ARI Research Note 95-32

Methods of Displaying Multiple Performance Measures from Simulator Exercises

Robert P. Mahan
University of Georgia

Armored Forces Research Unit
Barbara A. Black, Chief

Personnel and Training Systems Research Division
Zita M. Simutis, Director

April 1995

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

Edgar M. Johnson
Director

Research accomplished under contract
for the Department of the Army

University of Georgia

Technical review by

Theodore M. Shlechter

Accesion For
NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution /

Availability Codes

Dist	Avail and / or Special
A-1	

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)

2. REPORT DATE 1995, April

3. REPORT TYPE AND DATES COVERED FINAL 6/93 - 8/93

4. TITLE AND SUBTITL\
Methods of Displaying Multiple Performance Measures From Simulator Exercises

5. FUNDING NUMBERS
DAAL03-91-C-0034
63007
A793
2124
C01

6. AUTHOR(S)
Robert P. Mahan (University of Georgia)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-IK
5001 Eisenhower Ave.
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Note 95-32

11. SUPPLEMENTARY NOTES
Task was performed under Scientific Services Agreement issued by Battelle Research. Author affiliation: Department of Psychology, The University of Georgia, Athens, GA 30602

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words):

This report examined the notion of developing a summary display system that would extend the capabilities of the Unit Performance Assessment system (UPAS) in supporting after action reviews conducted in the Simulation Networking (SIMNET) environment. The report details the initial phase of an effort to produce displays that use integral display technology in supporting after action reviews. The report discusses the Cognitive Continuum Theory as a framework that may guide the development of the summary displays. In addition, the findings suggest that general categories of combat unit actions (Move, Shoot, and Communicate) may serve as a set of global dimensions that are well suited for summary display information. Finally, an empirical study is outlined that documents the proposed next step in the summary display system project.

14. SUBJECT TERMS
Intuitive cognition
Integral display
Cognitive continuum theory
Multiattribute utility theory

15. NUMBER OF PAGES 58

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18 298-102

ACKNOWLEDGMENTS

The author wishes to thank the Armored Forces Research Unit Chief, Dr. Barbara Black, as well as Drs. Bill Burnside and Theodore Shlechter for their assistance and support during this project. In addition, special thanks are due to Dr. David Bessemer for his patience in providing detailed and comprehensive feedback on an earlier version of the paper. His efforts are greatly appreciated. Finally, the author wishes to express sincere gratitude to the Simulation Network Armor Officer Basic Course (SIMNET AOBC) instructors whose participation and valuable recommendations were critical to the success of the project.

METHODS OF DISPLAYING MULTIPLE PERFORMANCE MEASURES FROM SIMULATOR EXERCISES

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Methods Of Displaying Multiple Performance Measures From Simulator Exercises

Introduction

The Army is continuing to develop distributed interactive simulation (DIS) as a viable cost efficient method for combined arms training and systems development. One objective for the Army is to use DIS as a way to represent virtual warfighting environments suitable for supporting the training and evaluation of military units, as well as the development of military weapons, standards, and organizations (UCF Institute for Simulation and Training, 1993).

DIS provides a system architecture that allows networking multiple warfare environments which are distributed at various sites and locations in the U.S. and elsewhere. The interactive simulation concept also allows many levels of virtual representations to be configured- from small entity platforms (e.g., weapons) to battalion level entities and larger. Future design goals include a seamless interconnection of all levels of the warfare environment resulting in full interoperability among the various electronic entity representations (UCF Institute for Simulation and Training, 1993).

Simulation technology appears especially useful in situations where the cost, safety, environmental, and political implications of field training are difficult to justify. Furthermore, it is believed that simulation-based training environments will help support many of the activities that are central to Army operations and mandates such as, combat development, system acquisition, and tests and evaluation. DIS allows combat system development and decisions regarding system acquisition to be made on the basis of evaluations conducted under realistic battlefield conditions. Evaluations made in this manner are likely to lead to very cost efficient methods for system development and improvement processes, as well as for system procurement practices in general.

General Objectives

The objectives of this report are to examine some important issues associated with building a summary display that, in theory, will complement and enhance the after action review procedures currently used by the Army. The report specifically defines some of the parameters for a display system that may be useful in extending the capabilities of the Unit Performance Assessment System for the simulated networking environment. Multiattribute utility theory is used to identify tactical information sources that expert trainers find useful when making holistic assessments of training performance. The results of the utility analysis are used to configure an empirical study that will test different display design concepts.

Simulation Networking (SIMNET)

Simulation Networking (SIMNET) represents an application prototype of distributed interactive simulation that is being used at the Mounted Warfare Simulation Training Center located at Fort Knox, Kentucky. Fundamental to SIMNET is the idea of networking multiple combat vehicles, both locally and in remote locations, which allows vehicle crews the ability to interact with one another in real time using a common terrain database. Information regarding vehicle position, movement, as well as a host of other variables, is transmitted over the network to all the vehicle simulators participating in a training event.

Performance measurement and evaluation in SIMNET is an important dimension in developing the device-based training concept. The ability to quantify performance effectiveness in training simulations using performance measures that exist in field training domains is necessary in order to evaluate crew effectiveness within the context of field doctrine, as well as evaluate the potential for transfer of training to field environments (Shlechter, Bessemer, & Kolosh, 1991).

Automated Data Collection System

The Unit Performance Analysis System (UPAS) is currently a measurement system supporting SIMNET (White, McMeel, & Gross, 1990; Meliza, Bessemer, & Hiller, 1994). UPAS collects data packets and translates these packets into data tables similar to the tables used at the National Training Center. These tables are then accessible through Structured Query Language (SQL) for various data analysis purposes. One of the specific goals of the UPAS system is to provide support for after action reviews (AARs) as well as take home packages (THPs). Both AARs and THPs represent important training aids for SIMNET.

After Action Review in Training Performance

The AAR is viewed by the Army as a way to provide soldiers with immediate feedback on relatively complex tactical issues that emerge during training. The AAR addresses three questions: (a) what happens, (b) why it happens, and (c) how to improve. It has become a consistent part of SIMNET training activities occurring immediately after a training scenario has been conducted. The AAR appears to possess a certain amount of face validity. For example, trainers feel that it is a very effective format in which to probe soldiers on their actions during training exercises, and this probing behavior helps to increase tactical knowledge and enhance performance.

Several military researchers are helping to clarify the scientific validity of AARs used as a training tool in distributed network simulations. For example, it appears that performance in the SIMNET environment is a function of the quality of feedback received on SIMNET training (Bessemer, 1990;

Shlechter, Bessemer, & Kolosh, 1991). Further, feedback of SIMNET performance after exercises is an important aspect of the transfer of training of SIMNET skills to field operations, and the AAR serves as an essential vehicle for this feedback (Kerins, Atwood, & Root, 1990; Bessemer, 1991; Shlechter, Bessemer, & Kolosh, 1991).

The research on the AAR concept seems to indicate that it is a way to support trainers with an unstructured approach to analyzing training events (Kerins, Atwood, & Root, 1991). Although, training objectives are based on defined standards of performance, the AAR is a flexible event-driven method that focuses on what occurred and how and why it impacted mission outcome. The AAR tends to be a dynamic interactive learning session. Participants are able to discuss why particular actions were taken and what alternate courses of action could have been more efficient. Thus, the interactive properties of the AAR distinguish it from other very formal training evaluations. However, it is this informal and flexible character of the AAR that best matches the complexity of many training exercises. The AAR seems to fit with the idea that many of the most meaningful assessments of training are configural in the sense that they capture not only the general dimensions of performance, but the interactions among dimensions. Here, feedback based on predetermined measures (e.g., objective physical indices) are often insufficient in revealing the network of variable relations that define dynamic training events. The AAR serves the purpose of focusing on the higher level general principles of combat performance.

Data Volume in Simulation Exercises

Perhaps one of the most significant problem areas associated with the simulation environment and automated data collection capabilities is the potential for generating overwhelmingly large volumes of information. The amount of data that is produced in computer simulations makes the process of summarizing and packaging this information for user feedback a key concern. Clearly, the performance measurements that are taken during a training event must be provided to the user in a friendly and informative manner if the data are to be used at all.

Data volume in combination with incomplete performance summary measures have become especially problematic for the AAR in SIMNET exercises which, in part, relies on data collected by UPAS for training evaluation. UPAS can be cumbersome and require very labor intensive analyses in order to effectively support SIMNET activities. However, a key dimension to the AAR is that it take place almost immediately after a training event is completed. This means that rapid data analysis methods must be developed in an effort to support conducting the AAR. Several graphic map displays have been developed to provide quick overviews of exercise events. Currently, however, the UPAS system also is configured to provide detailed quantitative system also is configured to provide detailed quantitative information on unit performance that requires a substantial amount of off-line data analysis. As a result, the AAR process has yet to be fully supported in current training exercises.

Work is progressing to develop data summaries that can be rapidly integrated into the AAR performance feedback process. Menu driven options that highlight AAR map displays, tables, and graphs are the focus of some of this work. The modification of the UPAS software to support AAR options is underway. This effort would minimize the need to create individual tables and graphs for each unique training event (necessarily an off-line process) and would essentially be formatted in the data analysis software itself.

