Air Threat Assessment: Research, Model, and Display Guidelines

Michael J. Liebhaber[1]
Pacific Science & Engineering Group, Inc.
6310 Greenwich Drive, Suite 200
San Diego, CA 92122

Bela Feher
Space and Naval Warfare Systems Center
Bldg. A33, Room 1401
San Diego, CA 92152-5143

Abstract

This paper briefly reviews a series of studies [Liebhaber, 1999; Liebhaber, 2000) that were undertaken to investigate the practice of U.S. Navy air threat assessment, describes a model of threat assessment that was created from the research data, and proposes guidelines for a threat assessment display within the context of an air defense decision support system (DSS). The studies provided a theoretical and applied basis for threat assessment by defining specific cue-data relationships and detailing the cognitive processes involved in air defense situation assessment. Those processes were incorporated into a proposed model of threat assessment that was successfully validated against threat ratings from experienced air defense decision makers. These data, combined with earlier work on DSS interfaces [Miller, 1992; Rummel, 1995] and recent discussions with air defense officers, were used to develop guidelines for displaying threat assessment data.

1.0 Introduction

Air defense decision-making is a complex task accomplished by a team of highly skilled personnel. Threat assessment is a fundamental, but poorly understood, component in that decision-making process. However, recent cognitive analysis of air threat assessment (e.g., Liebhaber, 1999; Liebhaber, 2000), combined with conclusions from earlier work (e.g., Miller, 1992; Rummel, 1995), has provided some insight into the data and processes used by experienced air defense personnel, and enabled the development of a threat assessment algorithm and a set of interface guidelines. This paper summarizes the key findings from recent studies of U.S. Navy air threat assessment, briefly describes a threat assessment model that was created from the research data, and proposes guidelines for a threat assessment display within the context of an air defense decision support system (DSS). The goals of this program of research were to understand the assessment process in order to present assessment support information to the user in a format that minimizes any mismatches between the cognitive characteristics of the human decision maker and the DSS.

1.1 Understanding Threat Assessment

Air defense decision-making has severe (possibly catastrophic) consequences for errors. It is a complex task accomplished by a team of highly skilled personnel. It requires mental integration of data from many sources. That integration requires a high level of tactical expertise, including

[1] The work reported here was performed for the Simulation and Human Systems Technology Division of SPAWAR Systems Center (SSC) San Diego by Pacific Science and Engineering Group, Inc. under contract number N66001-99-D-0050.

Report Documentation Page			*Form Approved* *OMB No. 0704-0188*
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.			
1. REPORT DATE **2002**	2. REPORT TYPE		3. DATES COVERED **00-00-2002 to 00-00-2002**
4. TITLE AND SUBTITLE **Air Threat Assessment: Research, Model, and Display Guidelines**			5a. CONTRACT NUMBER
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)			5d. PROJECT NUMBER
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Space and Naval Warfare Systems Center, San Diego,Bldg. A33, Room 1401,San Diego,CA,92152-5143**			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)			10. SPONSOR/MONITOR'S ACRONYM(S)
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**			
13. SUPPLEMENTARY NOTES **The original document contains color images.**			
14. ABSTRACT			
15. SUBJECT TERMS			

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**15**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

knowledge of the types of threats, ship's mission, Navy doctrine, and assessment heuristics built from experience.

Cognitive complexity is introduced by the multi-tasking requirements of the task [Chalmers, 1998]--Air defense personnel are often engaged in competing tasks. In addition to being responsible for all aircraft in their surveillance area, team members must maintain awareness of available resources, monitor audio and verbal messages, and prepare situation reports. To complicate matters further, a typical Combat Information Center (CIC) has over 20 consoles and up to 8 interface formats [Lyons, 2000], and critical data are manually recorded on a whiteboard or notepad. In this environment, it can be difficult for Air defense team members to notice or identify key pieces of information that may enable them to better understand the tactical situation. Due to the multi-tasking, tempo, integration demands, and short-term memory requirements, the task of the air defense decision maker can be characterized as cognitively challenging under usual conditions, and possibly worse under extreme conditions.

1.2 Providing a Cognitive Basis for Decision Support Systems

Tragic incidents in the late 1980s, such as the USS Vincennes incorrectly identifying a non-threat as a threat, and the USS Stark failing to identify a threat [Hutchins, 1996], made decision support issues a fleet priority. However, the threat assessment process, as it occurred in the operational environment, was not well understood or documented. Decision aiding was not based on cognitive decision-making research, and displays of that period were inadequate for integrating and conveying tactical information. Therefore, the Tactical Decision Making Under Stress (TADMUS) program was begun to evaluate Decision Support System (DSS) display concepts derived from current cognitive theory, most notably Naturalistic Decision Making [Zsambok, 1997]. A prototype DSS was developed and tested [Kelly, 1996]. The DSS was designed to support the cognitive strategies of tactical decision makers operating in highly complex, fast-paced littoral environments. The DSS augmented traditional displays with additional information and organized the information to support the decision problems dealt with by the commander, providing him with decision support tools that facilitated detailed, yet rapid analyses of tactical data. See Hutchins {1996}, Rummel [1995], and Morrison [2000] for reviews of the DSS.

