What Makes Decision Tasks Difficult?

Susan S. Kirschenbaum[1]
Naval Undersea Warfare Center Division Newport
1176 Howell St.
Code 2211, Building 1171/1
Newport, RI 02841-1708
Tel: (401) 832-3835
kirschenbaumss@csd.npt.nuwc.navy.mil

Abstract

Multi-method investigations of information gathering behavior for decision making in the submarine environment are reported. The two-pronged focus of these studies was classification of task difficulty and investigations of the effects of different difficulty classes on information gathering and decision making. Experimental methods included interviews, questionnaires, computer-assisted process tracing, verbal protocols, and interactive simulations. Results both help assess the strengths and weaknesses of each method and provide support for an information-clustering hypothesis. These results suggest a new approach for the design of complex decision support interfaces.

1. Introduction

A paradox exists in the decision support literature. Decision support for Command and Control (C^2) has been criticized for looking where the light is good, that is, for providing aid for the well understood problems. However, there is little documentation to guide the selection of appropriate candidates for aiding or for identifying the features that make a C^2 problem difficult. The two pronged focus of this paper is the classification of task difficulty (TD) and the investigation of the effects of different difficulty classes on pre-decisional information gathering in C^2 problems. Multi-method investigations of pre-decisional information gathering behavior in the submarine environment are reported.

Task difficulty (TD) interacts with almost every aspect of human performance. It is a major dimension of performance assessment. An easy task makes performance assessment impossible because it fails to discriminate among levels of success [Kirschenbaum, 1986b]. A difficult one may discourage maximum effort because the possibility of success is seen as low [Weiner, et al., 1972]. However, difficulty is a poorly understood and poorly defined construct. Most research that has used task difficulty as a manipulated variable has employed an arbitrary and unverified operational definition. These operational definitions have included information reliability [Suber, 1981], number of components [Wood, 1986], required knowledge (as in an achievement test), and behavioral requirements. The other approach has been to use task difficulty as a measured variable, based on either perceived workload or performance. Such an approach, while validated

[1] This research was supported by Naval Undersea Warfare Center's (NUWC) Independent Research (IR) Program as project #A43312

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **1999**	2. REPORT TYPE	3. DATES COVERED **00-00-1999 to 00-00-1999**

4. TITLE AND SUBTITLE **What Makes Decision Tasks Difficult?**	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) **Naval Undersea Warfare Center,Code 2211,1176 Howell St,Newport,RI,02841-1708**	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
1999 Command and Control Research and Technology Symposium

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**		**9**	

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

on the specific task, does not generalize to other tasks. In order to investigate the effects of situational variables on information gathering, it was necessary to devise some general systematic taxonomy. Classes of task difficulty are the elements of that taxonomy. Thus, one goal of this research was to define the dimensions of task difficulty, and validate them against examples drawn from each category.

As task difficulty is a significant influence on performance, another goal was to determine which situational elements of the submarine task make the problem difficult. To generate sources of difficulty, a series of progressively more structured interviews were held with experienced submariners and submarine instructors. Initially, these were unstructured probes for examples of difficult situations and critical incidents that the individuals had experienced or observed. Subsequently, more structured probes were used to elicit members of each of the hypothesized TD classes. Some of these examples were programmed into a simulation to create data for later experiments.

Initial interviews and behavioral observations led to four hypothesized classes of difficulty: complexity, rule incongruity, anomaly, and situational ambiguity. The complex problem is one with many elements, each of which pose little problem alone, but combined, tax the decision maker's ability. This may be due to memory limitations and/or workload capacity. Complexity is the most commonly cited cause of task difficulty [Wood, 1986]. Rule incongruity occurs either when well-learned, textbook-like rules-of-thumb conflict or when a low probability event occurs that is not covered by such rules. Initially, these frequently look like common events covered by common rules. Rules are designed to cover the most common situations. When a low probability event occurs, all but the best experts appear to stretch the rule to fit, and thus, miss-identify the situation. The ability to recognize that an event does not fit the rule, in some sense, defines the best experts [Arkes, 1990]. Anomalies, including human and machine errors, cause problems because they do not correctly match the operator's "mental model." When unrecognized, anomalies can combine to create true disaster (e.g., Three Mile Island). Lastly, humans are very uncomfortable with uncertainty. Thus, ambiguity in situational information is accommodated by extending the information gathering phase of decision making. After examining the data, a fifth class of difficulty time criticality, was added to account for the effects of situations where the response time is very limited. Table 1 defines all six classes of difficulty.

