

CHU-CHUN TANG

Variable Display Distributed Team Simulation: A Research Platform to Study Complex Team Decision Making

(Under the Direction of DR. ROBERT P. MAHAN)

The purpose of this study is to explore how display design and task complexity influences human decision-making in the context of team decision-making. The decision task is divided into three categories: adding, multiplying, and multiplying first then adding. The display has two formats: graphic and numeric/text. In the light of Wickens' proximity compatibility principle and Brunswik's lens model, this study tries to prove (1) that display properties are an essential link between task system and human cognition, and (2) that the size of display effects is moderated by the task complexity. The number of the participants is 251. Results confirm the hypotheses of this study. In addition, display and task complexity interacted with each other and caused different results in the staff and the team leader. Future research on display engineering is discussed.

INDEX WORDS: Display Design, Task complexity, Team composition, Decision making, Brunkswik, Lens model, Proximity compatibility principle, Cognitive continuum theory

VARIABLE DISPLAY DISTRIBUTED TEAM SIMULATION:  
A RESEARCH PLATFORM TO STUDY COMPLEX TEAM DECISION MAKING

by

CHU-CHUN TANG

B.S., The National Chengchi University, Taiwan, 1986

M.S., The University of Georgia, 1996

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial  
Fulfillment of the Requirement for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2001

© 2001

Chu-chun Tang

All Rights Reserved

VARIABLE DISPLAY DISTRIBUTED TEAM SIMULATION:  
A RESEARCH PLATFORM TO STUDY COMPLEX TEAM DECISION MAKING

by

CHU-CHUN TANG

Approved:

Major Professor: Robert Mahan

Committee: James Brown  
Philip Holmes  
Charlie Lance  
Stu Katz

Electronic Version Approved:

Gordhan L. Patel  
Dean of the Graduate School  
The University of Georgia  
December 2001

## ACKNOWLEDGEMENTS

It is a long way to get a doctor degree, longer than I thought. In this process I got married and finally have a son. I am turning into 40s in one and half years. My parents become older, too. I ask myself and some professors in the department about the meaning of life, and this degree did not answer my question. Although I did not find a life style that satisfies me, I am still grateful to have lots of family members and friends around me. They encouraged me while I was down. Without them I could not have this degree. Thus, I want to thank the following persons: my parents' and my wife Sophia's loving support, Dr. Mahan's advices and proof-reading, Dr. Brown's suggestions to my writings, and brothers' and sisters' prayers in the church. Especially to Sophia, my degree is a heavy load to her. She scarifies herself to take care of most housework and the education of my son. I cannot express enough thanks to her. Although I am still seeking a better way of living, I want to say I got this degree without regrets. Thanks to the Lord.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iv
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW.....	1
2 THE VARIABLE DISPLAY DISTRIBUTED TEAM SIMULATION (VDDTS).....	4
The Methodological Framework.....	4
Various Types of Display.....	5
Task Uncertainty Manipulation.....	5
Network Capability.....	6
Portability.....	6
3 THE THEORY BACKGROUND.....	8
Testing Multilevel Theory: Simulating A Distributed Team Task.....	8
The Display Engineering: The Proximity Compatibility Principle.....	13
A Unified Theory: The Cognitive Continuum.....	15
4 METHOD.....	20
Participants.....	20
Design.....	20
Procedure.....	25
5 RESULTS.....	27
6 DISCUSSION.....	34
Task Complexity.....	34
Display Formats.....	35
Display by Task Complexity Interactions.....	35

Team Composition.....	36
Theory Exploration.....	37
7 CONCLUSION.....	50
Future Work.....	52
REFERENCES.....	54

## CHAPTER 1

### INTRODUCTION AND LITERATURE REVIEW

For many years simulations have been used to improve humans' abilities to solve problems and enhance task performance, especially in the fields of engineering and military training. Along with the improvement of training techniques, powerful computing tools have been developed, that allow simulators to generate very realistic events. For example, pilots can learn to fly in a simulator before they actually fly a plane. Not only does using simulators reduce the cost and improve the safety of the training, but it also allows designers to represent complex scenarios that can never be produced in real environments, such as a nuclear power plant crisis. However, no matter how advanced a simulator is, or how many theoretical and practical advantages simulation may have, there remains one outstanding question: Is a particular simulator useful? To answer this question it is necessary to understand what makes a good simulator.

A simulator requires comprehensive evaluations to make it complete and useful, which includes (but is not limited to) task analysis, learning analysis, and instructional analysis (Flexman & Stark, 1987). The purpose of a task analysis is to provide the relevant task information for the operator to perform in the simulator. This includes information about how different task components are related and how operators interact with the task characteristics.

In the learning analysis the information for supporting learning, performance, and training is analyzed. Since learning is integral to training, a detailed and complete learning analysis facilitates transfer of training. Once both task analysis and learning analysis are conducted the final step is the instructional analysis. The instructional analysis is concerned with delivering information to the trainee so that the required skills

can be most efficiently learned. Modern simulators feature demonstration, coaching and guidance, performance monitoring and evaluation; all of which make learning easier and more efficient.

Although thorough analyses are useful to understand what is required to develop a simulator, as new technologies with higher fidelity are implemented, performance assessment becomes more difficult. Consider air fighter combat as an example: The critical measure is not the number of rounds of ammunition fired but the number of enemy fighters eliminated. However, the most important training issues are often largely based on the number of subtle tactics that require a series of complex decisions that are difficult to define and model. For example, the timing and choreography of an air-to-air target prosecution represents a myriad of decision nodes that are implicitly executed by an expert fighter pilot. The typical simulator, which only has operator's performance as the output, is useful for showing what an operator did but not what kind of decision strategy was used.

Furthermore, high fidelity simulations may hinder the development of the important component strategies during early training because the important task relationships are embedded in a complex environment. Under this circumstance the trainee may not attend to the most relevant information for acquiring the necessary decisional strategies for optimal task performance. Due to the limited mental resources human abilities to attend appropriate channels of information at the appropriate time is limited. For example, Wiener (1977) found that an aircraft pilot disregarded a critical diminishing altimeter reading while attending to a potential cockpit malfunction. Thus, an appropriate level of regulating information is critical to the success of the training (Holding, 1987).

In contrast with high fidelity simulations, simple tools can teach trainees to learn basic principles of task performance. For example, Kelley (1969) found that the learner performed better when a complex skill was decomposed into separate parts of easy subtasks in which critical principles for successful operations were revealed. Unlike high

fidelity simulations the important task relationships are revealed in the less complicated simulations, that helps the trainee attend to the appropriate channel of information and, therefore, may help them master the task. Thus, a simulator does not necessarily have to be high fidelity. It can be simple but contain all relevant information and skills to perform the task.

## CHAPTER 2

### THE VARIABLE DISPLAY DISTRIBUTED TEAM SIMULATION (VDDTS)

#### The Methodological Framework

One approach that may be useful in creating a simulation that allows subtle decision strategies to be tracked is to build the simulation around a decision measurement framework that constrains the simulator in a manner that captures target decision behavior. The methodological framework behind the VDDTS is based on Egon Brunswik's lens model.

From Brunswik's viewpoint, an organism itself cannot be separated from its environment when we try to understand its behaviors. The task information in the environment is presented to "the sensory array of the perceiving organism as proximal cues which are then processed centrally within the organism to yield some functional response" (Cooksey, 1996, p. 2). Because the proximal cues never perfectly represent the task in the environment, cues are only probabilistically related to task criteria. The notion of probabilistic relationship between the task criteria and cues can also be applied to how an organism processes the cue information and generates responses. A successful response requires that the relationship between cues and task criteria and the relationship between cues and response are strong. Brunswik presented this probabilistic relationship in the lens model (Brunswik, 1952).

In general terms, the lens model decomposes judgments through the parsing of correlation between judgments and criteria. By adopting the multiple regression technique, several indexes can be generated to indicate how well the operator's response relates to the ideal response. If an operator utilizes information in a valid way, the regression coefficients generated from operator's responses will be close to the ideal

model and the regression residual will be small. These regression coefficients can be used to interpret the operator's decisional policy and provide an objective standard to quantify decision-making. Therefore, one of the advantages of the proposed simulator is the ability to analyze an operator's decision policy while the operator is engaged in the decision task. This allows for monitoring the reaction of operator's strategy.

### Various Types of Display

In addition to on-line analysis capability, the proposed simulator presents cue information in several ways, varying from graphics to numbers and texts. This allows an investigation of how display format influences the quality of decisions. In addition, the simulator is capable of manipulating the uncertainty of the task by hiding some relevant information or displaying irrelevant information (see Mahan, Kirschenbaum, Jilg & Marino, 1998).

### Task Uncertainty Manipulation

One of the difficulties of processing information in natural environments is that its reliability is unknown. Internal sources of uncertainty include measurement errors, display errors, errors due to our limited perceptual capability, limitations of information processing capacity, and/or memory faults. There are also external sources of uncertainty such that relevant information may not be presented or the presentation of too much information prevents knowing what information is the most crucial to the task. All these factors combine to produce a system state with uncertain qualities. Research has shown that people respond differently in high uncertainty than in low uncertainty situations (e.g., Hammond, 1986). A key feature of the proposed simulator is the ability to manipulate different types of uncertainty. This is accomplished by integrating the ability to hide important cue information, display irrelevant cue information, and/or generate random errors mixed with the real cue information.

### Network Capability

In addition to the above features, the proposed simulator also has a networking function that allows team members to exchange or to discuss information in a distributed setting. After comparing the cost and technologies currently available, ethernet network communication is used as the vehicle for simulating distributed settings. This feature allows team members in different geographic locations to communicate with each other. Thus, the concept of team is no longer bound by a geographic or temporal limitation.

### Portability

Another valuable aspect of the proposed simulator is that it is portable. One of the problems of different computer systems, ( i.e., UNIX, PC, Mac), is that applications developed in one system cannot be used by another system. This means more money and time are needed to develop applications for different systems. After surveying several application development environments, JAVA, which is an internet object programming language developed by Micro System SUN company, is used as the tool to develop the simulator.

The proposed simulation contains three major parts: a setup utility, a server program, and a client program. The setup utility controls how the task is made including the number of roles involved, the interactions between information and roles, the interaction among roles (for team settings), the number of decisions (trials) in each session, and the session profile. The server program loads the configuration files, transfers information among clients, records the responses made by clients, and analyzes data. The client program displays cue information obtained from the server and serves as the medium for communication with other team members.

Although the proposed simulator appears promising, there are limitations. First, due to the nature of ethernet networking, the system performance might suffer when network traffic is heavy. Thus, network delay needs to be taken into account when applying this simulator in a wide area network. Second, although JAVA is supported by

many software developers, it is not mature enough to be a standard programming language. Constant updating makes future maintenance of this simulator potentially difficult. Third, some bugs are expected during the development of the simulator. Thus, a survey is given to the participant to ask his/her problems during experiments. Also, the experimenter tracks how the simulator is running during experiments.