One component of later versions of the DSS was called the Basis for Assessment (BFA) tool. It was a section of the DSS display that supported explanation-based threat assessment, in a manner similar to that described by Pennington and Hastie [1988], by providing system operators with a detailed list of evidence for and against the current assessment of the selected track (aircraft radar contact). The display was designed to present the relevant data necessary for a commander to evaluate all likely explanations for what a target might be, and what it might be doing. Its purpose was to reduce the likelihood of mistakenly mis-categorizing and engaging friendly or neutral tracks [Morrison, 1997]. While the DSS as a whole was successful, there was minimal theoretical and applied investigation of the threat assessment concepts. Feedback from experienced Navy personnel indicated that the BFA tool needed several design improvements behind the interface. The automated algorithm for assessing and rank ordering high threat tracks was deemed to be overly simplistic. A series of BFA studies was initiated to gain a better understanding of the human threat assessment process, and to develop algorithms and displays that better reflected that process.

2.0 Overview of Threat Assessment Research

A progressive series of in-depth studies were conducted to identify the cues (also referred to as data, attributes, or characteristics) that experienced air defense personnel use to evaluate the level of threat posed by a particular aircraft, specify the correspondence between aircraft behaviors and threat ratings, determine how air defense personnel assign a threat rating, and evaluate the impact of each factor on threat rating. Research methods included knowledge engineering and cognitive task analysis [Liebhaber, 1999], experiments [Liebhaber, 2000], and cognitive modeling (Liebhaber, 2001). Research participants were U.S. Navy personnel with an average of 3 ½ years of experience at-sea in one or more air defense (AD) roles within an Aegis Combat Information Center (CIC).

Earlier studies (e.g., Kaempf, 1992; Marshall, 1996; Miller, 1992; Schulze, 1999) have identified and studied cues within the context of air defense decision-making. However, our more recent series of studies identified a comprehensive set of cues, established the mapping between specific aircraft behaviors and threat ratings, and determined the effect of conflicting data on threat rating. In addition, our studies relied exclusively on U.S. Navy personnel who had air defense experience during one or more tours at-sea. Findings that are pertinent to the development of the threat assessment algorithm and display are briefly covered here. See Liebhaber and Smith [1999] and Liebhaber, Kobus, and Smith [2000] for more detail.

2.1 Cues Used In Threat Assessment

Although we identified a weighted list of 18 cues, participants typically used only 6 to 13 to assess aircraft. There was some overlap between the 18 cues and the cues reported by Miller [1992]. Cue weights were calculated from their frequency of use and position in the sequence used by participants. For example, high weight cues were used often and early in the evaluation process. The cues are listed in Appendix A. Six critical cues were identified based on their relative weights. They were, in order of importance: Origin, IFF Mode, Intelligence Report, Altitude, Proximity to an Airlane, and ESM (Radar Signature). The type and number of cues used by participants depended primarily on the type of aircraft that they thought they were evaluating. For example, participants in Liebhaber and Smith [1999] chose one set of cues to evaluate when they thought a track was a commercial aircraft (COMAIR), and a different, but overlapping, set of cues when they thought a track was a tactical military aircraft (TACAIR). These sets were labeled templates.

2.2 The Threat Assessment Process

Results indicated that experienced AD team members formulated a hypothesis about the current track (i.e., activated a template or schema that corresponded to a particular type of aircraft), evaluated the evidence (i.e., cues), and then generated a plausible assessment based on the support or contradiction of the data. This finding is consistent with some of the conclusions that grew out of the TADMUS program [Morrison, 2000]. The templates contained sets of cues with expected data values (i.e., behaviors), and a baseline threat rating that was consistent with a particular type of aircraft (e.g., commercial airliner). Participants appeared to evaluate cues in a relatively fixed sequence. However, there was evidence that cues were processed in sets or chunks (i.e., Origin & IFF Mode, then Intel, Altitude, & Airlane, and then ESM). Evaluation order was based on the weight or importance of each cue. Heavily weighted cues were evaluated earlier than low weighted cues. Perceived threat ratings were directly related to the degree of fit of aircraft data to template-driven expectations; aircraft that better fit the expected behaviors were assigned a lower threat rating. There was not any evidence that participants switched templates in the face of conflicting data. Instead, they accommodated conflicting data into their active template. Finally, participants appeared to be influenced by specific cues rather than the overall pattern of data. For example, participants were likely to change threat ratings if only one of the high weighted cues (i.e., one of the first six cues) contained data that conflicted with their expectations for the aircraft.

3.0 Threat Assessment Model

An overview of the proposed threat assessment model based on the above findings is shown in Figure 1. The process of threat assessment is contained within the *Assimilate Air Warfare Data* process. The model attempts to accurately incorporate the cognitive processes that are followed by experienced air defense personnel. It is not intended to be the most computationally efficient method, or to conform to any currently prescribed method for aircraft identification and assessment. It is presumed that the output of the model feeds into a decision making process such as those described by Endsley [1995] or Klein [1993]. As such, the proposed model may correspond to the early stages of situation awareness involved with perceiving and comprehension.

