the commit switch action, calculating the geometry for an intercept, conducting a search and rescue mission, etc. In this case, the information needs of the WD for each of these functions were considered in turn. We examined whether the existing system presently provides this information in an optimal format and location or whether improvements could be made. We recorded any display modifications that occurred to us at this point.

In terms of display recommendations, all of these sessions were treated as brainstorming opportunities; therefore technological limitations were not considered. Our goal was to target areas in which our displays could have a positive impact on WD performance. The set of display alterations generated was refined later in an iterative process during the storyboarding phase of the project. The themes or target areas identified during this phase of the project can be found in Table 2.

Based on this list, items were grouped into more general topic areas. These groupings do not have distinct boundaries--they overlap to a considerable degree, but each addresses a different aspect of the WD task. This aided us in considering the task and interface design from a number of different viewpoints as we moved into the storyboarding phase of the project. Below are listed the five general topic areas.

- **Helping the WDs focus attention appropriately**. As the WD task changes throughout a mission, an appropriate focus of attention ranges anywhere from the entire battle to an individual intercept. It is important that this transition be smooth and that the system provide critical information at all levels of focus.

- **Alleviating memory demands**. During periods of low to moderate workload, a WD can comfortably hold in memory information such as fighter availability, location of friendly assets, which portion of the map is water and which is land, etc. During high workload conditions, however, memory demands increase along with overall resource demands.

- **Aiding in the development of, and minimizing interruptions in, situational awareness**. WDs mentioned the importance of having good situational awareness in nearly every incident discussed. These incidents made it clear that distractions and interruptions have the potential to greatly jeopardize a mission.

- **Decreasing workload.** A WD's workload reaches near overload conditions during high traffic periods. This forces the WD to prioritize and to handle only those tasks that are most urgent, leaving other tasks until there is time to catch up. These conditions increase stress and degrade WD performance.

- **Supporting decision processes.** The toughest part of most decisions for a WD is assessing the situation. Often, standard operating procedures or the rules of engagement dictate appropriate response or action. It is thus important that the system provide the WD with critical information, in an easily understood form, so that the situation may be diagnosed quickly and appropriate action taken.

The requirements analysis phase provided the information needed to begin storyboarding. We were able to identify key features of the task itself along with the cognitive processes and strategies used by WDs. As we began the storyboarding process, we were equipped with both specific WD functions and general target areas to address in developing display modifications.

Table 2

Themes Derived from CDM Interview Data

- Loss of Situational Awareness (SA) during high activity periods
- Current screens are too cluttered. This leads to:
  - loss of tracks
  - inability to locate distressed tracks
  - inability to locate tankers
  - inability to locate enemy jammers
  - late detection of high fast flyers
  - loss of SA
- Radar dots are same color for enemy and friendly
- Cannot track who is who in furr-ball (often because of same color radar dots)
- Slow reactions to commit against enemy due to unknown availability of fighters. Availability is based on:
  - fuel
  - armament
  - mission
  - commit against other tracks
  - aircraft type
- Unable to differentiate boundaries (what is land, what is water)
- Cumbersome switch action panel
- Looking away at switch panel often leads to loss of SA
- Ambiguous information leads to difficulty with track identifications
- Looking down at tabular track info is often a contributor to loss of SA
- Must monitor tracks to determine intent
- Unable to remember fragmented air tasking order (FRAG)
- Blinking radar dots for track history do not provide enough information
- Timely communication with pilots, and understanding what they see, is essential
- WD mode changes throughout a mission among:
  - monitoring the radio
  - sorting targets for fighters
  - monitoring target sort by fighter pilots
  - allocating resources:
    - -- to tanker
    - -- against enemy
    - -- strike packages
    - -- combat air control (CAP) points
    - -- to search-and-rescue (SAR) efforts
    - -- to escort other aircraft
  - vector aircraft against aircraft
    - -- nobody trusts the computer geometry
    - -- they all figure the 3D geometry themselves
    - -- the geometry changes slightly with each type of intercept (stern, stern conversion, etc.)
- When WD mode changes, so does information to pilots regarding big picture
- Must input extra data in order to maintain network
- When at war, knowing ROE is easy. Other times it can be confusing, or forgotten

SECTION 4: STORYBOARDING

Storyboarding is a graphically based modeling technique which is based on requirements analysis and simulation methodology (Andriole, 1989). A storyboard is a sequence of displays that represent the functions that the system may perform when formally implemented. There are a number of storyboarding methods and little agreement about which technique works best. Some system developers believe that conventional flowcharting is sufficient, while others demand a "live" demonstration of the system-to-be. There are several viable modeling methods, including the development of narratives, the development of flowcharts, methods-based data-modeling and information-engineering approaches, and those that yield working prototypes.

In our view, the most useful model is one that allows users to view precisely what they can expect the system to do. Paper copies of screen displays are extremely useful, because they permit users to inspect each part of an interactive sequence. Bolt (1984) regards screen displays as acceptable "hybrid prototypes."

An interactive storyboard and its paper equivalent provide users with the best of both worlds. The computer-generated storyboard permits them to actually experience the system, while the paper copy enables them to record their comments and suggestions. Each "run" through the storyboard set becomes a documented evaluation session filled with information for the design team. The paper copies also comprise a permanent record of the iterative modeling process, providing an audit trail of the developmental process.

For this effort we produced both paper and computer-generated storyboards. The early paper versions provided us with a means to generate new ideas. The paper storyboards were easy to modify, we were not constrained by any computer software package. As our ideas became refined we transported our paper versions into computer-generated storyboards. These latter versions showed what the final system might look like and revealed areas which needed refinement. Presenting both the paper and the computer-generated storyboards to the software engineer throughout the entire storyboarding process proved valuable. We were able to confront system limitations and thereby side-step late programming problems.

Development of the Storyboards

The list of themes presented earlier in Table 2 was a main driver for the storyboards. We wanted to address each theme with at least one storyboard recommendation. In this section, we present the final 11 recommendations and show how they are linked to the common themes derived from the requirements analysis. We describe in detail how two of these recommendations, specifically the on-screen menu and symbology, evolved.

The Final 11 Recommendations

The common themes reflect the inability of the current display to support specific cognitive processes. Although the introduction of color to the AWACS system was a vast improvement over the initial monochrome version, WDs still find it difficult to maintain situational awareness. The cumbersome switch action panel, ineffective use of color, and abstractness of the symbols increase workload, memory, and attentional demands. It was clear that any modification to the display would need to take into account these user needs and capabilities. In a project such as this, it is also essential to remain cognizant of the limitations of, and opportunities provided by, the current technology used in the target community. To create "pie-in-the-sky" system modifications that could not be implemented on the AWACS aircraft would not serve the purpose of the project. It was our goal to see our recommendations in action in order to test their effectiveness in a high fidelity setting.

The final 11 modification recommendations follow. Roughly 40 recommendations were discarded due to technology limitations or our evaluation of their expected impacts. Throughout this process we consulted with experienced WDs to determine if our modifications were appropriate. Based on feedback from the user community we discarded or modified many of our initial ideas. Figure 3 represents how the 11 final modifications link to the five targeted cognitive processes; Table 3 represents how the set of 11 addresses each of the CDM-derived themes identified during in the requirements analysis.

Cognitive Processes

Recommendations	Situational Awareness	Workload	Memory	Attention	Decision Making
Symbology	X		X	X	X
On-Screen Menu	X	X	X	X	X
Color	X		X	X	X
Quasi-Automated Nominate	X	X			X
Hand Ergonomics	X	X			X
Colorize Radar Dots	X	X	X		X
Scorecard	X	X			X
Declutter	X	X			X
Vertical View	X	X			X
Animation	X				X
Automation		X			

Figure 3. Cognitive process areas and final recommendations.

Table 3

CDM-Derived Themes and Final Display Recommendations

- Loss of Situational Awareness (SA) during high activity periods
  Symbology
  Quasi-Automated Nomination (QAN)
  Declutter
- Current screens are too cluttered. This leads to:
  - loss of tracks
  Symbology
  Declutter
  - inability to locate distressed tracks
  Declutter
  - inability to locate tankers
  Declutter
  - inability to locate enemy jammers
  Symbology
  Declutter
  - late detection of high fast flyers
  Symbology
  - loss of SA
  Symbology
  QAN
- Radar dots are same color for enemy and friendly
  Colorize Radar Dots
- Cannot track who is who in furr-ball (often because of same color radar dots)
  Colorize Radar Dots
- Slow reactions to commit against enemy due to unknown availability of fighters
  QAN
  Availability is based on:
  - fuel
  - armament
  - mission
  - commit against other tracks
  - aircraft type
- Unable to differentiate boundaries (what is land, what is water)
  Color
- Cumbersome switch action panel
  On-Screen Menu
- Looking away at switch panel often leads to loss of SA
  On-Screen Menu
  Hand Ergonomics
- Ambiguous information leads to difficulty with track identifications
  Symbology
- Looking down at tabular track info is often a contributor to loss of SA
  Symbology
- Must monitor tracks to determine intent
  Symbology
- Unable to remember FRAG (Fragmented Air Tasking Order)
  Automation
- Blinking radar dots for track history do not provide enough information
  Symbology

Table 3 continued:

- **Timely communication with pilots, and understanding what they see, is essential**
  Automation (to provide checklists or reminders)
  Vertical View
- **WD mode changes throughout a mission among:**
  - monitoring the radio
  Automation
  - sorting targets for fighters
  Automation
  Vertical View
  - monitoring target sort by fighter pilots
  Vertical View
  - allocating resources:
    -- to tanker
    Automation
    Animation
    -- against enemy
    Automation
    Animation
    QAN
    -- for strike packages
    Automation
    Animation
    -- to CAP points
    Automation
    Animation
    -- to SAR efforts
    Automation
    Animation
    -- to escort other aircraft
    Automation
    Animation
  - vector aircraft against aircraft
    Automation
    Animation
    QAN
    -- nobody trusts the computer geometry
    Animation
    QAN
    -- they all figure the 3D geometry themselves
    Animation
    QAN
    -- the geometry changes slightly with each type of intercept (stern, stern conversion, etc.)
    Animation
    QAN
- **When WD mode changes, so does info to pilots regarding big picture**
  Automation
- **Must input extra data in order to maintain network**
  On-Screen Menu
- **When at war, knowing ROE is easy. Other times it can be confusing, or forgotten**
  Automation

The 11 recommendations are:

(1) Symbology features. We proposed that tracks of particular interest be highlighted by enclosing the track symbol in a circle. Our suggested modification included both high-threat tracks and high-value assets. It is important that the WD focus attention on or notice a high-threat track as soon as it is visible by radar in order to prevent the enemy from completing its mission. We suggested that the red circle around the red high threat symbol would allow the WD to discriminate those tracks from the other hostile tracks. They would also not have to remember the track number of the high-threat track, the circle identifies it for them. It is also important that the WD remember or be aware of the location of high-value assets, such as tankers, as they are often of high priority in terms of both protection and utilization. Both of these factors, awareness of the locations of high-threat tracks and of high-value assets, lead to a better situational awareness.

(2) On-screen menu. During the knowledge elicitation sessions, WDs explained that the need to look away from the scope in order to locate the correct switch on the panel often acts as a distractor and interferes with situational awareness. Our solution was to develop an on-screen menu. We initially chose the 24 most commonly used switches and incorporated them into a panel along the right side of the scope. This allowed the WD to select a switch action by moving the trackball pointer to the panel, or menu, and "clicking" (with the trackball select button) the desired function. This eliminated the need to look away from the scope. We also anticipated that making the switch actions more accessible would decrease workload and act as an aid in focusing attention appropriately.

(3) Color. Color was added to the map as a situational awareness aid. WDs must know the location of land and water in relation to the aircraft being controlled. While the present system provides the option of a background map, it is only a magenta outline on a black screen. This requires the WD to retain some knowledge of the local land and seascape in his/her working memory. We suggested that the water be represented in blue, so that even under rapidly changing, high workload situations, there would be no confusion as to which part of the map represents land and which represents water. Our primary aim here was to increase situational awareness, but we also believed that the modification would reduce memory demands and allow the WD to focus attention appropriately.

(4) Quasi-automated nomination (QAN) feature. At times a WD is faced with situations in which s/he is controlling multiple intercepts. That is, more than one friendly aircraft is being directed to intercept more than one enemy aircraft. This often creates a very high workload situation as s/he must be monitoring each intercept; feeding information to individual pilots concerning targets, surface-to-air-missile (SAM) sites, etc.; passing information via the data link; and calculating geometries. To reduce the workload level at such times, we suggested a new function termed "nominate." This feature would provide the WD with

recommendations for intercepts. The WD would input to the system which enemy fighters needed to be intercepted and the system would respond with recommended friendly aircraft with which to conduct the intercept. The recommendation would be based on: relative position (including altitude), speed, mission, aircraft type, fuel and armament available, etc. This feature would be activated via a new button in the on-screen menu. The system would allow the WD to either accept or cancel any recommendation. The QAN feature was added to reduce workload, aid in SA, and thus allow the WD to allocate more of his/her resources to making more informed decisions.

(5) Hand ergonomics. We wanted to allow the WDs to maintain their focus on the scope and eliminate the need to move their hands away from the trackball and keyboard to reach switches. We were exploring ways of activating key switch actions from the trackball and keyboard in order to reduce workload.

(6) Radar dots. We suggested color coding the radar dots to correspond with the track symbols. This modification was intended to reduce ambiguity and thus increase SA.

(7) Scorecard. In order to reduce memory demands, we explored methods of displaying the resources available to the WD (e.g., airborne and land-based fighters, tankers, etc.).

(8) Declutter. When attempting to find the closest available tanker, a WD must scan the display looking for particular types of track identifiers. This is not a simple task when the screen is cluttered with numerous tracks. For instance, it would be highly beneficial for a WD to request that the system only display tankers. The WD could quickly locate the desired track, and return the display to the original configuration. The decluttering could be applied to tankers, friendly aircraft, all enemy aircraft, all search and rescue (SAR)-qualified aircraft, all distressed tracks, all enemy jammers, etc. It would not only increase SA, but would lower the demands on memory and attention as well.

(9) Vertical view. The scope currently presents a god's-eye-view of the airspace. This can create a deceptive picture for the WD. In a situation where two aircraft are flying in the same position, but one is at a lower altitude than the other, the WD sees only one aircraft on the scope. We investigated implementing an additional vertical view of the airspace, so that the WD would have a better overall picture of the entire airspace.

(10) Animation. WDs describe the process of calculating when and where an intercept will occur as "mental gymnastics." We would describe it as mental simulation (Klein & Crandall, in press), where one begins with the present state of affairs and mentally plays events forward in time in order to develop expectancies and formulate a plan. This is an important skill in the development of good situational awareness but requires a fair amount of mental effort. We proposed that the system could perform these "mental gymnastics" and display the

potential intercept in an animation mode, thus saving the WD's mental resources.

(11) Automation. This was a precursor to the nominate function. We investigated methods of reducing the WDs' workload during high time pressure, high-stress situations. We were exploring an option that the individual WD would activate, in which specific portions of the job would be automated.

Because of limited resources, it was necessary that we determine which of the final 11 recommendations could have the greatest impact. It was determined that the first four mentioned above, symbology, on-screen menu, color, and the QAN, met our criteria. Therefore, these four comprise the revised system which was coded into the simulation facility system.

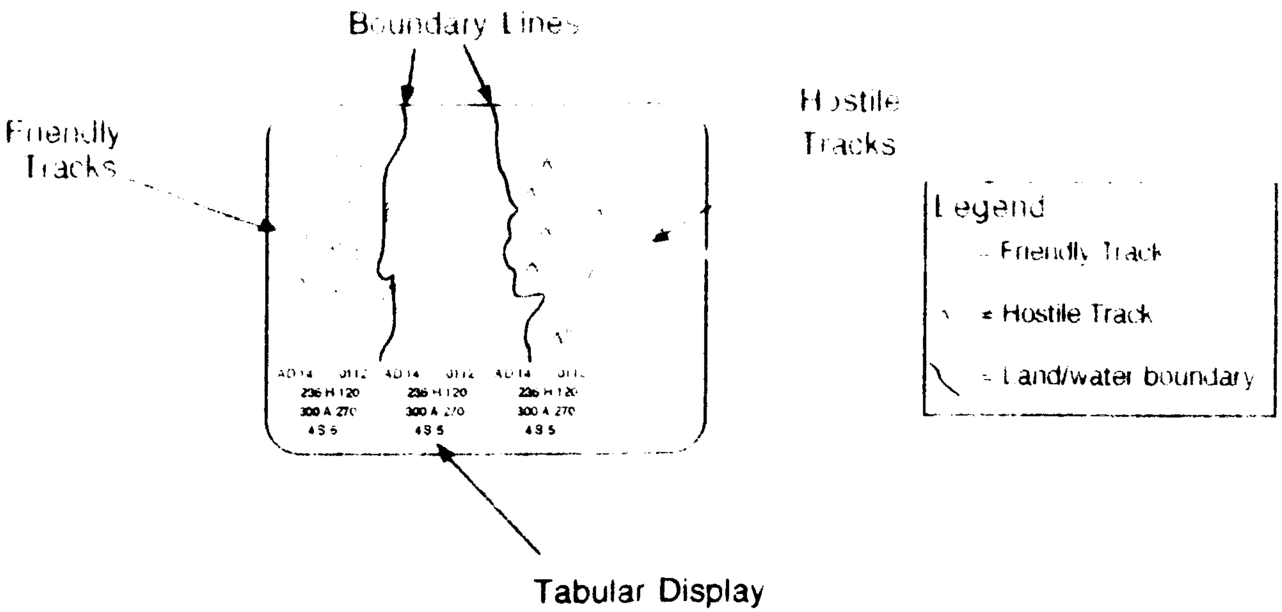
The Evolution of Two Modifications

In this section, we provide a detailed account by which we arrived at the on-screen menu and symbology modifications. Storyboarding/prototyping can be an arduous process in which the end result often appears obvious. The road taken is usually more interesting than the destination. Readers not interested in storyboarding may wish to move on to Section 5; readers interested in the process by which these two recommendations came to life may find the next section of interest.

Modification 1: The on-screen menu. The WD monitor is situated between the switch action panel and the feature category select panel. The WD must look away from the monitor, or scope, in order to input a system command (either via a switch action or feature category selection). The interview data consistently showed that this looking away from the scope was a major contributor to the loss of SA. Maintaining SA is the most important function for weapons directors. To help the WDs maintain SA, we needed to allow them to keep their eyes on the scope while executing important switch actions. To do this we sought to develop an on-screen menu which would contain the most-used switch actions.