## CHAPTER 3

### THE THEORY BACKGROUND

#### Testing Multilevel Theory: Simulating A Distributed Team Task

Over the past 20 years there have been rapid improvements in technologies that are changing the task demands of typical work environments. The result is that the accomplishment of some tasks cannot be handled by an individual and requires team operations. Teamwork is essential in business, military operations, political policy making, and other working environments.

Team decision-making is different from group decision-making because the former refers to the decision process undertaken by two or more interdependent individuals based on the information obtained from more than one source (Orasanu & Salas, 1993). Within this context team members share common goals and have defined roles and responsibilities. Team members make their unique contributions to the goal based on their task-relevant knowledge. The team leader receives input from the team members and makes the final decision. This working model is common in many organizations such as companies and military settings. In group decision-making the team dynamics are different. Although the members of the group share common goals, expertise is not differentially distributed among team members. Thus, team decision-making has a hierarchical structure, which is quite different from group decision-making.

Teamwork represents a way to improve performance in complex tasks by reducing the amount of information monitored by an individual team member, sharing tasks, and providing more inputs (knowledge or mental models, alternative viewpoints, observations, strategies, feedback etc.) for developing task solutions. Thus, in some

settings team performance should be better than an individual's performance. However, this is not always the case. Consider the decision made by the crew of USS Vincennes when they shot down an Iranian airliner. Although they were equipped with the most advanced electronic devices, teamwork did not prevent this disaster from happening. Therefore, teamwork does not necessarily result in optimal performance.

Research has shown that some factors influencing the quality of team decision-making are workload, stress or pressure (Urban, Weaver, Bowers, & Rhodenizer, 1996; Rogelberg & Rumery, 1996; Hollenbeck, Ilgen, Segó, Hedlunc, Major & Phillips, 1995), task cohesion (Zaccaro, Gualtieri, & Minionis, 1995), team structure (Urban et al.), decision support (Timmermans & Vlek, 1996) and team informity (Zaccaro et al.; Hollenbeck, Ilgen, Tuttle, & Segó, 1995; Gigone & Hastie, 1997; Schittekatte & Van Hiel, 1996; Gruenfeld, Mannix, Williams, & Neale, 1996).

Unlike other approaches to studying team decision-making, which assume that context variables alter team performance, the hierarchical team structure provides a formal analysis to understand how environmental task properties influence decisions by the multiple regression technique. This formal approach supports studying the cognitive response to very specific manipulations of the environment.

Individual decisions: The lens model. From the viewpoint of Brunswik's lens model, learning refers to the acquisition of a policy relating a perceiver to his/her environment based on the information sources (cues) available in the ecological context. Within this view the relationship between cues and the ecological criterion is fallible and uncertain. Through this probabilistic interpretation of the person/ecology relationship, the lens model, which was later extended and modified by Tucker (1964) as well as Hammond and his colleagues (Hammond, Stewart, Brehmer, & Steinmann, 1975; Hammond & Wascoe, 1980) provides both a normative and an ideographic method to quantify decision performance.

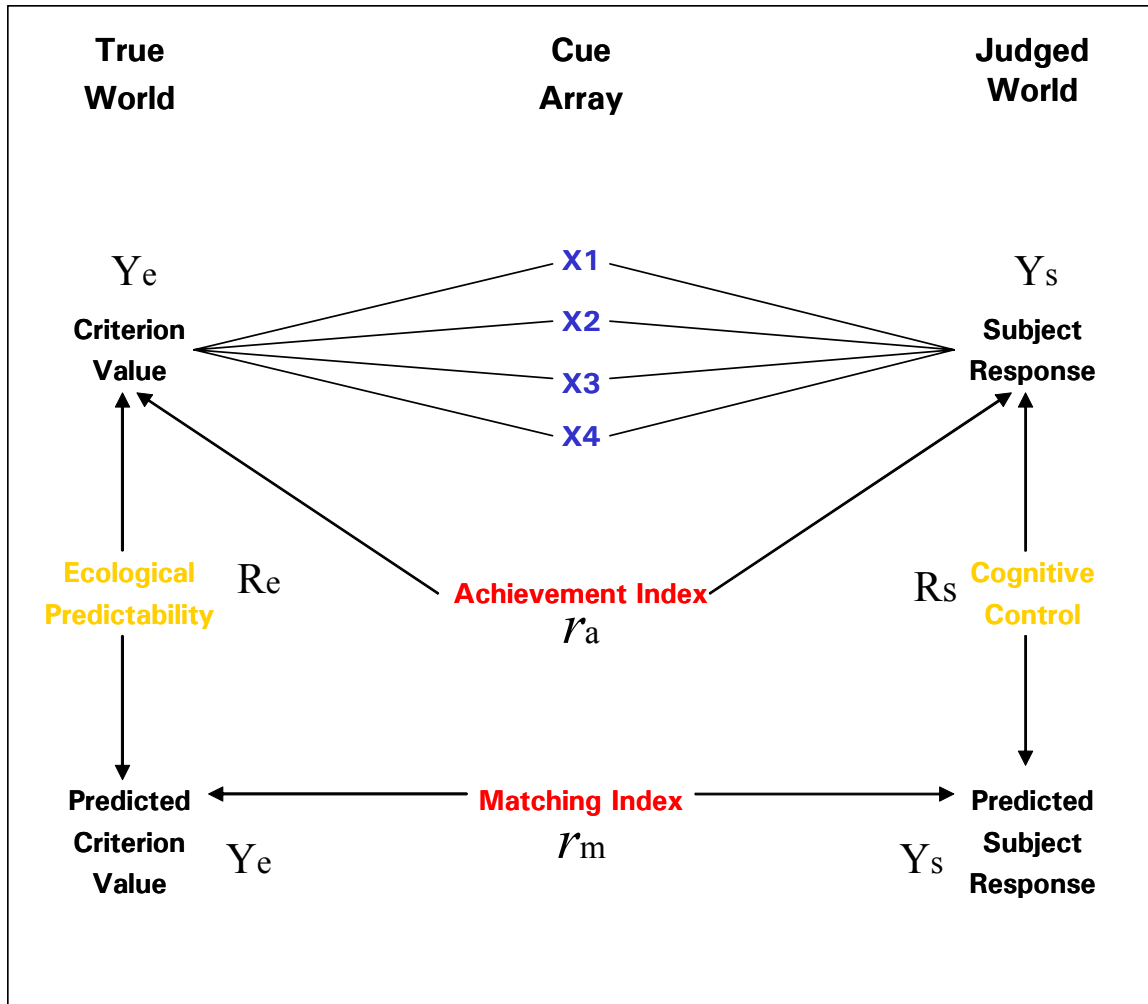


Figure 1. The Brunswik's lens model.

Figure 1 illustrates the relationship between decisions and cues in the ecological context. These cues may not be perfect indicators of the environmental events. Therefore, the correlation between cues and the events is the ecological validity, which is what Hammond later refers to as the ecological predictability. Ecological predictability can be used to index the uncertainty of the events such that the higher the ecological predictability the lower the uncertainty. The nature of learning requires people to utilize these cues to generate decisions. The correlation between cues and the decision defines cue utilization validity, which is called cognitive control by Hammond & Summers (1972). Performance requires both high ecological predictability and cue utilization

validity. A mathematical representation of the lens model for comparing judgment and task ecology is also shown in Figure 1.

First consider the right side of the lens model. Once observer's judgments ( $Y_s$ ) are made, a multiple regression can be used to capture how the subject utilizes the cues and generate a predicted judgment ( $\hat{Y}_s$ ).  $R_s$ , the cognitive control index, is the multiple correlation for the capturing policy between  $Y_s$  and  $\hat{Y}_s$ , which represents subject's control ability to execute his or her own judgment policy. The difference between  $Y_s$  and  $\hat{Y}_s$  is the error due to unmodeled or unsystematic random errors. In the same way, we can regress criterion ( $Y_e$ ) on cues and obtained a predicted criterion value ( $\hat{Y}_e$ ). The multiple correlation for the ecological system policy between  $Y_e$  and  $\hat{Y}_e$  is the ecological predictability  $R_e$ , which represents the degree that a set of cues predicts an event in the given ecology model. The lens model allows an analysis of how people utilize cues and how well the cues represent environment events by examining the  $R_s$  and the  $R_e$ , but also permits a comparison of the correlation between  $Y_e$  and  $Y_s$  (the achievement index  $r_a$ , which represents the agreement between the subject judgment and the criterion value),  $\hat{Y}_e$  and  $\hat{Y}_s$  (the matching index  $r_m$ , which stands for the extend to which the weights and function form of the ecology and subject judgment models agree).

Tucker (1964) proposed a modified lens model equation (LME), which is the standard form of the LME in decision-making research. The LME is:

$$r_a = GR_eR_s + C\sqrt{(1 - R_e^2)(1 - R_s^2)} \quad (1)$$

where  $G$  is the matching index, and  $C$  is the correlation between  $Y_e - \hat{Y}_e$  and  $Y_s - \hat{Y}_s$ . In Equation 1, the first term  $GR_eR_s$  is the linear component based on the results of the multiple regression, indicating how cues are related to ecological predictability and judgments and how predicted judgments and predicted criterion value are related. The second term  $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$  is the component of residual error, indicating unmodeled variance not explained by the linear regression. Theoretically, the residual variance should be random. But if the  $C$  coefficient is high, it may indicate that there is a systematic variation between two sets of residual error. A high  $C$  value can be due to (1) not including important cues, (2) a nonlinear relationship between cues and both sides of the lens model, and/or (3) chance. On the other hand, a low  $C$  value is rather ambiguous. It can happen when the residual of both sides is large but unrelated (Cooksey, 1996). Thus, when interpreting the coefficient  $C$  the experimenter should take the residual error of both sides into account.

The lens model provides a quantitative way to evaluate how people make judgments or decisions through four different kinds of indexes: cognitive control, ecological predictability, the achievement index, and the matching index. Furthermore, it also provides a way to analyze how people utilize cues. Through the coefficients of multiple regression, researchers can determine how important a specific cue is to an observer's judgment strategy. The lens model indices can be incorporated into the proposed simulator, which, unlike most other simulators, has the ability to model an observer's judgment policy in real time.