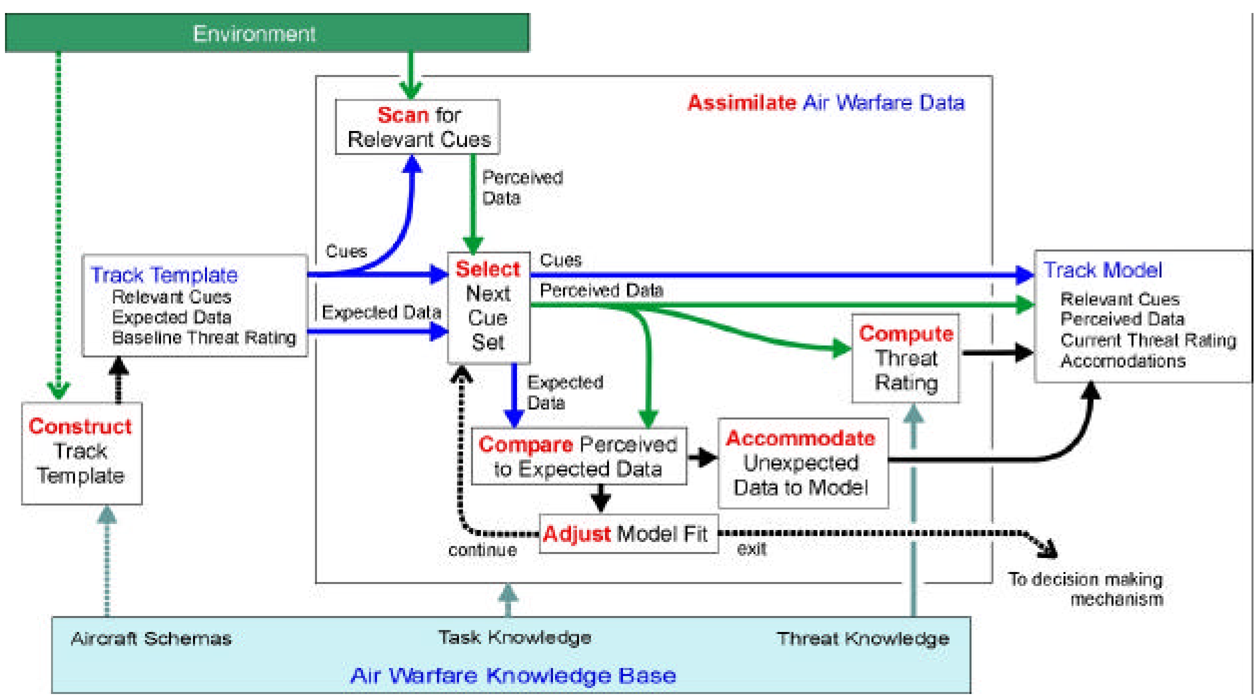


Figure 1. Threat Assessment Model

Information flow through the model begins with the construction of a template. Arrows indicate the direction of data flow. Dashed lines are used to indicate connections that were not studied. While we could infer from the data that track templates were being used by the research participants, we did not investigate how the templates were constructed or activated. Templates contain relevant cues for a particular type of aircraft, ranges of expected behaviors (data) for each cue, and a baseline threat rating.

Within the *Assimilate Air Warfare Data* process, the six critical cues are always evaluated first. The remaining cues are evaluated only if unexpected data are encountered in one or more of the six critical cues. Unexpected data are perceived data that do not fall within the range of expected values for the given cue, as defined by the active template. Not all available cues are evaluated; only those that are part of the template.

3.1 Scan and Select Cues

The environment (through CIC consoles, Intel reports, etc.) is scanned for cues (e.g., Altitude, Speed) that are relevant to the active template. The set of cues to be evaluated are selected from the input. The selection mechanism was inferred from data that indicated that participants were processing combinations of cues [Liebhaber, 2000].

3.2 Compare, Adjust Fit, and Accommodate

The perceived data (e.g., 10,000 ft.) are compared to the expected data (e.g., Altitude 20,000 ft.). If the perceived data are unexpected, then the fit of the model is reduced by the relative weight of the cue. An accommodation, typically an explanation or hypothesis, may be provided to reconcile the unexpected data to the template [Liebhaber, 1999]. Explanations attribute the data to another cause. Hypotheses attribute the data to a plausible inferred intent of the track. Accommodations were provided in about 36% of the cases where there was unexpected data.

3.3 Compute Threat Rating

The perceived data are also used to compute the current threat rating. The participants adjusted the baseline threat rating up or down, depending on the degree of threat associated with the current piece of data (e.g., for an aircraft in a littoral environment, a speed of 250 knots adds 0.2 to the current threat rating; a speed of 500 knots adds 1.8 to the current threat rating). Determination of the size of adjustments to threat level was done with data obtained from experienced participants. They provided information on the degree of threat posed by specific changes to a wide range of cue values. It did not appear that threat rating was simply the accumulation, or summing, of evidence. If this were the case, then one would expect that the number of cues that were evaluated would be related to threat rating (i.e., more cues lead to higher threat ratings); analysis reported in Liebhaber, Kobus, and Smith [2000] indicated that threat rating was independent of the number of cues examined.