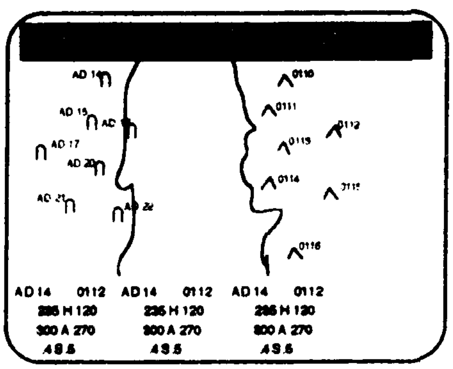
Current AWACS display. The screen represented next is a simplified version of the current AWACS configuration prior to any modifications.[2]

[2]In this format we are unable to represent the black background of the current display. Therefore, in the following storyboards, what is black is actually white, and what is white is actually black.



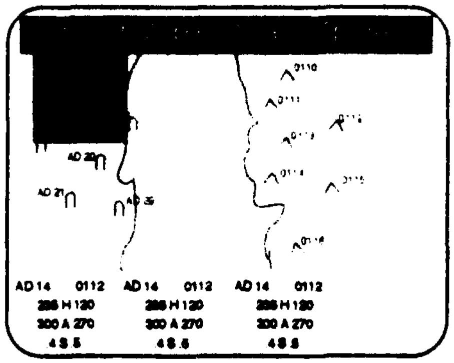
The Tabular Display (TD), located at the bottom of the screen, is where detailed track information is presented. The WD can select a track(s) of interest and the system displays the heading, altitude, and speed of that track. The system also calculates optimum geometries for intercepts. That is, after a WD has informed the system that a particular friendly track is to intercept any other airborne track (this is called pairing), the system calculates the optimum geometry for the interception of the two tracks and displays this in the tabular display.

Pull-down menu. The storyboard below represents our first attempt at an on-screen menu. From our interviews we determined that there were four major areas in which the WD inputs commands. These areas (Intercept, Display, Clean, Options) appear in the menu.



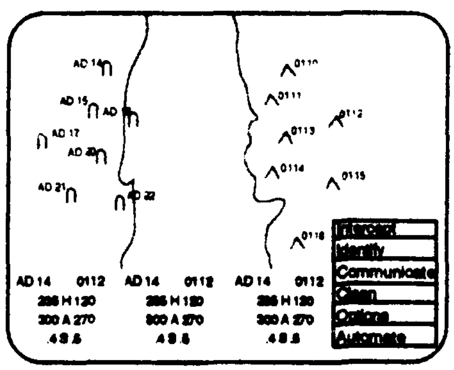
These main menu headings were selected based on the various tasks a WD performs during a given mission. The WD is typically conducting an intercept, selecting display features, cleaning up the display, or performing other options (like initiating an aircraft down point).

Expanded pull-down menu. With this display, the WD would select a menu heading, and a pull-down menu would appear. In the pull-down portion of the menu would be the specific switch actions which are appropriate for the selected mode. This menu would remain in the pull-down state until another menu heading was selected to enable the WD quick access to any subsequent switch actions.

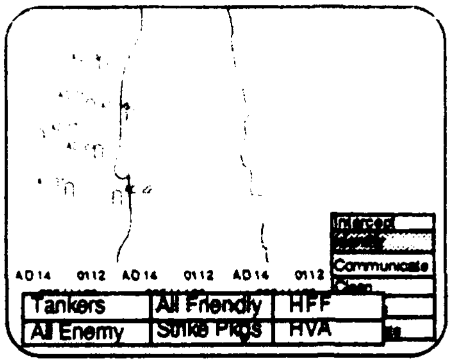


Constructing this menu structure helped us to understand the various switch actions and how they group together, but it was not an option we pursued any farther. The main problem with the menu was that when a WD selected a major heading, the pull-down portion of the menu would cover up part of the display. It was impossible to tell if the portion which was under the menu contained the track(s) of interest. This reason alone was enough to disqualify this structure, but we were also sensitive to the number of selections the WD had to make. For instance, to initiate a particular switch action the WD would need to select a menu heading, locate the desired action within the pull-down menu, and select it. This was not an improvement over the current system.

Exploding menu. We then modified the menu by placing it in the tabular display area and using an exploding, or overlaying, menu.



Expanded exploding menu. With this iteration we began to add new functions to the menu. For example, Identify could be used to "declutter" the screen. The WD could select Identify and an overlaid menu would appear. This menu would contain specific types of tracks which the WD could select and the system would temporarily display only those tracks.

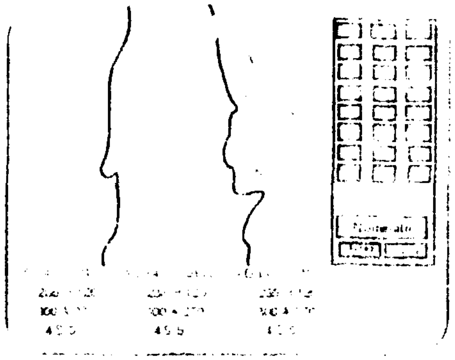


At this point we were still unsure as to whether the extra switch actions we proposed could be coded into the system, yet we continued to pursue a menu structure that would accommodate them. The above menu structure partially covered the TD when the exploded menu was present. This was not seen as a problem. We found during the requirements analysis that while selecting a switch action the WD is not attending to information in the TD. This menu structure also offered a great deal of flexibility. We were interested in new functions to the system (animation, automation, declutter, Quasi-Automated Nomination) and this menu allowed us to incorporate them. Also, certain switch actions are associated with others, as we found when developing the pull-down menu. The exploding option allowed us to maintain those groupings to better allow the WD to conduct follow-on switch actions.

Although we felt that the exploding menu was clearly a move in the right direction, it was not the answer. As was the case with the pull-down menu, the number of inputs to the system was higher for the exploding menu than for the current switch action panel. For any one input via the switch action panel, the WD was performing two with this menu structure. Again, this menu structure was eliminated based on the number of inputs necessary to perform switch actions.

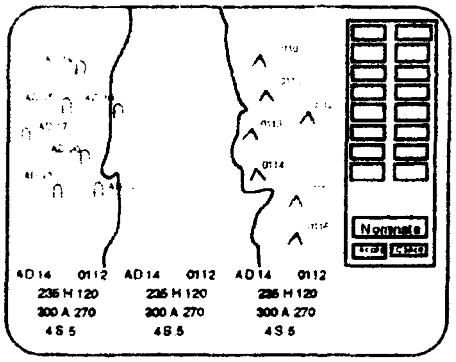
It was clear that any embedded menu structure would not suffice. We maintained our desire to place switch actions on the screen and make them accessible via the on-screen cursor. It should be noted that touch-screen displays are beyond the technology currently available aboard the AWACS aircraft. We were confronted with where on the screen to place all the essential switch actions, including some we added, and not take up essential screen space.

Initial 24 button on-screen menu. Our first step toward the final recommendation appears below. The on-screen menu contains 24 buttons representing the 24 most used switch actions. These 24 were selected by a subject matter expert (SME) and their placement within the menu was based on frequency of use. Note that at the bottom of the menu is a button labelled 'Nominate.' This is the button which activates the Quasi-Automated Nomination feature.



This menu structure, in theory, became the On Screen Menu modification. The buttons were decreased from 24 to 16, thus narrowing the menu, and the space between the top 16 buttons and 'Nominate' was utilized as a "working space" for switch actions that required follow-on inputs. To minimize training time, and decrease negative transfer, we used the switch action panel as a guide for the placement of the buttons relative to one another rather than trying for a more logical arrangement

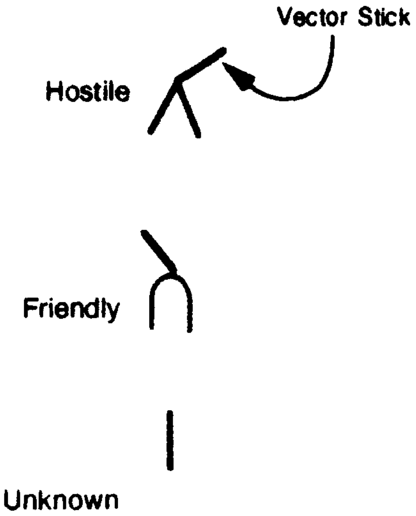
Final version of the on-screen menu.



Appendix B contains a full page storyboard of the final modification.

Modification 2: Symbology. We talked earlier about the road taken to a final display modification as being more interesting than the final product. This is certainly true for the symbology. The next few pages present the evolution of what could be the most necessary upgrade for the next generation AWACS system.

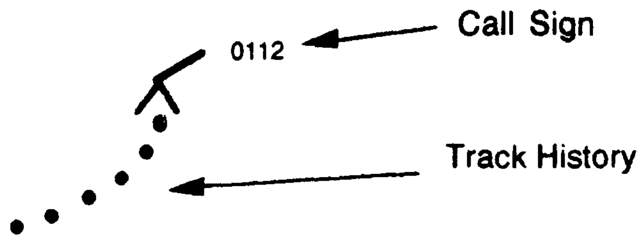
Current symbology. The current symbology appears next. Essentially, it conveys whether an aircraft picked up on radar is friendly, unknown, or hostile, the direction it's flying, and its relative speed.



Vector Stick - points in the general direction the aircraft is flying (heading) and its length gives a rough indication of the aircraft's speed (the longer the vector stick the faster the aircraft is flying).

Symbol - the color and shape represent whether the identified aircraft is considered hostile, unknown, or friendly.

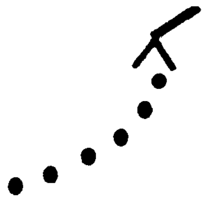
There are two important informational items that are associated with this symbology: the track call sign and the track history. The track call sign identifies the track via an alpha numeric string. This call sign is used to communicate with the friendly aircraft and to identify the enemy and unknown aircraft. The track history is a series of radar dots which display the last minute of track history. These dots, six in all, show the position of the radar dot for the last 6 ten-second intervals. These dots flash on the screen every two seconds. This may sound confusing, and it is. When aircraft converge (and their respective symbols converge) it appears as though there are flashing radar dots everywhere. The WDs refer to this as a "furr-ball;" it is nearly impossible to determine what is what.



It was our goal to increase situational awareness and decrease the demands on memory by placing more information on or near the symbol.

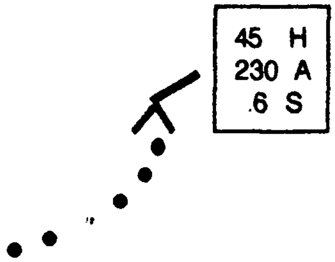
The following ideas are fairly independent. These were generated during the brainstorming sessions during and after the interviews. These began to lay the groundwork for the more sophisticated recommendations which were created later in the process.

Constant display of radar dots.



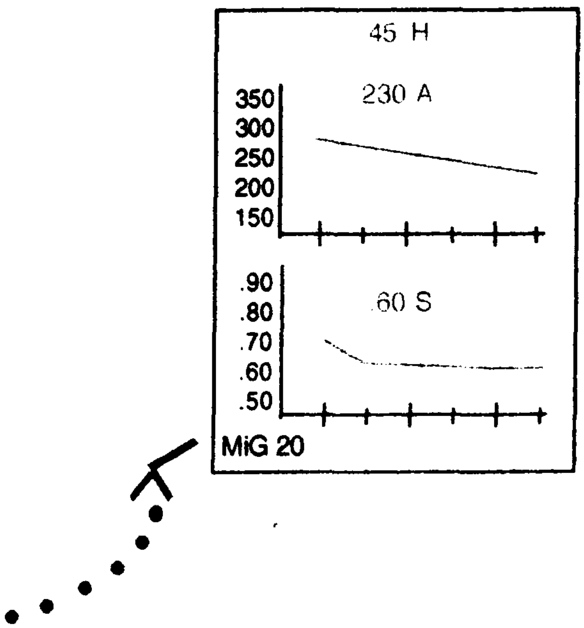
A rather easy fix would be to leave the radar dots on the screen instead of flashing them every two seconds.

Alpha numeric information box.



The heading, altitude, and speed information is currently located in the tabular display. We wanted to place it closer to the symbol to allow the WD to keep his/her eyes on the scope, not searching through the TD. The WD could "click," using the hook button on the track ball, on any track and an information box would appear next to the symbol. The box would remain on the screen for either a set duration or while the hook button is depressed.

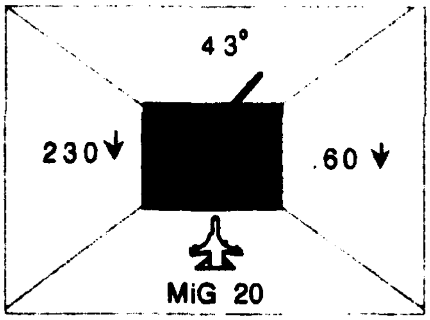
Graphical information box. A WD must attend to a particular track for some time to determine trends. The vector stick and track history dots do an inadequate job for relaying history information. Often the most important history is that of altitude. The current system does not present altitude history.



Graphically displaying the recent speed and altitude information would greatly increase the WD's SA. At this point we also included the aircraft type. Obviously, for enemy aircraft without a visual identification by a friendly pilot, this is hypothetical. Yet, this is another piece of information that will help the WDs make better allocation decisions.

Up to this point we had addressed the need for more information to be located next to the symbol. Allowing the WD to look next to a track for the current heading, altitude, and speed, as well as the recent history, greatly increases the amount of time s/he is attending to the track and not searching through the TD for current information. The next few steps were an attempt to place this information directly on the symbol.

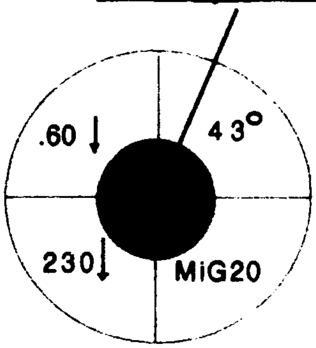
Box symbol.



Our first attempt at placing information on the symbol incorporated a box which contained the current symbology in the center and (clockwise from the top) the heading, speed, aircraft type with icon, and altitude. The arrows adjacent to the altitude and speed indicate the current trends of the track. The length of the arrow could be used to indicate the strength of the trend.

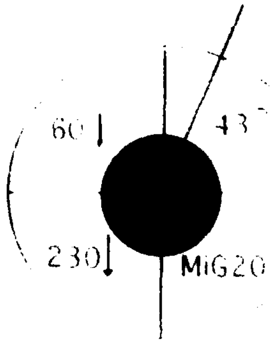
Although this iteration was an improvement, primarily due to incorporation of the icon of the aircraft and the arrows showing trend, this option possessed numerous shortcomings. First, it took up too much screen space. Within the square itself there was a large amount of unused space. Also, the vector stick protruding out of a square box could be misleading as it moves around the square. So, we continued to generate ideas. The most obvious was to change the square to a circle.

Circle symbology.



The same information except for the icon is contained in the circle symbology as was contained in the square, yet less screen space is consumed and any problems with the vector stick are resolved.

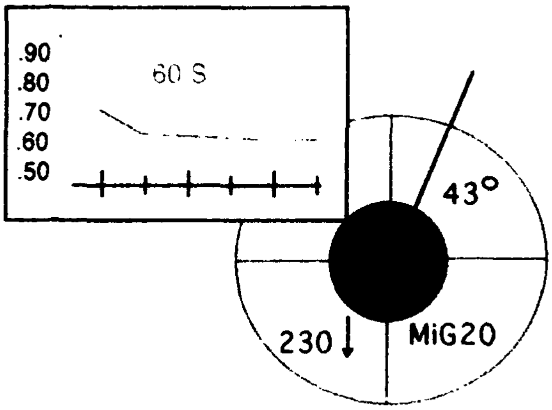
We began investigating other alternatives to this concept.



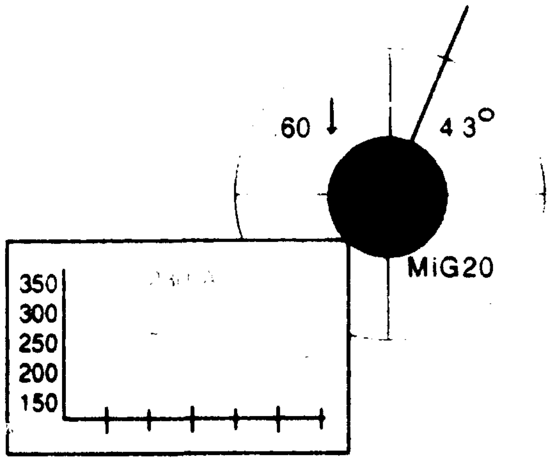
This version is similar to the previous one except that an icon of the aircraft has replaced the symbol. The extensive use of color to indicate hostile tracks makes the use of an abstract symbol redundant. Therefore, we replaced the abstract symbol with an icon of the hostile aircraft to improve situational awareness.

The above symbology represents to the user the current heading (via the vector stick and exact heading in the circle), the aircraft type (via the icon and identifier in the circle), the current altitude (with directional arrow to indicate recent trend), the speed (with directional arrow to indicate recent trend), and the track identification number (0112 in this case) of the aircraft. We still wanted to convey the recent trends of the track more effectively. Below is our final modification to the circle symbology.

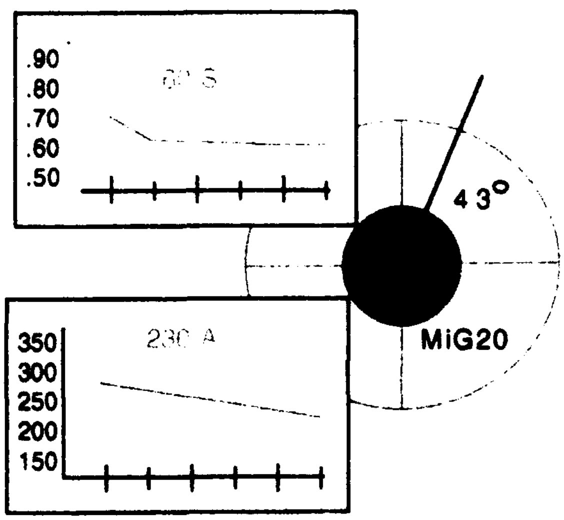
Circle symbology with icon and graphical track history. The symbol at rest would appear exactly as it does above. Yet, if the WD were to query the system for more information, the system would respond with a detailed description of recent track history.



Previously we showed how we could graph the data collected by the system for the last 60 seconds of track history. For the circle symbology the WD could "click," again using the hook button on the trackball, on any particular quadrant of interest. In this example, the current speed is shown as well as the recorded speeds of the track for the previous 60 seconds.



Another example using the graph to display track history. In this case, the WD has selected the altitude quadrant, and the system is displaying the corresponding altitude data for the previous 60 seconds

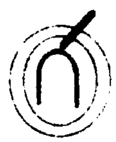


A final example to show how history data from multiple quadrants could be displayed. This data could stay on screen for a designated length of time. We did not investigate methods for either displaying the information for designated lengths of time, permanently, or while the hook button is depressed.