Multilevel decision-making model. Hollenbeck et al. (1995) proposed a hierarchical team decision-making model based on the lens model. In their model, team decision-making accuracy is determined by constructs that occur at one of four levels: the team level, the dyadic level, the individual level, and the decision level. At the team level the basic constructs are team informity, staff validity, and hierarchical sensitivity. Team informity refers to the number of relevant cues known by team members. The higher the

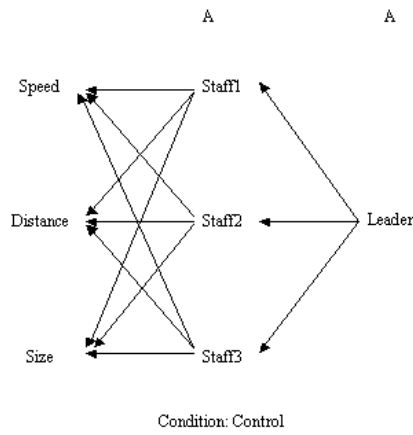
team informity the more the team knows the decision components. When other variables keep constant "team decision-making accuracy is higher for well-informed teams than for poorly informed teams" (Hammond et al., 1995). Staff validity refers to the quality of judgments made by team members in their areas of expertise. If team members can render correct judgments in their respective areas, then the staff validity will be high. However, high staff validity does not necessarily lead to a high level of team decision accuracy. Performance still depends on how well the team leader utilizes the information rendered by the team members. A leader with a high hierarchical sensitivity correctly weighs information provided by team members. When other variables are kept constant, a team with a leader who has high hierarchical sensitivity has a higher decisional accuracy. For example, condition E11 in Figure 3 shows that each staff member received three pieces of information to judge, then they made recommendations to the team leader.

In addition to the constructs occurring at the team level, Hollenbeck et al. (1995) cites three constructs below the team level: decision informity, individual validity, and dyadic sensitivity, which correspond to team informity, staff validity, and hierarchical sensitivity at the team level, respectively. Decision informity refers to the amount of information on a specific decision component; individual validity refers to how well any staff member's judgment predicts the team judgment; and dyadic sensitivity refers to the degree of similarity between how a team leader uses the information rendered by a specific staff member and the actual importance of that piece of information to the team goal. (For a review of his theory, please see Hollenbeck et al., 1995).

#### The Display Engineering: The Proximity Compatibility Principle

The manner in which information is represented is important for team performance. For example, over the past 10 years many studies have shown that analog displays are superior to verbal or digital displays in a visual monitoring task (Payne, Lang, & Blackwell, 1995); that organized displays can help people to acquire information (Endsley & Kiris, 1994; Schkade & Kleinmuntz, 1994); and that object-like displays can

be used to support information integration by exploiting the visual properties of the object. In contrast, when component values must be examined independently, numeric



A: Add  
 M: Multiply  
 M+A: Multiply first, then add

Leader		
Staff	A	M+A
M	E11	E12
M+A	E21	E22

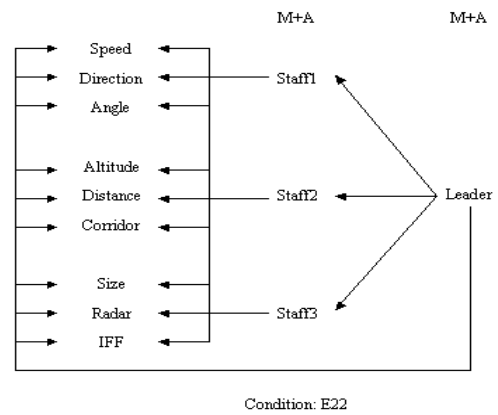
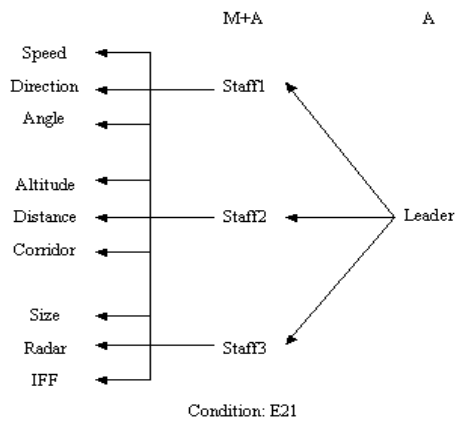
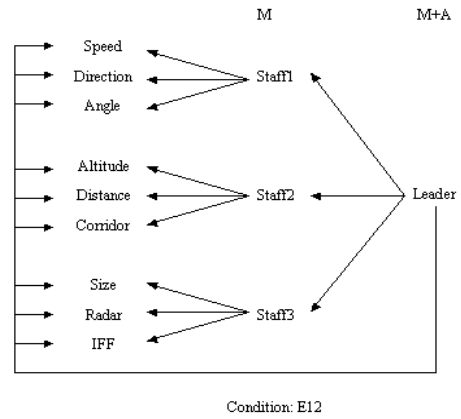
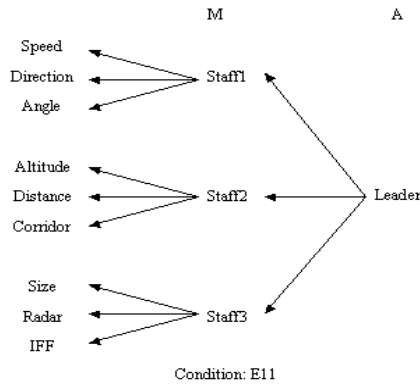


Figure 3. Experimental design.

displays are better (Bennett & Flach, 1992; Carswell, 1992; Elvers & Dolan, 1995; Spence & Parr, 1991; Wickens, 1986). All these studies suggest that task performance is improved when the display format matches the task demands (Coury & Pietras, 1989; Wickens & Carswell, 1995).

Wickens and his colleagues (Wickens, 1987; Wickens & Carswell, 1995) have developed the proximity compatibility principle (PCP) to relate display characteristics to task demands. According to the PCP, human performance is improved when the processing proximity matches the perceptual proximity. The perceptual proximity refers to the degree that information sources are perceptually similar in color, space, and/or code homogeneity. The processing proximity refers to the task demands. If the information has to be integrated in a holistic decision, then the processing proximity is high. In contrast, if information sources must be considered independently, then the processing proximity is low.

#### A Unified Theory: The Cognitive Continuum

The objective of the present study is to leverage and extend the cognitive continuum theory as the framework that predicts team performance. Evidence is strong that information integration tasks are best supported by graphic object displays (e.g., Wickens, 1987). However, it remains unclear how task complexity affects the display task mapping. That is, it is not well understood how task complexity alters the efficacy of object displays. Further, there is some evidence to suggest that as tasks become less complex the display advantage of object representations disappear (Mahan et. al, 1998).

The focus of the study was to test the notion that as an integration task becomes easier to perform it also becomes more analytical in nature. That is, in a simple integration task, because only a very few pieces of information are needed for a judgment, one can attend to each piece of information separately. This is to say, that low a complexity integration task can be analyzed. In contrast, a very complex integration task cannot be easily decomposed into separate information sources. Performance is

typically optimized by executing a holistic or intuitive cognitive process whose mechanism is parallel in process when combining information.

The cognitive continuum Theory (CCT) views human cognition as a continuum that is anchored by intuition and analysis. In the same way, tasks also range on the intuitive-analytical continuum and task properties determine the location on this continuum. Maximum task performance is obtained when the task and human cognition are matched. Hammond (1986) grouped task properties into three categories: complexity of task structure, ambiguity of task content, and form of task presentation. This study manipulates the number of cues and cue function forms as to the task complexity while varying the form of task presentation. The hope is to understand how team members differ in decision performance under different levels of task complexity on graphic and numeric task presentations (display formats).

This study is relevant for several reasons. First, although different displays may influence human performance, how this influence is generated is still not clear. Is there any kind of mediator needed in order for different display formats to be influential? How do other task properties or individual differences interact with the display format? In addition to the unknown underlying mechanism, the results are not congruent on the findings of display studies. For example, consider displays for problem solving tasks. Schaubroeck & Muralidhar (1991) reviewed several studies and found that the graphic display is superior to the tabular in some studies and inferior in other studies. Since there are discrepancies among the findings, the current study provides another opportunity to reexamine the display effects under different task demands.

Second, this study links the proximity compatibility principle (PCP) with the cognitive continuum theory (CCT) (See Hammond, 1990b for details). The CCT asserts that task properties induce a particular cognitive mode. On the other hand, according to the PCP, perceptual properties influence the way people process information. But the PCP does not mention how the perceptual properties and the task complexity impact human cognition. Clearly it is only possible to perceive and process the task when it is

presented in a medium. A display is a medium through which a task influences our cognition. By combining the PCP with the CCT together, the compatibility between task and display mode becomes important, as does the compatibility between the display and cognition mode. Also, the task complexity allows the experimenter to leverage the congruence between task and cognition along the continuum. Thus, human performance on a task can be maximized when there is congruence between the task, the display, and the mode of cognition.

Third, this study investigates how team dynamics are altered by different displays. Most studies investigating display effects on human performance are focused at the individual level. A research question that is currently receiving a lot of attention is whether effects identified at the individual level apply at the team level. As mentioned before, team work reduces the amount of information processed by each individual but requires more communication between team members. Does the leader require the same display configuration as staff members?

Three variables are manipulated in the present study: task complexity (high and low), type of display (graphic and numeric) and role playing (staff and the leader). The task requires the participants to integrate information and to make judgments for which a true value exists. Task complexity refers to the difficulty of integrating information to form a decision. In terms of the lens model, task complexity is determined by the number of cues, the number of interactions among the cues, and the number of cues involved in the interactions. Due to the limitation of memory capacity, people perform poorly on integrating information analytically (i.e., to follow a particular rule) when the number of the cues increases (Hammond, 1986). Especially when there are interactions between cues, this further taxes working memory and inhibits the ability to integrate the information. Thus, when task complexity is high (i.e., there are many cues and/or interactions involved in the decision-making), people respond intuitively to the task. More interestingly, the robust nature of linear models allows people to approach the ideal answer while failing to use the computationally defined algorithm. However, the errors

generated under this operation mode are normally distributed. This means that intuition generates relatively less precise answers than analysis but that the error is small and normally distributed.

Two types of display formats are used in the present study: graphic and numeric. In the graphic display, information sources are represented as objects. In contrast, in the numeric display, information sources are presented as a list of numbers. As mentioned before, object-like displays help people integrate information. Furthermore, if a display is a mediator between task and cognition, the display effect will be moderated by the task properties (i.e., the effect of display will become more salient when performing a complex task than a simple task). This display effect has its limitation and cannot be applied to every type of task environment. For example, if a task is very simple, then an operator expends little mental effort. Thus, a display cannot be helpful since the task is easy. In contrast, a very complex task requires a lot of mental resources and effort. If there are not enough mental resources for the operator to parse information out of the display or to maintain the currently acquired information, then the operator cannot benefit from the display and task performance will be degraded.

When considering role playing within different task complexities and display settings, another interesting point emerges. In some cases, the tasks for the staff and the leader are different. Since the effect of a display may be moderated by the task properties, there may be the cases in which the staff do not benefit from a particular display but the team leader does or vice versa. Thus, research that examines the interactions between the display, the individual, and the team at different levels of task complexity is required.

The hypotheses of this study are as follows.

Hypothesis 1a: Task performance drops as the complexity goes high. This is simply because the easy task takes less effort and mental resources to perform.

Hypothesis 1b: In general, display properties mediate task performance.