3.4 Continued Processing

In keeping with empirical findings, unless unexpected data were encountered the model will stop after the first six cues. Otherwise, the model will continue reading, comparing cues, adjusting fit, and computing threat rating until either the fit of the model returns to 100% or there are no more cues to process. If the perceived data are within the range of expectations then the fit of the model is increased by the relative weight of the cue. Otherwise fit is reduced, prompting assessment of further cues, as before.

3.5 Threat Assessment Algorithm

A ruled-based threat assessment algorithm was developed from the model. It analyzed relevant data and computed a threat rating for each air track. The output of the algorithm was compared to that of 17 experienced[2] participants in Experiment 2 of Liebhaber, Kobus, and Smith [2000]. The output was obtained by running the algorithm on a data set from Experiment 2. Threat Ratings of the algorithm were significantly correlated with those of the experts (r = .708, p = .022), and it performed within the range of human expert variability.

Performance of the algorithm was also compared to previous research findings to determine if it rated the threat level of Friendly and Enemy tracks in a manner similar to the human experts. Multiple comparisons indicated that the threat ratings that were produced by the algorithm were similar to those produced by human experts in the earlier studies ($p < .05$). Like its human counterparts, the algorithm computed significantly higher threat ratings for Enemy tracks than for Friendly tracks.

4.0 Threat Assessment Interface Guidelines

Several capabilities have been associated with the threat assessment display from its conception in the TADMUS DSS (e.g., see Kelly, 1996 or Morrison, 1997). Primary among them has been the need to support explanation-based reasoning through the use of evidence lists. The BFA research program has focused on investigating the nature of the evidence and how it contributes to the threat classification of the human decision makers.. The conclusions of our research have been combined with the recommendations of Miller [1992] and Rummel [1995] to create a set of desirable operational capabilities for a threat assessment system. A threat assessment system should be able to

- Compute and display the threat rating of tracks from the evidence (data),
- Support explanation-based reasoning,
- Avoid user errors caused by framing, anchoring, and confirmation biases, and
- Generate a track priority list that matches highly with human-generated lists.

These capabilities were used as a framework for developing display guidelines.

There is a large body of information on displays that is potentially relevant to a threat assessment interface. Liebhaber & Feher [2002] reviewed a number of studies that compare verbal, numeric, and graphic presentation of probabilistic data. Their primary application of their findings to threat assessment displays is that they suggest that depicting probabilistic information graphically leads to less bias [Anderson, 1998; Gigerenzer, 1995; Kirschenbaum & Arruda, 1994] and decreases risk-taking behavior [Stone, 1997]. In addition, air defense officers indicate that they have less confidence in numerical readouts of aggregated information, like a threat

[2] Mean = 2.9 years of at-sea experience in an air defense role.

rating, because numbers imply a degree of precision that is not present in what is essentially a subjective piece of information (personal communication, 2001).

4.1 Display Recommendations from TADMUS DSS Studies

This section briefly reviews recommendations for DSS displays from studies that extensively evaluated the TADMUSS DSS. Only those recommendations that apply to threat assessment are included here.

Miller [1992] proposed eight display enhancements for the TADMUS DSS based on in-depth interviews with air defense officers. While their ideas covered many aspects of the DSS, those that most directly pertain to threat assessment were:

- To use base rates to indicate the typicality of certain events,
- To permit the operator to compare alternative hypotheses,
- To indicate the perceived level of threat posed by a particular track (and provide a rationale for its assessment), and
- To locate all relevant information near the track and on-screen.

Rummel [1995] also developed several display recommendations for the TADMUS DSS based on extensive interviews with experienced air defense officers. Some of his recommendations pertinent to air warfare were to provide:

- Raw data to enable verification of information,
- A comprehensive list of evidence that supports the current hypothesis regarding the track [but avoiding] alphanumeric format, and
- A track priority list that takes a wider range of data into consideration.

4.2 Threat Assessment Interface Guidelines

The following guidelines were developed from concepts derived from the BFA research program discussed above, evaluations of the TADMUS DSS interface [Miller, 1992; Rummel, 1995], and recent discussions with air defense officers. See Liebhaber and Feher [2002] for more information on the interface. Based on the available data, the content of a threat assessment display should have the following features.

1. **Display a threat assessment window on-screen when a track is hooked.** The window should contain an indication of threat rating, threat history, and a comprehensive list of cues. Air defense officers indicated that they would prefer this type of data instead of the standard character readout (CRO) (personal communication, 2001). They also preferred that the window be on the Geoplot, close to the track. Participants in Miller [1992] also suggested on-screen access to track data.

2. **Compute and display the threat rating of tracks.** Display the threat rating in graphic format to avoid a false sense of precision. Experienced air defense officers felt that perception of threat was a fuzzy concept and should not be indicated with a number (personal communication, 2001). Numbers, in their opinion, implied a false sense of accuracy. Their opinion was supported by Kirschenbaum and Arruda [1994], who found that graphic formats were better for tasks that required a value judgment (e.g., more or less threatening). A graphic format may also promote less risky decisions [Stone, 1997]. Threat ratings should be displayed with verbal descriptors (High, Medium, and Low) rather than numbers or percentages. The air defense officers felt the descriptors were potentially less confusing than numbers, and that three levels of discrimination were sufficient. Their statements concur with the findings of Wallsten, Budescu, Zwick, and Kemp [1993] regarding verbal descriptors being easier to understand than numbers.