Summary Display Aid For Platoon-Level AAR

There has been an interest by users of UPAS (most notably course instructors and unit leaders acting as SIMNET trainers) to develop summary displays that can exploit some of the formatting modifications of UPAS currently proposed and mentioned above. One particular interest has been to produce a display that globally summarizes a training exercise. For example, a quick look at such a display would give a good overall assessment of a platoon's general performance for a given training event. Such a display would likely present information on training dimensions common to many kinds of SIMNET exercises. Implicit in the summary display concept is that by using this display trainers would be able to quickly form general impressions of performance, and rapidly deliver these impressions to the soldiers. The information necessary for these "summarizing activities" would need to be packaged in a manner that would facilitate a very rapid set of judgments by trainers as to the general quality of performance by a platoon.

Display Requirements

This report examines several key aspects associated with the design of a summary display system which would complement AAR activities currently conducted by the Army. Any display design project must meet the basic criterion of acceptance by the intended user group. For example, if the display offers little in the way of face validity for the user group (i.e., fails to present the most sensible information given the task at hand), is inefficient in terms of how it communicates with the user (i.e., poorly packages the information), and is too cumbersome to use (i.e., requires great amounts of mental resources in order to be useful), then it simply will not be found acceptable by the users. Thus, when one considers creating an information display, one must simultaneously consider three interrelated components. The first component deals directly with the structure of the task or the nature of the system whose measured characteristics represent the information to be used in the display. The second component of the display process deals with the manner in which this information is packaged and displayed to the user. The third component deals directly with the user of the information and the decision making or evaluation process itself. This last issue references how relevant task-specific data are extracted from the information display and processed by the user. As will be shown, this processing component is highly dependent on the first two components described.

User acceptability means that the task environment, display format, and cognitive processing functions of the user must be fully examined if a new display is to be found useful by the user community. In addition, the dependent nature of the above system components creates an interface infrastructure that calls for addressing the complex interconnections among the components in order to fully elucidate some of the emergent properties of the display system. Emergent properties are viewed as a product of the interactions among these system components.

The requirements for a summary display may first appear straight forward and simple. However, as will be shown, the intelligent design of such a display is a substantial undertaking. For example, the display must use information that is generally linked in valid ways to performance in many different training scenarios, as well as to specific situational training exercise standards defined in the Army Training and Evaluation Program (ARTEP) Mission Training Plan (MTP) documents. Locating the "right" performance dimensions to use in order to meet these criteria will be crucial to the successful production of a summary display system. In addition, because the AAR process takes place almost immediately after a SIMNET training scenario is completed, the data must be easily packaged for presentation. Further, the AAR calls for an information display format where the most relevant information is easily identified and extracted by the trainer. This is a key concern for the AAR. The rapid dissemination of performance feedback demands that the multidimensional performance information be easily manipulated by trainers (i.e., identified, extracted, and combined) to produce general quality evaluations. Clearly, if the only possible way to generate these training quality evaluations is by using computational devices to assist the user with information processing, then the AAR will no longer serve its intended purpose - be a timely process for SIMNET performance feedback.

Specific Objectives

The present report outlines an effort to develop a summary display that can be used by trainers as a training tool to facilitate the rapid dissemination of critical performance feedback in the AAR. Building such a display requires several steps. The first step, which represents the initial goal of the present work, was to define a group of performance measures that are globally linked to a variety of training scenarios. Identifying a set of measures that commonly predict performance in a wide range of training activities is a fundamental aspect of tactical training environments. Further, determining the relative importance of these measures in predicting the outcome of various training events was necessary. Detailed interviews with trainers conducted within a multiattribute utility framework were used to establish the most relevant performance measures for the summary display, as well as their diagnostic importance (weighting) in judgments of the quality of platoon training performance.

The second step of the project will include empirical work in order to determine what display formats are most effective in supporting the rapid dissemination of training information necessary for the AAR process. An outline of the empirical work is included in Appendix B of the present report. The final step of the summary display project is the construction of the display and interface that will use the UPAS data collection system software.

Cognitive Aspects of Display Design

Developing AAR aiding design concepts, especially those associated with visual data displays, requires a detailed understanding of the cognitive foundations of display theory. This understanding is crucial if a display system is to be of any value to those assigned to use it. Haphazard development of information training displays will certainly lead to poor acceptance by the SIMNET community. Further, as will be discussed below, haphazard display design has the risk of a display supporting the wrong assessments of training exercises by trainers leading to erroneous training feedback.

Trainers as Intuitive Experts

It is clearly necessary to describe the intended user of this system. Although, a guiding engineering principle in computer display design is to build for a wide range of users, the nature of the task (supporting the AAR) means that the efforts will be directed toward assisting Army trainers with performance assessment. Trainers, whether unit leaders, school instructors, or dedicated simulator observer/controllers, are a special population of soldiers with unique skills, knowledge, and most importantly, expectations. Our goal is to produce an information display system that is valuable to soldiers acting as training instructors in SIMNET. Our general user profile is that of a highly trained commissioned or noncommissioned officer that is experienced with ARTEP MTP standards at the battalion level and below. Most, if not all complex training assessments made within the SIMNET environment are made by trainers executing multidimensional judgments concerning a variety of training criteria. Whether the judgments are made in real-time during an actual simulation, or in an off-line context using performance checklists and other types of performance data, the trainers' judgments are viewed as the manifestation of superior understanding including a large domain of military standards and practices, and therefore, qualify as expert judgments.

Research in cognitive science has clearly demonstrated that experts and non-experts are different on nearly every dimension of cognitive functioning, from memory and problem solving to learning and reasoning. Two basic themes have emerged from the cognitive literature that define the uniqueness of experts that are relevant to the current efforts underway here. First, expertise appears domain-specific. Any special skills of expertise are lost when one moves out of the expert's specialization. This essentially implies that the expert's problem solving skills are specifically tailored to the problem area of specialization. Non-experts have a tendency to reason in reverse from unknowns to givens. Experts on the other hand reason forward using stored functional units of information from givens to goals (Larkin, 1979). This forward reasoning aptitude only seems to develop in specific domains. Slatter (1987) suggests that the thought processes of experts become domain adapted.

Secondly, the thinking of experts relies more on automated processes (Shiffrin & Schneider, 1977). These processes are most certainly parallel and functionally independent, and are similar to pattern recognition components of visual perception. Non-experts tend to manifest very controlled processing that is linear-sequential in nature, and very similar to the protocol required for deductive reasoning (Larkin, McDermott, Simon, & Simon, 1980). As they gain experience, experts tend to rely less on analysis or analytical/deductive thinking and more on pattern recognition-like or intuitive thinking.

The notion of automated parallel processing that characterizes expertise deserves significant elaboration. It is here that experts demonstrate their superior skills and understanding of complex problems. From a decision support viewpoint one would be most interested in developing a technology that exploits the natural tendency of experts to organize information in a manner that supports automated parallel processing. Further, because parallel processing is very fast, this kind of cognitive behavior would fit nicely with the AAR mandate for rapid performance assessments for immediate training feedback. The design process for decision support will have clearly failed if instead of helping the expert toward automated processing, it drives the expert toward a laborious sequential and linear processing of information. We want to use caution in our display design process to avoid the risk of creating a non-expert out of our expert! Thus, packaging information must be compatible with (a) the way experts prefer to operate on information, and (b) the nature of AAR task requirements.

Intuitive Cognitive Functioning

Experts are unique in the sense that they are highly intuitive within a specific problem domain. Understanding how to format and display information for use by a domain expert requires knowledge of some of the properties of intuitive problem solving that are manifestations of automated parallel processing.

It is reasonable to suggest that the activity of rapidly forming general assessments concerning complex performance would be seen as an important quality for a trainer conducting an AAR. In fact, it is this quality, the ability to quickly and effortlessly extract and combine the most relevant information from a large information array that distinguishes the superior abilities of experts over those less experienced. But, what does it mean to be able to look at an information array and see meaningful patterns which reveal the latent structure in the information, and then make judgments on the characteristics of those patterns? Equally important is whether it is possible to support, encourage, and/or enhance this kind of processing through efficient display design techniques.

Intuitive Problem Solving

The notion of intuitive cognition is often presumed to be distinguishable from analytically-based modes of cognition in the sense that intuition tends to be a kind of seamless form of cognition that manifests parallel processing characteristics (i.e., quick, effortless, single stage). Kahneman, Slovic, & Tversky (1982) characterize it as an informal reasoning process that lacks the application of formal methods of calculation. Most researchers agree that intuitive cognition is fundamentally inferential, where, for example, several propositions regarding relationships about a given phenomenon are combined with general knowledge to yield a verbal conclusion (Johnson-Laird, 1990; Evens, 1990). Some have been more pointed in differentiating between these forms of cognition. Johnson-Laird (1990) notes when characterizing the distinction between analysis and inference: "it is one thing to multiply two numbers, and quite another to draw a picture, to compose a sonata, or to write a poem" (p. 156). Further, he goes on to suggest that inference may be associated with the same cognitive mechanisms that are brought to bear when one "uses one's imagination". Although, intuitive cognition has typically been difficult to operationalize, it clearly has been conceptually differentiated in the past from analytical modes of cognition (e.g., Beach & Mitchell, 1978; Brooks, 1978; von Winterfeldt & Edwards, 1986; Kahneman, Slovic, & Tversky, 1982; Garner, 1981).

In contrast to intuitive cognition, analytical cognition is typically conceived as being a highly proceduralized and deductive serial process that conforms to the basic canons of a particular system of logic. Confirmation to the canons of logic insure the validity of solutions. Hammond (1981) notes that the logic system serving as the framework for supporting analytical cognition can be; "(a) mathematical in nature (e.g., simultaneous equations in word algebra problems), (b) statistical (e.g., Bayesian or Fisherian statistics for uncertain events), (c) propositional logic (e.g., modus tollens, modus ponnens), (d) problem solving strategies (such as opening and end games in chess), (e) scientific laws (as in physics and chemistry) and (f) ideological maxims (in political debates)", (pp. 15).