Table 1: Classes of Task Difficulty

Code	Class name	Description
1.	Control:	Textbook situation
2.	Criticality:	Short available response time, e.g., own ship is target of weapon
3.	Complexity:	Many individual elements, e.g., many targets
4.	Ambiguity:	Signal unclear, e.g., high background noise
5.	Anomaly:	Unexpected, unexplained event, e.g. human error
6.	Rule incongruity:	Cues do not match common rules, e.g., low probability event

2. Study I: Task Difficulty Questionnaires

Several studies were undertaken to test the hypothesized difficulty elements generated by the interviews. Two of these employed structured questionnaires. For economy, these are reported together. One was an individual questionnaire and the other sought group judgments. Unlike the usual questionnaire, a multi-method approach was used to compare judgment with performance.

2.1.Method

2.1.1. Subjects.

Subjects for these studies were instructors in the tactical training department of the U.S. Submarine School at Groton, CT. They were chosen as subjects because they have the opportunity to observe COs and crews during training and to simultaneously know truth about the scenario. This is because they use a sophisticated simulator/trainer that emulates the world outside the submarine attack center, while providing the actual equipment within. The trainer allows instructors to control all aspects of the "world" including ocean conditions and characteristics and actions of other platforms. All the instructors responded to the first questionnaire. A panel of five responded to the second.

2.1.2. Procedure.

The first questionnaire provided simulated data on five scenarios and asked the subjects to indicate their understanding of the situation, responses that they would make, and rate each for difficulty. This was administered in three sessions of three to five subjects in each. After completing the questionnaire, a discussion was held with each group to elaborate on individual responses. In the second questionnaire, the effects of individual variables were investigated and group judgments sought on how accurately and quickly the situation could be assessed.

2.1.3. Variables.

Four classes of situation have been hypothesized to be causes of difficulty. These are complexity, rule incongruity, ambiguity, and anomaly. The scenarios exemplifying these classes were (1) a maneuvering target (complexity), (2) a bow null (rule incongruity), (3) an overlead (also rule incongruity), and (4) a high sea state (ambiguity). Anomaly was simulated by adding a realistic human error to one of the above scenarios. The scenarios were created by selecting a control (easy) scenario from recent at-sea exercises and modifying it appropriately. Realistic data for experimental use were created in the Combat System Evaluation and Effectiveness Laboratory (CSEAL) high fidelity simulation. The second questionnaire separated such environment and geometry dependent variables as range, aspect, target type, tracker, sea state, and sound velocity profile.

2.2.Results

At the end of each leg the subjects rated each scenario for difficulty, drew a line-of-sight diagram (LOS), and indicated their intended action (see Table 2). The means and standard deviations for the difficulty ratings for each scenario can be found in Table 2. The results validate the control scenario. It was judged as easy, with relatively little variation among subjects. Furthermore they were uniformly able to draw the LOS diagram and in somewhat more agreement about the maneuver than with most other scenarios. The high ambiguity (high sea state) situation was judged the most difficult. The high ambiguity (high sea state) situation was judged the most difficult. Interestingly, although one of the two rule incongruity scenarios (overlead) was not judged as very difficult, there was a large variance among the subjects in that judgment. This implies that there may be a large variation among how well various crews perform in this situation. The high ambiguity (high sea state) situation was judged the most difficult.

Table 2. LOS accuracy and maneuver consistency among subjects

	LOS % correct	Maneuver % consistency
Ambiguity (high sea state)	88	43
Complexity (zigging target)	86	50
Rule incongruity (bow null)	38	57
Rule incongruity (overlead)	77	86
Control	100	66

In each of the four difficult scenarios, there were inconsistencies among the three measures. For example, although the high sea state was judged as most difficult, it had the second most accurate LOS diagrams. In contrast, the overlead had a low difficulty rating, but the highest standard deviation. Although LOS accuracy was only 77%, action recommendation consistency was the highest for any scenario, even the control. The bow null, the other rule incongruity scenario, was judged as only moderately difficult, but showed the lowest correct LOS drawings and only chance consistency in action recommendations. The uncoupling of the three measures is somewhat disconcerting. It indicates inconsistencies in the judgement process and argues strongly for a behavioral investigation of these scenarios. The subjects reported observing problems (during attack center exercises) with narrow aspect and opening contact geometries and with a thermal layer environmental situation. Additionally, the situation in which an anomaly occurs was noted as problematic. Scenarios containing these conditions will be considered for further investigation.

Table 3. Mean subjective ratings of difficulty

	Mean*	SD
Ambiguity (high sea state)	4.1	1.07
Complexity (zigging target)	3.2	0.66
Rule incongruity (bow null)	3.0	0.93
Rule incongruity (overlead)	2.7	1.20
Control	2.1	0.78

* 5-point scale

The second questionnaire, that was specifically designed to investigate the effect of sensor, range, and aspect on detectability and localization difficulty, indicated that aspect, target type, and range can affect the difficulty of detecting a target. These results have been used to refine the scenarios used in later phases of this program.