3. **Show threat rating history**. This would enable a better sense of track history, a DSS enhancement noted by Miller [1992]. It may also enhance explanation of the data by promoting story coherence and consistency [Pennington, 1988], and was requested by users giving feedback on a proposed threat assessment interface (personal communication, 2001).

4. **Provide a list of all assessment cues**. The corresponding data values (e.g., Speed = 230 knots) should also be displayed. These must be the same cues that are used in the assessment algorithm that computes the threat rating. Providing a comprehensive list avoids several biases, and is consistent with user preferences regarding verification and confidence [Miller, 1992; Rummel, 1995]. In addition, the full list should help avoid over-reliance on only a few cues.

5. **Order cues by importance to the decision maker.** List from most to least important. This helps the display conform to user expectations and facilitates building a coherent story that explains the evidence, and thus facilitates decision-making [Pennington, 1988]. Importance of the current set of cues was determined empirically in Liebhaber, Kobus, and Smith [2000]. Presentation order has been shown to affect judgments of U.S. Navy [Perrin, 2001] and U.S. Army [Adelman, 1996] air defense officers.

6. **Show the impact of each cue on overall threat rating.** Each cue should have a graphic frequency indicator that shows how far the data value deviates from the cue’s expected value. To support explanation-based reasoning, use two panels to display the indicator bars: Supporting and Counter Evidence. This method of display would help avoid familiarity bias, over-reliance on a subset of cues, and was preferred by experienced air defense officers over two panels of alphanumeric data only (personal communication, 2001). Displaying cues in order of preferred use would not overcome user’s reliance on the first few cues, or the influence of a change to one of the high-weighted cues. Therefore, it would be helpful to show the impact of each cue on overall threat rating. Doing so could avoid potential over-reliance on the initial few cues, especially when they have little impact on the overall threat rating.

7. **Provide a track priority list.** The list would be separate from the threat assessment window. The purpose of the priority list is to reduce cognitive load of air defense personnel by giving them the ability to grasp mission-critical information at a glance. It would show tracks in order, from most to least threatening, as computed by the threat assessment algorithm. It would be the basis for Contacts of Interest (COI) and Critical Contacts of Interest (CCOI) that are maintained by the air defense team. Based on personal communication (2001), the priority list should provide the following essential data: Track number, bearing, platform, and threat rating. A priority list generated by an assessment algorithm would use a broader range of data than current systems, and would facilitate the creation of situation reports, a feature requested by DSS users [Rummel, 1995].