According to the PCP, when the task requires integration, a high proximity display

benefits the task performance. Thus, the graphic display will be better than the text/numeric display on task performance for all team members.

Hypothesis 1c: Task complexity moderates the display effects, meaning that display effects are small when the task is easy while the effects are enlarged when the task is difficulty. As the task complexity moves from high to low, the induced cognition moves toward the analytical anchor, that makes the graphic display becomes less salient than in the high complexity situation.

Hypothesis 2a: Team performance is maximized when the display format matches the task performed by participants. That is, a graphic representation facilitates the task requiring integration of the information. Again, according to the PCP, the graphic display will have better performance than the text/numeric display on the integration task because the former has the higher proximity than the latter.

Hypothesis 2b: The combination effects of display, task complexity, and role playing influences the hierarchical/diadic sensitivity and the staff/individual validity.

## CHAPTER 4

### METHOD

#### Participants

Participants for this study were recruited from the University of Georgia research pool. By participating in this study, the student fulfilled course requirements. Participants were required to be familiar with computers. All participants were treated in accordance with the "Ethical Principles of Psychologists and Code of Conduct" (American Psychological Association, 1992).

#### Design

A 2 x 2 x 2 mixed experimental design was proposed, in which task complexity (high vs. low) and role playing (staff vs. the leader) were the between factors, and the display format (graphic icon vs. numeric matrix) was the within factor. A control group is used as the basis of comparison with experimental groups.

Synthetic task simulation. The task was adapted from Hollenbeck's (1995) design to represent a Navy air traffic control scenario, which requires operators to simulate the AWACS operation including monitoring aircraft activities shown on the screen, assessing the situation, and making judgments about whether or not the aircraft is threatening. The goal of the task is to protect the aircraft carrier from being attacked by hostile aircrafts. In the present study, a team contains three staff members (weapon director 1, 2, and 3) and a leader (the senior weapon director). A set of rules was given to participants in advance to define the role and how to utilize cues to make a judgment. Possible judgments were to ignore, review, monitor, warn, ready, lock-on, and defend. The judgment of each participant is recorded, as well as response time.

Task complexity manipulation. Task complexity was manipulated by nature of the rule. As defined in the previous section, the task complexity was determined by the number of cues and the relationship among cues (or cue function forms). In the high complexity condition the participant was asked to integrate 9 pieces of cue information. First, the participant was shown three groups of information with 3 pieces of cue information in each group at a time, then, according to the threat level to the aircraft carrier, multiplied cue information within each group, added information across three groups, and, finally, made a judgment. Thus, the task first required multiplication then addition (the M+A task). Table 1 shows the group and its related cue information for each task complexity condition. The ecological decision policy for the high complexity group is as follows.

$$Y = speed * direction * angle + altitude * carrier * distance + size * radar * iff \quad (2)$$

In contrast with 9-cue information in the high complexity task, the low complexity task only integrated 3 pieces of cue information at a time. In this condition the participant only saw a group of information and was asked to multiply the information shown on the screen (the M task). An example is given below.

$$Y = speed * direction * angle \quad (3)$$

In addition to the low complexity tasks, another type of integration task was introduced. In this condition the participant was also asked to add three pieces of information, one from each group, based on the threat posted to the aircraft carrier (the A task). The ecological model for the correct answer was defined as below.

$$Y = speed + distance + size \quad (4)$$

Table 1  
Area of Expertise and Its Related Cues

Expertise & Its Related Cue	Complexity		
	A	M	M+A
<b>Movement</b>			
Speed	*	*	*
Direction		*	*
Angle		*	*
<b>Distance</b>			
Altitude		+	*
Distance from the Carrier (Distance)	*	+	*
Distance from the Corridor (Corridor)		+	*
<b>Hostility</b>			
Aircraft Size (Size)	*	@	*
Radar Type (Radar)		@	*
Identification of Friend or For (IFF)		@	*

Note: The participant will see only one category of cue information in the M condition.

In terms of the level of the integration requirement and the induced cognition along the continuum, the M+A task posted a very high demand to the participant and induced the cognition toward the intuitive mode. Contrast to the M+A task, the M task was less complicated, and, therefore, the induced cognition was less intuitive on the

continuum. Unlike the other two tasks, the A task was much easier to perform and cue information could be processed in sequence and separately. Thus, it induced more analytical-like responses.

Display formats. Two types of displays were used in the present study: graphic icon and numeric matrix. Different numbers of cues were presented to the participant according to his/her level of task complexity. For example, in a high complexity tasks, the leader saw 9 pieces of object information plus input from staff's judgments. The examples of graphic and text display are shown in Figure 2.

Client Station: SD

**Properties**

speed altitu size

direct angle distan corrid radar iff

WD1 WD2 WD3

**Air Combat Threat Assessment**

- Introduction
- AWACS Composition
- Your Mission

Introduction

**W**elcome on board!

The AWACS is a reconnaissance plane equipped with advanced electronic devices and radar systems. It is usually sent out to collect information on any military activities on the ground and/or in the air. Because of its excellent detecting abilities, it becomes the eyes of the naval carrier group and other types of military operations. In this study you will become a part of an AWACS team. You are going to show us your abilities to integrate information.

**My Decision**

Ignore  
Review  
**Monitor**  
Warn  
Ready  
Lock-on  
Defend

SEND

**Correct Answer**

N/A

**Hitting Rate**

Hit	Near Miss	Miss	Miss 1	Miss 2

Start M... 3:30 PM

**Client Station: SD**

Properties			
Attribute	Value	Min	Max
speed	574.10956	100.0	800.0
direction	11.861729	0.0	40.0
angle	-13.718691	-15.0	15.0
altitude	19516.799	5000.0	35000.0
distance	53.798683	1.0	200.0
corridor	34.64094	1.0	50.0
size	29.909702	10.0	65.0
radar	7.953434	1.0	9.0
iff	1.4332381	0.2	1.6

Other Decisions	
Player	Decision
WD1	Ready
WD2	Monitor
WD3	Warn

## Air Combat Threat Assessment

- [Introduction](#)
- [AWACS Composition](#)
- [Your Mission](#)

Introduction

**W**elcome on board!

The AWACS is a reconnaissance plane equipped with advanced electronic devices and radar systems. It is usually sent out to collect information on any military activities on the ground and/or in the air. Because of its excellent detecting abilities, it becomes the eyes of the naval carrier group and other types of military operations. In this study you will become a part of an AWACS team. You are going to show us your abilities to integrate information.

Start

**My Decision**

Ignore  
Review  
Monitor  
**Warn**  
Ready  
Lock-on  
Defend

SEND

**Correct Answer**

The Correct answer for entity 20 is Monitor

**Hitting Rate**

Hit	Near ...	Miss	Miss 1	Miss 2	Miss 3	Miss 4
15.0	55.0	25.0	5.0	0.0	0.0	0.0

Figure 2. Examples of the graphic and numeric display.

Experimental design. Figure 3 presents the configuration of experimental design after combining the task complexity with the team configuration (the staff vs. the team leader). The word above staff and the leader indicates how the decision is made: A for adding, M for multiplying, and M+A for multiplying first then adding. One control and four experiment groups were proposed. In the control group both staff and the team leader performed the A task. However, leaders made judgments based on their staff's judgments on the aircraft while the staff based on the cues provided by the experimenter. As to the experimental groups, staff performed either the M task or the M+A task while the leader performed the A task or the M+A task. Thus, the possible combinations for the experimental groups were staff-M/leader-A, staff-M+A/leader-A, staff-M/leader-M+A,

and staff-M+A/leader-M+A. These four groups were denoted as E11, E21, E12 and E22, respectively.

By looking at the difference between the group E11, the group E21 and the control group, the display effects on different types of tasks (i.e., multiplying, adding cue information, or both) for staff will be shown. By examining the difference between the condition E11 and E12 as well as the difference between the condition E21 and E22, the display effects under different task complexity for the leader were shown.

### Procedure

Participants were randomly assigned to the control group or any of the experimental groups. All participants made judgments on both the graphic and numeric display formats. Before the experiment began, each participant, depending on what role they played, received instructions on how to perform the task. Each participant had 40 training trials, 20 trials for each type of display, then 60 experimental trials, 31 trials for graphic display and 29 trials for the numeric display. The order of the display was randomly generated by the computer. There was no time limit on making the response. However, participants were encouraged to make their judgments as quickly and accurately as possible. The judgment and its response time were recorded, and a feedback (hit, near miss, miss, miss 1, and miss 2) was shown on the screen after each trial. In the end of the experiment the achievement index score was calculated. Only cases where the achievement index scores of the both displays were significant at the 0.05 level were taken into analysis. That is, only the data generated by the participant who showed their ability to master the task were taken into final analysis.

Due to the limited number of participants at an experimental session, it was difficult to collect data on the team basis. Thus, the data collection was divided into two sections. First, collecting the data of the staff members, then team leaders. There were 10 teams for A, M, and M+A task conditions. Only participants of whose achievement index score was significant at the 0.05 level were qualified to become a staff member. Once the

staff members were formed, their decision behaviors were replicated by the simulator which presented their decisions to the team leader. Each team got at least two to four qualified team leaders to represent the team decisions.

## CHAPTER 5

### RESULTS

The task demands for each group was adding (A), multiplying (M), or multiplying first then adding (M+A). Eight groups were generated based on the team composition and task configuration. These groups were staff-control, staff-M, staff-M+A, leader-control, leader-E11, leader-E21, leader-E12, and leader-E22 where all control groups performed the A task.

After obtaining the participant's judgments, the matching index, the cognitive control index, and the achievement index were generated. Table 2 shows the descriptive statistics of these indexes across different task complexities and displays.

*Hypothesis 1a: Task performance drops as the complexity goes high.*

When considering only task complexity and ignoring team composition, the main task complexity effect was significant,  $F(2, 248) = 165.35, p < 0.01$ . If team composition is taken into account, the overall task complexity was still significant,  $F(7, 243) = 76.81, p < 0.01$ . The data supported the hypothesis that task complexity influenced participant's performance in the direction that participants in the high complexity condition performed less well than in the low complexity condition.