The testing and modifying of the symbologies could easily be a project by itself. The abstractness of the symbol, the vector stick, and the blinking radar dots do a very poor job of relaying information to the WD. Our cognitive task analysis pointed out that during high traffic periods it is more common for WDs to lose SA. This loss of SA can often be traced to the symbology. The WD must look away from any track of interest, search through the TD to find that track number, and then interpret the alpha numeric data that are displayed. It doesn't take long to get behind and thus forget what they are doing. Short-term memory just cannot keep up. Once SA is lost the current symbology does little to help the WD regain it. The placing of more information, displayed in a more effective way, would aid the WD in regaining "the picture."

Final version of symbology. Our final recommendation was rather extensive. We recommended radical changes to the current system, completely discarding the abstract geometric symbol. Within the time frame and budget limitations of this effort, we were unable to incorporate our preferred modification into the simulation system. Therefore, we utilized the knowledge gained during the Cognitive Task Analysis (CTA) and determined that at a minimum we needed to call the WD's attention to the most important friendly tracks and the most threatening enemy tracks. This would allow the WD to quickly distinguish these tracks from others, and therefore respond more quickly to protecting the high-value friendly assets and thwarting the mission of the high-threat enemy aircraft This could not have been accomplished had we not pursued other alternatives to the symbology. As we progressed through the above alternatives, we began to understand more about the WD tasks and how they could benefit from extensive symbology upgrades. The graphic below represents the bare minimum which needs to be done to increase the SA of already over-saturated users.

Enemy High-Threat Track



Friendly High-Value Asset

Following the usability testing of the above modification, the number of circles around the symbology was reduced to one. See Appendix B for a full page storyboard of the final modification.

Our intention was to increase situational awareness, lower workload, allow for better allocation of attention, decrease the demands on memory, and therefore provide for better decision making. The on-screen menu and symbology modifications were major steps toward this goal.

SECTION 5: USABILITY TESTING WITH A COGNITIVE PERSPECTIVE

The requirements analysis provided the basis for identifying the HCI features which could have the greatest impact. Throughout the storyboarding phase we revisited the user community to solicit feedback regarding the proposed changes. But, this type of feedback is really only educated guess work; no paper or computer storyboards can take the place of actual user testing. That is, any design team must test the proposed system in its fully coded state with knowledgeable users prior to any evaluation. Failure to take this step could leave good display ideas lying on the evaluation room floor as a result of poor implementation.

As with any system, users will find new ways to employ the options given to them. It is important not only to document these unanticipated uses, but to document the errors as well. Errors and innovative uses are often the key to determining if a design is even headed in the right direction. If the users simply find ways to bypass the new design options in order to utilize the old ones, then the entire design needs to be rethought. If the errors occur during use of the new design, these errors need to be addressed and modifications made.

Two pilot studies were conducted to test if the experimental design was sound, if the system was stable, and if the proposed system was achieving the desired effects. The pilot studies were conducted two weeks apart; the second pilot study took place four weeks prior to the actual evaluation to provide a reasonable amount of time in case major system modifications were necessary.

Pilot Study Number One

The initial pilot study showed that more training was necessary on the revised system than had been planned. The experimental design called for equal training on both systems, but it became obvious that the WDs needed more time using the revised system during a simulated mission. Although they picked up the revised system rather quickly, our original training schedule was far too optimistic.

The training and mission sessions were highly interactive. We used the pilot studies as an opportunity for on-line knowledge elicitation sessions. When we noticed the WDs having difficulties, we would step in and help. All problems and comments were documented so they could be addressed later.

The initial pilot study was extremely informative and highlighted certain problem areas, yet it also showed us that we were on the right track. The menu, we believed, was a good idea, the current implementation of it was not. It is important to note that had we simply relied on user feedback as our guide, we would have discarded the menu altogether. None of the three participants in the initial piloting felt the menu was of any value. We also learned that we needed to tracks without causing them to be distractors.

It should also be noted that one of the major modifications to the system could not be implemented prior to the first pilot study. The Quasi-Automated Nomination (QAN) feature could not be properly coded prior to the pilot study. Therefore, no user testing could be performed on the nominate procedure. The QAN feature was partially coded prior to the second pilot study, yet proper testing still could not be performed due to the incompleteness of the feature. In the end the QAN feature was incorporated into the final design without adequate user testing.

Modifications Following the First Pilot Study

We determined that three modifications to the menu needed to be accomplished prior to the second pilot study in order for it to be effective:

- The number of switch actions in the menu needed to be decreased (to speed up learning and thus decrease time spent looking for switch action).

- A mark needed to be placed on the scope at the last pointer position prior to entering the menu (to provide a memory and attention aid for the WD when coming out of the menu).

- The pointer needed to enter the menu at the same position each time. This would increase the rate of learning for button location (the same motions would be used each time a particular button was selected).

The first of these was easy. We polled WDs as to which switches they would most like to have in the menu, and which switches they used the most. On this basis, we narrowed the number of buttons to 16 (two rows of eight). To increase the rate learning, we used the current switch action panel as a guide for the placement of the buttons relative to one another.

The ability to leave a marker on the screen prior to entering the menu seemed to be the most difficult. How could we know where the WD was looking when s/he decided that a switch action was needed? We determined that we had to somehow ask them to leave a marker on the screen in the area of interest. We determined that the best way to utilize the menu, and place a mark on the scope, was to allow them to "pop" into the menu. In other words, when they determined that a switch action was necessary, they could activate the menu from their current pointer position. To do this we utilized the middle mouse button on the track ball. The WD would simply press the middle mouse button to activate the menu and a mark would be placed on the scope at the last pointer position. After the desired switch action was selected, the WD would press the middle mouse button again to "pop" the pointer back to the marked location. This method not only allowed for a rapid return to the desired location on the scope, it also addressed the third problem area. The pointer was now "popping" into the menu at the same point each time (middle, left edge). Therefore, the relative location of the buttons to the pointer position was the same each time the menu was activated.

Prior to the second pilot study we also reduced the number of circles around the symbologies of the high-threat and high-value tracks to one. This satisfied our criterion of making the tracks stand out, and made them less distracting.

Pilot Study Number Two

The second pilot study indicated that the modifications described above greatly increased WD acceptance of the revised system. As stated earlier, the pilot studies were scheduled to test the experimental design, system integrity, and prototype usability. After the second pilot study, it was determined that the system and experimental design were sound.

The useability testing provided by the pilot studies was critical to the overall success of the project. Up to this point, we had no way to determine how users would utilize the system during an actual mission. We could not simulate the time pressure, track saturation, or other real-life elements which confront WDs during actual missions. By observing the users and listening to their comments we were able to feature the revised system prior to the actual evaluation. These modifications were a central reason for the success of the revised system.

SECTION 6: EVALUATION

This project was designed to evaluate the CSE approach taken, not simply to design and construct an interface. The evaluation was an integral part of the project from its inception, and was one of the primary reasons for selecting the AWACS WD station because it afforded us the opportunity to conduct a carefully controlled evaluation using a high fidelity simulation.

The major evaluation goal was to examine whether the revised interface out performed the current interface where it counts–in outcome measures such as keeping hostile aircraft away from their targets, shooting down more hostile aircraft, and having fewer friendly aircraft shot down.

There was also an important secondary evaluation goal, to see whether we had accomplished our objectives of reducing workload and increasing situational awareness. The concept of a Cognitive Systems Engineering approach is to identify interface design objectives in terms of the cognitive processes that should be supported. We had identified one objective as increased situational awareness, and another as the reduction of workload by diminishing memory and attentional requirements. Situational awareness was improved by providing red circles around threats and green circles around important assets, and by using color to show water and land masses. Memory demands were reduced by using an on-screen menu so that the WD didn't have to look away from the screen so often and then try to reorient to the ever-changing radar screen. The evaluation was explicitly designed to obtain measures of situational awareness and of workload. These measures were indirect and unobtrusive, embedded within the tasks the WDs were performing.

In this section, we first describe the evaluation methodology and then present the findings and our interpretations.

Methods and Design

All participants were certified AWACS WDs from Tinker AFB, Oklahoma. Their simulator and flight experience ranged from 266 hours to 4300 hours.

Participants. Eight groups of three Weapons Directors participated in the study. The data from six of the WDs could not be used due to base-wide computer system failures at Brooks AFB which disrupted the experimental sessions for two groups. An additional WD was dropped from the study when it became clear to the experimenters that this WD was not motivated learn and operate the new system. This left us with 17 WDs.

Apparatus. The experiment was conducted during June and July 1992 at the AESOP facility. The facility is described in detail in Schiflett et al. (1990). The AESOP facility has four crewstations configured as AWACS WD consoles. These consoles have high resolution graphics displays, modular switch panels with programmable switch function, communication panels, QWERTY keyboards, and trackballs. Several high fidelity, high resolution video terminals serve as consoles for simulation pilots, ground controllers, and investigators. The AESOP computer systems consists of: a cluster of two VAX 11/780s, two MicroVAX IIIs, and a VAXstation III/GPX; four high resolution, color graphics Silicon Graphics 4D/50 workstations, and multiple disk drives, tape drives, and printers. A 10-node communication network provides audio communication during simulations.

The WD consoles are configured to closely resemble the WD station aboard the AWACS aircraft. The simulators differ from the actual AWACS in four areas: (1) the use of a standard QWERTY alpha numeric keyboard instead of the AWACS crewstation ABCD keypad, (2) the use of a smaller trackball, (3) the use of a radically different communication panel, and (4) the availability of only the most commonly used switch actions.

The physical arrangement of simulator equipment also mimicked the AWACS aircraft. The WD consoles were placed next to each other, so that WDs were able to communicate with each other visually as well as verbally. The SD console was also in close proximity, directly behind the WDs. Simulation pilots were located in a soundproof room separated from the WD stations.

The simulation was interactive in that the WDs' actions early in the scenario had direct impact on future events. For example, if many of the WDs' fighter aircraft were destroyed at the beginning of the scenario, there would be fewer resources left to fight as the scenario progressed and the battle became more intense.

Experimental procedures. Each participant spent two days at the AESOP facility. On Day 1, participants were trained on both the current and the revised systems. Training always began with the current system interface. The differences between the simulator and the actual AWACS WD crewstation were discussed and the participants were given a 2.5 hour session in the simulator to practice using the current system. A ten-minute break was scheduled 1.5 hours into the simulation and a half-hour debrief was held after the session to provide ample opportunity for questions about and discussion of the current system interface as represented in the simulator.

After a lunch break, training on the revised system interface began. First, members of the research team described the four modifications presented in the revised interface using paper representations. After participants understood these changes, they were given an opportunity to practice using the new system interface. This half-hour practice session focused primarily on the step-by-step use of the on-screen menu and QAN procedure. Participants were then given a three-hour scenario in which to practice using the revised system. Participants were repeatedly told to practice using all the features of the new system, because when they became caught up in the simulation, they often failed to do so. Often participants would become more concerned about performance (winning the war) than learning to use the revised system. In an effort to counteract this tendency, a scheduled break mid-way through the scenario was spent discussing the features of the revised system interface and attempting to ensure that all participants knew how to use them. In addition, the half-hour debrief after the practice session was used to address problems participants had in using the revised interface features. Even so, there were differences in participants' ability to use the on-screen menus and, particularly, the Quasi-Automated Nomination procedure.

On the second day of the experiment, performance was assessed on both systems, one in the morning session and the other in the afternoon. Order of presentation was counter-balanced. After a 30-minute warm-up session, the WDs were presented with a 3.5 hour simulation in which the situation escalated from peacetime to war. The NASA-Task Load Index (TLX) (Hart & Staveland, 1988), a subjective measure of overall workload, was administered immediately following the simulation. A 15-minute debrief was held and the WDs were released for lunch. The afternoon session in the simulator was identical to the morning session with a slightly altered scenario for the simulation. At the end of the afternoon session, the Subjective Workload Dominance technique (SWORD) (Vidulich, 1989) and a questionnaire developed by Klein Associates, Ratings of Revised System Impact, were administered in addition to the NASA-TLX. SWORD is a measure of subjective workload that allowed us to examine components of the WD job. Participants were asked to rate and compare the workload of three specific tasks: reinitiating symbology, pairing air defense fighters (ADF), and conducting an intercept. The Ratings of Revised System Impact questionnaire addressed specific areas of cognitive processing (e.g., attention, memory). A final debrief was held to answer any remaining questions and record any comments or insights the WDs offered.

Both scenarios used during the 3.5 hour test simulation, named Saturn and Krypton, were high workload Defense Counter Air (DCA) scenarios. The situation escalated rapidly from peacetime to war. The primary mission of each WD was to defend the friendly airbases by directing friendly fighters to intercept any hostile aircraft in the area. Over the course of the scenario, four waves of hostiles were presented to each WD. It was expected that by the time the fourth wave arrived, the WDs would be in a near overload situation. The Saturn scenario was derived from the Krypton scenario by rotating the Krypton scenario, changing the names of scenario players, and using a different background map. Both scenarios were essentially the same. The purpose of changing the appearance of the scenario was to lessen the impact of practice.

We originally intended to counterbalance the time of day and the display type with respect to scenario used. However, base-wide computer system failures resulted in the loss of data for two groups of WDs, thus disrupting the counter-balancing scheme. Table 4 displays the ordering of the scenarios and system's presentation.

Table 4

Order of Presentation of the Scenarios and Systems

Display/Scenario	# of A.M. Participants	# of P.M. Participants
Revised system/Saturn	6	2
Revised system/Krypton	6	3
Current system/Saturn	3	6
Current system/Krypton	2	6

Note that more participants were tested on the current system during the afternoon session. As we would expect the effects of practice to be greater in the afternoon sessions than the morning sessions, this results in a bias against the revised system.

Dependent Measures

The hypothesis at its most general level was that the alternative system would improve cognitive functioning and, in turn, performance. We selected a battery of dependent measures that enabled us to assess the effects of the display revisions. These measures were on several dimensions: outcome performance, cognitive processing, workload and situational awareness. Each of these is described below. Perhaps the most straightforward dimension to measure is performance.

There are behavioral measures accepted by the WD community as valid indicators of performance. These measures provide a bottom-line indicator of who is winning the war. The 11 outcome measures employed here were adapted from this set.

In addition, we utilized process indices. We identified four tasks which would be considered embedded tasks in order to measure workload. We also included four measures that we believed would give us some indication of situational awareness.

Two subjective measures of workload, NASA-TLX computer version and a paper and pencil version of SWORD, were administered. Finally, Ratings of Revised System Impact, a questionnaire concerning specific aspects of the alternative system, was used. See Table 5 for the full set of dependent measures recorded during the testing sessions.

Additionally, expert ratings of WD performance, based upon the outcome measures described above, were done after all testing sessions had been completed.

Table 5

Set of Measures

Outcome Measures	Process Measures
Hostile strikes completed	• Workload Measures
Number of hostile penetrations	Response to SD inquiry:
Penetration depth - penetrators only	% correct
Penetration depth - all	% only acknowledged
Hostile shot down	% no response
Friendlies shot down	avg. response time
Hostile to friendly kill ratio	Response to visual alerts
Friendlies lost to low fuel	% responded to
Friendlies shot by friendlies	average response time
Total friendlies lost	Intercept approach specified
% fired missiles that missed	• Situational Awareness Measures
	Recorrelations - % correct
	Time symbology incorrect
	Air refuelings
	AC return-to-base
	Airborne order/scrambles ratio
Workload Ratings	**Ratings of Revised System Impact**
NASA TLX - overall workload	
SWORD	
reinitiating	
symbology	
pairing air defense fighters	
conducting an intercept	

Outcome measures. Most of the outcome measures were taken from a larger set developed by the AESOP facility during previous experiments with WDs. These are intended to measure task outcomes in contrast to processes. They provide a bottom-line indicator of who is winning the war. Descriptions of the 11 performance measures and hypotheses regarding revised system impacts are listed below:

Measure/Description:	Desired Impact of Revised System:
1. Hostile strikes completed: The number of times an enemy aircraft was successful in completing an airstrike on a friendly base.	Fewer hostile strikes completed.
2. Number of hostile penetrations: The number of times an aircraft penetrated friendly airspace.	Fewer hostile penetrations.
3. Penetration depth--penetrators only: Of those hostile aircraft that did enter friendly airspace, the average depth of penetration.	Less average depth of penetration.
4. Penetration depth--all: The sum of the distances penetrated by each hostile aircraft, divided by the total number of hostile aircraft (whether they had penetrated friendly airspace or not).	Less average distance penetrated by all hostile aircraft.
5. Hostiles shot down: The number of hostile aircraft destroyed by friendly aircraft.	More hostiles destroyed by friendly aircraft.
6. Friendlies shot down: The number of friendly aircraft destroyed by hostile aircraft.	Fewer friendlies destroyed by hostile aircraft.
7. Hostile to friendly kill ratio: The ratio of hostiles destroyed to friendlies destroyed.	Better kill ratio (more hostiles destroyed, fewer friendlies destroyed).
8. Friendlies lost to low fuel: The number of friendly aircraft destroyed by fuel depletion.	Fewer friendlies destroyed by fuel depletion.
9. Friendlies shot by friendlies: The number of friendly aircraft destroyed by friendly fire.	Fewer friendlies destroyed by friendly fire.
10. Total friendlies lost: The total number of friendly aircraft lost.	Fewer total friendly aircraft lost.
11. Percent fired missiles that missed: The percentage of missiles fired that did not reach the intended target.	Lower percentage of missiles fired that missed.

Process measures: Workload. The rationale behind the use of embedded tasks to measure workload is that, as workload increases, performance is likely to degrade on tasks that fall lower on the priority list. For example, a WD is occasionally asked to update the Senior Director regarding specific information about his/her lane. While it is important to respond to these requests for information, it is more important to be monitoring an intercept and communicating with pilots. As workload increases, we would expect to see more instances of the WD responding to such inquiries with an acknowledgment of the question (putting the answer off until later), incorrect answers, and, in instances of extreme workload, no response at all. We hypothesized that the revised system would decrease workload, and thus expected to see improved performance on the embedded tasks with the revised system. The three workload measures and associated hypotheses are listed below:

Measure/Description:	Desired Impact of Revised System:
1. Response to SD inquiry: Over the course of each test scenario, WDs were asked ten questions by the SD. We recorded the reaction time and whether the response was correct, incorrect, no response, or acknowledged but not answered immediately.	More timely, correct responses.
2. Response to visual alerts: Visual alerts in the form of time checks, distressed tracks, and messages from the Senior Director. We recorded the reaction time and the percentage of visual alerts to which the WD responded. In this case we did not attempt to capture accuracy of response, due to the difficulty in defining accuracy. Reaction time was measured from the time at which the alert was displayed to the time at which the WD cleared the alert.	Greater percentage of responses and shorter reaction times.
3. Intercept approach specified (e.g., cutoff, stern conversion): The WDs were informed that they were in a data link, requiring them to enter the approach specified after every commit switch action. This involves an additional switch action that is often neglected during periods of high workload. We recorded the percentage of times the approach was specified when the commit switch action was used.	Greater percentage of commit switch actions with the approach specified.