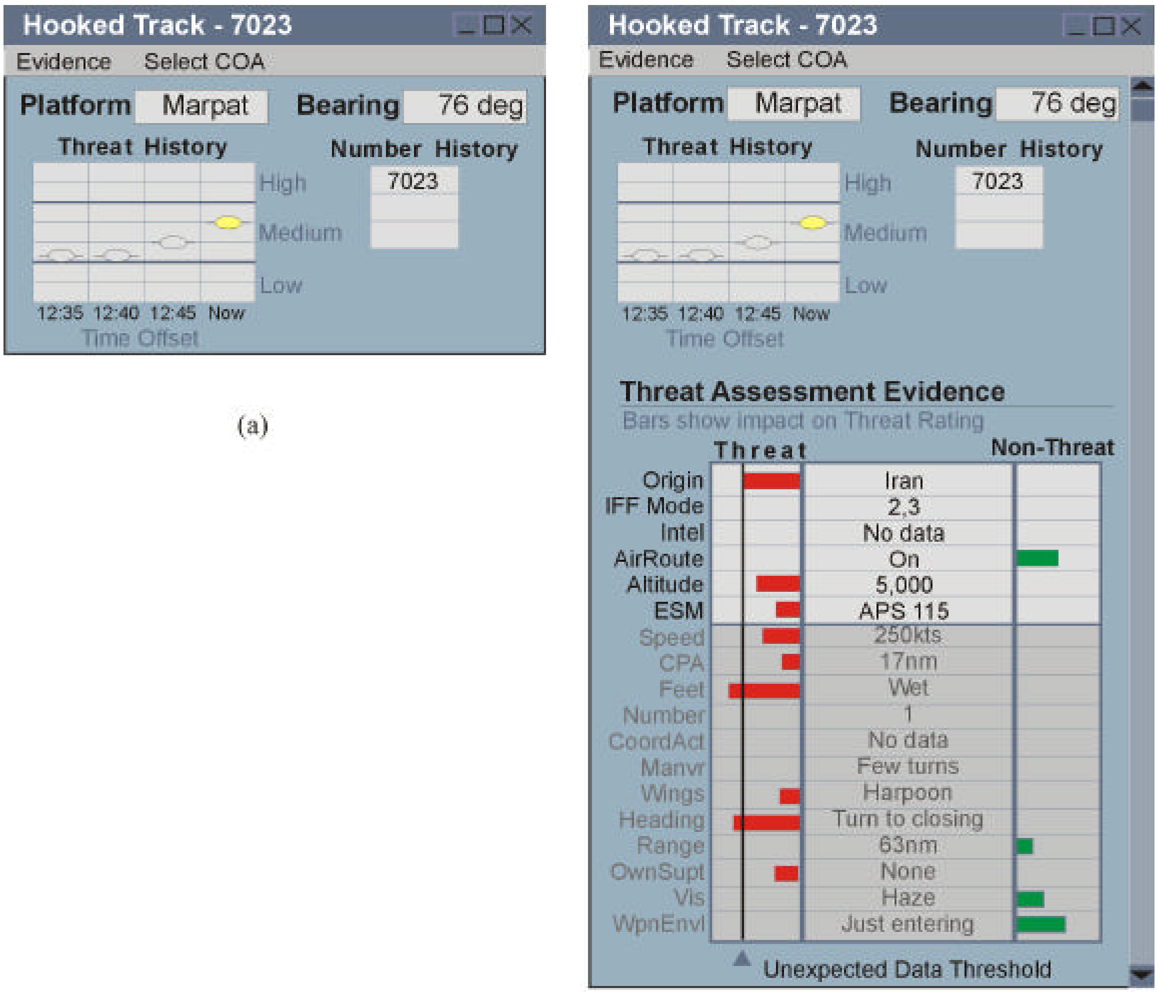
4.3 Proposed Threat Assessment Display

This section describes a proposed threat assessment display (called a Threat Window) and a Priority List Window. It is envisioned that these windows would be part of a DSS similar to the TADMUS DSS. The threat assessment system that drives the windows would continuously receive data on air tracks (i.e., radar contacts) from ship sensors and other information sources, compute or recompute a threat rating for each track, and produce a list of tracks, prioritized from highest to lowest threat, for display. For hooked (i.e., selected) tracks, the system would display the evidence (i.e., data) that went into the threat calculation.

4.3 Threat Window

The Threat Window was developed from the empirical data gathered in our experiments to implement the concepts and guidelines discussed above. The Threat Window is shown in Figure 2 as a window labeled “Hooked Track”. Figure 2a shows how the threat window looks when the system operator hooks a track. Only essential track information appears in this reduced window: Track number, platform (i.e., type of track), bearing, threat history, and track number history. Earlier versions of the interface showed only current threat rating vice threat history. Threat history and track number history were requested by experienced air defense officers who provided feedback on versions of this interface (personal communication, 2001). Threat history shows changes in threat rating over time. Track number history is useful in situations where the same aircraft is assigned more than one track number. History was not studied in BFA research program, so the details await future investigation. Although the threat assessment algorithm computes threat ratings on a range from 1 (Low) to 7 (High), threat ratings are described only with High, Medium, and Low verbal labels (Note, however, that the Threat History display has 7 horizontal lines to conform to the full range of ratings).

Figure 2b shows the expanded threat window. All of the evidence that goes into computing threat rating is shown in the evidence portion of the threat window. Evidence is divided into a feature list and a data box. Features are listed, in order of decreasing importance, along the left side of the data box. The value of the six highest weighted features is highlighted by reducing the font brightness of the other features. The data box is divided into three panels. There is a brightness difference for the box background and text for the high weight features. The middle panel contains the raw data associated with each feature. The impact of a given piece of evidence on threat level is shown in the left and right panels of the threat box. Bars on the left show increases to threat rating. Bars on the right show decreases to threat rating. The bars show the degree of change or impact; longer bars mean more impact. For example, while the Altitude and ESM of Track 7023 in Figure 2b increase its threat rating, its Altitude has a larger effect on the increase.



(a) (b)

Figure 2. Threat assessment window. 2a. Only essential data are visible. 2b. All data are visible.

4.4 Priority list

A proposed Priority List Window is shown in Figure 3. Tracks are listed in order, from most to least threatening, as computed by the threat assessment algorithm. In current operations, Contacts of Interest (COI) and Critical Contacts of Interest (CCOI) are maintained manually by the air defense team, either on a status board or on paper. The contents and layout of the priority window approximates the manual format.

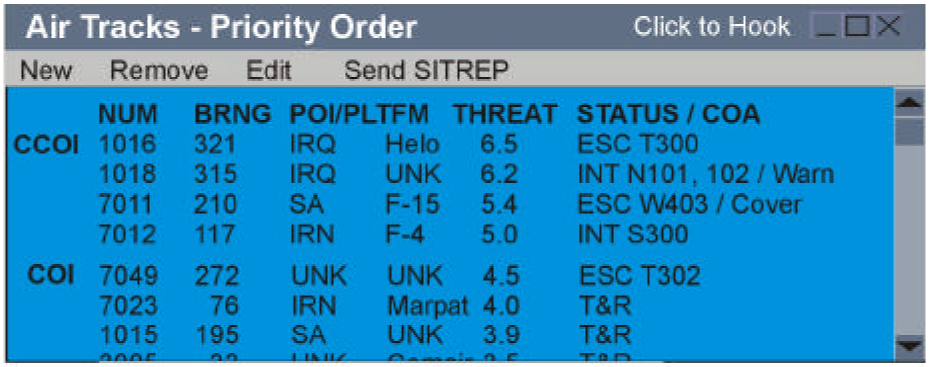


Figure 3. Track priority window.