3. Study II: Task Difficulty Paired Comparisons

In the paired comparison methodology, a number of elements are compared along a single dimension. In this case, 15 scenario elements were compared for difficulty. The resulting data provide a mean ranking of the elements, some indication of clustering, and an indication of difference in rankings by subject category.

3.1.Method

3.1.1. Subjects.

The groups of subjects were composed of battle stations parties from two submarines. All levels of experience and all battle stations jobs were represented in the sample.

3.1.2. Variables.

Fifteen possible scenario events were provided. Three examples were drawn from each of five categories of TD: control, complexity, anomaly, and rule incongruity. These are listed in Table 3 according to expected class membership. Relevant subject variables were type of ship (SSN or SSBN), rank of the subject, job he held, and years of experience.

Table 3. A priori difficulty classes and scenario element codes

Control	
O	Narrow band, direct path contact at 15kyd.
C	Broad band contact with bearing rate of 1-2 degrees/minute.
G	Narrow band, bottom bounce contact at bearing 080.
Rule Incongruity	
K	First leg, overlead situation.
E	Sudden, broad band contact with very low bearing rate.
M	Sudden broad band contact with bearing rate of greater than 5 degrees/minute.
Ambiguity	
I	High traffic, narrow band target.
D	High sea state (noisy environment), possible quiet target.
N	Two narrow band targets passing each other, track correlation problem.
Anomaly	
B	Sonar signal appears to split into two.
J	Failure of system solution to match sonar bearings, reason unclear.
F	Loss of signal, reason unknown.
Complexity	
H	More than 10 surface contacts, you want to come to periscope depth.
L	Melee situation, close, maneuvering target.
A	More than 10 sub-surface contacts, all may (or may not) be hostile.

3.1.3. Procedure.

The battle stations party from a single ship was assembled in a conference room and the experiment explained. Subjects were given a three-page response booklet. Page one contained instructions and a list of scenario elements. Page two contained a matrix of all pairs of elements. Subjects were asked to mark the more difficult of each pair. Difficulty was defined in terms of typical approach and attack scenarios. Page three requested demographic information such as rank, job aboard ship, years of experience, and type of ship and equipment.

3.2.Results/Conclusions

A cluster analysis (Figure 2) confirmed most of the a priori class memberships, but element K clustered with the easy, control elements. This may have been because the experimenter identified the condition, rather than just giving symptoms, as she did with the other elements. Although element C did not cluster with the control class, it was included in that group because of its ranking.