*Hypothesis 1b: Display properties mediate task performance.*

Table 2

Means and Standard Deviations of the Matching, Cognitive Control and Achievement Indexes for Task Complexity and Type of Display

## (1) Matching Index

Task Complexity		Display				
		N	Graphic		Numeric	
			M	SD	M	SD
Staff	C (A)	30	0.885	0.184	0.856	0.150
	M	30	0.956	0.075	0.869	0.137
	M+A	30	0.689	0.199	0.519	0.296
Leader	C (A)	33	0.989	0.015	0.989	0.020
	E11 (A)	32	0.965	0.068	0.972	0.057
	E21 (A)	32	0.979	0.034	0.973	0.039
	E12 (M+A)	29	0.661	0.197	0.587	0.258
	E22 (M+A)	35	0.795	0.137	0.564	0.249

## (2) Cognitive Control Index

Task Complexity		Display				
		N	Graphic		Numeric	
			M	SD	M	SD
Staff	C (A)	30	0.750	0.179	0.743	0.106
	M	30	0.873	0.048	0.800	0.124
	M+A	30	0.739	0.102	0.773	0.080
Leader	C (A)	33	0.932	0.037	0.927	0.062

E11 (A)	32	0.824	0.137	0.847	0.113
E21 (A)	32	0.883	0.064	0.886	0.133
E12 (M+A)	29	0.764	0.087	0.769	0.110
E22 (M+A)	35	0.777	0.076	0.728	0.113

## (3) Achievement Index

Task Complexity	N	Display			
		Graphic		Numeric	
		M	SD	M	SD
C (A)	30	0.655	0.214	0.615	0.173
Staff M	30	0.795	0.108	0.652	0.146
M+A	30	0.501	0.206	0.372	0.188
C (A)	33	0.903	0.054	0.902	0.072
E11 (A)	32	0.783	0.159	0.808	0.147
Leader E21 (A)	32	0.837	0.093	0.844	0.144
E12 (M+A)	29	0.480	0.157	0.467	0.188
E22 (M+A)	35	0.795	0.137	0.564	0.249

Note: C stands for control, A for adding, M for multiplying, and M+A for multiply first then adding.

On average subjects performing integrative tasks defined by all rules were best supported by the graphic display,  $F(7, 243) = 23.31, p < 0.01$ . As predicted by the PCP, the graphic display is better than the text/numeric display. The same information packaged in a graphic or in a numeric/text display led to the difference on performance. Thus, the display is the mediator between the information and task performance. The task in this study was to integrate information and make a judgment about the threat level

indicated by the information. A high proximity display in this study was the graphic display. The result shows that the graphic display had equal or better task performance than the numeric/text display. Thus, this hypothesis was supported by the data presented in this study.

*Hypothesis 1c: Task complexity moderates the display effects, meaning that display effects are small when the task is easy while the effects are enlarged when the task is difficult.*

Task performance was influenced by the task complexity. That is, the performance dropped as the task became more complex,  $F(7, 243) = 76.81, p < 0.01$ . Furthermore, the display by task complexity interaction effect on the achievement index is significant at 0.01 level,  $F(7, 243) = 5.79$ . Table 3 shows the multiple comparisons of the display effect across different task complexity settings and all comparisons are significant at the 0.05 level. The evidence from matching and achievement indexes strongly supports the hypothesis that task complexity moderates the display effect. The complexity of adding, multiplying, and multiplying first then adding tasks ranges from low to high. Table 5 shows that as the task increases in complexity, the advantage of the graphic display over the numeric/text displays increases.

*Hypothesis 2a: Team performance is maximized when the display format matches the task performed by participants. That is, when the display is congruent with induced cognition by the task, the participant's performance is better. This increases total team performance.*

Table 3

Matching Index: Comparison of the Display by Task Complexity Interaction

	DF	MS	F
Staff-Control (A) vs. Staff-M+A	1	0.298	5.68*
Leader-Control (A) vs. Leader-E22	1	0.906	17.25*
Leader-E11 (A) vs. Leader-E22	1	0.940	17.91*
Leader-E21 (A) vs. Leader-E22	1	0.845	16.09*
Leader-E12 vs. Leader-E22	1	0.388	7.40*
Staff-Control (A) vs. Leader E22	1	0.656	12.49*
Staff-M vs. Leader-E22	1	0.334	6.36*
Staff-M+A vs. Leader-Control (A)	1	0.456	8.69*
Staff-M+A vs. Leader-E11	1	0.483	9.20*
Staff-M+A vs. Leader-E21	1	0.418	7.96*

\*  $p < .05$

For the team leader only, the display effect on the achievement index is significant at the 0.01 level,  $F(1, 156) = 7.77$ , with the graphic display outperforms the text/numeric display. Since the nature of task was to integrate information, and the graphic display had a higher proximity than the numeric/text display, the graphic display was expected to have a better performance than the numeric/text display. The result confirmed this viewpoint. Thus, when the display format is matching the nature of task, individuals as well as the whole team benefit from the display.

*Hypothesis 2b: The combined effects of display, task complexity, and role playing influences the hierarchical/diadic sensitivity and the staff/individual validity.*

According to Hollenbeck's model, three constructs determine decision performance. They are informity, validity, and sensitivity at the team level and at the individual level. Since all participants receive all pieces of information in order to make the correct decision, the team informity is the same for each team. Thus, there is no way to compare team informity across teams in this study. However, at the individual level, the informity (i.e., decision informity) can be tested by comparing Staff-M+A with Leader-E12 and Leader-E22. These three groups performed exactly the same task. The only difference was the leader had three extra pieces of information from their staff. However, this extra information did not generate significant differences between the three groups. This is contrary to Hollenbeck's assertion that higher informity will lead to better performance if other factors are kept constant. This discrepancy with Hollenbeck's assertion can be the fact that the task was so difficult that the participant did not have enough mental resource to utilize all information. However, when the display effect was taken into account, differences existed between Leader-E12 and Leader-E22, Staff-M+A and Leader-E12. With the aid of staff's input the leader's task became easier than the staff's.

For the matching index the MANOVA repeated measures test shows that the display effect is significant at the 0.01 level,  $F(1, 243) = 25.73$  with the graphic display better than the other display, the task complexity effect is significant at 0.01 level,  $F(7, 243) = 72.28$  with the low complexity task better than the high complexity task, the display by task complexity interaction is also significant at 0.01 level,  $F(7, 243) = 4.73$ . In addition to matching and achievement indexes, response time (RT) is also used as an indication of the task performance. Results show that overall graphic display has responded to faster than the numeric/text display,  $F(1, 243) = 12.04$ ,  $p < 0.01$ . The low

complexity task has less RT than the high complexity task,  $F(7, 243) = 16.28, p < 0.01$ . It also shows a display by task complexity interaction effect,  $F(7, 243) = 16.26, p < 0.01$ .

## CHAPTER 6

### DISCUSSION

The main purpose of the proposed study was to investigate how displays influence hierarchical teams on integrating information under different task complexity conditions. Response time and the lens model indexes were examined. While all hypotheses were sustained, the underlying decision making behavior was more complicated than its explicit outcome. As discussed in the previous section, the lens model equation provides a way to examine these underlying processes. The lens model equation defines that the achievement index is the product of the matching index, the cognitive control and the ecological validity. Since the ecological validity was predetermined by the model set by the experimenter and the focus of this study was not on the errors of the human decision behavior, the following discussion focuses on how the remaining three indexes, i.e., the matching index, cognitive control, and the achievement index as well as response time differed in the experimental manipulation.

#### Task Complexity

As reported in the results section, the achievement index dropped when the task became complex. It also showed that the matching index and cognitive control presented the same pattern. To reflect the change of task properties, the simple task also had less response time than the complexity task. The relationship between task complexity, and display effect across team composition is shown in Figures 4, 5, 6, and 7 for the matching index, the cognitive control, the achievement index, and the response time, respectively. All these indexes combined consistently confirmed the hypothesis 1a that task performance drops as the task complexity becomes high.

### Display Formats

The influence of the display format on task performance showed a significant difference in the achievement and matching indexes while no difference was observed for cognitive control. Response times were shorter in the graphic display than in the text/numeric display. These data suggest the display format manipulation had less impact on cognitive control. Since cognitive control shows the ability to reliably execute the participant's mental policy toward the task, they constantly executed the policy regardless of the display formats once they mastered the task. In this study participants were screened based on their performance on the task. Thus, they basically showed their ability to master the task. Thus, their cognitive control index was relatively high with the minimum value of 0.75. Under this circumstance, the display format manipulation did not reach significance. As predicted by the PCP, the high-proximity graphic display performed better than the numeric/text display in the integration task. Thus, the hypothesis 1b that display properties mediate task performance was supported.

### Display by Task Complexity Interactions

The display by task complexity interaction showed a similar pattern to the display effect. That is, significant differences were observed on the display by task complexity interaction in the achievement index, the matching index, and the RT while no significant difference was found for cognitive control. Further, Figure 4 and 6 showed that the difference between graphic and numeric/text display varied across team/complexity settings. In fact, the difference was small when the task was easy while the difference was large when the task was difficult. Thus, the data supported hypothesis 1c that task complexity moderates display effect.

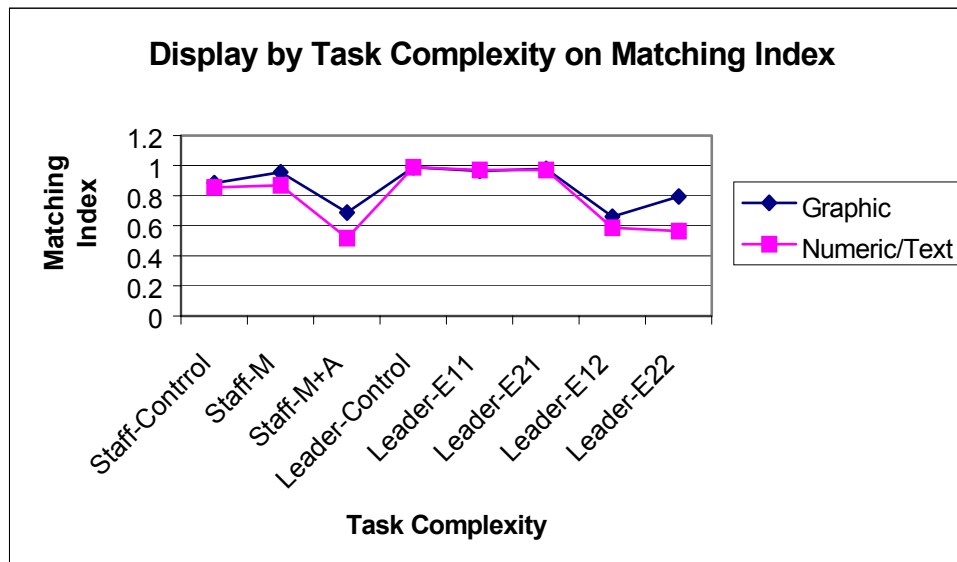


Figure 4. Display by Task Complexity Interaction on Matching Index.

### Team Composition

The differences between staffs and leaders were found in all four measures: the matching index, cognitive control, achievement index and RT. Although staffs and leaders performed the same type of task, leaders always showed higher achievement, matching, and cognitive scores and less RT than staffs. A possible explanation for this could be due to the meaning of the information. Staffs had to integrate raw information. Staffs first converted each piece of information into a meaningful representation in terms of threat level, then added and/or multiplied them to generate a final threat score for the object, then decided a proper action to take. Compared to the staffs, leaders processed the information that had already been digested by staffs. That is, the information carried an explicit meaning with itself such as warning. Since leaders did not spend time and effort to label raw information with a meaning, which could be the most difficult part of the whole process, leaders outperformed staffs in four measures. The detailed comparisons between staffs and leaders are discussed in a later section.