While the nature of intuitive cognition is often challenging to specify precisely, von Winterfeldt and Edwards (1986) are somewhat more intrepid then others at detailing their opinions on what intuitive cognition is, or perhaps, what it is not. They argue that there are multiple forms of intuitive cognition beyond the simple dictionary definition of "immediate and effortless understanding". For example, they believe an important form of intuitive cognition is observed in problem solving when one verifies a particular solution arrived at via analytical means. They suggest that this latter example of intuition applies to analytical problems whose answers are evaluated on the level of their reasonableness. Here, the notion of reasonableness is similar in nature to what Simon (1976) argues in his satisfying principle, or Klienmuntz's (1985) idea of approximation. This particular kind of cognition is most likely the processing mode induced by difficult analytical problems where the strategic goal in mind is that of achieving a solution that is "good enough" (Simon, 1976, 1978). Analytical problems that induce satisfying or approximation are those problems that are typically beyond a person's ability or willingness to mentally perform or follow the computational procedures necessary for the correct solution. Instead, one evaluates the answer on the grounds of its being reasonable given the parameters (i.e., the values in the case of arithmetic) of the problem (von Winterfedlt and Edwards, 1986), and the criteria established for goal achievement (Simon, 1976).

Researchers typically acknowledge (a) intuition is a mode of cognition that is used daily by people in natural decision making environments and (b) that it is qualitatively different from analytically driven problem solving. The majority of attempts at elaborating on the distinction between analytic and intuitive forms of cognition have done so by illustrating the profound shortfalls inherent in relying on intuitive problem solving skills in decision making. Tversky and Kahneman (1974) and later Kahneman, Slovic and Tversky (1982) present an extensive review of the errors and biases related to intuitive cognition that take the form of heuristic strategies in decision making. However, very few studies have adequately defined what intuition is, nor how an individual's intuitive processing compares with his/her analytical skills. The latter point is best dramatized by Hammond, Hamm, Grassia and Pearson (1987) who cogently dispute that virtually all major research efforts have benchmarked a person's skill at exercising intuitive judgment with an analytically derived algorithm or axiomatically defined rule system. They have stressed that using a prescriptive model criterion index is unsatisfactory, and at best, an indirect method of characterizing the efficacy of intuitive cognition. Some have taken a different approach in disputing the shortfalls in complex human decision making. For example, Anderson (1986) argues that many of the notorious and alleged decision biases reported in decision making research are simply manifestations of poor experimental/methodological techniques, fallacious experimental interpretations and artifactual laboratory findings. Finally, situational factors such as time pressure, risk and uncertainty can affect judgment and decision making in unique ways.

The rationale used by Hammond and others (Hammond, Hamm, & Grassia, 1986; Wright & Murphey, 1984) for questioning comparisons between intuitive cognition and person-independent operations, such as computational procedures in normative performance models, is that this methodology restricts the assessments of the reliability and precision of intuitive judgment in several ways. First, the benchmark used to assess performance is arbitrary in the sense that different normative standards, ranging from propositional logic models to statistical algorithms and so on have been employed without benefit of describing how the standards are different and what it means to not achieve the standard of performance set by the particular analytical model under consideration (for a historical perspective on the normative decision making issue, see; Slovic, Fischhoff, & Lichtenstein, 1977). Secondly, the performance standards typically receive all the relevant information necessary for deductive analysis and computation, and the actual solutions to these problems are derived without error. The availability of decision relevant information and operations on that information distinguish the standard (normative) model from models used by people in natural decision making environments. Finally, standard performance models provide the uppermost level for accuracy and thus cannot fail to demonstrate that these models are equal or superior to intuitive models of cognition (Hammond et al., 1987; Wright & Murphy, 1984).

Cognitive Continuum Theory

In order to more fully characterize the essential properties that represent intuitive and analytical forms of cognition, Hammond (1980) has elected to concentrate on principles of information organization and their precise relationship to the properties of a given task. From this organizational perspective, Hammond has developed the notion of a cognitive continuum that is anchored by analytical and intuitive modes cognition (Hammond, 1980, 1981; Hammond et al., 1987). The cognitive continuum view has allowed for the development of a theoretical framework offering specific predictions of task-driven cognitive behavior that is likely to provide a more detailed examination of human cognition over the more traditional approaches of applying normative standards in assessing cognitive efficiency. Hammond defines specific methods for testing various organizing principles, and this fact alone distinguishes the theoretical potential of the cognitive continuum from other conceptions of information organization that do not lend themselves well to operationalization and empirical verification (e.g., most notably, the notion of "schemata", Schank & Abelson, 1977; and "production systems", Newell, 1973; Anderson, 1976). Further, by predicting cognitive behavior in a manner that is independent of task definitions, Hammond avoids the circular logic in the assumption that nonanalytic processing is being manifested because an individual is performing an analytical task poorly (Garner, 1981).

A fundamental substantive contribution made by the cognitive continuum theory lies in the notion that cognitive efficiency, and thus performance, is in part, a function of the congruence between the properties of the task and the cognitive organizing principles employed by the decision maker . The theory essentially describes a system of two contiuua: (a) one associated with the task, and (b) the other associated with the cognitive disposition of the individual. Oversimplifying the theory (see Hammond, 1981; Hammond et al., 1987), the assertion is that various properties of the task "induce" a particular mode of cognition lying somewhere between the analytical and intuitive poles on the cognitive continuum. For example, a simple, highly structured deterministic task (e.g., simple mental arithmetic) is likely to induce a mode of cognition (i.e., organizing principle) at the analytical end of the continuum. Here, the psychological and or behavioral consequences of such an organizing principle is that a very proceduralized set of operations are executed, at a somewhat methodical pace with a relatively high degree of accuracy, where the subject is highly aware of the organizing principle. In contrast, a complex, ill-structured and ambiguous task is likely to induce a mode of cognition in the person that is closer to the intuitive end of the continuum. The intuitive cognitive mode would tend to be associated with a holistic organizing principle, executed quickly and with lower overall accuracy when compared to some normative standard, and where the subject would manifest less awareness of the actual organizing principle being used in performance. In effect, the closer the congruence between the properties of the task and the optimal mode of cognition given the structure of the task, the more efficient cognition is likely to be. Thus, for example, an individual utilizing analytical skills to solve a very complex ill-defined problem will likely display incongruence between the organizing principle that is analytic in this case, and the properties of the task, which call for a very different organizing principle that is intuitive in nature. This incongruence will ultimately lead to inefficient cognition.

Tasks can also require quasi-rational organizing principles (Brunswik, 1956). Thus, a task can be defined as possessing both analytical and intuitive properties. The location of the task on the task continuum would be somewhere between the two cognitive poles. The implications for performance on such a task would be that an individual must use both analytical and intuitive skills to adequately perform the task. The nature of a quasi-rational organizing principle is consistent with other cognitive formulations for goal directed decision making, most notably Simon's (1957) bounded rationality principle (see Hammond, 1981).

Ecological Context and Task Structure

The fundamental edifice of the cognitive continuum theory is the notion that the task is responsible for structuring (i.e., inducing) a particular form of cognition. The theory has been developed within the context of the Brunswikian notion of probabilistic functionalism and 'vicarious mediation' in visual perception (Brunswik, 1956). Brunswik's view was that visual perception is the activity characterized by a perceiver interacting with his/her ecological environment; an environment whose tendency it is to distribute or "scatter its effects". With this notion of probabilistic functionalism, he viewed the important ecological dimensions (or cues) of the environment as being probabilistic and not fully reliable or dependable. The fact that the environment presents the perceiver with redundant information in the form of correlated cues (i.e., the environment is vicariously mediated), means that the perceiver must wisely select and use the cues most diagnostic of a given behavioral or perceptual goal. A rather good functional example of the meaning of the probabilistic nature of environmental cues is taken from Gordon (1990) on Brunswikian Psychology:

"Suppose we are searching for an edible fruit. Let us assume that edible fruit is (a) darker, (b) redder, (c) softer and (d) sweeter. Obviously, darker and redder are visual cues, softer is tactile, sweeter is gustatory: the environment is scattering its effects. And these cues, the only ones available, are all imperfect: all carry some risk. Not all ripe fruit is red, nor is all red fruit edible. Sweetness often indicates edibility, but some poisonous fruits are sweet. Some fruit is less edible when soft, some fruit will be rotten." (pp. 101).

As a functionalist, Brunswik's basic perceptual theme was adaptive in nature. That is, in order to survive the perceiver must deal in risk and uncertainty by acting like an intuitive statistician. Thus, the perceiver must be able to a) select meaningful cues from a plethora of ecological information, b) factor the riskiness of the situation, c) combine the cues and risk factors, and d) render a judgment leading to action (e.g., avoid the thicket of trees else risk being eaten by a tiger). Brunswik (1952, 1956, 1957) was responsible for introducing a formal systems approach to the study of human cognition. Brunswik declared that:

"Both organism and environment each with properties of its own......Each has surface and depth, or overt and covert regions. It follows that much as psychology must be concerned with the texture of the organism.....it must also be concerned with the texture of the environment" (1957; pp. 5)

From his probabilistic perspective on the relationship between a perceiver and his/her environment came Brunswik's lens model of behavior, which defined the structural characteristics of the person/ecology relationship. This unique model, with both normative and descriptive features, defines the complex multidimensional representation of human behavior within an ecological context (Brunswik, 1952, 1956). It has been since modified and expanded in order to represent a general model of human judgment and decision making (Tucker, 1964; Hammond & Adelman, 1976; Hammond, McClelland, & Mumpower, 1980; Brehmer & Joyce, 1988).