Process measures: Situational awareness. Situational awareness is an ambiguous construct that has grown out of the aviation domain. This term has been used to describe a state in which one is aware of the overall situation, and at the same time incorporate incoming information, form expectancies about the situation, and react in a timely and appropriate manner. In an attempt to capture this, we chose measures that would provide an indication of how well the WD tracked the situation and planned ahead appropriately. Many of the display changes were made specifically to better support the WDs' situational awareness. Therefore, we would predict that the revised system better supports situational awareness and thus expect to see improved performance on these measures with the revised system. Measures of situational awareness and associated hypothesis are presented below:

Measure/Description:	Desired Impact of Revised System:
1. Reinitiating symbology: Radar dots and their corresponding symbology become "disconnected" as an aircraft deviates from its predicted course. It is important that the WD notes this and "reinitiates the symbology" to the radar dot, thus recorrelating the track. This involves a switch action and the use of the track ball to drag the symbology back onto the correct radar dot. We recorded the percentage of correct recorrelations and the total time that tracks were uncorrelated.	Greater percentage of correct recorrelations and less time with uncorrelated tracks.
2. Refueling: In general, it is more efficient for aircraft to refuel in the air (provided they are equipped with sufficient armament). In order for the WD to make a decision whether to refuel an aircraft in the air or return it to a base, s/he must have an awareness of what other resources are available in terms of airborne fighters, fighters available on the ground, location of the tanker and the base relative to the aircraft in question, etc. We recorded the number of aircraft instructed to refuel in the air and the number of aircraft that were sent to the base.	Greater number of air refuelings.

3. Airborne orders/scrambles ratio: It is more desirable to issue airborne orders (inform the ground that additional aircraft will be needed at a specified time in the future) than it is to scramble aircraft from the ground (inform the ground that additional aircraft are needed immediately). In order to issue airborne orders, however, the WD must plan at least five minutes into the future.	Higher ratio of airborne orders to scrambles.

Workload ratings. We chose to administer two subjective workload measures in order to get a more complete understanding of the effects of the revised system. Subjective instruments are believed to be more sensitive than performance or embedded task measures of workload in some situations. Whereas outcome measures are sensitive only under overload conditions, subjective workload measures reflect increases in effort or capacity expenditure in both overload and nonoverload conditions. The two measures utilized here, NASA-TLX and SWORD, provide very different approaches to the measurement of subjective workload. The strengths of each are described below.

NASA-TLX emphasizes the multidimensional aspect of workload. People are asked to consider components of workload rather than an overall global estimation of workload. Each of these dimensions or components is rated on a subscale. The subscales are then combined using a weighted average, thus providing an overall workload score. Experimenters then have access to ratings both for the individual dimensions of workload and an overall workload score. NASA-TLX utilizes absolute judgment in that people are asked to estimate the workload of a specific task without a baseline or referent task.

SWORD is a unidiminsional technique in which people are asked to estimate workload relative to another task. This allows the operator to decide which elements or dimensions are relevant factors in the overall workload of a specific task. The resulting workload ratings represent workload on a ratio scale relative to all other rated tasks. In the present study, participants were asked to rate the workload of three specific tasks with each system: reinitiating symbology, pairing air defense fighters (ADF), and conducting an intercept.

Ratings of Revised System Impact. Additionally, we administered a questionnaire developed by Klein Associates, Ratings of Revised System Impact, that addressed specific modifications included in the revised system. This questionnaire was intended to directly assess the impact of the revised system on the specific areas of cognitive processing identified in the requirements analysis phase of the project: attention, memory, situational awareness, workload, and decision making.

Expert ratings. In complex tasks and at higher levels of expertise, component measures of performance often cannot convey the whole picture. In spite of the numerous objective and subjective measurement techniques utilized in this study, we still lacked an indicator of whether we had improved overall WD performance. None of the individual measures of performance, of situational awareness, or of workload convey that overall effect. To address this question we obtained a subjective appraisal of WD performance from an SME.

A highly experienced WD (a former WD instructor) was asked to rate each WD's performance on both systems. For each of the 17 WDs, the scores on the trial using the current system were printed on one sheet of paper, and the scores using the revised system were printed on a separate sheet of paper. These 34 sheets contained no information concerning the WD's experience level or which system had been used. The sheets were randomized and the rater assessed the performance using a five-point scale where 1 = excellent and 5 = poor.

Results and Discussion

Table 6 presents means and standard deviations of WD performance on the current and revised systems, for all measures.

Our original plan for data collection and analysis was based on assumptions of adequate numbers of participants, a fully counterbalanced design, and use of analysis of variance methods to (ANOVA) examine system differences. As we entered the data analysis phase of the project, we realized it would be necessary to reevaluate that plan.

Simulator malfunctions resulted in the loss of almost one-third of the sample and inevitable counterbalancing problems. In addition, examination of the distributions revealed that a number of the dependent variables were distributed non-normally. In many cases, variances were as large or larger than the associated means, making significance tests based on measures of central tendency of questionable value. Distribution tests revealed that for over 50% of the variables, distributions deviated significantly from normal. We also found that for approximately 20% of the variables, performance variability was significantly greater on one system than on the other. However, we did not find that greater variability was consistently associated with either the current or revised system.

Given the reduced N and distribution problems, we decided to abandon the ANOVA tests. Instead, we calculated individual difference scores and used these to test whether the degree of change in performance from the current to the revised system was significantly different from zero. The statistical approach is based on the t distribution and is commonly used in studies in which the same individuals are assessed under different conditions or treatments (Spence, Underwood, Duncan, & Cotton, 1968). We also calculated percent of change from the current to the revised system, based on the average performance on each system. Although somewhat crude, the percent change offers a good metric for

Table 6

AWACS WD Performance on Current and Revised Systems: Means and Standard Deviations

Measures	Current System Mean	Current System (S.D.)	Revised System Mean	Revised System (S.D.)
Expert Ratings (1 = High)	3.77	(1.25)	2.82	(1.33)
Outcome Measures				
Hostiles strikes completed	0.65	(0.86)	0.53	(0.94)
Number hostile penetrations	6.76	(4.16)	7.41	(4.47)
Penetration depth (penetrators only)	55.99	(31.41)	39.29	(14.30)
Penetration depth (all)	16.06	(10.84)	14.26	(11.10)
Hostiles shot down	19.47	(0.87)	19.76	(0.66)
Friendlies shot down	5.47	(2.79)	4.64	(2.03)
Hostile to friendly kill ratio	5.06	(4.39)	5.52	(4.06)
Friendlies lost to low fuel	0.53	(0.62)	0.71	(0.99)
Friendlies shot by friendlies	0.06	(0.24)	0.06	(0.24)
Total friendlies lost	6.06	(2.99)	5.41	(2.15)
% fired missiles that missed	8.95	(7.13)	5.79	(7.67)
Process Measures				
Workload				
Response to SD Inquiry				
% correct	51.87	(20.02)	56.15	(18.20)
% only acknowledged	28.87	(16.46)	25.66	(15.80)
% no response	5.88	(7.83)	5.35	(6.47)
Avg. response time	69.81	(40.27)	83.41	(37.70)
Response to visual alerts				
% responded to	46.47	(14.98)	47.06	(22.30)
Avg. response time	38.60	(7.35)	37.17	(11.00)
Intercept Approach Specified	15.62	(11.61)	12.91	(10.60)
Situational Awareness				
Recorrelations - % correct	90.95	(5.78)	93.14	(4.70)
Time symbology incorrect	2.37×10^4	(2.00×10^4)	2.43×10^4	(2.70×10^4)
Air refuelings	1.24	(2.05)	2.18	(2.72)
AC return to base	6.35	(5.06)	5.18	(3.79)
Airborne orders/scrambles ratio	0.87	(1.70)	0.92	(1.54)
Workload Ratings				
Overall (NASA TLX)	64.86	(13.09)	72.07	(9.97)
Specific Tasks (SWORD)				
-reinitiate	0.10	(0.06)	0.24	(0.12)
-pair ADF	0.10	(0.05)	0.18	(0.09)
-intercept	0.22	(0.06)	0.16	(0.06)
Ratings - Revised System Impact (1 = High)	N/A	N/A	2.61	(0.89)

the overall impact of the revised system on performance of the WD task. Table 7 presents the difference scores and the percent change for each of the measures.

Order and scenario. Due to the difficulties encountered in the counterbalancing scheme, more subjects used the current system second and the revised system first. Based on our observations we felt this might bias the results against the revised system. The two test scenarios were identical except that they had been rotated (the enemy came from the north in one scenario and from the east in the other), the names of the airbases were changed, and the landmass was different. Our concern was not that one scenario was easier than another, but that the WDs noticed that they were identical and began to predict events.

We found that there was no significant effect for order or for scenario. Although the WDs began to predict certain events, they did not perform better in the second session than the first. We had hypothesized that, since the WDs were using a system they were more familiar with and making some predictions regarding events, they would perform better in the second session when using the current system. They did not.

Outcome measures. The outcome measures can be considered overall "win the war" measures. These measures indicate how well the WD is handling the battle. Are hostile aircraft penetrating into friendly airspace? More importantly, how far are the hostiles penetrating and are they bombing friendly airbases? How many friendly aircraft are placed in the direct line of fire of the hostile aircraft and are then shot down? These measures provide an overall picture of how well the WD is performing in the areas of battle management and air defense.

WDs using the revised system shot enemy penetrators down farther away from friendly assets than they did when using the current system [($t(16) = -1.86$, $p < .01$).][3] Also, more hostiles were shot [($t(16) = 1.34$, $p < .10$)] and fewer friendlies were shot by hostiles [($t(16) = -1.33$, $p < .10$)] with the revised system than with the current system. These two measures resulted in an increase in the mean change for kill ratio of .47 (9%). Furthermore, considering that 3.2% fewer missiles were fired per WD that missed with the revised system [($t(16) = -1.55$, $p < .10$)], it is clear that the WDs were doing a better job of efficiently and effectively intercepting enemy fighters. More enemy aircraft were downed, fewer friendlies were lost, and less armament was wasted.

Examination of mean differences for the other outcome measures also indicates improved performance by WDs when using the revised system. Not only were 20% fewer hostile strikes completed, but fewer WDs had hostile strikes completed against them (5 vs. 8) when using the revised system. A hostile strike completed represented at least one friendly airbase bombed. Although there were

---

[3]It was our hypothesis that performance would improve when using the revised interface. Therefore, one-tailed tests were used. One-tailed tests are appropriate when a directional hypothesis can be specified.

Table 7

Comparison of AWACS WD Performance on Current and Revised Systems: Mean Difference Scores and Percent Change

	Avg. Difference Revised-Current System	% Change	Desired Impact of Revised System
Expert Ratings (1 = High)	0.94 ***	26%	+[1]
Outcome Measures			
Hostiles strikes completed	-0.13	20%	-
Number hostile penetrations	0.65	9%	-
Penetration depth (penetrators only)	-16.70 ***	-30%	-
Penetration depth (all)	-1.81	-11%	-
Hostiles shot down	0.29 *	3%	+
Friendlies shot down	-0.82	-15%	-
Hostile to friendly kill ratio	0.47	9%	+
Friendlies lost to low fuel	0.18	25%	-
Friendlies shot by friendlies	0.00	0%	-
Total friendlies lost	-0.65	-11%	-
% fired missiles that missed	-3.20 *	-36%	-
Process Measures			
Workload			
Response to SD inquiry			
% correct	4.20	8%	+
% only acknowledged	-3.20	-11%	-
% no response	-1.00	-17%	-
Avg. response time	1.02	16%	-
Response to visual alerts			
% responded to	1.00	2%	+
Avg. response time	-1.44	-37%	-
Intercept Approach Specified	-0.03 *	-17%	+
Situational Awareness			
Recorrelations - % correct	2.18 *	3%	+
Time symbology incorrect	51.60	2%	-
Air refuelings	0.94 ***	76%	+
AC return to base	-1.24 *	-18%	-
Airborne orders/scrambles ratio	0.05	5%	+
Workload Ratings			
Overall (NASA TLX)			
Specific Tasks (SWORD)			
- reinitiate	0.13 ***	140%	-
- pair ADF	0.08 **	80%	-
- intercept	0.03	-27%	-
Ratings-Revised System Impact (1 = High)	N/A	N/A	N/A

* $p < .10$; ** $p < .05$; *** $p < .01$, one tailed t test

[1]Indicates the desired direction of revised system impact.

slightly more friendly aircraft lost due to fuel depletion with the revised system (0.53 vs. 0.71), there were still 11% fewer total friendlies lost with the revised system. This indicates that WDs were doing a better job of winning the war when using the revised system.

Process measures. Process measures provided a set of indices of the revised system's effect on workload and SA. We hypothesized that the revised system would lower workload and increase SA.

Workload. Each WD was asked approximately 10 questions by the Senior Director (SD) over the course of each scenario. Typically these questions addressed the current state of the aircraft in the WD's lane. For each question, we recorded whether the WD responded correctly, acknowledged the request, or failed to respond. The rationale here is that if the WD is experiencing a manageable level of workload, s/he would be more likely to respond quickly with a correct answer than to merely acknowledge the question (putting the answer off until later) or to fail to respond to the question. We found that the WDs had 8% more "correct" responses (mean difference of 4.2) when using the revised system. The revised system produced an average of 3.2 fewer "acknowledgements only" responses (an 11% change), and one less "no response" (a 17% change). Taken together, the findings indicate that the WDs were responding more often with accurate information with the revised system. However, the average response time to SD inquiries was slightly less for the current system (1.02 seconds, a 16% drop), indicating a trend in the opposite direction. In operational terms, this indicates that the WDs were doing a better job of supplying the SD with thorough, accurate information with the revised system, but they were taking slightly more time to do so. However, none of the effects were significant. In that there is no clear indication of better performance on either system, no difference in workload was detected with this measure.

Our measure of WD responses to visual alerts showed similarly inconclusive results in terms of workload. These visual alerts ranged from notifications of friendly aircraft in distress to time checks. In this case, the number of visual alerts responded to and the reaction time were recorded. Again, the rationale is that if the WD is experiencing a manageable level of workload, s/he would be more likely to respond quickly to any visual alert. When using the revised system the WDs responded to one more visual alert, no real difference at all from the current system. Their responses were, on the average, 1.44 seconds faster with the revised system. From an operational standpoint, one might say that pilots aboard distressed aircraft were receiving attention faster with the revised system. Again, however, the lack of a clear indication of better performance on either system indicates that this measure did not detect a differences in workload between the interfaces.

A third embedded task used to measure workload involved specifying intercept approaches. The WDs were informed that they were in a datalink, requiring them to enter the 3-D approach of the friendly aircraft to the enemy aircraft after every commit switch action. This involves an additional action that is often neglected during periods of high workload. The WDs informed the system of the type of approach more often ($t(16) = -1.36$, $p < .10$) when using the current system, indicating an increase in workload with the revised system.

Viewed together, these three measures do not indicate a strong difference in workload between the two systems. There is a weak trend in favor of the current system. This is not surprising however, if one considers that participants received only 4.5 hours of training on the revised system, yet had years of experience with the current system. It is quite impressive that even though the WDs may have been experiencing a higher level of workload due to their unfamiliarity with the revised interface, their performance on outcome measures ('n terms of winning the war) still improved.

Situational awareness (SA). There were four measures of SA. These measures ranged from how well the WDs maintained the tracks on their scope to how well they were thinking ahead in their use of the fighters assigned to them.

Often, the track symbology becomes disconnected from the radar dot. When this occurs the WD must recorrelate (called "reinitiating") the symbology back to the correct dot. The accuracy of placing the correct symbol on the correct dot is extremely important and indicates how aware the WD is of which track is which and where on the scope it belongs. A WD without good SA can forget which symbol belongs to which dot, causing recorrelations to become a deadly game of chance. Placing a friendly symbology on a hostile radar dot can be disastrous. The WDs recorrelated their tracks more accurately with the revised system than with the current system [($t(16) = 1.43$, $p < .10$)]. The total time that tracks were uncorrelated favored the current system by an average of five seconds for each four-hour session. Again, when using the current system they were responding slightly faster but with potentially disastrous results.

The remainder of SA measures indicate the ability of the WDs to predict future events and plan for them. It is more efficient to refuel an aircraft that has armament than it is to return it to base and call up another fighter. Therefore, a WD who was aware of the situation and was properly utilizing his/her assets would have a higher number of refuelings and fewer aircraft returned to base (RTBs). When using the revised system the WDs had more refuelings [($t(16) = -3.24$, $p < .01$)] and fewer RTBs [($t(16) = 1.51$, $p < .10$)] than when using the current system. These two findings together support the idea that when using the revised system the WDs were better able to plan ahead and efficiently use the resources at their disposal. Related to this is the ability of the WDs to determine when in the future they will need more fighters to enter the battle. A WD who is thinking ahead can issue an airborne order (more than five minutes into the future); but if caught off guard, fighters must be scrambled (needed in less than five minutes). The ratio of airborne orders to scrambles increased, but not significantly by 5% for the WDs using the revised system. That is, more airborne orders and fewer scrambles were issued with the revised system, indicating that the WDs were able to plan ahead more and scramble less.

Workload ratings. There were two measures of workload, the NASA-TLX (Task Load Index) and the Subjective Workload Dominance (SWORD) technique. The NASA-TLX provides an overall measure of workload, whereas the SWORD technique allows for analysis of various tasks within the overall frame. The WDs rated the revised system higher in overall workload [($t(16) = -2.29$, $p < .05$)] with the NASA-TLX. Using the SWORD technique the WDs compared the workload associated with each system for three tasks: reinitiating symbology, pairing air defense fighters, and conducting the intercept. They rated the revised system as higher in workload for reinitiating symbology [($t(16) = -4.19$, $p < .01$] and pairing air defense fighters [($t(16) = -3.38$, $p < .01$)]. They rated conducting the intercept lower for the revised system, but this was not significant. Again, this is not surprising given that they were only trained for 4.5 hours on the revised system. These findings confirm the trend seen with the embedded measures of workload. The WDs subjective rating of their workload contradicts the improved performance indicated by the outcome measures. It is likely that workload was higher when using the revised system for the lack of training alone, yet the WDs were still able to perform many WD tasks better with the revised system.