### Theory Exploration

Although data supported the hypotheses of this study at the global level, the micro-dynamics between experiment manipulations revealed more theory-related information. Thus, paired-comparison of main effects and interactions were conducted to try to better understand how task complexity, display format, and team composition affect human decision process.

The Matching Index. Figure 1 clearly shows that the matching index is the correlation between the predicted subject responses and the predicted criterion values. As mentioned in the introduction section, the index shows the level of agreement between the participant's mental policy and the ecological policy set by the experimenter. A high score is obtained when participant's cue weighting and cue relationship are close to the ecological model. In other words, the matching index shows the participant's knowledge about the task.

For staffs, no difference was found between the A and the M tasks. As mentioned in the introduction section, the A task was more analytical in nature than the M task. In terms of task difficulty, the A task was also easier than the M task. If the CCT is correct, one should expect to see a difference between these two tasks. However, this was not the case. One possible explanation could be that the matching index is not sensitive to the cue function form but to the number of cues that need to be integrated. In other words, both A and M tasks required participants to integrate three pieces of information and that did not post enough loading to the participant's information processing system. Thus, the participant still had enough mental resources to complete the task requirement although the M task required more mental operations than the A task.

On the other hand, no significant difference between the M and the A tasks can be the result of the incompatibility between the display format and the induced cognition by the task. That is, the graphic display used in this study implied a multiplication relationship between information which was not congruent with the A task. As asserted

by the PCP, optimal performance is obtained when there is a congruence between display and task properties. The CCT asserts that the optimal performance is obtained when the task properties match the induced cognition. Therefore, the A task is best served with the display that implies addition relationship among cues but, in fact, the A task was shown a graphics that implied multiplication in this study. Data suggested that as shown a graphic display the M task ( $\underline{M} = 0.956$ ) outperformed the A task ( $\underline{M} = 0.885$ ),  $F(1, 58) = 3.83$ ,  $p = 0.055$ . However, the difference became less significant when showing a numeric/text display.

As task complexity increased, the matching index started to drop. This relationship is shown in Figure 4. This same pattern was found when investigating the leader in the different tasks. No significant difference was found between the groups of Leader-Control, Leader-E11, and Leader-E21 since they all performed the A task. However, the difference occurred when comparing the low complexity task with the high complexity task,  $\underline{F}(1, 159) = 347.35$ ,  $p < 0.01$ .

Although the dynamics in the leader groups shared a lot of similarity with the staff groups, one finding that cannot be accounted for was that Leader-E12 had a lower matching score than Leader-E22. Because leaders in both groups performed the exact same M+A task and received the same type of display, no significant difference was expected between the groups. However, the results showed that the difference was significant,  $\underline{F}(1, 62) = 4.13$ ,  $p < 0.05$ . One possible explanation for this result may be due to the fact that the task performed by the staff members was different. That is, while all leaders performed the M+A task, the staff members performed the M task in the group E12 in contrast with the M+A task in the group E22. Given the fact that the leader knew the type of the task performed by his or her staff members, Leader-E22 knew not to rely on their staff's input too much because their staff were asked to integrate complex cue information and their judgments might not be tenable. Thus, the leader-E22 had to spend more time integrating information than the leader-E12,  $\underline{F}(1, 92) = 19.3$ ,  $p < 0.01$ . As a result, the leader-E22 had better idea of the task properties, that led to a higher matching

score for the leader-E22 than the leader-E12. In other words, this difference can be the result of the level of processing information. This is to say, the information is processed more deeply and the memory trace is deeper as well. Thus, the leader-E22 was better in understanding the cue relationship than the leader-E12. However, the evidence showed that the advantage of the level of processing was limited and could not compete with the complexity of the task.

When comparing the staff's and leader's performance, the results showed there were differences for all staff-leader comparisons except Staff-M+A vs. Leader-E12. Although both groups performed the M+A task, which required integrating nine cues, the leader also saw the judgments of their staff members. These extra three pieces of information complicated the original task further and made the task more difficult to perform. As a result, one should expect to see differences between the staff and the leader. However, the data showed the contrary result. Three possible factors might have contributed to this discrepancy. First, it may have taken less mental resources for the leader to integrate the staff's explicit input than for staff to integrate raw data shown on the screen. Second, Leader-E12 may have strategically attended to and put more weight on, the staff's input because they knew that the staff performed the M task, which was easier than their own task. Third, the staff's input may have provided a clue to reveal the cue relationship, altering the original task properties. To test whether the leader was really different from staff in performing the high complexity task, a comparison between Staff-M+A with Leader-E22 and Leader-E12 was performed. The result shows a statistical difference at 0.05 level,  $F(1, 91) = 4.00$ .

For the display effect, as mentioned in the results section, the graphic display had a higher matching index than the numeric/text display. When investigating the display by task complexity interaction, a few interesting points were found. First, the advantage of the graphic display over the numeric/text display increases with increasing task complexity. In Figure 4 both graphic display and numeric display are about the same in the condition Staff-Control (A), then graphic display is slightly better than the numeric

display in Staff- M, and much better in Staff-M+A. At the team leader level the graphic display is about the same as the text display in the condition Leader-Control (A), Leader-E11, and Leader-E21, then the graphic is slightly better than the text in Leader-E12, and much better in Leader-E22.

Second, as to the display by task complexity interaction, Table 3 shows that most comparisons between the A and the M+A tasks were significantly different. Only a few comparisons between the M and the M+A tasks reached a significant level. This evidence suggests that the task complexity moderated the display effect.

Table 4

Matching Index: Comparison of the Display Effect Among Task Complexity After Ignoring Team Composition

	DF	MS	F
A vs. M	1	0.155	2.95
A vs. M+A	1	1.231	23.44 *
M vs. M+A	1	0.0003	0.01

\*  $p < .05$

As reported before, no significant difference was found between the A and the M+A tasks on the main effect of the matching index. Was this finding also observed in the display by task complexity interaction? To test this relationship further, a comparison was made between the A task with the M task and the M+A task. The results confirmed the relationship and are reported in Table 4. However, Table 4 also shows that the difference between the M and the M+A task was much smaller than between A and the M tasks. That is, although the task complexity moderated the display effect, the advantage of graphic display over the numeric/text display was limited and soon reached ceiling. The evidence from the main effect test showed that the A and the M tasks were

close in nature but were very different from the M+A task. If task complexity moderates the display effect in an unlimited way, one should expect to find a big difference between the M and the M+A tasks but this was not found. The display effect increased as task complexity moved from A to M and a small increase from M to M+A. Although neither increase from A to M or M to M+A was significant, the overall display effect increase A to M+A was significant,  $F(1, 219) = 23.44, p < 0.01$ .

Third, although there was no difference between Staff-M+A and Leader-E12 or Leader-E22, Table 3 shows that the difference between Leader-E12 and Leader-E22 was significant. In Figure 4 it clearly shows the display effect had a big increase from Leader-E12 to Leader-E22. As mentioned before, this difference may be due to the leader in the E12 condition adopting their staff's judgment input thereby reducing the task complexity. As discussed, task complexity moderated the display effect. Figure 4 shows that the graphic display contributed to the increase of display effect from E12 to E22 while the performance of the text display was about the same for both groups. This further confirms the notion that optimal performance is obtained when task properties, display properties, and the induced cognition are all matched. The high proximity configuration of the graphic display produced better performance on the highly integrated M+A task in the E22 group than in the E12 because the induced cognition had a better mapping with the task properties for the E22 than for the E12.

Cognitive Control. Cognitive control is the correlation between the participant's true responses and the predicted responses. A high score indicates that the participant is able to reliably execute his or her decision policy. Figure 5 shows the cognitive control over different task complexity settings. Data showed that the overall display effect was not significant. That is, the participant had about the same ability to execute his or her own policy on both types of displays. However, as the task complexity changed, the results showed that the effect was significant at the 0.01 level,  $F(7, 243)=22.15$ . Regardless of task complexity, the smallest value for this index was around 0.75, which indicates the participant's policy toward the task was very stable. As shown in Figure 5

leaders performing the A task had the highest scores. This can be the result of the simple task plus explicit text representation of the display. The next highest score was the staff performing the M task. The rest of groups had about the same score. The reason why the staff-control had a lower score than the staff-low group,  $F(1, 58)=17.63$ ,  $p<0.01$ , can still be due to the incompatibility between the display configuration and the task. Summing up the findings for the cognitive control index, data suggest that once the participant was trained, only task complexity influenced the participant's task performance regardless of display format.

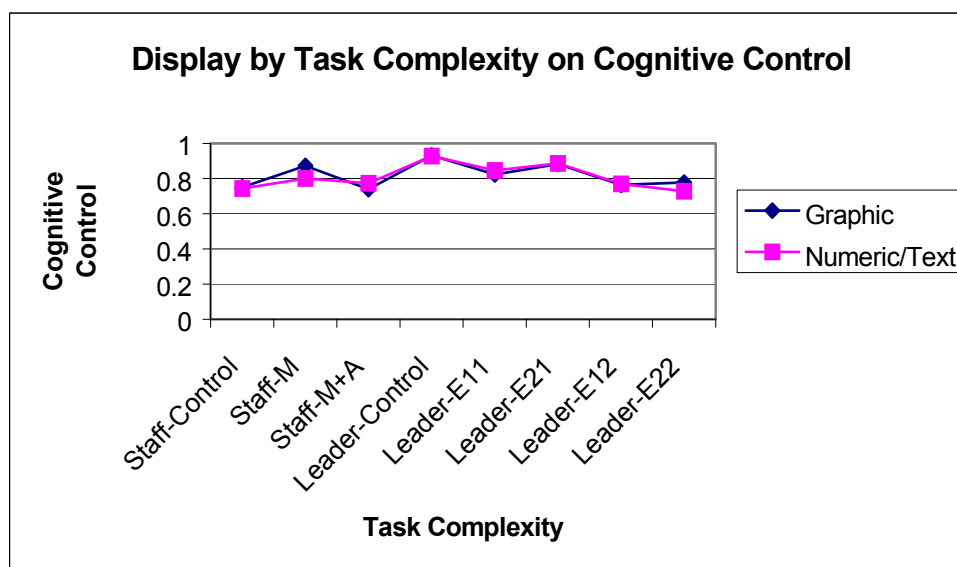


Figure 5. Display by Task Complexity Interaction on Cognitive Control.

The Achievement Index. The achievement index explains the relationship between the subject response and the criterion value. The value indicates the extent of the agreement between the judgments and criterion values which is the product of the matching index and cognitive control, although the participant may not utilize the cue information the same as the rule set by the experimenter. Figure 6 demonstrates how display formats interacted with task complexities on the achievement index.