5.0 Discussion

This paper reviewed the key findings from our recent studies of U.S. Navy air threat assessments, described a threat assessment model that was created from the research data, presented proposed guidelines for displaying air threat assessment information, and presented a prototype display implementation based on our studies. The key findings from our recent studies provide new contributions to the study of threat assessment specifically, and situation assessment/awareness in general. Those findings include:

- User-created templates define which cues will be evaluated and the permissible range of data for each cue.
- Cues were:
    - Evaluated in a fairly consistent order;
    - Weighted differentially;
    - Processed in sets or chunks reflecting their weights and information value.
- Air defense threat evaluators:
    - Did not rely on all data (only data associated with the cues in their active template);
    - Did not change templates in the face of conflicting data;
    - Were influenced by conflicting data in specific cues rather than the overall pattern of data.
- Perceived threat level:
    - Was related to the degree of fit of observed data to expected data ranges in the evaluator's active template;
    - Was not related to the number of cues that were evaluated during threat assessment.

The findings were used to build a threat assessment model and algorithm. The model closely followed the cognitive processes employed by air defense personnel. In general, the algorithm appeared to compute threat ratings in a manner that was within the range of variability of human experts, and its performance was congruent with expectations from previous research.

Finally, this paper discussed the development of guidelines for displaying threat information to decision makers within the CIC. The proposed guidelines and a prototype threat assessment display conform to the expectations of tactical decision makers, a critical feature of effective decision support tools. It displays the data that they need, in the order in which they use it, thereby contributing to their rapid assimilation of the information. The proposed interface (or information presentation) also allows users to weigh the evidence. However, in addition to simply showing supporting and counter evidence, indicator bars show the strength of that evidence. All of these features will help system users avoid common decision-making biases, and reduce the likelihood of misses and false alarm errors.

This paper, building on other recent studies, summarizes new insights into the threat assessment process and the means for displaying useful information to air defense decision makers. However, additional research is needed to confirm these findings. Our understanding of the user's templates needs further refinement, including basic research on their formation, content, and activation. The sequential processing of cues also warrants additional investigation. The observed ordered processing may be due to task demands, but data from Liebhaber, Kobus, and Smith [2000] clearly indicate that cues are weighted. Processing of cues in sets, or chunks may reflect working memory limitations or latent associations. The mechanisms for adjusting template fit during continued processing of cues are also unclear. Indications of these activities were observed, but further investigation is needed to specify their nature and the influences on the adjustment processes. Finally, the proposed display guidelines only address the content of a threat assessment interface. A human-system interface usability study would be needed to evaluate and optimize the exact look and feel of the interface.

6.0 References

[Adelman, 1996] Adelman, L., Bresnick, T., Black, P.K., Marvin, F.F., & Sak, S.G. (1996). Research with Patriot air defense officers: Examining information order effects. *Human Factors, 38*, 250-261.

[Anderson, 1998] Anderson, J.L. (1998). Embracing uncertainty: The interface of Bayesian statistics and cognitive psychology. Conservation Ecology, 2. [Online] http//www.consecol.org/vol2/iss1/art2.

[Chalmers, 1998] Chalmers, B.A. (1998).On the Design of Computer-Based C2 Decision Support for a Modern Frigate. In *Proceedings of the 4th International Command and Control Research and Technology Symposium*. Nasby Park, Sweden: U.S. Department of Defense, C4ISR Cooperative Research Program.

[Endsley, 1995] Endsley, M.R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

[Gigerenzer, 1995] Gigerenzer, G. & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684-704.

[Hutchins, 1996] Hutchins, S.G. (1996). *Principle for intelligent decision aiding* (Tech. Rep 1718). San Diego, CA: Space and Naval Warfare Systems Center.

[Kaempf, 1992] Kaempf, G.L., Wolf, S., Thordsen, M.L., & Klein, G. (1992). Decision making in the AEGIS combat information center (Tech. Rep.). Fairborn, OH: Klein Associates, Inc.

[Kelly, 1996] Kelly, R. T., Hutchins, S. G. & Morrison, J. G. (1996). Decision processes and team communications with a decision support system. In *Proceedings of the Second International Symposium on Command and Control Research and Technology*. Monterey, CA: National Defense University.

[Kirschenbaum, 1994] Kirschenbaum, S.S., & Arruda, J.E. (1994). Effects of Graphic and Verbal Probability Information on Command Decision Making. *Human Factors, 36*, 406-418.

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

[Liebhaber, 1999] Liebhaber, M.J. & Smith, C.A.P. (1999). Naval Air Threat Assessment. [CD-ROM]. In *2000 Command and Control Research and Technology Symposium*. Naval Postgraduate School, Monterey, CA.

[Liebhaber, 2000] Liebhaber, M.J., Kobus, D.A., & Smith, C.A.P. (2000). *Empirical Investigation of Air Threat Assessment* (Tech. Rep.). San Diego, CA: Pacific Science and Engineering Group.