Revised system ratings. A 31-item questionnaire was developed to assess the perceived effect of the revised interface on WD performance. The questionnaire was administered to participants following the activities of the test day. The questions were fairly specific and asked about the effect each of the modifications had on: identifying tracks, locating tracks, planning ahead, memory, situational awareness, attention, judgment, decision making, and workload.

Overall, the revised system received a rating of 2.61 on a five-point scale (1 = revised system made it much easier, 5 = revised system made it much more difficult). The rating for each of the modifications was as follows: Symbology (2.16), Color (2.57), On-Screen Menu (2.78) and QAN (2.94). This rating gives an overall indication of user acceptance of each of the modifications. It is not surprising that the QAN feature received the lowest rating, given that it had not been user-tested prior to the evaluation.

Expert ratings. Despite the amount of data collected on a large number of measures (28), it was still difficult to determine the overall effect of the revised system on WD performance. One way to obtain these data was to ask a subject matter expert (SME) . This SME, who had previously served as a WD instructor, was asked to rate the performance of each WD without regard to system. This was accomplished by preparing "profile" sheets of each WD using each system. These profiles, totaling 34, contained all of the measures collected for each subject. The SME rated the WDs using the revised system higher in overall performance than the WDs using the current system [($t(16) = 3.57$, $p < .01$)]. Based on the SME ratings the revised system improved overall WD performance by a mean difference of 0.94, nearly one full rating point on a five-point scale.

Effect of Experience

During the experiment we noticed that some WDs didn't seem to be performing as well as others. We wondered if this variation in performance was based on experience. After the experiment, it was determined that the participants could be classified into three experience levels based on their combined flight and simulator hours. WDs with less than 500 combined hours were placed in the low-experience group, WDs with between 500 and 1000 combined hours were placed in the middle-experience group, and WDs with more than 1000 hours were placed in the high-experience group. This allowed for six participants in both the low and high groups, and five participants in the middle group. We anticipated that the least-experienced WDs were performing most poorly, and that the revised system would improve performance for the low-experience group the most. We also believed that the high-experience group was performing at an optimum level and would, therefore, not improve as much as the other two groups when using the revised system.

Contrary to our hypothesis we found that the revised system improved performance for the high-experience group the most, with the low-experience group second. The middle-experience group didn't perform as well as the other two groups when using the revised interface. Interestingly, analyzing the mean differences for all three groups, the middle-experience group didn't perform as well as the other two groups when using the current system either. Appendix C presents mean standard deviations and percent change for the experience groups.

Possible explanations for this result include:

- The middle-experience group consists of WDs who are at a point in their careers when they are still attempting to mentally organize larger WD concepts. They are good WDs who are simply modifying how they operate in the AWACS environment. This transformation from novice to expert is difficult. Concepts are still being formulated regarding aircraft vectoring, fighter flow, and communication. Doing things by the book, as the low-experienced WDs were, works to a point. But as real-world experience increases and some of what was taught in school is either forgotten or purposely discarded, new concepts and constructs are created in order to become proficient. The middle-experience group performed well on the outcome measures, but they just didn't have the resources to do it all. Whether they knew it or not, they decided that winning the war was more important than keeping their scope clean and maintaining proper fighter flow.

- The middle WDs do not have as much recent experience with the WD tasks as the other two groups. The low-experience group was recently out of training and the high experience group was composed almost entirely of instructors. These two groups have more recent time in the simulator and therefore, are more familiar with the WD tasks. The middle-experience group performed well on the outcome measures ("winning the war") but they did not perform well on the other measures. They were oversaturated.

We believe the most likely explanation is the second one. The middle WDs simply don't have as much recent simulator and flight experience as either the WDs fresh out of school or the instructors who are teaching in the simulators. Examination of the data with regard to experience levels reveals that when using the revised system the high-experience level WDs showed improvement on 86% of the measures, the low group showed improvement on 60% of the measures, and the middle-experience group showed improvement on 52% of the measures.

The revised system certainly had an impact on WD performance. The overall effect, based on the SME rating, was quite positive. We predicted that the revised system would have a positive effect for the SA measures and workload measures, and hypothesized that by supporting the WDs processes, the revised system would also result in better performance for the outcome measures. With the exception of workload, the revised system performed as predicted.

SECTION 7: CONCLUSIONS

This project demonstrated the application of Cognitive Systems Engineering (CSE) to the redesign of the AWACS WD interface. This perspective allowed us to pinpoint, and design a system that supported, the cognitive processes of the users. This explicit identification of the essential user processes was maintained throughout the evaluation of the final product. Major findings and conclusions of the project include:

- It is possible to achieve significant improvements in operator performance via upgrades/retrofits of system features. In the present study, overall WD performance improved 26% (based on the SME ratings).

- Success of such efforts depends on use of focused, theoretically guided strategies to direct system changes, rather than simply throwing new technology into an existing system.

- When the operator's job involves perceptual judgments, assessment and diagnostic components, problem solving, and decision making, analysis of the operator's cognitive tasks is critically important. Interface and other system modifications must not only take these into account, they must also actively support the cognitive elements of the user tasks.

Revised System Impacts

We have shown that the revised system improved performance for a number of outcome and process measures. But, how well did it perform in an operational sense?

- WDs improved on 73% of the measures for which winning the war could be inferred. These included:

  - A 20% decrease in hostile strikes completed
  - A 15% decrease in friendly assets shot down
  - A 3% increase in hostiles shot down
  - A 9% increase in kill ratio
  - A 36% decrease in missiles fired that missed the target
  - An overall 11% decrease in total friendly aircraft lost

- WDs improved on 83% of the measures associated with responding to visual alerts and Senior Director inquiries. These included:

  - A 37% faster reaction time to aircraft in distress and time checks
  - An 8% increase in correct responses to the Senior Director inquiries
  - A combined 28% fewer instances of only acknowledging or not responding to the Senior Director inquiries

• WDs improved on 80% of the measures associated with SA. These included:

- A 76% increase in air refuelings
- Consequently, an 18% decrease in aircraft returned to base
- A 5% increase in the ratio of airborne orders to scrambles
- A 3% increase in correct recorrelations

It should be noted that we were prepared for there to be a ceiling effect on performance, especially for the more experienced WDs. That is, we had been told that the WDs were performing at such an optimum level that little or no improvement could be expected. This may be the case, and WDs may have been performing as well as they could given the limitations of the current system. However, the revised interface allowed them to achieve appreciably higher levels of performance.

Cognitive Systems Engineering

One of the major aspects of our CSE approach was the use of Cognitive Task Analysis. Following some of our earlier work (Klein, Calderwood, Clinton-Cirocco, 1986; Thordsen, Wolf, & Crandall, 1990; Wolf, Klein, Thordsen, & Klinger, 1991), we used Concept Maps and the Critical Decision method to achieve a deep understanding of how the WDs were thinking about their tasks, how they were making decisions, and how they were drawing inferences and learning how to perceive their screens. All of the team members involved in interface design believe that the Cognitive Task Analyses, particularly the Critical Decision method data, were of central importance in understanding where the bottlenecks were and what types of cognitive processes needed to be better supported. Considering that we only had time to conduct 13 two-hour interviews, this was a very large return on investment. Furthermore, the output of these interviews directly identified the context of the tasks. We could use the incident accounts to suggest ideas for improved interfaces and mentally simulate what might happen if these suggested features had been present during an incident--how it might have helped and where it might have interfered. In short, the critical incident data, overlaid with the cognitive probes for what the WD was thinking about, provided us with a context-rich account that was a platform for making suggestions and a testbed for evaluating them.

The Critical Decision method gets directly at the most difficult tasks, in contrast to conventional methods such as behavioral task analysis, which seek to decompose complex tasks into basic elements, perform careful and systematic analyses of these basic task elements, and then find some way to reassemble the elements to draw conclusions about the higher level tasks.

Lessons Learned

The application of CSE is relatively new, therefore one would expect to learn a great deal about the process each time it is applied. The goal of this section is to provide future CSE teams with some guidance so they can avoid some of the traps and repeat the successes.

An evaluation facility with full-fidelity simulation capabilities is essential if the impact of a full system is to be determined. In this instance, the selection of the AESOP facility with its experienced, professional staff, was the primary reason we were able to determine the full impact of our modifications. The AESOP facility contains a system whose interface can easily be modified and into which any definable measure can be programmed and collected. The value of this facility cannot be stressed enough.

Many design teams will place their best data collectors on one team, their best designers on another team, and their best evaluators on yet another team. We have found that far too often valuable knowledge is lost when the information is transferred from one team to the next. In this project, although various individuals' strengths were utilized, all members played integral roles throughout the project. The importance of this "shared understanding" notion surfaced during the user-testing portion of the project. The observers of the pilot studies had been involved in both the design and interview phases and could quickly see where WDs were having problems. These team members had a thorough understanding of the WD domain which allowed for rapid modifications that worked. Had these individuals not been involved in the interview phase, this knowledge of the subtle aspects of the WD tasks would not have been available and the system would certainly not have achieved such positive results.

Additional Opportunities

This project was limited in scope and resources; much information was gained during the requirements analysis that either could not be considered because of resource constraints, or did not fit into the scope of the project. We offer these findings as additional benefits of the project.

Ground controllers. In many of our interviews, we were told that WDs who had previous experience as ground controllers are typically the best WDs. Some of the reasons we were given for this phenomenon include:

- Ground controllers work with the same set of pilots for extended periods of time. They develop relationships and trade stories. Not only does this develop camaraderie with the pilot community, but it allows the controller to get direct feedback regarding his/her methods of communication. The WDs, on the other hand, seldom if ever interact on a personal basis with pilots. Feedback regarding their communication skills is rare. WDs simply don't know the people they are talking to.

- Ground controllers use a grease pencil to track the path taken by the aircraft within their area. This means they are personally involved in the data. They have recorded it, they remember it, they own it. It is easier for them to notice trends and remember information.

- Ground controllers understand the radar system better. They are more aware of how accurate, and inaccurate, the system is. This understanding of the system allows them to provide more accurate information to the pilots.

Senior director's needs. During the requirements analysis we interviewed WDs, Senior Directors (SDs), and WD Instructors (IWDs). It became apparent that the tasks involved with being a WD are not the same as those for an SD. Yet, the interface and console are the same. The WD monitors and vectors specific tracks, while the SD monitors the overall battle. The WD communicates with pilots often, the SD not as frequently. Yet, the positions are not drastically different. The SD often performs WD tasks, s/he may take over a particular track from a WD, or s/he may direct a search-and-rescue mission. There is a great deal of information the SD must contend with and s/he must do so with a WD interface. It needs to be determined where these positions overlap and where they diverge. The respective interfaces need to reflect those commonalities and differences.

Standardize communications. More experienced WDs, especially the former ground controllers, commented that most WDs provide too much information to the pilots. They clog up the airways with chatter that interferes with the communication that must take place among the pilots themselves. Often, the pilots simply stop listening to the WDs. This evolves into mutual mistrust among the pilots and the WD. With some standardized training regarding communication with the pilots, the overall effectiveness of the battle group would likely improve.

Summary

This project was initiated to evaluate the impact of Cognitive Systems Engineering on interface design. We wanted to examine the level of effort needed to perform a CSE approach, and the resulting effect on operator performance.

The revised interface showed clear-cut improvements over the current AWACS Weapons Director interface. The global performance rating was more than 25% higher for the revised interface. The specific measures of outcomes, such as hostile strikes completed, hostile aircraft shot down, missiles that missed their targets, and degree of penetration allowed by enemy aircraft all illustrated the superiority of the revised interface.

These results were obtained despite a number of factors that favored the current interface. The WDs had received an average of 1180 hours with the current interface after completing their training programs. In contrast, they received only 4.5 hours with the revised interface. This was sufficient time to learn to operate the new features, but not enough time to become proficient in using these features. If the WDs had been given an additional 20 to 40 hours, the results would probably have been even more striking.

Nevertheless, the findings are sufficient to make the point that a CSE approach is a cost-effective way to retrofit existing systems and improve performance. The redesign effort took a relatively brief period of time, 10 months, and resulted in rates of effectiveness that would have been difficult to achieve through additional training or more powerful equipment.

The three facets of Cognitive Systems Engineering all played a role in the design process. The use of recent technological opportunities such as color and rapid threat identification made it possible to help the WDs quickly notice the dynamics of the situation. The use of naturalistic perspectives on cognition and decision making allowed us to support the WDs' needs for maintaining situational awareness while performing different sub-tasks, for judging the level of threat posed by enemy aircraft, and for locating critical assets such as tankers. The use of Cognitive Task Analysis was perhaps the most important aspect of CSE, revealing the way the WDs needed to interpret cues and make decisions. Together, these three components enabled us to design an interface that centered around the difficult decisions and judgment tasks, rather than around the information available through the system.

The project did not result in a formula for CSE. The use of technological reviews, reviews of cognitive science, and Cognitive Task Analysis, all entered into the design process but did not define any sequence of steps for generating design solutions. We do not feel that it is possible to standardize the design process, particularly for difficult functions such as those performed by WDs. Members of design teams will need to identify what makes the judgment and decision-making tasks difficult for each type of job, and how the interface can support the operators. Cognitive Systems Engineering does offer a strategy for identifying and supporting the most difficult aspects of a job, and for making these the central factors driving the interface design process.

REFERENCES

Andriole, S. J. (1989). Storyboard prototyping: A new approach to user requirements analysis. Wellesley, MA: QED Information Sciences, Inc.

Baron, J. (1988). Thinking and deciding. NY: Cambridge University Press.

Beach, L. R. (1990). Image theory: Decision making in personal and organizational contexts. West Sussex, England: John Wiley & Sons Ltd..

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

Bolt, R. A. (1984). The human interface: Where people and computers meet. Belmont, CA: Lifetime Learning Publications.

Calderwood, R., Crandall, B. W., & Baynes, T. H. (1988). Protocols of expert/novice command decision making during simulated fire ground incidents (KATR-858-88-02Z). 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. W., & Klein, G. A. (1987). Expert and novice fireground command decisions (AD-B154 224). Alexandria, VA: Defense Logistics Agency.

Crandall, B. (1989). A comparative study of think-aloud an critical decision knowledge elicitation methods. ACM SIGART, 108, April.

Crandall, B., & Calderwood, R. (1989). Clinical assessment skills of experienced neonatal intensive care nurses. Yellow Springs, OH: Klein Associates Inc. Final Report for the National Center for Nursing, NIH under Contract No. 1 R43 NR01911 01.

Crandall, B., & Gamblian, V. (1991). Guide to early sepsis assessment in the NICU. Fairborn, OH: Klein Associates Inc. Instructional manual prepared for the Ohio Department of Development under the Ohio SBIR Bridge Grant program.

Crandall, B., & Klein, G. A. (1988). 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.

Gowin, D. B., & Novak, J. D. (1984). Learning to learn. NY: Cambridge University Press.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. IEEE Transactions on Systems, Man, and Cybernetics, SMC-17(5), 753-770.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.), Human mental workload. Amsterdam: North-Holland, 139-183.

Isenberg, D. J. (1984, November/December). How senior managers think. Harvard Business Review, 80- 90.

Kaempf, G. L., Wolf, S., Thordsen, M. L., & Klein, G. (1992). Decisionmaking in the AEGIS combat information center. Fairborn, OH: Klein Associates Inc. Prepared under contract N66001-90-C- 6023 for the Naval Command, Control and Ocean Surveillance Center, San Diego, CA.

Keeney, R. L., & Raiffa, H. (1976). Decisions with multiple objectives: Preferences and value tradeoffs. NY: Wiley & Sons.

Klein, G. A. (1989). Recognition-primed decisions. In W.B. Rouse (Ed.), Advances in Man-Machine System Research, 5, 47-92. Greenwich, CT: JAI Press, Inc.

Klein, G. A. (in press). Naturalistic decision making--Implications for design. Fairborn, OH: Klein Associates Inc. Prepared under contract DLA900-88-D-0393 for CSERIAC, Dayton, OH.

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

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. IEEE Transactions on Systems, Man, and Cybernetics, 19(3), 462-472.

Klein, G. A., & Crandall, B. W. (in press). The role of mental simulation in naturalistic decision making. In J. Flach, P. Hancock, J. Caird, and K. Vicente (Eds.), The ecology of human-machine systems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, G. A., Orasanu, J., Calderwood, R., & Zsambok, C. E. (in press). Decision making in action: Models and methods. Norwood, NJ: Ablex Publishing Corporation.

Klein, G. A., & Thordsen, M. (1988). Use of progressive deepening in battle management. Proceedings of the 11th Biennial DoD Psychology Conference, Colorado Springs, CO.

Lipshitz, R. (1989). Decision making as argument driven action. Boston University: Center for Applied Social Science.

McFarren, M. R. (1987). Using concept mapping to define problems and identify key kernels during the development of a decision support system. Master's thesis, School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, OH.

McNeese, M. D., Zapf, B. S., Peio, K. J., Snyder, D. E., Duncan, J. C., & McFarren, M. R. (1990). An advanced knowledge and design acquisition methodology: Application for the pilot's associate. Final Report AAMRL-TR-90-060. WPAFB, OH: Armstrong Aerospace Medical Research Laboratory.

Noble, D., Boehm-Davis, D., & Grosz, C. G. (1986). A schema-based model of information processing for situation assessment. Vienna, VA: Engineering Research Associates. (NTIS No., ADA163150).

Orasanu, J., & Connolly, T. (in press). The reinvention of decision making. In. G. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok (Eds.), Decision making in action: Models and methods. Norwood, NJ: Ablex Publishing.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.

Pennington, N., & Hastie, R. (1981). Juror decision making models: The generalization gap. Psychological Bulletin, 89, 246-287.

Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. IEEE Transactions on Systems, Man and Cybernetics, SMC-15(2), 234-243.

Schiflett, S. G., Strome, D. R., Eddy, D. R., & Dalrymple, M. A. (1990). Aircrew evaluation sustained operations performance (AESOP): A triservice facility for technology transfer. Prepared for U.S. Air Force School of Aerospace Medicine, Brooks AFB, TX.

Shanteau, J., & Phelps, R. H. (1977). Judgment and swine: Approaches and issues in applied judgment analysis. In M. F. Kaplan & S. Schwartz (Eds.), Human judgment & decision processes in applied settings. NY: Academic Press.

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

Soelberg, P. O. (1967). Unprogrammed decision making. Industrial Management Review, 8, 19-29.

Spence, J. T., Underwood, B. J., Duncan, C. P., & Cotton, J. W. (1968). Elementary statistics, 2nd edition. New York: Appleton-Century-Crafts.

Thordsen, M. L., & Calderwood, R. (1989). Protocol analysis of flight crew decision making during simulated malfunctions. Yellow Springs, OH: Klein Associates Inc. Prepared under Order No. A72145C for the NASA/Ames Research Center, Moffett Field, CA.