The result of the achievement index was very similar to the matching index and Figure 6 shared a lot of similarities with Figure 4. The display effect, the task complexity effect, and the display by task complexity interaction were all significant. By examining the task complexity effect alone, the achievement index dropped when the task became complex. It also shows that the leader had a higher score than staff even when performing the same adding task. This difference can be due to the power of explicit judgments from the staff members. Because leaders received staff members' judgments either in the plain

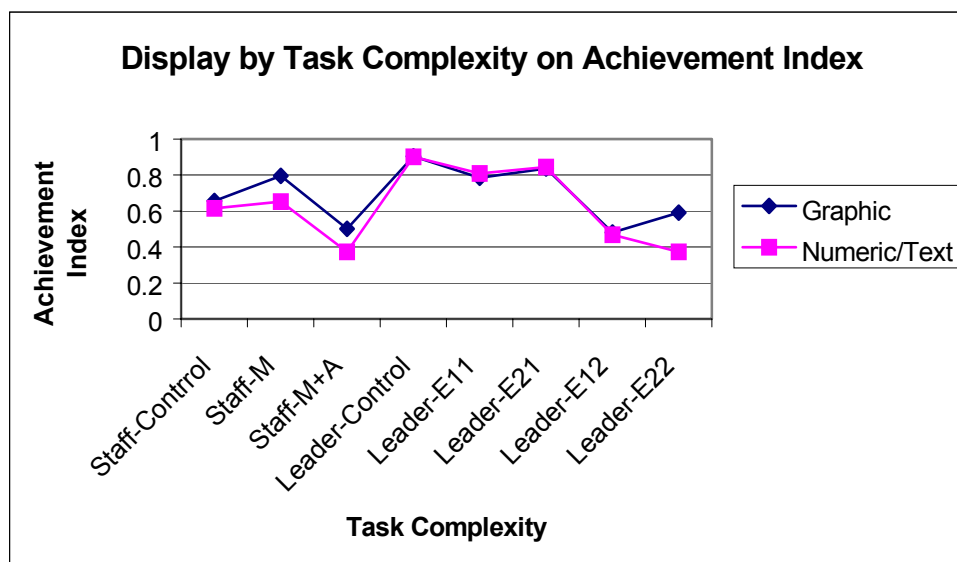


Figure 6. Display by Task Complexity on Achievement Index.

text format such as lock-on or in the graphic format in these groups, it took less mental resources for leaders to convert this meaningful information into a final judgment than for staff members to convert the raw information to a judgment. Thus, leaders had the advantage of these explicit inputs during information integration and this resulted in better performance than the staff.

On the other hand, leaders performed equally well in these groups regardless of display formats. If PCP is correct, one would expect the leader to perform better in the graphic display than the text format, but the results did not support this viewpoint. Two

possible explanations may account for this discrepancy. First, the task was so easy that the participant accomplished it without sacrificing performance. However, the evidence from RT showed that the average RT for the text format was much faster than the graphic display,  $F(1, 94) = 58.54, p < 0.01$ . If different task formats were the same to participants, their RT for both formats should have been close. But this is not what was found in this study. Thus, the first explanation cannot be accepted. Second, the explicit and meaningful text input changed the induced cognition. That is, the explicit and meaningful input was already integrated into the human memory system and thus it took less time to combine the information. Contrary to this, the participant in the graphic format needed to decode graphic information, decide the threat level suggested by the staff then evaluate the total threat and make a final judgment. There are more steps to go through for the graphic display than the text display and the RT data confirmed this explanation. Thus, the advantage of the high over the low proximity display disappeared in this integration task. This result seems to be contrary to what the PCP predicted (Wickens & Carswell, 1995) but it might explain why Schaubroeck and Muralidhar (1991) found the discrepancy of the graphic and tabular effects on different tasks.

One thing worth mentioning here is that a significant difference was found between Staff-Control (A) and Staff-M,  $F(1, 58) = 8.61, p < 0.01$ . Although both groups processed three pieces of cue information, the nature of the task was quite different. In fact, the M task was more difficult than the A task. Thus, it was expected that the Staff-M should get a lower score than the Staff-Control. But the result was opposite. One possible explanation might be due to the fact that the graphic display in the Staff-Control was counter to participant's mind operations. The task for Staff-Control was to add. However, the graphic display presented cue information in area, which implied a multiplication relationship among cues. Due to the incompatibility between the format of the display and the nature of the task, subject performance might suffer. This result was coincident with the findings from the PCP (Wickens, 1987). However, this incompatibility did not

influence the matching index, which suggests that the number of cues determines the level of matching.

Figure 6 also shows the pattern that the display effect increased when the task became complex. To test this, a comparison between all A tasks with M, and M+A tasks was conducted. Results are shown in Table 5. It turned out that all comparisons were significant at the 0.05 level. This finding was not predicted by the PCP but was by the CCT. As the task became complex, the participant's cognition leaned toward intuitive mode. Since people tend to process information perceptually at the intuitive mode of the cognition, the graphic display matched with human cognition to a greater extent and therefore the performance of the graphic display was better than the numeric/text display. This was confirmed in this study.

Table 5

Achievement Index: Comparison of the Display Effect Among Task Complexity After Ignoring Team Composition

	DF	MS	F
A vs. M	1	0.477	10.83 *
A vs. M+A	1	0.753	17.1 *
M vs. M+A	1	0.212	4.82 *

\*  $p < .05$

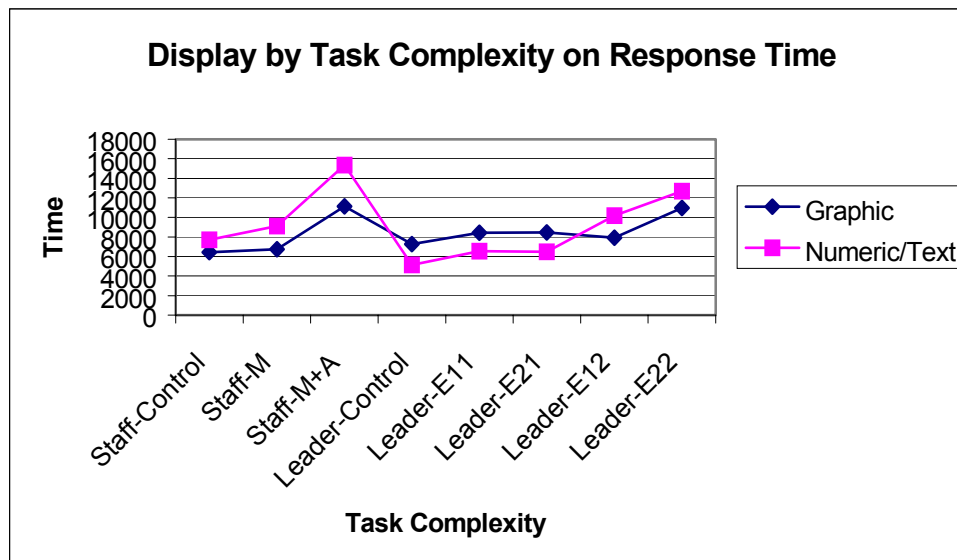


Figure 7. Display by Task Complexity on Response Time.

Response Time. When comparing within the same team composition, i.e., within staff or within the leader, the response time (RT) showed the pattern that displays had their effects on how people responded to stimuli across different tasks. Figure 7 shows that staff responded more quickly on the graphic display than on the numeric display. The finding is supported by the PCP because a high proximity display performed better in the integration task. However, this advantage disappeared when the leader was performing the adding task. In fact, the text display had less RT than the graphic display. As mentioned in the previous section, this result could be due to the fact that the explicit and meaningful information has fewer steps to integrate. Under this circumstance the leader's task is only to take the average of staff's explicit input. In contrast to the text display the graphic display is somewhat ambiguous and uncertain to the participant. Thus, it makes sense that the leader responded more quickly on the text display than on the graphic display. But when the leader performed high complexity tasks (M+A), the pattern was reversed again and was coincident to the findings in the staff that the graphic display was better than the text display.

The overall RT analysis showed the pattern that increasing task complexity increased RT. When considering the display by task complexity interaction other interesting points emerged. For example, compare Staff-Control with Leader-Control, E11, and E21. The participants in these groups performed the A task and the overall RT among these groups was not different from each other. However, the display effect between staff and leader was significant. The result is shown in Table 6. Again, the reason behind this could still be the format of numeric or text display. What staff saw in the text format was the values of three cues, which were real numbers. When performing the assigned task staff needed to convert these numbers into threat levels, then convert the threat into a proper judgment. On the other hand, the leader only saw staff's judgment in the explicit text format. Because it was more difficult for staff to convert numbers into judgment than for the leader to convert staff's plain judgments into his or her own judgment, the difference between the leader and staff on RT could be realized. But when the leader was performing the M+A task, the display effect between staff, who still performed simple adding task, and the leader disappeared. In fact, the interaction was not significant (see Staff-Control vs. Leader-E12 and Leader-E22 in Table 6). The leader in the high complexity group could be distracted by cue information and the graphic display became dominant. Although the leader in the high complexity groups was overwhelmed by cue information, they still took advantage of the staff's input because their RT was less than the staff in the high complexity group,  $F(1, 92)=19.3, p<0.01$ . Without someone's judgment input the cost to integrate complex information is so high that the RT for the staff in the high complexity group dramatically increased. Quite different from the Leader-E12 group, nevertheless, the Leader-E22 group did not rely on staff's input as much as in the Leader-E12 because their RT started to increase. Thus, the display effect started to increase as the function of task complexity.

Table 6

Pair Comparisons of Display by Task Complexity Interaction on Response Time

	DF	F
Staff-Control vs. Staff-M	1	1.62
Staff-Control vs. Staff-M+A	1	11.55 *
Staff-M vs. Staff-M+A	1	4.52 *
Leaders-Control vs. Leader-E11	1	0.11
Leaders-Control vs. Leader-E21	1	0.04
Leaders-Control vs. Leader-E12	1	26.70 *
Leaders-Control vs. Leader-E22	1	22.54 *
Leaders-E11 vs. Leader-E21	1	0.02
Leaders-E11 vs. Leader-E12	1	23.10 *
Leaders-E11 vs. Leader-E22	1	19.11 *
Leaders-E21 vs. Leader-E12	1	24.34 *
Leaders-E21 vs. Leader-E22	1	20.29 *
Leaders-E12 vs. Leader-E22	1	0.42
Staff-Control vs. Leader-Control	1	16.35 *
Staff-Control vs. Leader-E11	1	13.60 *
Staff-Control vs. Leader-E21	1	14.56 *
Staff-Control vs. Leader-E12	1	1.28
Staff-Control vs. Leader-E22	1	0.28
Staff-M vs. Leader-Control	1	28.58 *
Staff-M vs. Leader-E11	1	24.81 *
Staff-M vs. Leader-E21	1	26.10 *
Staff-M vs. Leader-E12	1	0.02
Staff-M vs. Leader-E22	1	0.63

Staff-M+A vs. Leader-Control	1	56.59 *
Staff-M+A vs. Leader-E11	1	51.00 *
Staff-M+A vs. Leader-E21	1	52.85 *
Staff-M+A vs. Leader-E12	1	5.00 *
Staff- M+A vs. Leader-E22	1	8.99 *

\*  $p < .01$

As mentioned before, the leader in the E22 group tended not to trust his or her staff's input, i.e., the leader thought that the validity of the staff's judgment dropped when the task became complex. Thus, hypothesis 2b seems to be partially validated. Unfortunately, a separate test on validity and sensitivity was hard to conduct in this study because the staff validity and the hierarchical sensitivity varied across teams. A further study is called to investigate the relationship between sensitivity and validity.