[Liebhaber, 2001] Liebhaber, M.J. (2001). *Description and Evaluation of an Air Defense Threat Assessment Algorithm* (Tech. Rep.). San Diego, CA: Pacific Science and Engineering Group.

[Liebhaber, 2002] Liebhaber, M.J & Feher, B. (2002). Guidelines and Proposed Interface for Displaying U.S. Navy Air Defense Threat Assessment Data (Tech. Rep.). Pacific Science and Engineering Group, 6310 Greenwich Drive, Suite 200, San Diego, CA 92122.

[Lyons, 2000] Lyons, R. (2000, January). *Joint Tactical Terminal (JTT) Operational Analysis Report*. San Diego, CA: Space and Naval Warfare Systems Center.

[Marshall, 1996] Marshall, S.P., Christensen, S.E., & McAllister, J.A. (1996). Cognitive differences in tactical decision making. In *Proceedings of the 1996 Command and Control Research and Technology Symposium* (pp. 122-132). Monterey, CA: Naval Postgraduate School.

[Miller, 1992] Miller, T., Wolf, S., Thordsen, M.L., & Klein, G. (1992). *A decision-centered approach to storyboarding anti-air warfare interfaces* (Tech. Rep.). Fairborn, OH: Klein Associates, Inc.

Morrison, J.G., Kelly, R.T., Moore, R.A. and Hutchins, S.G. (1997). TADMUS Decision Support System document. *1997 IRIS National Symposium on Sensor and Data Fusion*. Lexington, MA: MIT Lincoln Laboratory, 14-17 April.

[Morrison, 2000] Morrison, J.G. (2000). *Tactical Decision Making Under Stress (TADMUS) Command 21: Decision Support for Operational Command Centers (DeSOCC)* [On-line: http://www-tadmus.spawar.navy.mil]. San Diego, CA: SPAWAR Systems Center, San Diego.

[Pennington, 1988] Pennington, N. & Hastie, R. (1988). Explanation-Based Decision Making: Effects of memory Structure on Judgment, *Journal of Experimental Psychology, 14*, 521-533.

[Perrin, 2001] Perrin, B.M., Barnett, B.J., Walrath, L., & Grossman, J.D. (2001). Information order and outcome framing: An assessment of judgement bias in a naturalistic decision-making context. *Human Factors, 43*, 227-238.

[Rummel, 1995] Rummel, B.K. (1995). *Subjective evaluation of human-computer interface options for a tactical decision support system* (Tech. Rep. 1698). San Diego, CA: Naval Command, Command and Ocean Surveillance Center.

[Schulze, 1999] Schulze, K.G., Achille, L.B., Schmift-Nielsen, A., Deubner, B., & Stroup, J. (1999). *A networked Wizard-of–Oz experiment to study the effect of communication on decision-making performance* (Tech. Rpt. NRL/FR/5510-99-9901). Naval Research Laboratory, Navy center for Applied Research in Artificial Intelligence, Information Technology Division, SPAWAR, 4301 Pacific Highway, San Diego, CA 92110-3127.

[Stone, 1997] Stone, E.R., Yates, J.F., & Parker, A.M. (1997). Effects of numerical and graphical displays on professed risk-taking behavior. *Journal of Experimental Psychology: Applied, 3,* 243-256.

[Wallsten , 1993] Wallsten T.S., Budescu, D.V., Zwick, R., & Kemp, S.M. (1993). Preferences and reasons for communicating probabilistic information in verbal or numerical terms. *Bulletin of the Psychonomic Society, 31*, 135-138.

[Zsambok, 1997] Zsambok, C.E. & Klein, G.A., (Eds.). (1997). *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

7.0 Appendix A. Threat Assessment Cues

Cues are listed in alphabetical order.

Attribute	**Description**
Airlane	A published or otherwise known commercial air route.
Altitude	Approximate feet above ground or an indication of change (e.g., climbing).
Coordinated activity	Track is communicating with, or nearby, another track.
Course	Heading - Exact compass heading or indication of heading relative to own ship (i.e., opening or closing).
CPA	Closest Point of Approach - Estimated distance that track will pass by own ship if the track and own ship remain on their current courses.
ESM/Radar	Electronic Support - Electronic emissions from the track (typically indicates the type of radar system the track is using).
Feet Wet/Dry	A *Feet Dry* track is flying over land. A *Feet Wet* track is flying over water.
IFF Mode	Identify Friend or Foe. Signals from a track that indicate if it is a friendly, or perhaps neutral, aircraft.
Maneuvers	Indicates the number of recent maneuvers, or if the track is following the ship.
Number/Composition	Number of aircraft in the formation.
Origin/Location	Indicates the country from which the track most likely originated.
Own Support	Availability of nearby friendly ships or patrol aircraft (CAP)
Range/Distance	The track's distance from own ship.
Speed	Approximate airspeed or an indication of change (e.g., increasing).
Visibility	Approximate number of miles, or an indication of atmospheric conditions (e.g., haze).
Weapon envelope	The track's position with respect to its estimated weapons envelope.
Wings Clean/Dirty	A track without weapons is designated *Wings Clean*. A track with weapons is designated *Wings Dirty*.