Thordsen, M., Klein, G. A., Michel, R., & Sullivan, E. (1988). Methods for providing direct feedback about decision processes for command-and-control classroom exercises (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. L., Klein, G. A., & Wolf, S. (1990). Observing team coordination within Army rotary-wing aircraft crews (Final Technical Report). Yellow Springs, OH: Klein Associates Inc. Prepared under contract MDA903-87-C-0523 for the U.S. Army Research Institute, Aviation Research and Development Activity, Ft. Rucker, AL.

Thordsen, M. L., Wolf, S., & Crandall, B. (1990). User-defined requirements for the JSTARS self-defense suite. Final report prepared under contract F33615-88-D-0536 in conjunction with the UNISYS Corporation for the Armstrong Aerospace Medical Research Laboratory, Human Systems Division, Air Force Systems Command, Wright-Patterson AFB, OH. Dayton, OH: UNISYS Corporation.

Vidulich, M. A. (1989). The use of judgment matrices in the subjective workload dominance (SWORD) technique. In Proceedings of the Human Factors Society 33rd Annual Meeting, pp. 1406-1410. Santa Monica, CA: Human Factors Society.

von Winterfeldt, D., & Edwards, W. (1986). Decision analysis and behavioral research. Cambridge: Cambridge University Press.

Wohl, J. C. (1981). Force management decision requirements for Air Force Tactical Command and Control. IEEE Transactions on Systems, Man, and Cybernetics, 618-639.

Wolf, S., Klein, G., Thordsen, M., & Klinger, D. (1991). Factors affecting the decision making of fire direction officers, (Final Technical Report). Yellow Springs, OH: Klein Associates Inc. Prepared under contract DAAA15-90-C-1054 for the U. S. Army Ballistics Research Laboratory.

APPENDIX A

Incidents and Analysis

Here are four incidents derived from the CDM interviews. These incidents were selected based on a number of criteria:

- they are representative of the sample
- 2 were with Senior Directors (SDs), 2 were with WDs
- 2 were Search-and-Rescue (SAR)
- 2 were live intercepts
- 1 was a war-time intercept
- 1 was a peace-time intercept
- 1 SAR was with an SD
- 1 SAR was with a WD

Following each incident is an analysis of the entire interview.

Incident 1

It started with the detection from a base in northern Iraq of two aircraft taking off. It was pretty easy to detect them since we were keyed into the area and there really wasn't that much activity. If we saw something that looked like it was flying, we all looked at it. So this time we saw the two aircraft taking off and heading south. I immediately made calls to the northernmost CAP fighters that there were enemy aircraft in the air. As they continued to head south I committed the CAP fighters against them. The consideration for me at this point was the role of the CAPs. They are to be defensive counter air, so to send them out far into Iraq isn't necessarily their role and it could potentially leave open the role they are supposed to be performing. If the bad guys had launched anything f¨rther south, towards Saudi, those guys wouldn't have been there. We had other fighters that were on tanker that could theoretically fall forward and fill that role, but we were still leaving ourselves a little vulnerable. There was also the fact that there was a lack of air threat from the Iraqis, they simply weren't flying much. So I figured that it was safe to send those guys up there.

Soon after I committed the northern CAP, the enemy element broke out into four aircraft. Two groups of 2 in lead trail formation. There were now two elements. I committed the backup CAP against the second element, the farthest north group. So at this point we had two simultaneous 2 v 2's; the initial commit, which was the northern CAP against the first group, and the second commit, which was the backup CAP against the second group. The spacing in the groups made it impossible for me to only commit out one group against both targets.

I vectored the first group in until they called *radar contact*. I was then in a monitoring mode for that group while I continued to provide vector information for the second group. The first group moved in on the bad guys and called their standard calls. Then they called their kills. The second group called radar contact, closed on the enemy, but only called one kill. One of the enemy aircraft turned and headed back toward its base. Our guys gave chase but couldn't catch him.

Analysis of Incident 1

One of the first areas we discussed following the construction of events was that of resource allocation. We were interested in why the WD had selected those particular tracks to commit against the enemy. In this instance the CAP (a term for aircraft that are assigned to a particular point in space, typically there to guard against enemy attack) were selected, and then the back up CAP. As it turned out, everything was fine. But, if the CAP fighters had not been successful there was no protection for the E3 (the AWACS aircraft). Essentially this WD utilized all available resources to thwart the enemy and did not reserve any for protection. This is not standard practice. As we found out, this WD became focussed on the intercept itself, and not until it was completed were replacement CAP requested.

In this incident, the WD seemed to change his/her mode of communication at various times. For instance, there was a mode change as the friendly fighters closed in on the enemy. The WD stopped talking to the fighter pilots and began to monitor the radios instead. We were interested in when this occurs, and why. Also, what information was necessary to relay to the pilots before the switch and what was necessary after.

Another interesting item that came from this interview was the determination of the degree of threat of the enemy. The WD spent a fair amount of time attempting to determine the degree of threat of the enemy aircraft. Were they high fast flyers? Were they jammers? Were they standard enemy fighters? To determine this the WD had to monitor the track information. This information is displayed at the bottom of the scope. The WD must keep a close eye on the track's rate of climb, altitude, and speed. The enemy aircraft in this situation maintained a steady altitude, speed, and heading. The speed and altitude were such that they had to be fighters, and their direction seemed to indicate hostile intent. But, in monitoring the tracks for this information, some situational awareness for other air traffic was lost. Hence, no other aircraft were called for protection of the E3, and the two aircraft that were on tanker were not used.

And lastly, what were the parameters that meant that two friendly groups needed to be used. As it turned out, the determining factor was distance between the enemy aircraft. How did this WD know these parameters? The computer does not provide any of this information.

Incident 2

I was in Iceland and we got a call that there were some Russian aircraft heading our way. We got the scramble order and were airborne within the required time. After getting airborne we performed the required check outs and the system checked out fine. We received more information regarding their position during this time.

After completing my check out, two friendly aircraft checked up on my frequency. I was instructed that an intercept take place only if the Russian aircraft penetrate Icelandic airspace. I then had to find the boundaries for the airspace. After I determined the boundaries, I vectored the friendly aircraft in the direction of the enemy and informed them that they are not to cross into international airspace. As the two elements began to come together, the Russian aircraft turned parallel to the boundary. I figured they did this when the friendly fighters radar picked them up. We call this painting. The Russian aircraft knew that the friendly aircraft's radar system had picked them up, so they turned to avoid a conflict. I instructed our fighters to turn, as well, and to maintain distance between the Russians and themselves. So, what's happening is the Russians are flying parallel to the Iceland boundary, as are the friendly fighters, with the boundary between them. After a period of time the Russians turn in toward the boundary, I instruct our fighters to do the same. The Russians then turn out again, apparently after they are picked up by the friendly fighters radar. This occurs two or three more times. A cat-and-mouse game that I am told goes on all the time. Finally, the friendly aircraft pull in behind the Russian aircraft and conduct an intercept. They are simply shadowing them. Only now the friendly aircraft are slightly outside the Icelandic boundary. Almost immediately the lead pilot calls and tells me that they are going to need fuel.

At about this time two British Tornadoes and a British tanker check up on my frequency. Also, at about this time, the Russian aircraft turn toward home. They appear to be heading back the same way they came. So, I pull the friendly aircraft off of the intercept and vector them toward a tanker. I vector the British fighters to come in behind the Russian fighters to escort them. After the British fighters pull in behind the Russians, they report that they are having mechanical difficulties and are returning to their base. At this time the British tanker gets pretty aggressive and pulls in behind the Russian fighters. He makes some standard fighter calls to me, and I vector him in as if he were a fighter. He then "paints" the Russian aircraft with his radar and escorts them out of the area.

So in the end, a British tanker intercepted two Russian fighters and escorted them out of the area. We don't think that the Russians ever knew that it wasn't a fighter on their tails.

Analysis of Incident 2

The probes for this incident varied. Initially, we discussed the geometry of an intercept. This interview revealed a common theme, no one trusts the geometry that the computer recommends. There are several reasons for this. One, the computer doesn't always know the type of intercept being conducted (stern, stern conversion, cut-off, or pursuit). Second, the information the computer is using to tabulate the geometry is not always accurate. Third, what the scope shows now could be the state of the world 10 seconds ago. That could be a very important 10 seconds. Fourth, there is a general mistrust for any computer tabulations. These include things like fuel levels, armament states, etc.

Many of the intercept incidents we collected involved training or simulation (Red Flag, etc.) missions. In this active intercept there was more uncertainty, the stakes were higher, and the time pressure was real. These variables helped to increase the stress level of the WD, and rapid decisions were necessary. This allowed us a window into how WDs make decisions and when they must defer to a Senior Director (SD).

We discovered that unless it is war-time the WD typically does not determine which friendly assets are committed against enemy tracks. This information often comes from the SD or the MCC (Mission Crew Commander). But, if the dynamics of a rapidly changing situation warrant, the WD has authority to act accordingly. In this case, the WD decided to utilize the British fighters, and then the tanker, to complete the intercept.

This WD discussed another common theme: *the loss of situational awareness*. Initially, this WD was only monitoring two enemy targets, two friendly aircraft, and one tanker. As the incident unfolded, three more friendly tracks (the British fighters and tanker) became available, but at about that time the two initial friendly tracks were sent to tanker. This is not considered high traffic, yet this WD spoke about slight deteriorations in SA. Therefore, something else was contributing to this loss.

We discovered that, for WDs, looking away from the scope to input switch actions was a major contributor to this loss of SA. To execute a switch action, the WD must take his/her eyes off of the scope. The WD looks to the right, where a panel that contains about 100 switches is located. Once the desired command is located, the WD presses that button until a light appears, indicating that the system has recognized the action, and more information is input to the system via the keyboard or the trackball. Experienced WDs do not take much time to do this, maybe three to five seconds. But, in a high stress, fast paced situation three to five seconds can be a long time. SA is easily lost when the WD must think about where a switch is, find it, press it, and then input more information. Looking away from the scope even once a minute in order to execute a switch action will begin to compound into a total loss of SA. But, to consider that during high activity periods the WDs are executing up to 6 switch actions a minute, it is easy to see that SA deteriorates rapidly. They simply cannot fit all the pieces together in a coherent picture any longer.

Incident 3

I guess the one that stands out in my mind the most was a search and rescue (SAR) mission in Iraq. I'm watching my scope when I hear a pilot call out that he lost his partner. His wing-man had gone down. This is a pretty typical way of finding out because all that happens on the scope is that the radar dot goes away. We never had a single aircraft out there alone. As is typical with this type of situation, the wing-man flew back over the downed site and said something like "I'm over him right now." That allows me to put a downed aircraft symbol on the spot. This spot on the scope narrows it down to about a five-mile range. At the time we didn't know too much about the area, but the downed pilot began calling us on his portable radio. He told us that there were Iraqi troop movements everywhere.

So, I made the radio call that there was a downed aircraft. In about a minute or two this guy calls up on my frequency and says that he is SAR qualified. Not everyone was, but luckily we had a guy up who was. This really shortened the time it took us to get someone on the way. Typically it takes about 10 minutes before anyone heads toward the downed site. In this case, it was about two minutes. We were also somewhat lucky because of this guy's location. He was in central Iraq which is pretty wide open and unpopulated. Had he been in eastern Iraq, western Kuwait we would have had a much more difficult time. In central Iraq there were very few SAM (surface-to-air missile) stations and a lot less air traffic.

Even with all this going for us it was pretty touch and go. The downed pilot had thousands of screaming guys with guns coming at him, so he buried himself in the sand. He actually dug himself into the ground like a post, just barely sticking out so they couldn't really see him. It was smart on his part. It kept him cool and camouflaged.

Our first objective was to remove the hostile element from the area so that we could bring in some support helicopters. I noticed that we had some slow moving helicopters to the south, so I informed them that there was a SAR effort going on and that we may need their help.

What ended up happening was the SAR aircraft rolled into the area and wreaked havoc on the enemy. They were dropping and firing everything they had in order to back the enemy off. A lot was happening. There were planes everywhere and the downed guy was radioing to us the enemy locations. He was that close. The SAR aircraft continued to pound on the enemy troops until we could get a helicopter in there. Finally, a helicopter located him and we got him out.

The enemy would have gotten him if it weren't for really quick thinking on our part. Those SAR aircraft had to use everything they had on the convoy, but that's part of the game. In war there are no rules.

Analysis of Incident 3

Again, one of the primary areas we probed during this interview was that of situational awareness. This WD seemed to have a pretty good idea where aircraft were in his/her particular lane of defense. But, as the incident unfolded, SA deteriorated for all other aircraft except those involved in the effort.

We also talked at length regarding symbologies and markers. The system simply drops the radar dot of any aircraft that falls below the radar. This does not alert the WD to a problem. Understandably, aircraft often fly below the radar level and any obtrusive form of notification would be a hinderance. On the other hand, if a spot had been placed by the system at the exact location in which the last radar contact was made, this WD would have had a better idea of the downed location. If an aircraft is lost due to enemy fire, one of the last things one of his/her wing-men want to do is fly back over the spot so that a WD can mark the scope.

This WD also spoke at length about switch actions. We discussed how it is often difficult to locate the desired switch action on the panel and that the system does not always accept the action. This compounds the situational awareness "breaks" alluded to in the previous incident. With these breaks occurring once or twice a minute, SA can quickly be lost.

Incident 4

I was notified that a Marine Harrier had been shot down. I notified the command center of the estimated location and began writing down information that I may need later (intelligence info, terrain info, etc.). I designated a special point on the scope, so everyone would know the estimated position and I designated a radio frequency for the SAR effort. At this time I needed to maintain my regular SD duties as well.

After about 30 minutes the SAR unit (two aircraft) checked in. We immediately attempted to go to secure radios. It was unsuccessful. There had been numerous problems with the radios on these type of aircraft before, so we attributed our inability to go secure on that. We found out later that the problem was actually a miscommunication on our part and that the radios were fine. We were just too busy to investigate. Since we couldn't go with secure radios, we had to use codes in which to relay locations. This meant I had to find my paperwork, and so did the pilots.

After finding my paperwork and informing the aircraft of the downed position, I called the tanker controller to request a tanker for support. I was still continuing to perform the typical SD duties. Things were pretty hectic.

The tanker controller called back to inform me that he would have a tanker available soon. I told him the frequency of the SAR effort (so that the tanker could switch to it) and to turn the tanker north and hand him over to me. Prior to the tanker checking in, I vectored the two SAR aircraft to the general area of the downed aircraft. When the tanker checked in, I vectored it as far north as possible, in position for quick refueling but out of immediate danger. My role then changed to that of a monitor. I was still maintaining my other functions as an SD and was simply monitoring the effort. I was mainly listening for when they (the 2 SAR aircraft) would need fuel so that I could vector them toward the tanker. At this time I was simply coordinating with the MCC about the position of the E3, status of strike packages, status of CAP, etc.

The two SAR aircraft began communicating with the downed pilots via their portable radios. They discovered a common point of reference (a soccer field) and they seemed to be getting closer together. I heard a call from the SAR aircraft that they were in need of fuel. I vectored them to the tanker and informed the tanker that they are on the way (he should have known because he was monitoring the radio, but you can never be too sure).

After they came off of the tanker, they headed back to the downed sight. They did not need me to revector them since they knew where they were going. I did need to monitor for any air traffic that may have affected them, either friend or enemy.

This particular effort was extremely frustrating because we never found the downed pilots. The SAR aircraft talked with them several times and they felt as though they had a common reference point, but they never found them.

Analysis of Incident 4

This was one of the later interviews, and it served to validate much of what we had heard previously. For instance, the procedures associated with a Search-and-Rescue (SAR) mission were discussed previously. Yet this time we were offered a slightly different perspective because this interviewee was with a Senior Director (SD). As an SD, this individual was responsible for all WDs. Therefore, a single search-and-rescue mission had to be prioritized with all the other missions that were being controlled by this particular group. This increase in workload surfaced when the SD and the SAR aircraft could not go to secure radios. The SD simply did not have the time to troubleshoot, so s/he selected the lowest level to communicate with the SAR pilots (the codes). S/he looked for an explanation which fit the problem, found one (previous problems with the aircraft), and determined a course of action (using the codes). There was no time to search for other explanations

This WD also confirmed:

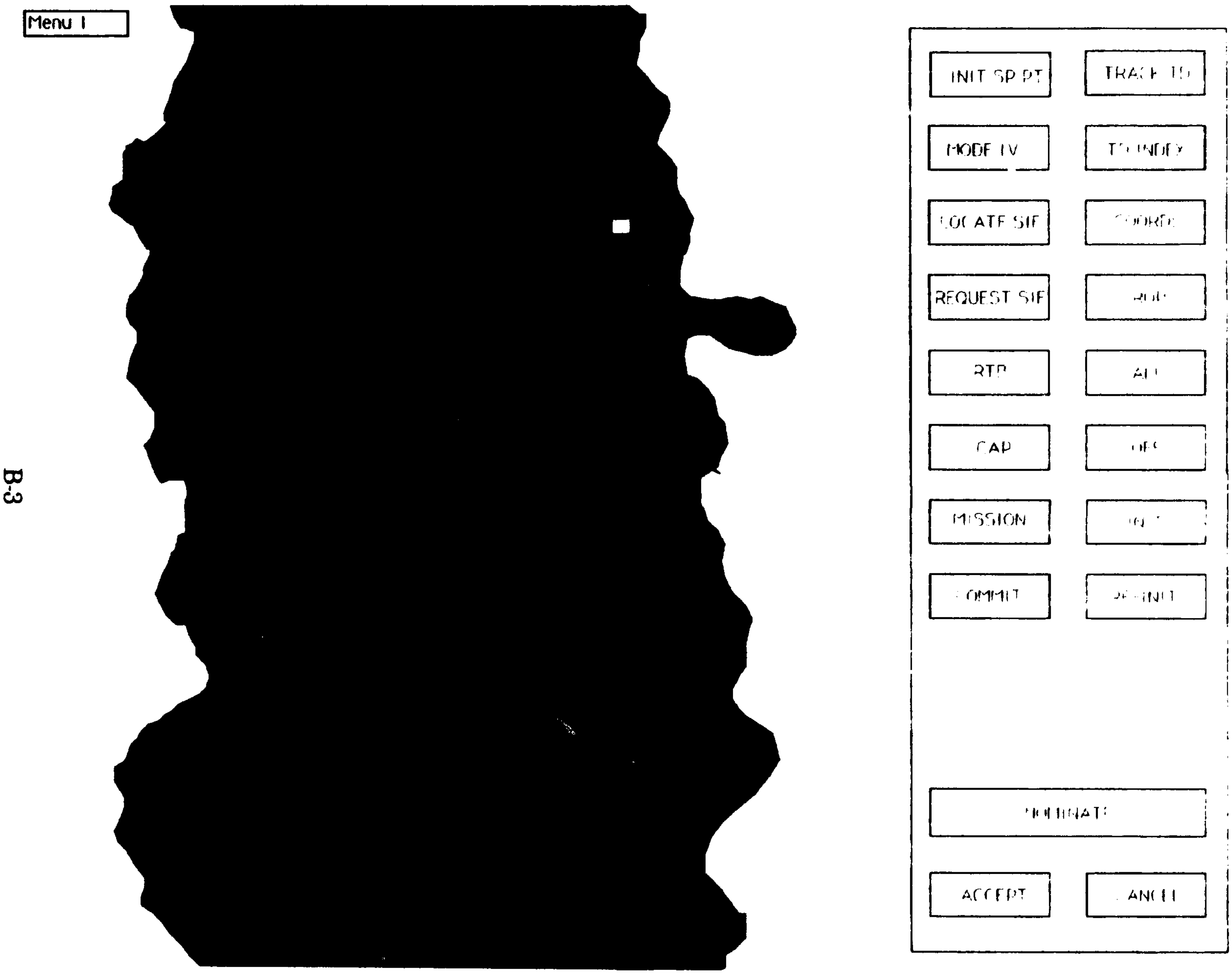
- loss of SA during high switch action periods
- no one trusts the computer geometry for vectoring of fighters
- sometimes, especially at lower expansion levels, it is difficult to know if the aircraft are "feet wet or feet dry," over water or land
- it is often difficult to locate friendly high value assets
- it is easy to attend to a particular area of the scope and forget about the other aircraft