Figure 1 illustrates that the lens model essentially distinguishes between an object or condition that is defined by various information sources (cues), and the psychological representation of the object or condition which is defined through a particular judgment policy. The lens model portrays the environment as a series of cues whose relationships with the environment are less than perfect. A decision maker is viewed as interacting with his or her environment through a 'lens' which is often distorted because of this imperfect and uncertain relationship. The relationship between the cues and the environment is typically characterized by "ecological validities" that, in theory, can range in absolute value from 0 to 1.0. Ecological validity represents the predictive importance of each cue. The manner in which a decision maker uses particular cues can be modeled by a regression equation that predicts an individual's judgment of an object from a linear combination of cue weights. The degree to which a decision maker accurately assesses the characteristics of an object or condition in the environment is expressed by the correlation between the object's true values and those predicted by the decision maker (Hammond & Wascoe, 1980).

The lens model provides a formal means for quantifying the influence of various task features on human cognitive behavior. As can be seen in Figure 1, the lens model provides the means for manipulating various properties of the task, and provides a network of task and cognate behavioral descriptive terms that can be useful in locating a person's cognitive activity on the continuum.

It appears that cognitive research efforts are responding to the importance of ecological context by acknowledging the pivotal role of task properties in cognition (i.e., what is going on outside the head). For example, in his detailed review of decision making, Payne (1982) asserted that decision making in general is very much contingent on the demands of the task. Simon (1978) also makes this point when addressing issues associated with adaptive systems by indicating that it is

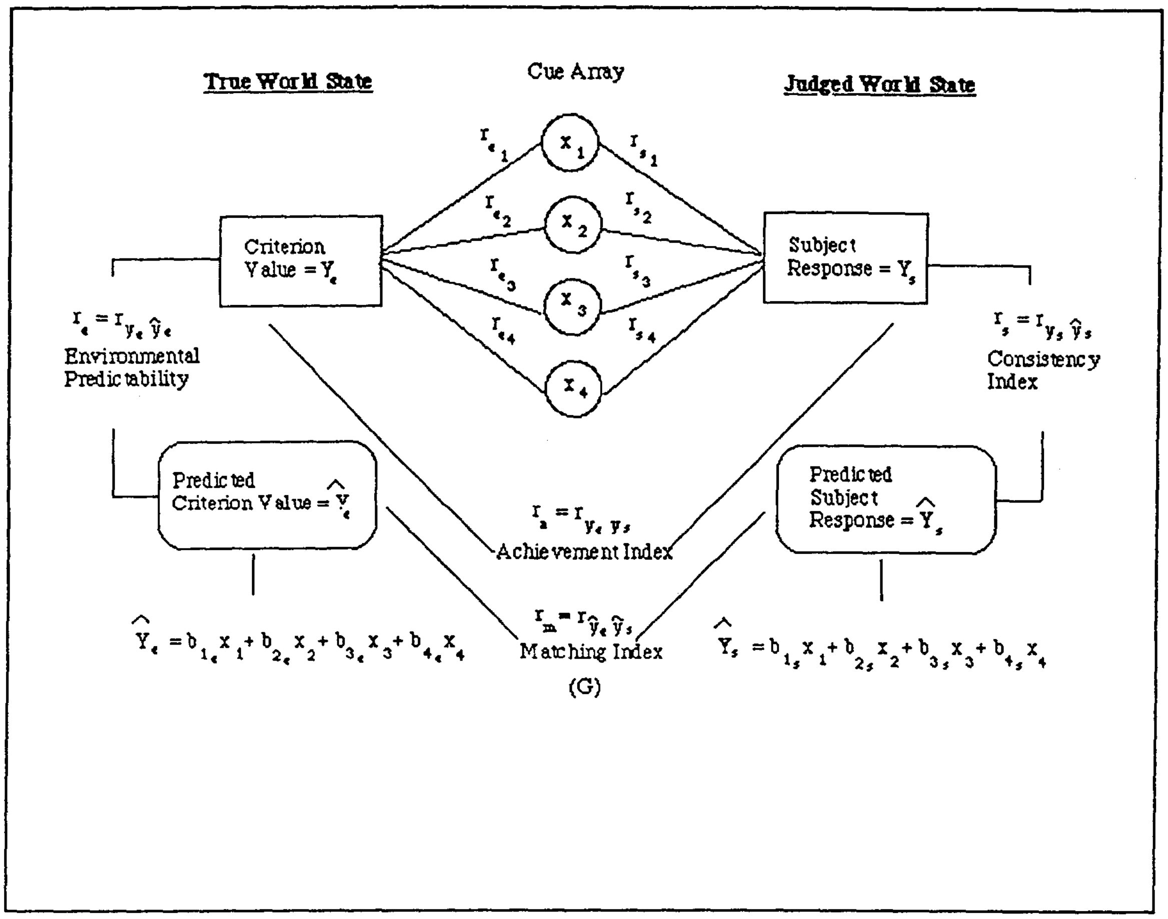


Figure 1. Lens model of judge and environment interaction.

features of the task that strongly influence and guide overall behavior. Newell (1973) also suggests the importance of task ecology when discussing various architectural theories of cognition noting that it is environmental properties which largely determine cognitive hardware requirements or architectural structure. Garner's (1974, 1981) highly cited work on object perception and the principles of "configurality" and "emergent features" underlying unanalyzed perception (i.e., holistic perception) bears directly on his arguments for the importance in specifying properties of the stimulus. Garner essentially argues that task properties alone serve to mediate the mode of object perception. Thus, it is task ecology that induces either a holistic mode of perception, or a perceptual strategy that is based on an analytic decomposition of the object attributes (Garner 1981).

Garner's work on object perception has led to a number of current debates in applied psychology. For example, a controversial topic is associated with the question of what kinds of data displays are actually superior at inducing holistic processing (cf., Carswell & Wickens, 1987; Sanderson, Flach, Buttigieg, & Casey, 1989). Wickens' proximity compatibility hypothesis follows from Garner's work and essentially defines the importance of compatibility between the properties of the task, and in this case, how it is presented and displayed to the operator in order to induce optimal cognitive processing. The proximity compatibility hypothesis "attempts to relate the processing of the displayed information to the nature of the task information processing characteristics" (Wickens & Andre, 1990, pp. 62).

Wickens' arguments for compatibility between display structure and cognition bears, in part, on the work documenting the search for veridical and nonveridical mental representations of physical events and attributes (i.e., how people think about the physical world) (see Gentner & Stevens, 1983 for review). The interest in this work for engineering psychology is to build displays that are most congruent with the way in which people naturally organize and use information. Moreover, the work on mental representations raises performance measurement issues of the kind Hammond et al., (1987) point out in that it seems reasonable to first understand how people think about a given phenomenon before rendering claims concerning the efficiency of human cognition that is based upon a specific performance modeling methodology.

The implications of the above discussion on display design are simple and direct. Performance depends, at least in part, on the properties of the task. This perspective is echoed in the cognitive continuum theory framework where the task itself is viewed as being instrumental in structuring the mode of cognition used in judgments about objects, things or behavior. Cognitive continuum theory asserts that if the task is complex, multidimensional, and possesses uncertain qualities, the most efficient approach for decision making will necessarily be intuitive in nature. If on the other hand, the task is very well defined, supporting absolute and precise statements about task parameters, an analytical mode of processing may be called for. In most cases, however, both analysis and intuition are parts of an expert's judgment. That is, tasks often call for both kinds of processing. In this case the task would induce a quasi-rational mode of cognition that lies somewhere in between the end points on the cognitive continuum. As we shall see below, the task of assessing simulator-based tactical performance in a timely manner favors an intuitive approach to judgment. This is because the problem area is highly complex and contains various degrees of uncertainty depending on the particular level of assessment one wants to address. Further, the intuitive approach favors the manner in which a trainer's expertise is organized. Expertise is often best served by allowing trainers to respond holistically to patterns in tactical information.

An additional implication of the above discussion is the notion of information format. If one accepts, at least in part, that the task is responsible for structuring and inducing a particular form of cognition, then it follows that the structure of the task must be preserved in the physical display itself. This means that the transformation from the task to the display of task parameters must be invariant so as to induce the appropriate organizing principle called for by the structural properties of the task. From this perspective, it is clearly possible to have a task calling for a particular cognitive mode, yet this mode not be supported by the physical properties of the display interface. This disconnect between task structure and display structure will likely lead to ineffective cognitive performance. The physical properties of the display itself are crucial for preserving task structure.

Display Techniques for AAR Decision Support

The idea of a display that induces a user toward an intuitive (holistic) mode of information extraction and integration as opposed to a detailed and time consuming functional analysis of information leads us into newly emerging concepts in research on decision support technology. It is important to discuss some of the developments in display technology that are currently seen as supporting intuitive judgments, as well as the theoretical frameworks that are guiding some of this work.

The basic principle behind developing a summary display that offers SIMNET trainers a method for providing immediate feedback useful in the AAR process relies on the idea that complex information can be displayed in a manner that can be quickly understood with very little mental effort needed for this understanding. Thus, the goal is to select a display format where the meaning of the information presented is highly evident and does not require a detailed and intensive off-line analysis.