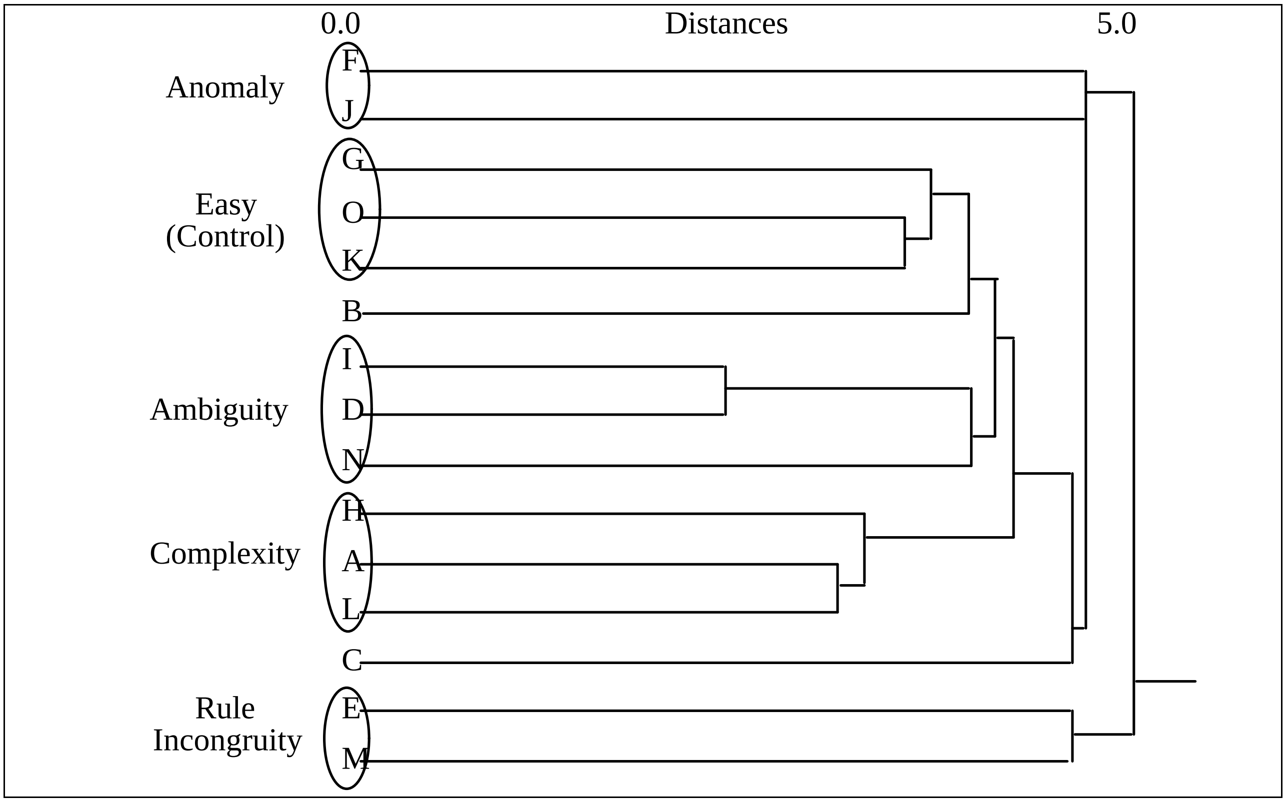

Figure 2. Cluster tree for scenario elements. Codes are found in Table 3.

Mean difficulty rankings can be found in Table 4. It was found that (1) complexity scenario elements were ranked as most difficult by all classes of officers; (2) scenarios containing anomalies (especially F and J) were ranked as more difficult depending on the attack center job; and (3) rule incongruity elements (E and M) were ranked as more difficult by the more experienced officers. Element K was apparently easy and should have been included in the control group. Control element C was seen as more difficult by senior officers, although it could not be grouped with any other element type. To further investigate the relationship between subject characteristics and rankings, the results were partitioned by job, rank, and years of experience. The variation of ranking on anomaly events was accounted for by job type. In the case of rule incongruity, variation was due to experience level of the respondents. Interestingly, the more experienced respondents saw these events as causing more difficulty. Note that a similar surprising finding was reported by Arkes [1990] in an investigation of hindsight bias among physicians.

Table 4: Mean rankings for element difficulty

Element type and code	Mean ranking*	Standard Deviation
Control		
K	2.2	2.1
O	2.8	2.0
C	3.1	3.7
G	3.8	2.4
Rule Incongruity		
E	6.0	4.1
M	7.7	4.3
Ambiguity		
I	7.0	2.2
D	7.2	2.7
N	7.9	2.3
Anomaly		
B	6.2	2.7
J	8.7	3.8
F	8.8	3.8
Complexity		
H	10.4	2.6
L	10.9	2.4
A	11.8	2.1

*Rankings range from 1, easiest, to 15, most difficult.

4. Discussion and conclusions

Although time does not permit discussion here, subsequent research supported the task difficulty classifications described above and investigated the effects of each on decision making behavior [Kirschenbaum, 1990]. Results confirmed that information gathering behavior was related to difficulty classes as follows. Complexity caused a narrowing focus as the decision maker shed tasks. Rule incongruity resulted in collection of the wrong data by the less experienced subjects. Time criticality shortened data collection time and led to more decision errors. Ambiguity resulted in to more time spent gathering information. The effects of anomaly errors are difficult to characterize, except that they were disruptive if undetected.

Dedicated support for command and control must focus on those areas where the problem is most difficult. Research has shown that pre-decisional situation assessment is made more difficult by specific classes of events; complexity, ambiguity, rule incongruity, time criticality, and anomaly, and that these classes of difficulty lead to different kinds of information gathering behavior and decision errors. The challenge now is to provide support that alleviates these errors. The strengths of powerful digital information systems must be matched by a strong, new decision support concept, grounded in as much knowledge of the human as there is now of the algorithms and technology.

5. References

[Arkes, H., August, 1990] *Very Practical Judgmental Decision Making.* Invited address, 98th Annual Convention of the American Psychological Association at Boston, MA.

[Kirschenbaum, S. S., 1990] *Command Decision Making: Lessons Learned*. (NUSC TM No. 902149). Newport, RI: Naval Underwater Systems Center.

[Kirschenbaum, S. S., 1986b] *Acquisition of Self-efficacy Expectations as a Function of Task Difficulty, Experience, and Social Comparison Information* (Doctoral dissertation, University of Rhode Island, 1985). *Dissertation Abstracts International, 46,* 4064B.

[Sheridan, T. & Ferrell, W. R., 1974] *Man-machine systems: information, control, and decision models of human performance*. Cambridge, MA: MIT Press.

[Suber, C. F., 1981] Effects of information reliability in predicting task performance using ability and effort. *Journal of Personality and Social Psychology, 40,* 977-989.

[Weiner, B., Heckhaussen, H., Meyer, W., and Cook, R., 1972] Causal ascriptions and achievement behavior: A conceptual analysis of effort and reanalysis of locus of control. *Journal of Personality and Social Psychology, 21,* 239-248.

[Wood, R. E., 1986] Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37,* 60-82.