APPENDIX B

Storyboards of Final Display Recommendations

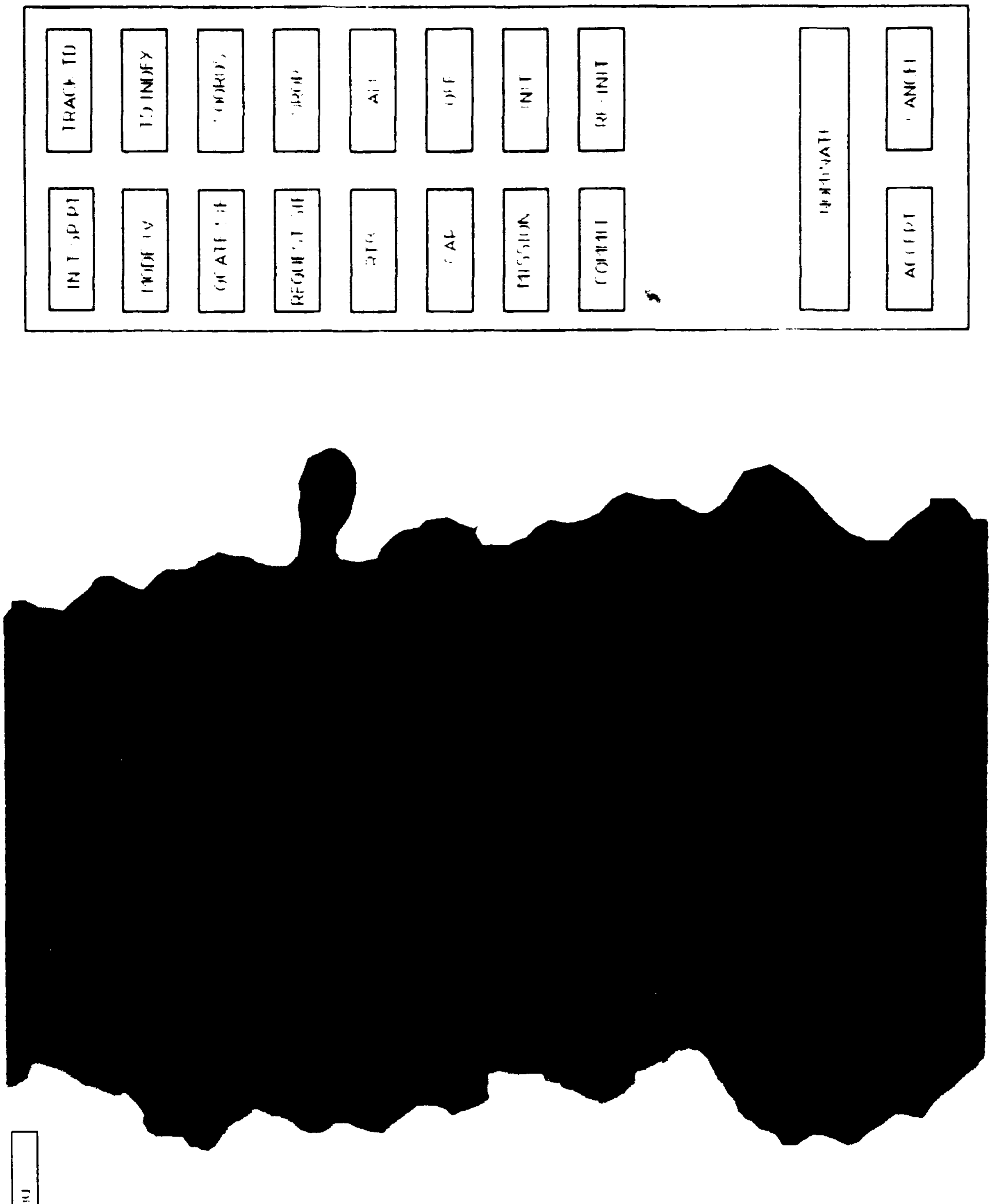
Menu 1

This storyboard represents the display at rest. The on-screen menu is obvious at the right, the water has been shaded blue, and the yellow square represents the current cursor position. The menu is not active at this time.


Menu I
INIT SP PT
MODE IV
REQUEST SIF
RTP
CAP
MISSION
ACCEPT

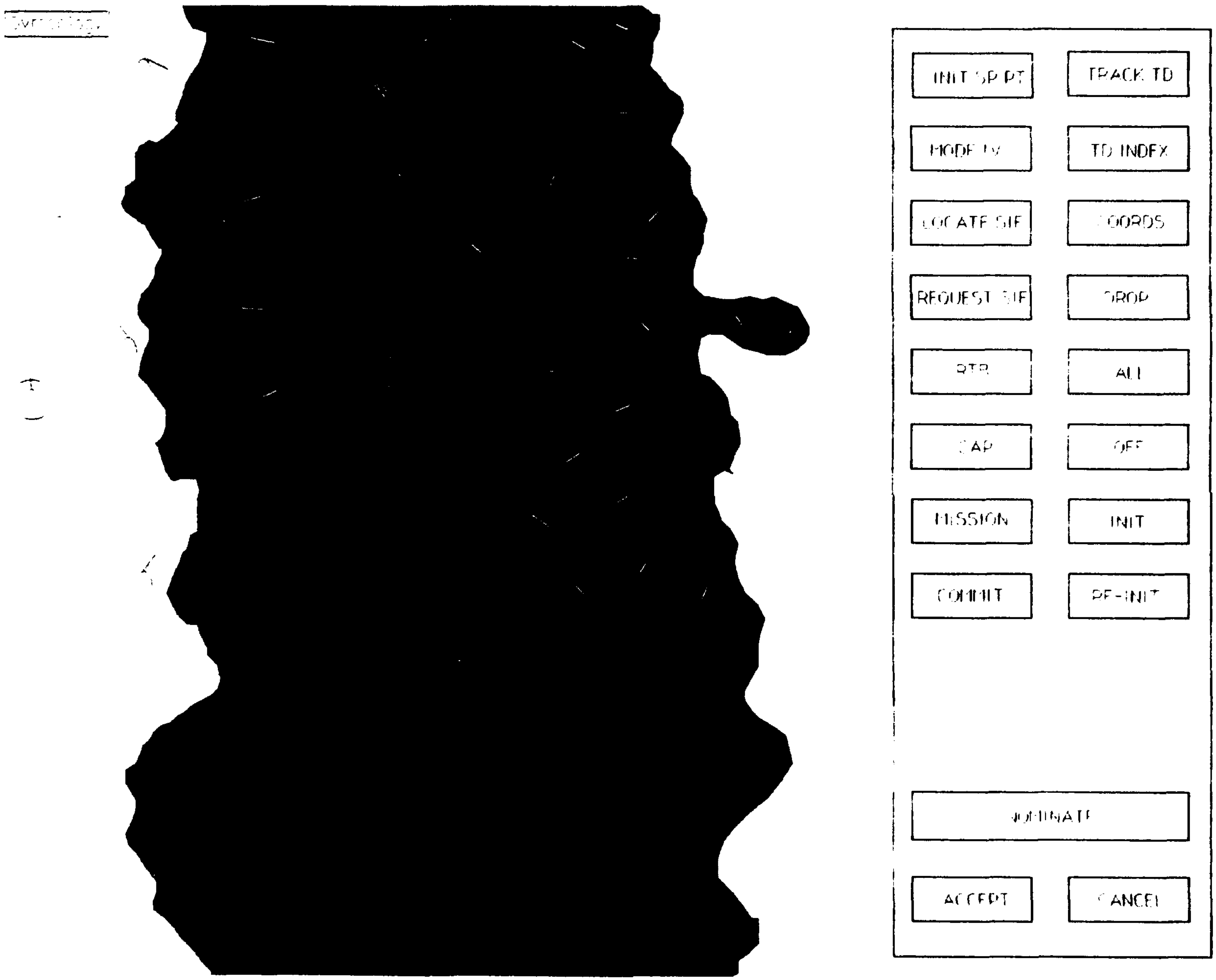
Menu 2

This storyboard is simply a repeat of Menu 1, but the menu is now active. Note the color of the menu has changed from white to yellow, the current cursor position is now represented by an arrow in the menu box, and the previous cursor position is being held at the intersection of the two red lines. At this point the WD would select the desired action by moving the menu cursor onto the desired "button" and pressing the hook button on the trackball. By depressing the middle trackball button, the cursor would return to its previous point, and the menu would deactivate. If the WD "rolls" the cursor out of the menu, the cursor returns to the conventional square representation and the two intersecting lines remain on the screen for five seconds. This allows the WD to return to the previous place on the screen and does not penalize him/her for what could have been a mistake (rolling out).


TRACK TD
TD INDEX
RTB
OFF
MISSION
INIT
COMMIT
RE-INIT
ACCEPT
CANCEL

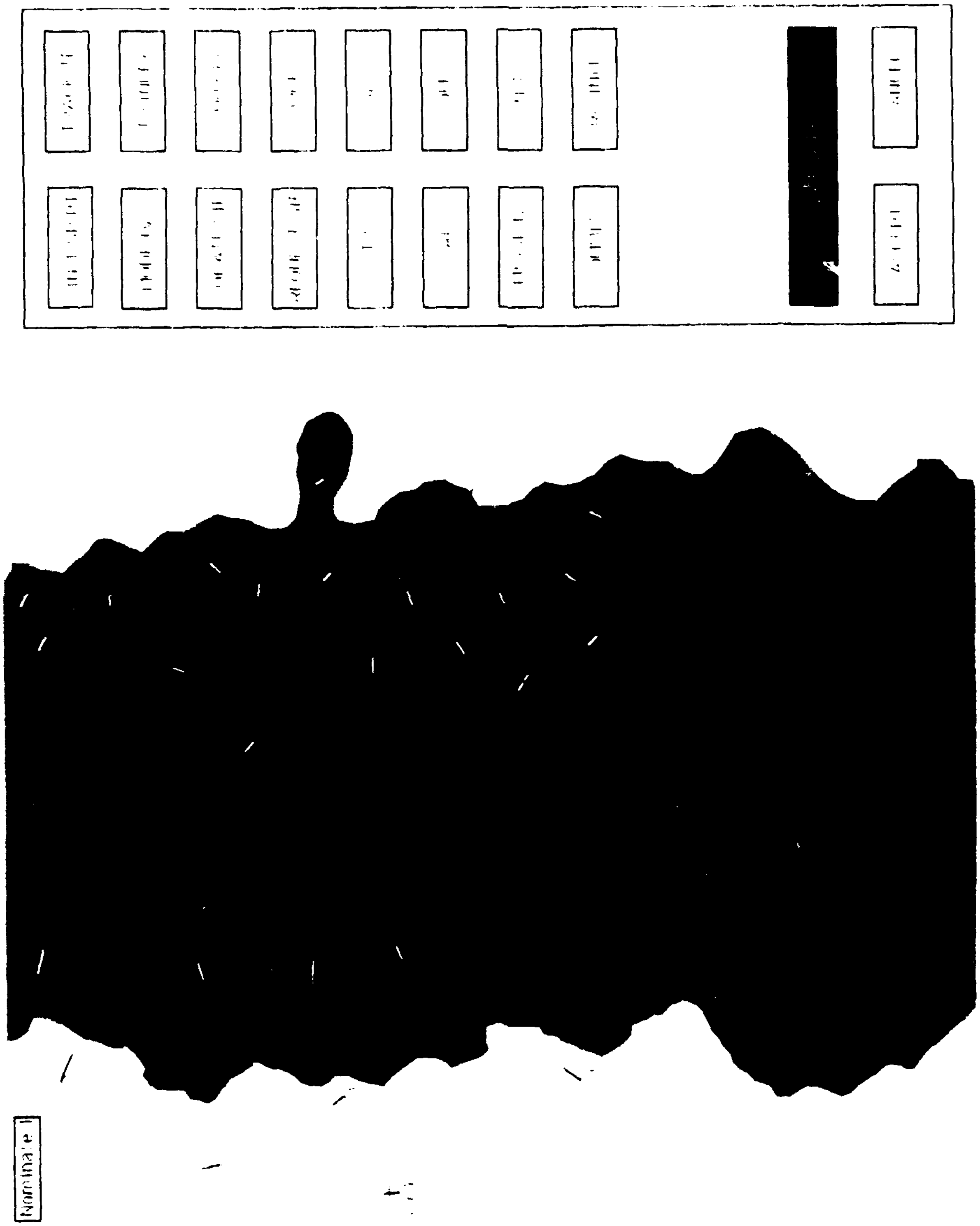
Symbology

This storyboard represents the symbology modifications. The friendly high-value assets and enemy high-threat tracks have been enclosed in a circle. The friendly high-value asset is typically a tanker or the AWACS aircraft itself. The enemy high-threat track is either a high-fast flyer or a jammer.


INIT SP PT
TRACK TD
MODE IV
TD INDEX
LOCATE SIF
COORDS
REQUEST SIF
DROP
RTR
ALL
CAP
OFF
MISSION
INIT
COMMIT
RE-INIT
NOMINATE
ACCEPT
CANCEL

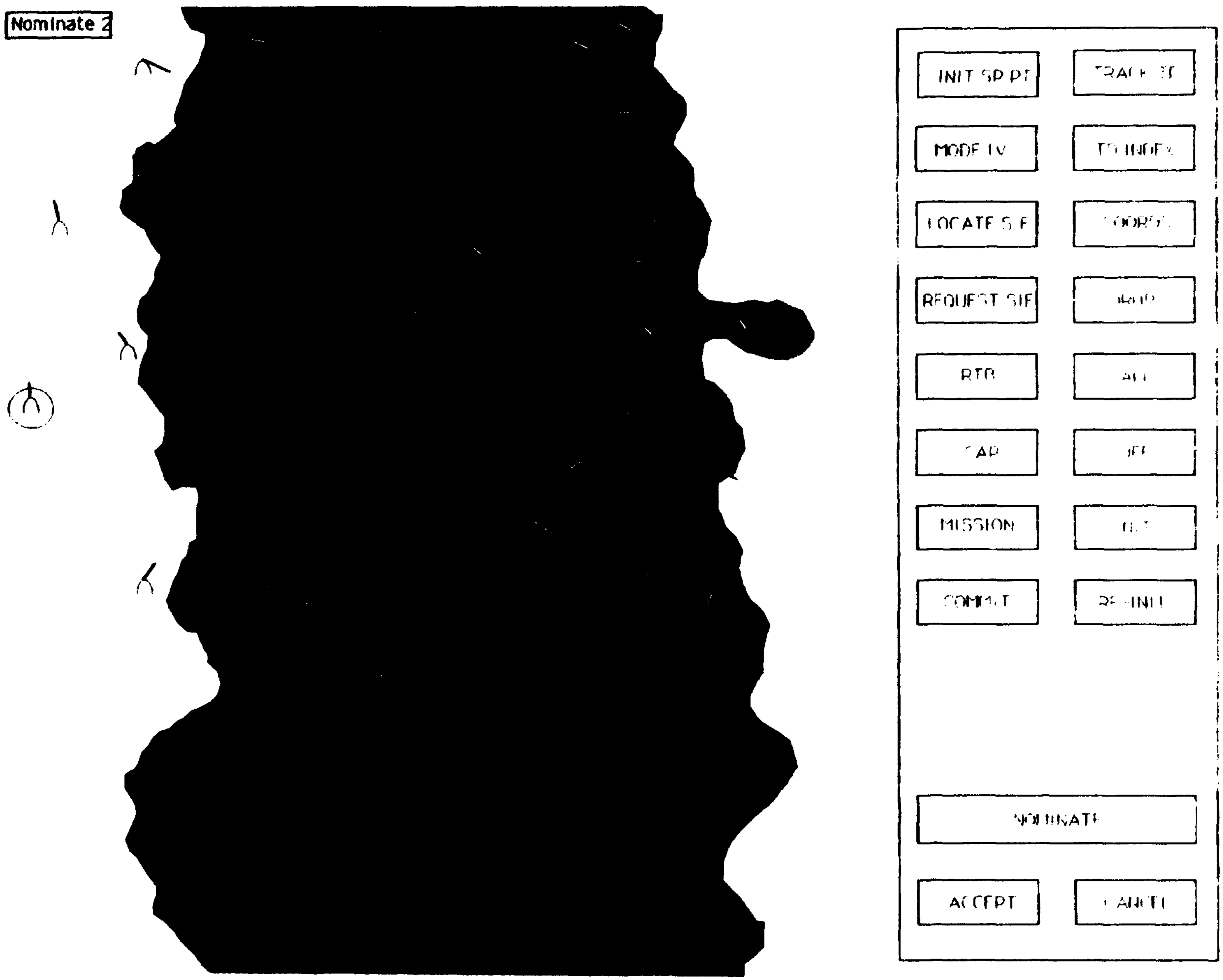
Nominate 1

This storyboard represents the display with both friendly and enemy tracks. You will note that a friendly track also is encircled. This represents a high-value asset, most often a tanker or the actual AWACS aircraft. In this instance the WD is selecting Nominate from the On-Screen Menu. The function of Nominate is to allow the WD to ask the system for recommended fighter pairings. Essentially, the WD is asking the system for help with resource allocations.