Visual Displays

Display theory research has also come to many of the same conclusions described above for modes of cognitive functioning. However, these conclusions have, in most cases, been based on empirical investigations of visual perception. As we have seen, information processing within complex decision environments has elicited the work conducted in perception field (e.g. Brunswik, 1956). A large amount of research in perception exists on the perceptual processing of multidimensional stimuli (see Pomerantz, 1981 for a comprehensive review). Historically the interest in this area has focused on the characteristics of the processing system and its functioning. However, more recent attention has been placed on investigations in the inherent structural properties of stimuli and their effect on the perceptual process. Garner (1970) stimulated interest in this area by arguing that before one can understand the details of perceptual processing, one must understand the details associated with the structure of the stimulus.

A central issue in the research on stimulus structure has been the nature of the relation between dimensions which represent a multidimensional stimulus. Garner (1974) has discussed two major ways in which stimulus dimensions can be related: (a) integral and (b) separable. Psychologically, integral dimensions appear as an integrated whole, whereas separable dimensions are seen as distinct and separate. Further, dimensions tend to be integral if the existence of one dimension depends on the existence of another dimension. An example of integral dimensions would be the hue and saturation of a color (Garner, 1970). Any one of the dimensions of color (i.e., brightness, hue, and saturation) can not exist without values on the other dimensions.

The study of integral and separable dimensions has traditionally been limited to relatively simple stimuli and elementary perceptual processes. Stimuli are generally composed of physical attributes such as color or geometric form and are often limited to two binary dimensions. In discrimination and identification studies, the task is to evaluate stimuli on the basis of one dimension while the stimuli vary along another dimension. The rule relating stimuli to responses is based on physical attributes of one of the stimulus dimensions. Since these tasks are generally easy to perform, speeded responses are obtained and reaction time serves as the primary dependent variable. Classification and similarity scaling do not specify what aspects of the stimuli the subject should use in forming classes and assignment ratings. Instead, the interest is on identifying stimulus structure by the way in which the subjects globally view the stimuli. With integral dimensions, classification is based on the overall similarity structure of the stimuli; with separable dimensions, subjects group stimuli on the basis of a single dimension.

Integral dimensions appear to show an increase in the speed of processing when two dimensions are correlated, and interference when the two dimensions are orthogonal (Garner, 1974; Foard & Kemler, 1989). Stimuli composed of integral dimensions produce a Euclidean metric in direct distance scaling, facilitate the discrimination of stimuli on one dimension when another dimension varies in a correlated manner, and inhibit the discrimination of stimuli on one dimension when another dimension varies in an orthogonal manner. Further, selective attention to a single dimension of an integral stimulus appears difficult to achieve because perception is dominated by the overall similarity structure, or the emergent features corresponding to holistic processing (Garner, 1970).

In contrast, separable dimensions are those that do not show any improvement in the speed of processing when the dimensions are correlated, nor interference when they vary orthogonally. Separable dimensions (e.g., separate vertical bars) produce a city-block metric in direct distance scaling and produce neither facilitation with correlated dimensions nor interference with orthogonal dimensions in a discrimination task. This is to say, that mapping separable dimensions onto integral forms possessing correlated dimensions such as an object will not facilitate processing. This is because separable dimensions support focused attention on specific features of the stimulus (Garner, 1974; Pomerantz, 1981).

Taken together, the above research indicates that the structure of the stimulus plays an important role, not only in elementary perceptual operations, but also in higher order cognitive processes. Specifically, stimuli composed of integral dimensions appear to facilitate performance across several types of perceptual and cognitive tasks. Recent interest has focused on the role of integral dimensions for facilitating performance on information integration tasks.

Integral vs. Separable Visual Displays

Many recent studies have addressed the relative merits of different display formats on human performance. The results of many of those studies support the use of integrative, object-like displays for enhancing a user's ability to assimilate complex information (Carswell & Wickens, 1987; Coury et al., 1989). Such displays appear to be especially effective in situations including multidimensional task variables where decision values are intercorrelated. There is substantial evidence to suggest that object displays are superior to alphanumeric displays in many applications in which identification of system state (or perhaps training state) requires integrating data from several information sources (Carswell & Wickens, 1987; Casey & Wickens, 1986; Wickens, 1984).

The superiority of integral displays has been attributed to the perceptual cues and redundant information inherent in such representations (see Garner, 1974). An operator can use the redundancy in perceptual cues to simplify classification of system data by associating a unique object configuration (e.g. "a happy face") to a specific system state category. Mapping objects or features to a state category occurs when the values of system variables are correlated with a particular state category and the physical representation of those values creates a configuration with a unique size, shape and orientation. Thus, the user need not attend to specific values of the system variables but can rely on rapid, perhaps holistic, integral processing of an object-like configuration of features to determine the state of the system. In terms of multiple-resource theory (Wickens, 1984), object displays produce a spatial code that allows integral processing of system data. This spatial information is associated with what is sometimes called "emergent qualities" of the display. The term "emergent quality" comes from the phenomenological experience of the spatial information "popping" out of the display (but see Carswell & Wickens, 1987).

Alphanumeric displays and tabular information, on the other hand, require the user to attend to each individual system variable and to serially process information as a verbal code. Because these separable display formats require mental manipulation of numerical values to determine category membership, the underlying correlational structure of the system is not readily apparent (at least in Garner's 1970 sense). Presumably, because separable display formats require focusing on each system variable dimension in order to interpret and integrate across system variables, the separable display requires more processing time than the integral display. Consequently, many researchers have concluded that the appropriate display format depends upon the underlying statistical properties of data in a task (Wickens & Andre, 1990; Mahan, 1994).

Discussion on Visual Displays

The above discussion on display configuration implies that there exists a very close link between the judgment/evaluation process and the manner in which information used for this process is physically presented to the user. Oversimplifying the very brief discussion on integral and separable displays, if the task is one of making judgments about a very complex system which is composed of correlated information dimensions, an integral display will be most compatible with the task at hand and will support the mode of cognition most compatible with task structure. In contrast, if the problem domain is well defined and composed of orthogonal information dimensions, a separable display may be more appropriate. Clearly, selecting the wrong display for the task will be counterproductive in the sense that the SIMNET community will not find it useful. The worse case scenario, of course, is that a inappropriate display will lead to wrong kinds of feedback and the risk of poorly prepared and trained soldiers.

Having made, in part, a theoretical case for qualitative differences in the modes of cognition that are induced by the properties of the task, and the display characteristics that preserve the properties of the task, it is necessary to examine a content area that can potentially serve as the source of information for summary display judgements in the AAR.

Summary Display For Platoon-Level AAR

The tactical dimensions Move, Shoot and Communicate, are viewed as forming the tripartite division of action in armor ground-based military doctrine. While, mission, enemy, terrain, troops, and time available (METT-T) define the initial conditions, force structure, and/or tactical context under which battles are fought, performance within METT-T constraints is often evaluated in terms of the movement, fire, and communication exhibited by a given military unit. For example, if a platoon is successful at moving to an objective, engaging and disabling the enemy when it is encountered, and efficiently communicating developments and situations to command as they unfold, the platoon will have succeeded in its mission. Clearly, these are very general dimensions of action, and in some situations are inappropriate dimensions on which to base performance. However, in many tactical scenarios, both offensive and defensive, doing well on these dimensions will often translate into successful tactical outcomes.

Move, Shoot and Communicate are thus considered high level concepts directly linked to tactical performance that may be useful in describing or summarizing the effectiveness of a platoon conducting a number of training exercises. Because these are general dimensions, they should support feedback on general characteristics of performance over a wide range of military scenarios. In the present context of developing efficient AAR techniques that are compatible with SIMNET and UPAS technologies, it seems reasonable to consider general information of the type found in the dimensions above as a possible source of information for creating an AAR summary display.

Method

Multiattribute Utility Analysis

Appendix A presents multiattribute utility analysis (MAU) procedures which are viewed as being useful in structuring a decision problem in a manner that allows a group of judges to choose among several decision options. The details of MAU are introduced in Appendix A by examining a decision problem associated with selecting a city location for work. In addition to using these procedures as a method for selecting a preferred decision option, MAU procedures are often used as input to other assessment methodologies, such as those associated with task analyses (Edwards, 1980). In the present report, the primary utility of MAU is in aiding with the elicitation of the weights used by judges in their evaluation of the various training performance dimensions.

Structuring the Value Tree

A value tree represents the weighted decomposition of global constructs into lower order attributes. Since our value tree represents the interaction between decisions and military objectives its construction will necessarily rely on the expertise of trainers who routinely evaluate training performance. As Appendix A indicates, it is customary to begin building the tree with a set of general constructs that are assumed to be linked in some manner to quality of performance. Moving to the right in the tree from each general concept, the expert decision makers list subcomponents that they believe define these concepts. That is, the experts parse these general concepts into their primitive components (low level attributes). Frequently, these subcomponents can themselves be divided into even more fundamental, and lower order attributes producing a multi-branched and complex value tree. One terminates this content refinement process when attributes can no longer be parsed or refined by the experts.

Determining Weights

There are several estimation procedures that are used in establishing the weights used in MAU. Direct ratings, curve drawing, ratio estimation and category estimation are versions of numerical estimation in which experts are presented with some anchored scale and asked to rate or otherwise estimate the importance of the stimulus relative the anchors.

One form of direct rating has experts distribute 100 points over a series of decision attributes so that the point allocations reflect the importance of the individual attributes. Often, average weights are derived by combining the point allocation totals of several experts. The weights are normalized so they add to 1.0. This particular method has proven to be an efficient method for the value measurement process in MAU and was used in the present study (see Boudreau, 1991).

Subject Matter Experts

The decision experts used in the present study were 12 Armor Officer Basic (AOB) instructors who were actively supervising platoon-level training at the Mounted Warfare Simulation Training Center located at Fort Knox, Kentucky.