## CHAPTER 7

### CONCLUSION

The evidence from this study supports that the display mediates the interaction between the task properties and cognition. Different types of displays led to different performance on the task. In this study the participant was asked to integrate information. From the prediction of Wickens' proximity compatibility principle, the graphic display, which is high in proximity, should have a better performance than the numeric/text display on the integration task. The result from this study is consistent with Wickens' principle.

This study also finds that the task complexity moderates the display effect, i.e., the graphic display had a larger advantage over the numeric/text display when the task complexity increased. This is consistent with the assertion of Hammond's cognitive continuum theory (CCT). According to the CCT, the task performance is maximized when task properties match with the cognitive organization principles employed by the participant. Given the fact that other factors are kept constant, the increase in task complexity will push the nature of task to move from analytical zone to the intuitive zone. In this study the task fell into three categories: adding three cues (A), multiplying three cues (M), and multiplying three cues first then adding the result from three subcategories (M+A). The difficulty of the tasks ranging from low to high was A, M, then M+A. From Wickens' viewpoint, a task to integrate cue information is more intuitive than a task to ask the value about a particular cue. As discussed in the previous paragraph, the graphic display was better than the numeric/text display on the current task. In addition, since the cognitive operation induced by the graphic display was more intuitive than by the text display, it is expected that the advantage of the graphic display

over the text display increased as task complexity increased. The results from this study support this assertion. Furthermore, this study also finds that participant's performance dropped when the display did not match with the nature of the task. In one condition, the participant was shown a display that implied multiplication relationship among cues. But in fact the task was to add cue information. The cognitive mode invoked by the display was not congruent with the task demand, as the result, the performance of the participant dropped. Summing up the findings in this study, the maximum performance was reached when task properties, display properties, and the cognitive mode were all congruent with each other.

This study also finds that the staff and leaders responded differently although they all performed the same type of task. On the adding task, the leader outperformed the staff in the achievement and the matching index. Although the overall RT was about the same among these groups, the display effect in the leader groups was opposite to the staff group. This difference was due to the fact that the text display in the leader groups explicitly showed the decision of the staff. Thus, the leader spent less effort to convert these judgments into their final judgment. Compared to the leader group, the staff saw three numbers and needed more effort to convert these numbers into a threat level and then a proper decision. As a result, the leader spent less time on making decisions and yielded a better performance. Again, the difference between the leader and staff is a good example on how information is presented influences our decision quality.

On the high complex (M+A) task, both staff and the leader needed to integrate nine pieces of information. In addition, the leader viewed three extra pieces of information from staff decision input. If the processing time for each cue is the same, one should expect that the RT for the leader should be larger than staff's. However, this was not found. In fact, the leader had about the same time (the E22 group) or less time (the E12 group) than staff. Thus, the staff input changed the nature of the task and made a complex task much easier to handle. Although the overall performance (the achievement index) was about the same, display effects in these groups were different. As suggested in

the previous sections, task complexity moderates the size of the display effect. Because the display effect in the E12 group showed the same pattern as the simple task while the E22 group showed the same pattern as the high complexity in the staff group, it suggests that the induced cognition for E12 and E22 groups was not the same. One possible explanation is that the leader in the E12 group gave more weight to the staff input than the E22 group because the leader knew the validity of the staff input was not the same. This might be an indication of how staff validity may influence team performance. However, there is no hard evidence to prove what leaders really thought about their staff. In terms of testing of Hollenbeck's team performance indexes, informity, validity, and sensitivity, the design of this study could not systematically answer the relationship among them. A further study is called for to explore the relations.

#### Future Work

To improve a better understanding of how display, task complexity and team composition influence our decision, some suggestions are made for future studies. First, different subjective responses were found from participants regarding display. Some claimed the graphic display was better while others said it was worse. The fact is that some people seem good at handling graphic information while others are good at processing numeric or text information. Their abilities to handle different formats of information may confound with the display effect. Thus, future studies may want to take their individual difference abilities as the covariate and hopefully generate clearer evidence of how display design works in the task environment.

Second, in the current study, task complexity for the leader groups only falls into two classes, adding versus multiplying first, then adding. A further study may include multiplying only in the task bank so that both staff and the leader can be compared with each other across different task complexity settings.

Third, participants in this study have ample time to make their decisions. Adding time pressure to the experimental paradigm may magnify the display effect and further test how the PCP works with the CCT.

Finally, although there was no time limit for participants to answer the question, apparently the high complexity groups needed more time to finish the task. In addition, the participant usually becomes impatient at the end of the task because their matching index started to drop. Thus, how to keep the participant motivated and performing consistently needs to be taken into account in future studies.

## REFERENCES

- Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focused attention, and problem solving. Human Factors, 34(5), 513-533.
- Carswell, C. M. (1992). Choosing specifiers: An evaluation of the basic tasks model of graphical perception. Human Factors, 34(5), 535-554.
- Cooksey, R. W. (1996). The methodology of social judgment theory. Thinking & Reasoning, 2(2-3), 141-173.
- Coury, B. G., & Pietras, C. M. (1989). Alphanumeric and graphic displays for dynamic process monitoring and control. Ergonomics, 32(11), 1373-1389.
- Elvers, G. C., & Dolan, N. J. (1995). A comparison of the augmented bar display and the object display. Ergonomics, 38(4), 777-792.
- Endsley, M. R., & Kiris, E. O. (1994). Information presentation for expert systems in future fighter aircraft. International Journal of Aviation Psychology, 4(4), 333-348.
- Flexman, R. E., & Stark, E. A. (1987). Training simulators. In G. Salvendy & et al. (Eds.), Handbook of human factors, p.1012-1038. New York: John Wiley & Sons.
- Gigone, D., & Hastie, R. (1997). Proper analysis of the accuracy of group judgments. Psychological Bulletin, 121(1), 149-167.
- Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision-making: How member familiarity and information distribution affect process and performance. Organizational Behavior & Human Decision Processes, 67(1), 1-15.

Hammond, K. R. (1986). A theoretically based review of theory and research in judgment and decision-making (Report No. 260). Boulder, CO: Center for Research on Judgment and Policy, University of Colorado.

Hammond, K. R. (1990b). Intuitive and analytical cognition: Information models. In A. P. Sage (Ed.), Concise encyclopedia of information processing in systems & organizations (pp. 306-312). Oxford: Pergamon Press.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social Judgment theory. In M. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes (pp. 271-312). New York: Academic Press.

Hammond, K. R., & Summers, D. A. (1972). Cognitive control. Psychological Review, no, 1, 58-67.

Hammond, K. R., & Wascoe, N. E. (1980). Realizations of Brunswik's representative design. San Francisco: Jossey-Bass.

Holding, D. H. (1987). Concepts of training. In G. Salvendy (Ed.), Handbook of Human Factors (pp. 939-962). New York: John Wiley & Sons.

Hollenbeck, J. R., Ilgen, D. R., Segoe, D. J., Hedlund, J., Major, D. A., & Phillips, J. (1995). Multilevel theory of team decision-making: Decision performance in teams incorporating distributed expertise. Journal of Applied Psychology, 80(2), 292-316.

Hollenbeck, J. R., Ilgen, D. R., Tuttle, D. B., & Segoe, D. J. (1995). Team performance on monitoring tasks: An examination of decision errors in contexts requiring sustained attention. Journal of Applied Psychology, 80(6), 685-696.

Kelley, C. R. (1969). What is adaptive training? Human Factors, 11, 547-556.

Mahan, R. P., Kirschenbaum, S. S., Jilg, J. M., & Marino, C. (1998). Supporting diagnostic judgment during delayed-feedback decision making through reliability feedforward technologies. International Journal of Cognitive Ergonomics, 2(3), 217-242.

Orasanu, J., & Salas, E. (1993). Team decision-making in complex environments. In G. A. Klein & J. Orasanu & et al. (Eds.), Decision-making in action: Models and methods, 327-345. Norwood: Ablex Publishing Corp.

Payne, D. G., Lang, V. A., & Blackwell, J. M. (1995). Mixed versus pure display format in integration and nonintegration visual display monitoring tasks. Human Factors, 37(3), 507-527.

Rogelberg, S. G., & Rumery, S. M. (1996). Gender diversity, team decision quality, time on task, and interpersonal cohesion. Small Group Research, 27(1), 79-90.

Schaubroeck, J., & Muralidhar, K. (1991). A meta-analysis of the relative effects of tabular and graphic display formats on decision-making performance. Human Performance, 4(2), 127-145.

Schittekatte, M., & Van Hiel, A. (1996). Effects of partially shared information and awareness of unshared information on information sampling. Small Group Research, 27(3), 431-449.

Schkade, D. A., & Kleinmuntz, D. N. (1994). Information displays and choice processes: Differential effects of organization, form, and sequence. Organizational Behavior & Human Decision Processes, 57(3), 319-337.

Spence, R., & Parr, M. (1991). Cognitive assessment of alternatives. Interacting with Computers, 3(3), 270-282.

Timmermans, D., & Vlek, C. (1996). Effects on decision quality of supporting multi-attribute evaluation in groups. Organizational Behavior & Human Decision Processes, 68(2), 158-170.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hirsch, Hammond, and Hirsch, and by Hammond, Hirsch, and Todd. Psychological Review, 71(6), 528-530.

Urban, J. M., Weaver, J. L., Bowers, C. A., & Rhodenizer, L. (1996). Effects of workload and structure on team processes and performance: Implications for complex team decision-making. Human Factors, 38(2), 300-310.

Weiner, E. L. (1977). Controlled flight into terrain accidents: System induced errors. Human Factors, 19, 171-185.

Wickens, C. D. (1986). The effects of control dynamics on performance. In K. R. Boff & L. Kaufman & et al. (Eds.), Handbook of perception and human performance, Vol. 2: Cognitive processes and performance, 1-60. New York: John Wiley & Sons.

Wickens, C. D. (1987). Information processing, decision-making, and cognition. In G. Salvendy & et al. (Eds.), Handbook of human factors. See p.72-107 (xxiv, 1874pp. | Print (Paper) ed., pp. pages 72-107). New York: John Wiley & Sons.

Wickens, C. D., & Carswell, C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. Human Factors, 37(3), 473-494.

Zaccaro, S. J., Gualtieri, J., & Minionis, D. (1995). Task cohesion as a facilitator of team decision-making under temporal urgency. Military Psychology, 7(2), 77-93.