Nominate 1

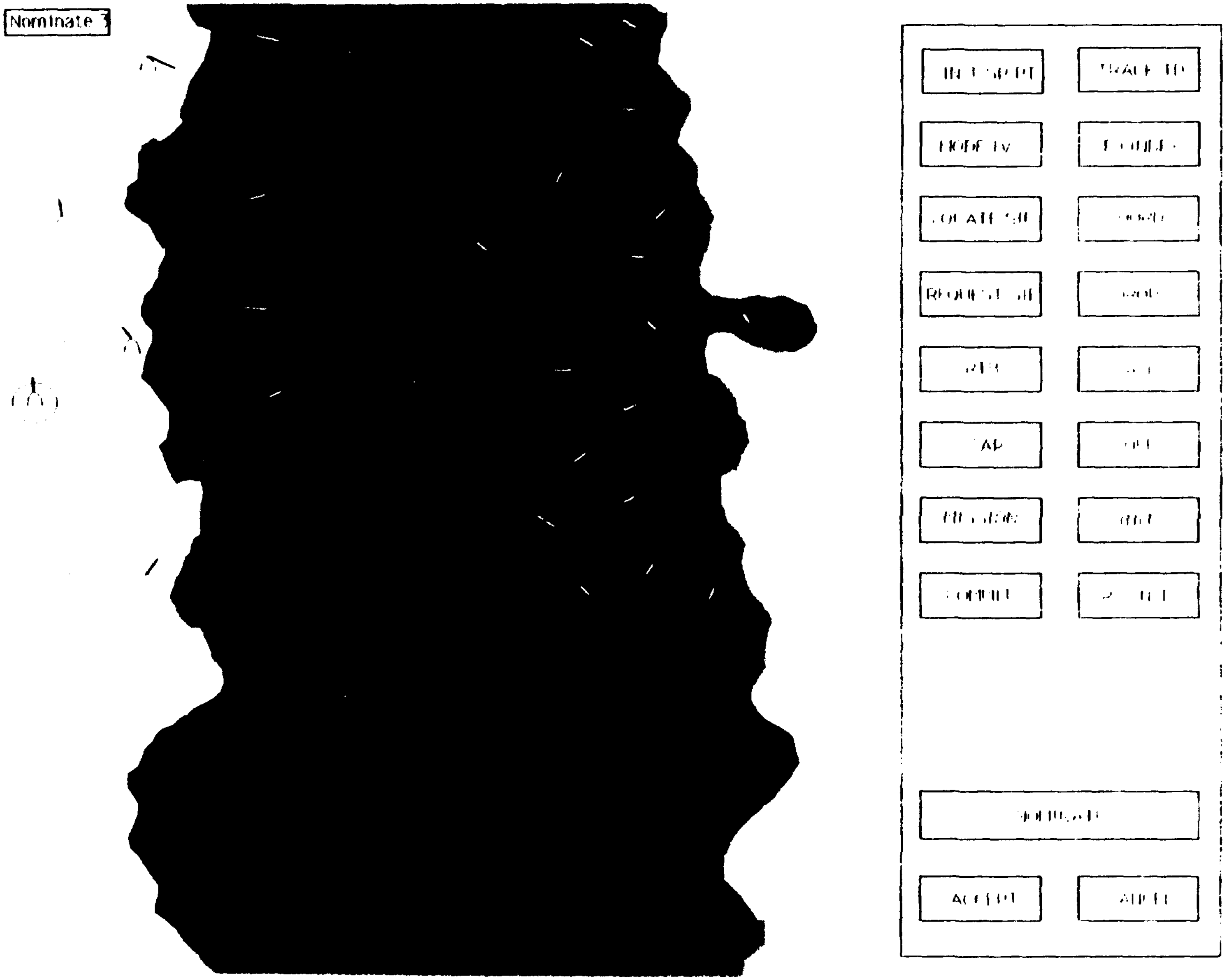
Nominate 2

This storyboard represents the tracks for which the WD has requested help. The system is quickly determining the optimal friendly resources for suggested intercept pairings.


Nominate 2
INIT SP PT
MODE IV
LOCATE SIF
RTB
MISSION
ACCEPT

Nominate 3

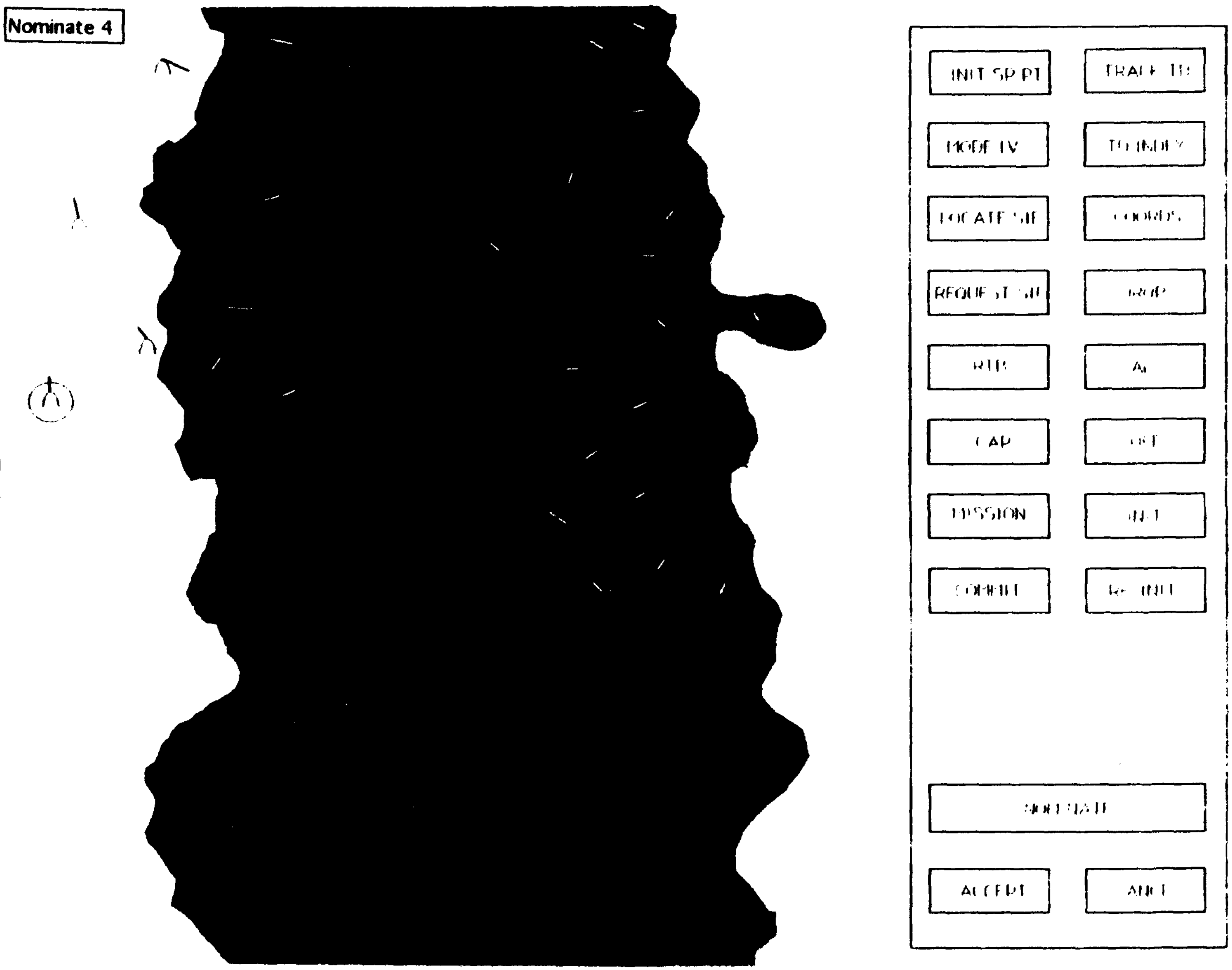
This storyboard represents the system's suggestions. The system has placed green squares around the friendly assets, red squares around the enemy targets, and intercept lines to show the pairings. At this point the WD can Accept any or all of the pairings, Cancel any or all of the pairings, or monitor the situation until a decision can be made. If the WD Accepts the suggested pairings, the system executes a Commit switch action for the pairing. To Accept, the WD simply moves the cursor to either of the tracks in the desired pairing and presses the hook button on the trackball. The system then executes the Commit switch action and provides the WD with the necessary intercept information. To Cancel selected tracks, the WD simply selects Cancel from the menu, and moves the cursor to either of the tracks in the pairing to be canceled and presses hook on the trackball. To Cancel all pairings, the WD simply "double-clicks," using the hook button on the trackball on Cancel from the menu.


Nominate 3

Nominate 4

This storyboard simply represents a different system solution for Nominate given different friendly aircraft locations.

Nominate 4



APPENDIX C

Descriptive Statistics for Low, Medium, and High Experience Groups

Table A

AWACS WD Performance on Current and Revised Systems: Means and Standard Deviations for Low, Medium, and High Experience Groups

LOW EXPERIENCE

Measures	Current System		Revised System		Difference Score
	Mean	S.D.	Mean	S.D.	
Expert Ratings (1 = High)	3.17	1.47	2.33	1.51	0.83 **
Outcome Measures					
Hostiles strikes completed	0.83	1.07	0.50	0.84	-0.33
Number hostile penetrations	5.67	6.09	6.17	4.12	0.50
Penetration depth (penetators only)	73.01	47.09	35.54	13.38	-37.47 *
Penetration depth (all)	15.81	16.67	11.17	10.15	-4.63
Hostiles shot down	19.50	1.22	19.50	0.84	0.00
Friendlies shot down	4.17	3.66	4.50	2.43	0.33
Hostile to friendly kill ratio	7.94	6.60	6.83	6.62	-1.11
Friendlies lost to low fuel	0.50	0.55	0.67	0.82	0.17
Friendlies shot by friendlies	0.17	0.41	0.17	0.41	0.00
Total friendlies lost	5.08	3.32	4.83	4.07	0.50
% fired missiles that missed	8.93	7.76	4.49	6.74	-0.04 **
Process Measures					
Workload					
Response to SD Inquiry					
% correct	6.17	2.04	6.67	2.25	0.05
% only acknowledged	2.33	1.37	2.00	1.10	-0.03
% no response	1.00	1.10	0.67	0.82	-0.03
Avg. response time	7.92	4.75	8.25	5.00	3.33
Response to visual alerts					
% responded to	48.33	11.69	51.67	13.29	0.03
Avg. response time	37.83	6.56	33.30	4.33	-45.33 **
Intercept Approach Specified	17.03	9.64	7.40	6.59	0.10 ***
Situational Awareness					
Recorrelations - % correct	93.32	3.66	92.84	6.70	-0.005
Time symbology incorrect	$2.05x10^4$	$2.29x10^4$	$2.83x10^4$	$4.10x10^4$	$7.83x10^3$
Air refuelings	101.33	2.34	102.17	2.56	0.83 *
AC return to base	106.00	4.56	103.67	2.88	-2.33 *
Airborne orders/scrambles ratio	1.58	2.77	1.42	2.33	-0.16
Workload Ratings					
Overall (NASA TLX)	61.83	9.05	72.90	8.88	11.07 *
Specific Tasks (SWORD)					
-reinitiate	0.06	0.04	0.18	0.17	0.12 **
-pair ADF	0.08	0.02	0.12	0.27	0.05 *
-intercept	0.28	0.13	0.27	0.14	-0.01
Ratings-Revised System Impact (1= High)	N/A	N/A	2.63	0.52	N/A

* $p < .10$; ** $p < .05$; *** $p < .01$, with one tailed t test

Table A (continued)

MEDIUM EXPERIENCE

Measures	Current System		Revised System		
	Mean	S.D.	Mean	S.D.	Difference Score
Expert Ratings (1 = High)	4.40	0.89	3.40	1.52	1.00
Outcome Measures					
Hostiles strikes completed	0.60	0.55	1.20	1.30	0.60
Number hostile penetrations	7.80	1.64	9.60	3.78	1.80
Penetration depth (penetators only)	44.06	14.86	41.90	8.70	-2.17
Penetration depth (all)	17.28	7.55	20.36	9.87	3.08
Hostiles shot down	19.40	0.89	19.60	0.55	0.20
Friendlies shot down	6.40	2.51	4.40	2.07	-2.00
Hostile to friendly kill ratio	3.47	1.44	5.08	1.69	1.16 *
Friendlies lost to low fuel	0.80	0.84	1.00	1.41	0.20
Friendlies shot by friendlies	0.00	0.00	0.00	0.00	0.00
Total friendlies lost	7.20	2.28	5.40	1.95	-1.80
% fired missiles that missed	8.36	5.88	11.47	9.91	+0.03
Process Measures					
Workload					
Response to SD Inquiry					
% correct	5.20	1.79	5.60	2.30	0.04
% only acknowledged	4.20	2.59	3.60	2.19	-0.05 **
% no response	0.40	0.55	0.60	0.89	0.02
Avg. response time	6.98	2.63	7.88	3.41	8.92
Response to visual alerts					
% responded to	48.00	17.89	46.00	27.92	-0.02
Avg. response time	37.36	9.65	37.74	12.63	3.80
Intercept Approach Specified	16.73	13.17	17.31	10.93	0.01
Situational Awareness					
Recorrelations - % correct	87.37	7.35	91.17	1.28	0.04
Time symbology incorrect	2.62×10^4	2.00×10^4	2.57×10^4	1.09×10^4	-4.84×10^2
Air refuelings	100.00	0.00	100.00	0.00	0.00
AC return to base	108.20	7.26	108.20	4.02	0.00
Airborne orders/scrambles ratio		0.17	0.25	0.18	0.16-0.01
Workload Ratings					
Overall (NASA TLX)	62.40	21.37	73.10	11.93	10.70 **
Specific Task (SWORD)					
-reinitiate	0.06	0.03	0.21	0.08	0.09 *
-pair ADF	0.10	0.04	0.25	0.10	0.14 *
-intercept	0.21	0.16	0.17	0.07	-0.20
Ratings-Revised System Impact (1 = High)	N/A	N/A	2.74	0.48	N/A

* $p < .10$; ** $p < .05$; *** $p < .01$, with one tailed *t* test

Table A, continued

HIGH EXPERIENCE

Measures	Current System		Revised System		Difference
	Mean	S.D.	Mean	S.D.	Score
Expert Ratings (1 = High)	3.83	1.17	2.83	0.98	0.82 ***
Outcome Measures					
Hostiles strikes completed	0.50	0.55	0.00	0.00	-0.50 **
Number hostile penetrations	7.00	3.69	6.83	5.34	-0.17
Penetration depth (penetators only)	48.92	14.59	40.88	19.71	-8.04
Penetration depth (all)	15.31	7.06	12.26	12.84	-3.05
Hostiles shot down	19.50	0.55	20.17	0.41	0.67 *
Friendlies shot down	6.00	1.79	5.00	1.90	-1.00 *
Hostile to friendly kill ratio	3.49	1.02	4.59	1.79	1.10 **
Friendlies lost to low fuel	0.33	0.52	0.50	0.84	0.17
Friendlies shot by friendlies	0.00	0.00	0.00	0.00	0.00
Total friendlies lost	6.33	2.16	5.50	2.07	-0.83 *
% fired missiles that missed	9.47	8.62	2.38	3.98	-0.07 **
Cognitive Process Measures					
Workload					
Response to SD Inquiry					
% correct	5.67	2.88	6.17	1.72	0.05
% only acknowledged	3.17	1.17	3.00	1.79	-0.02
% no response	0.50	0.84	0.50	0.55	0.00
Avg. response time	7.02	3.89	8.82	3.26	18.00
Response to visual alerts					
% responded to	43.33	17.51	43.33	27.33	0.00
Avg. response time	40.42	6.99	40.55	14.39	1.33
Intercept Approach Specified	13.29	13.75	14.77	12.66	0.01
Situational Awareness					
Recorrelations - % correct	91.56	5.49	99.09	3.95	0.04
Time symbology incorrect	$2.49x10^4$	$2.02x10^4$	$1.90x10^4$	$2.52x10^4$	$5.97x10^3$
Air refuelings	102.17	2.14	104.00	2.90	1.83 ***
AC return to base	105.17	2.99	104.00	3.35	-1.17
Airborne orders/scrambles ratio	0.74	0.62	1.04	1.09	0.30
Workload Ratings					
Overall (NASA TLX)	69.93	7.28	70.40	10.97	0.47
Specific Tasks (SWORD)					
-reinitiate	0.10	0.06	0.24	0.12	0.14 **
-pair ADF	0.10	0.05	0.18	0.09	0.08 *
-intercept	0.22	0.06	0.16	0.06	-0.06
Ratings Revised System Impact (1 = High)	N/A	N/A	2.52	0.53	N/A

* p < .10; ** p < .05; *** p < .01, with one tailed t test

Table B

Percent Change in Performance for Low, Medium, and High Experience Groups

		Experience		
		Low	Medium	High
Measures	Desired Impact	% Change	% Change	% Change
Expert Ratings (1 = High)	+	27%	23%	27%
Outcome Measures				
Hostiles strikes completed	–	-40%	50%	– [1]
Number hostile penetrations	–	9%	23%	-2%
Penetration depth (penetrators only)	–	-51%	-5%	-16%
Penetration depth (all)	–	-29%	17%	-20%
Hostiles shot down	+	0%	1%	3%
Friendlies shot down	–	7%	-32%	-17%
Hostile to friendly kill ratio	+	-14%	46%	32%
Friendlies lost to low fuel	–	33%	25%	33%
Friendlies shot by friendlies	–	0%	0%	0%
Total friendlies lost	–	9%	-25%	-13%
% fired missiles that missed	–	50%	37%	-75%
Process Measures				
Workload				
Response to SD Inquiry				
% correct	+	8%	8%	9%
% only acknowledged	–	-13%	-14%	-5%
% no response	–	-33%	33%	0%
Avg. response time	–	4%	13%	25%
Response to visual alerts				
% responded to	+	7%	-4%	0%
Avg. response time	–	-12%	1%	0%
Intercept Approach Specified	+	-56%	3%	11%
Situational Awareness				
Recorrelations - % correct				
Time symbology incorrect	–	30%	-2%	24%
Air refuelings	+	63%	0%	85%
AC return to base	–	-39%	0%	-23%
Airborne orders/scrambles ratio	+	-10%	6%	41%
Workload Ratings				
Overall (NASA TLX)	–	12%	17%	1%
Specific Tasks (SWORD)				
-reinitiate	–	69%	69%	58%
-pair ADF	–	37%	58%	44%
-intercept	–	-2%	-21%	-36%
Ratings-Revised System Impact (1 = High)	N/A	N/A	N/A	N/A

[1]No hostile strikes were completed against the high experienced WDs when using the revised system.