Procedure

Interview No. 1.

Each AOB officer was interviewed separately at three different time periods. The first interview was given to the officers in order to disaggregate the primary dimensions (i.e., Move, Shoot and Communicate) into low level attribute lists. An unstructured approach at conducting the interviews was used. The unstructured nature of the interview supported a reasonably efficient way to probe the officers about these attributes. Several versions of the question "Within the rubric of Move, Shoot and Communicate, what kind of activity do you look for as reflecting effective performance?" were asked. The basic emphasis with this line of query was to solicit responses from the officers on how they considered all three dimensions together as a single (complex) structure. The nature of these questions forced the officers to think about interrelated characteristics. The second part of the first interview asked the officers about the primary dimensions independently of one another. Questions of the sort "what does it mean to communicate effectively during a mission?"; or "how is movement related to the effective execution of orders?". These questions provoked discussions that led to the decomposition of the primary dimensions into relevant attributes. The interviews concluded when the officers were unable to extend their attribute lists.

The discussions about the attributes that composed the primary dimensions were necessarily informal. In several cases, attributes were stated as being important by the officers, yet the officers were unable to convey that they indeed had meaningful preferences for different degrees of the attribute. For example, several responses by officers were related to motor and perceptual skills of the soldiers in a platoon. These skills were considered, at least initially by some, as influencing the ability of the platoon to conduct various maneuvers and missions. However, those officers who promoted this idea found it difficult to say much about these attributes other than "more is better". Their real preferences may have been related to the effects of motor and perceptual skills on something like gunnery tasks, which often call for very refined motor control. These particular attributes did not manifest themselves as being value relevant because the officers could not articulate preferences for different quantities of the attributes.

The two part initial interview also allowed for identification of judgmentally dependent attributes, or those that were highly intercorrelated. For example, setting a trigger line and determining the timing of an engagement often mean the same thing; knowing at what range the enemy will be engaged. Removing the redundancy of attributes was possible by using an informal approach at eliciting the attribute lists.

Results

Analysis of Attributes. Once the initial attribute lists were generated that defined the higher order dimensions-- Move, Shoot, and Communicate, the lists were analyzed to reveal those attributes that were common to all judges. An attribute was considered common if every judge listed it as an important component of the higher order values. Table 1 displays the lists of those attributes.

Table 1

Example Definitions Used by AOB Instructors For Low-Level Attributes

Dimensions and attributes	Definitions
Move	
Timing	crossing phase lines; arriving at checkpoints
Density	platoon formation and inter-vehicle distances
Security	executing overwatch maneuvers
Speed	velocity of platoon movement
Navigation	negotiating route objectives

Table 1 continued

Shoot	
Firing Plan	organizing who in the platoon shoots where
Trigger Lines	set timing and range of enemy engagement
Kill Zones	determining suitable engagement area
Concealment	using camouflage, stealth, and position
Communicate	
Timely	prompt reports of events and terrain features
Comprehensive	reports contain all relevant information
Accurate	valid reports
Succinct	reports are brief and quickly delivered using standard operating formats

Interview No. 2-- Valuation. The common attribute lists were shown to each judge. The judges were asked to generate weights for the attributes within each primary dimension. For example, each judge was asked to consider the attributes associated with the primary dimension Move. Each judge was instructed to list these attributes in order of their relative importance to the notion of platoon Movement on the battlefield. The next step of the interview asked each judge to assign relative numerical weights to the attributes. The judges were asked to distribute 100 points over the attributes so that the points reflected the relative "share" of importance. The numbers were then normalized to add to 1.0. The same procedure was followed for attributes under the other primary dimensions. Finally, the primary dimensions were weighted by each judge following the same weighting procedure described above.

Interview No. 3-- Revaluation. After each of the 12 judges had made their personal assessments regarding attribute and primary dimension weights, each judge was shown the assessments made by the other judges. Each judge was asked to examine and discuss the individual differences inherent in the assessments, and afterward each judge was asked to once again produce a set of weights. While there still remained individual differences, the range of differences was now much smaller. The final set of weights was produced by averaging the results of the second weighting across all judges. The average values were accepted as representing the value system, and are shown below in Figure 2.

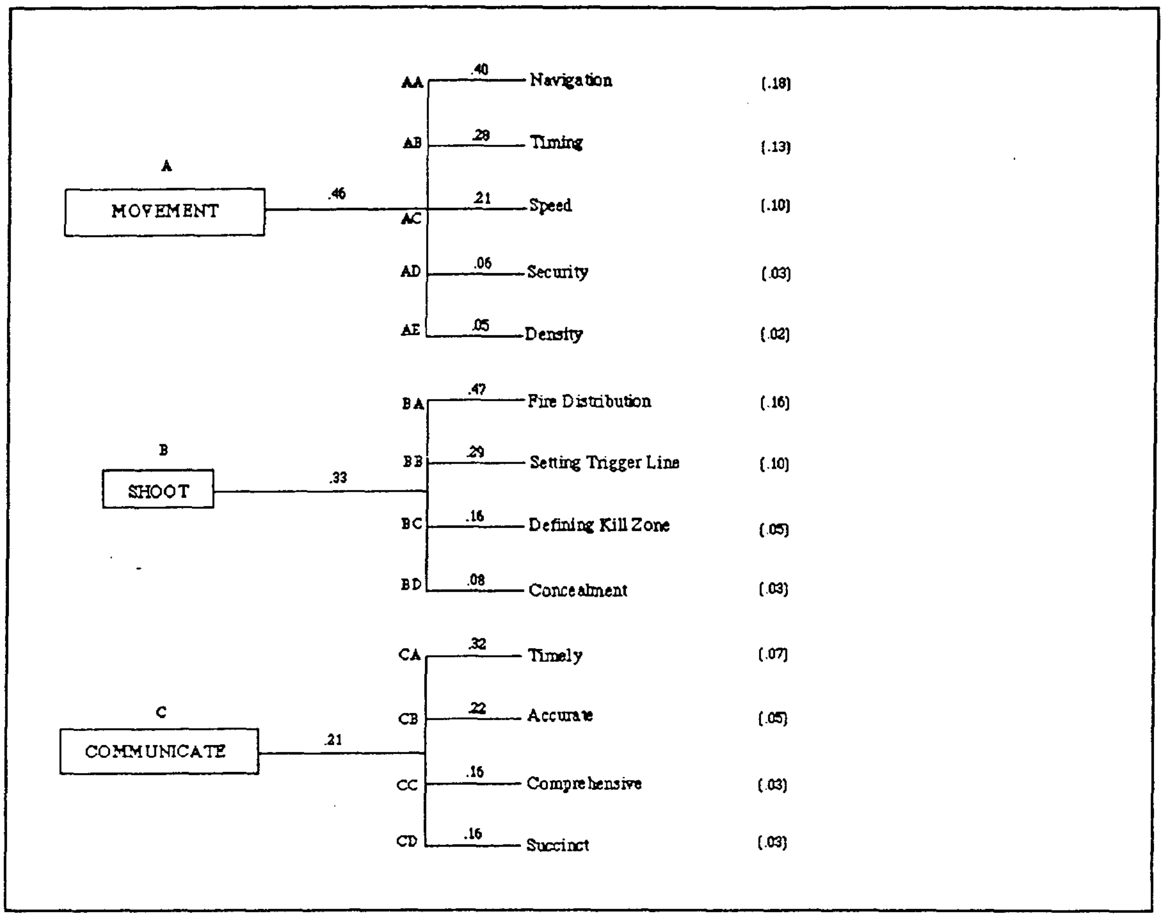


Figure 2. Value tree for Move, Shoot and Communicate.

General Discussion

The value system for Move, Shoot and Communicate described above defines lists of attributes that were found to be common in interviews with expert officers training junior officers on tank platoon mission tactics and procedures. The results of this interview and scaling process should not be viewed as definitive or absolute. The aim of the project was to establish some of the accepted measures by which instructors judge and evaluate tank platoon training activities. It is clearly recognized that METT-T must always be carefully considered in virtually any Armor combat operation. Further, no one set of tactical performance measures can define every tactical situation. However, the elements contained in the primary dimensions of Move, Shoot, and Communicate are relevant to all types of terrain and many different tactical scenarios.

It is not surprising that the movement of a platoon was viewed by instructors as the most important general characteristic of mission success, nor is it surprising that speed of movement was weighted as being one of the most important attributes, second only to knowing where one was going (i.e., navigation). With the number and variety of antitank weapons now available, and the range at which direct fire begins, being able to maneuver at speed is essential for survival. It is also viewed as being essential to conducting many offensive operations. The U.S. Army Field Manual 100-5, Operations, cogently characterizes the concept of speed for Armor Forces: "Speed is absolutely essential to success; it promotes surprise, keeps the enemy off balance, contributes to security of the attacking force, and prevents the defender from taking countermeasures" (1986; pp. 97).

The attribute lists generated by the officers for the other primary dimensions- Shoot and Communicate, were equally sensible. Knowing who in a platoon will be shooting where and at what time is critical for organizing the most efficient execution of massed direct fire, as well as achieving other collateral objectives such as maintaining stealth and concealment. In addition, platoon security is also served by a consistent and efficient fire plan. Similarly, it is necessary to execute efficient communications during any mission and these communications must be quickly delivered, accurately made, and timely in nature.

It is also not surprising that the attribute lists generated do not, as a rule, directly correspond to ARTEP MTP subtasks which define mission standards associated with platoon movement, fire, and communication in situational training exercises (STXs). First, ARTEP MTPs define very specific STXs, and the subtasks within the STXs necessarily tend to be tailored to particular mission operations (e.g., conducting a deliberate attack). The attribute lists generated by the officers in the present study define much broader and general dimensions of performance. The objective was to select primary dimensions that were common to many STXs. Thus, the lists of attributes presented here are meant to correlate with many subtasks which define MTP standards for different kinds of events. For example, conducting a tactical road march and conducting a deliberate attack share some of the same attributes contained in the primary dimensions of Move, Shoot, and Communicate. Here movement and communication are essential for a successful outcome in either exercise.

A second departure from ARTEP MTP standards is due to the fact that a special criterion was used for screening the attributes listed by the instructors. The attributes included in the present analysis had to be scalable on some preference dimension. In some cases the scaling criterion was direct. For example, some of the attributes used by instructors to evaluate platoon performance are linked directly to objective indicants such as time, velocity and so on. These tend to be easily scaled by the instructors on preference (e.g., the more time- the poorer the performance; the faster the platoon- the better the performance). In other attributes, however, the preference dimension was more latent (e.g., defining a kill zone). Instructors had to be able to defend an attribute on the grounds that they considered different degrees of achievement or quality with respect to the attribute. In ARTEP MTP standards the subtasks are evaluated on either a "go" or "no-go" basis. This simple dichotomy prevents the application of a utility analysis which relies on scaling the attributes. The scaling process is necessary for mapping these dimensions to a display of some kind that in turn allows communication of a number of different training states to be made. Presumably, the go/no-go criterion method severely restricts an expert trainer's aptitude to generate much finer detailed evaluations concerning training performance.

Summary

It seems reasonable to expect that SIMNET training aids will move toward multi-media driven formats in future systems, and that UPAS, or more advanced systems evolved from UPAS, will provide the analytical tools for instructional support. Interconnecting various aspects of training events with a variety of presentation media is likely to make the AAR even more effective than it is today. Developing summary AAR aids that can be used to provide immediate feedback to units is a step in this direction.

The present project was aimed at the possibility of creating an AAR summary display that facilitates a very rapid dissemination of complex training feedback. The report focuses on the relationship between the properties of the AAR and the elements of an optimal presentation format. Several tasks were identified that appear well suited in representing criteria for assessing a wide range of SIMNET activities. Most notable were the primary tasks- Move, Shoot, and Communicate, as well as a host of respective subtasks. The next phase of this research will examine several different displays to determine the most likely candidate display system for supporting AAR activities. The details of this phase of research are presented in Appendix B.

References

Anderson, J. R. (1976). Language, memory, and thought. New Jersey: Erlbaum.

Anderson, N. H. (1986). A cognitive theory of judgment and decision. In B. Brehmer, H. Jungermann, P. Lourens, & G. Sevon (Eds.), New directions in research on decision making (pp. 63-104). Amsterdam: North-Holland.

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

Bessemer, D. (1991). Transfer of SIMNET training in the Armor Officer Basic Course (ARI Tech. Rep. 920). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A233 198)

Brehmer,B. (1976). Subjects' ability to find the parameters of functional rules in probabilistic inference tasks. Organizational Behavior and Human Performance, 17, 388-397.

Brehmer, B., & Joyce, C. R. B. (1988). Human judgment: The SJT view. New York: North-Holland.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch, & B. Lloyd (Eds.), Cognition and categorization. New York: Erlbaum.

Brunswik. D. (1952). The conceptual framework of psychology. In International encyclopedia of unified science, Chicago: University of Chicago Press.

Brunswik. E. (1956). Perception and the representative design of psychological experiments (2nd ed.). Berkeley: University of California Press.

Brunswik. E. (1957). Scope and aspects of the cognitive problem. In H. Gruber, K. Hammond, & R. Jessor (Eds.), Cognition: The Colorado symposium (pp. 5-31). Cambridge, MA: Harvard University Press.

Burnside, B. L. (1990). Assessing the capabilities of training simulations: A method and simulation Networking (SIMNET) application (Research Rep. 1565). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A226 354)

Carswell, C. M., & Wickens, C. E. (1987). Information integration and object display: An integration of task demands and display superiority. Ergonomics, 30, 511-527.

Casey, E. J., & Wickens, C. D. (1986). Visual display representation of multidimensional systems (Tech. Rep. CPL-86-2). Urbana-Champaign: University of Illinois, Psychophysiology Laboratory.

Coury, B. G., Boulette, M. D., & Smith, R. (1989). Effect of uncertainty and diagnosticity on classification of multidimensional data with integral and separable displays of system status. Human Factors, 31, 551-570.

Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. IEEE Transactions on Systems, Man and Cybernetics, SMC-7, 326-340.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley.

Edwards, W. (1980). Reflections on and criticism of a highly political multiattribute utility analysis. In C. L. Cobb & R. M. Thrall (Eds.), Mathematical frontiers of behavioral and policy sciences. Boulder, CO: Westview Press.

Edwards, W., & Newman, J. R. (1982). Multiattribute evaluation. Beverly Hills, CA: Sage.

Evens, J. St. B. T. (1990). Deductive reasoning in human information processing. In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale, & K. V. Wiles (Eds.), Modelling the mind (pp. 255-269). Oxford: Clarendon Press.

Foard, G., Kemler, A. T., & Nelson, L. (1989). Human performance modeling in display design (Tech. Note 3-88). Aberdeen Proving Ground, MD: Human Engineering Laboratory.

Garner, W. R. (1970). The stimulus in information processing. American Psychologist, 25, 350-358.

Garner, W. R. (1974). The processing of information and structure. New Jersey: Erlbaum.

Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), Perceptual organization. New Jersey: Erlbaum.

Genter, D., & Stevens, A. L. (1983). Mental models. New Jersey: Erlbaum.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin.

Gordon, I. (1990). Theories of visual perception. New York: Wiley.

Hammond, K. R. (1966). The psychology of Egon Brunswik. New York: Holt, Rinehart, and Winston.

Hammond, K. R. (1980). The integration of research in judgment and policy (Rep. No. 226). Center for Research on Judgment and Policy.

Hammond, K. R. (1981). Principles of organization in intuitive and analytical conditions (Rep. No. 231). Center for Research on Judgment and Policy.

Hammond, K. R., & Adelman, L. 91976). Science, values, and human judgment. Science, 194, 389-396.

Hammond, K. R., Hamm, R. M., & Grassia, J. (1986). Generalizing over conditions by combining the multitrait-multimethod matrix and the representative design of experiments. Psychological Bulletin, 100, 257-269.

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

Hammond, K. R., McClelland, G. H., & Mumpower, J. (1980). Human judgment and decision making: Theories, methods and procedures. New York: Praeger.

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

Johnson-Laird, P. N. (1975). Models of deduction. In R. Falmagne (Ed.), Reasoning: Representation and process. New Jersey: Erlbaum.

Johnson-Laird, P. N. (1990). Human thinking and mental models. In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale, & K. V. Wiles (Eds.), Modelling the mind (pp. 155-170). Oxford: Clarendon Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Keeney, R. L., & Rafia, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs.* New York: Wiley.

Kerins, J. W., Atwood, N. K., & Root, J. T. (1990). *Concept for a common performance measurement system for unit training at the National Training Center (NTC) and with simulation networking* (ARI Research Rep. 1574). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A230 129)

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science, 31,* 680-702.

Larkin, J. H. (1979). Information processing and science instruction. In J. Lockhead & J. Clement (Eds.), *Cognitive process instruction.* Philadelphia: Franklin Institute Press.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. D. (1980). Expert and novice performance in solving physics problems. *Science, 208,* 1335-1342.

Mahan, R. P. (1994). Stress induced strategy changes toward intuitive cognition: A cognitive continuum framework approach. *Human Performance, 7,* 82-130.

Meliza, L. L., Bessemer, D. W., & Hiller, J. H. (1994). Providing unit training feedback in the distributed interactive simulation environment. In R. Holz, J. Hiller, & H. McFann (Eds.), *Determinants of effective unit performance: Research on measuring and managing unit training readiness.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* (pp. 283-308). New York: Academic Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* New Jersey: Prentice-Hall.

Payne, J. (1982). Contingent decision behavior. *Psychological Bulletin, 92,* 382-402.

Pomerantz, D. (1981). Object perception. In M. Kubovy & J. R. Pomerantz (Eds.), Perceptual organization. Jew Jersey: Erlbaum.

Sanderson, P. M., Flach, J. M., Buttigieg, M. A., & Casey, E. J. (1989). Object displays do not always support better integration task performance. Human Factors, 31, 183-198.

Schank, R. C., & Abelson, R. P. (1977). Scripts, plans, goals and understanding. New Jersey: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Shlechter, T., Bessemer, B., & Kolosh, K. (1991). The effects of SIMNET role playing on the training of prospective platoon leaders (ARI Tech. Rep. 938). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A224 913)

Simon, H. A. (1957). Models of man: Social and rational. New York: Wiley.

Simon, H. A. (1976). Administrative behavior: A study of decision making processes in administrative organizations (3rd ed.). New York: Free Press.

Simon, H. A. (1978). Rationality as a process and product of thought. American Economic Review, 68, 1-16.

Slatter, P. E. (1987). Building expert systems: Cognitive emulation. Chichester: Ellis Horwood.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. Annual Review of Psychology, 28, 1-39.

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

Tversky, A. (1973). Elimination by aspects. Psychological Review, 79, 281-299.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

UCF Institute for Simulation and Training. (1993). Distributed interactive simulation development: Operational concept 2.3 (IST-TR-93-25). Orlando, FL: Author.

U.S. Army Training and Doctrine Command. (1989). Army training 2007 (TRADOC Pam 350-4) Fort Monroe, VA: Author.

U.S. Department of the Army Field Manual 100-5 (1986, May). Operations (Washington, DC: 97 pp.)

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

White, S., McMeel, K., & Gross, W. (1990). Unit performance measurement and training program design for networked simulators (PRT-4013-1-90). Woodland Hills, CA: Perceptronics.

Wickens, C. D. (1984). Engineering psychology and human performance. Glenview, IL: Scott, Foresman.

Wickens, C. D., & Andre, A. D. (1990). Proximity compatibility and information display: Effects of color, space, and objectness on information integration. Human Factors, 32, 61-78.

Wright, J. C., & Murphey, G. L. (1984). The utility of theories in intuitive statistics: the robustness of theory-based judgments. Journal of Experimental Psychology: General, 113, 301-322.

Appendix A

Multiattribute utility (MAU) models have been developed to support complex decision making. They have been traditionally used as a set of tools for describing, forecasting and explaining decisions (see Edwards, 1977; Keeney & Rafia, 1976). All MAU models share common features and requirements for application to decision making situations. In order to use these models one must typically:

1. Identify a set of decision options that represent the alternative courses of action that are open to the decision maker and that are under consideration.

2. Identify a set of dimensions or attributes that capture the characteristics of the decision options in that they represent what the decision maker sees as the important things to consider.

3. Measure the level of each attribute produced by each option by using a scale for each attribute in question.

4. Combine the attribute values for each decision option using a payoff function which reflects the combination of attribute weight and utility (importance) to generate a total utility value for each decision option.

The significance of MAU is that it allows for a logical decomposition of the decision problem into constituent parts that can be quantified in terms of the magnitude of both attribute levels and the importance (utility) of these attributes to decision makers. March and Simon (1958) note that MAU can overcome the limits on the rationality of decision makers. Here, rationality refers to the ability of people to act like statistical algorithms.

Decision Attributes

An essential component of MAU is associated with parameterizing the decision space. A common example given is that of deciding where to locate a business (see Edwards & Newman, 1982; Milkovich & Boudreau, 1991). The location selected for a given business may depend on a range of variables that provide, when numerically combined, a multidimensional index of location quality. Typically, many locations are screened for potential consideration. Many of these locations are eliminated on the grounds that they violate various hard criteria, such as zoning laws, excessive public resistance, inability to satisfy legal requirements, as well as basic cost constraints. Once, a set of options are generated for serious consideration, the next step is to identify the relevant attributes that define location quality.

There are numerous approaches used to define the attributes in a decision problem. Very often one simply interviews subject matter experts in an effort to reveal the major dimensions of a given problem. Once a collection of problem dimensions have been elicited from the experts, one then selects a reasonable number of dimensions that are common across expert judges.

Other, statistical approaches are also used. For example, a detailed questionnaire can be constructed which contains many questions that, in theory, measure the domain associated with the decision problem. However, instead of asking the subject matter experts to verbally indicate what dimensions they feel are most relevant to the decision problem, the dimensions are derived using correlational techniques that reveal latent structure in the experts' ratings of individual questions. Ratings on effectiveness, efficiency, quantity, quality, cost and other scalable variables are then evaluated to determine which questions tend to cluster with each other. These groups of questions form factors that are taken to define the decision problem.

While the statistical approach is an elegant method for identify the attributes in a complex decision problem, it is only as good as the original set of questions on which it is based. Thus, while the internal validity of the approach is high, the external validity may be low. That is, the questions selected by the researcher may not be representative of the decision problem. On the other hand the interview approach guards against the external validity issue by eliciting the attributes directly from the experts. The drawback here is that the internal validity may be compromised by asking the experts to produce the attributes.

Value Measurement

There are several numerical estimation procedures that are used in quantifying decision attributes. Typically, stimuli used to elicit the quantities for MAU analysis can be riskless in nature, or presented to the experts as gambling scenarios. Riskless outcomes are used when one is interested exclusively in value measurement, while gambles are used in the more complex utility measurements where combining value with utility in a cost-benefit pay-off framework is used.

Direct ratings, curve drawing, ratio estimation and category estimation are versions of numerical estimation in which experts are presented with some anchored scale and asked to rate or otherwise estimate the importance of the stimulus relative the anchors. These approaches are all used in riskless outcomes. Numerical estimation procedures such as bisection and "difference standard sequence" are used in gambles (see Boudreau, 1991). Both bisection and difference standard sequence rely on combining probabilities with utilities in determining value.

However, the most widely used estimation procedure is direct rating. As a version of a frequently cited example, consider a new Ph.D just leaving graduate school in search of an academic position. One variable that is important to this individual is the location of the institution. Location is associated with proximity to friends and family, climate, culture, entertainment potential, as well as the proximity to locations used for scientific conferences. This individual clearly has preferences for living in different locations in the country and further, these preferences can be explicitly made.

In this example, adapted from Edwards (1982), we ignore many of the attributes that would clearly play a role in the candidate's selection of a particular institutional location. Here the location alternatives are:

1. Ann Arbor
2. Boston
3. Chicago
4. Los Angeles
5. Atlanta

The interviewer first asked the decision maker to select both the worst and best city on the basis of all the characteristics that may make a city a good or poor choice (at the same time ignoring other variables of the job selection problem like the prestige of the institution and the salary offered). The responses generated from the candidate are:

Best: Atlanta
Worst: Los Angeles

Its important to have the decision maker articulate why it is he/she selected those cities as the best and worst of the set of cities considered. It may be that Atlanta has mild climate, is relatively less expensive to live in than the other cities and is smaller in size. On the other hand, Los Angeles is viewed as perhaps being too big, which means traffic and noise is a negative attribute to living in Los Angeles and too expensive. These are qualitative values are thought of as defining an underlying value scale for the decision maker.

Once the other alternatives are ranked between the extreme selections, the decision maker is given the chance to modify the rankings. That is, it is important to let the decision maker consider the set of rankings in an effort to fine tune the rank ordering of the cities. As Edwards (1984) notes, one must be cautious in forcing a decision maker to be consistent in the preference process because the meaning of the underlying value scale is still being defined in some sense by the decision maker. Thus, allowing the decision maker to verbalize the preference rankings and consider altering the order of the options is tantamount to generating more precise and stable value measures. Let us assume that the final ordering of the cities from best to worse is as follows:

1. Atlanta
2. Ann Arbor
3. Chicago
4. Boston
5. Los Angeles

The second series of steps is necessary for assigning numerical quantities to this qualitative information. A very simple approach that has been frequently used is assigning a "0" as the anchor for the worst city and a "100" as the anchor for the best city. Essentially, the upper endpoint is arbitrary. The lower may be arbitrary unless ratios are to be formed from the ratings, in which case a 0 must be used. The remaining cities are rated in between.

Consistency is established by allowing judges to modify their rankings. One approach is tho have judges rank items (cities in this case) that are not in the original list. Essentially, this is a form of cognitive feedback which provides a way to enrich the original scale (i.e., revalidate the judges' conception of the scale) by asking he/she to use the original cities as anchors for further rankings. This process is typically followed by some sort of revising by judges on the adequacy of the ranking process.

The scale construction process stops when the judges are comfortable with their assessments. While, the assessments do not need to be overly precise, it is nevertheless important the judges feel comfortable with the relative spacings and that the scale makes sense to the judges. The basic idea behind one kind of direct ratings is seen in the following array of cities where both the best city (Atlanta) and the worst city (Los Angeles) sever as the anchors for the quality scale.

1. Atlanta (arbitrary assignment)
2. Ann Arbor (judgment, relative arbitrary assignment)
3. Chicago (judgment, relative arbitrary assignment)
4. Boston (judgment, relative arbitrary assignment)
5. Los Angeles (arbitrary assignment)

Defining Multiattributes

It is often useful to decompose the scale into more detailed components that can be scaled themselves. For example, the preference for a particular academic job in the above example simply collapsed the various dimensions on which a chosen job was based into an overall quality scale. If a judge has trouble placing items on a given scale because the meaning of the scale shifts from one inter-item comparison to the next, it means that the problem is multidimensional and that multiple scales are needed to fully address the decision making problem. That is, the problem has multiple value dimensions. The assessment process requires individually addressing component attributes before a quality decision is made. Thus, scaling such attributes as proximity to family, climate and so on, prior to arriving at a decision regarding the overall quality of a given location would be necessary.

A Multiattribute Example For Job Preference

Structuring. A thorough analysis of the job preference problem would surely call for a more detailed assessment of the attributes that compose the quality metric in the above example. Accepting a job offered from an institution located in a particular city would depend on a number variables. Even after a prescreening process that eliminates jobs on the basis of hard criteria (lack of relevant degree programs, budgeted time for research, and so on) a number of options still remain.

The analytic, or top down approach to building value trees has been detailed by Keeny and Raffia (1976). It naturally begins with having the judges define the attributes that are reacted to a particular problem area. In the present case, the job candidate is interviewed in an effort to identify the relevant dimensions that are associated with the job selection problem. These interviews tend to be partially unstructured. That is, a set of standard items or questions may be asked of the judge (or judges) on the basis of a general problem analysis, but the analyst also is knowledgeable enough about the interview method to ask the judge for more detailed responses if he or she feels that this will facilitate structuring the problem and identifying all relevant attributes.

Quantification

Valuation of the attributes starts from an explication of the judges most general values relevant to the problem. Thus the judge is asked to state the general values that are associated with the primary dimensions of the problem that were elicited from the judge in the structuring process. The analyst asked the judge to explain what the initial value categories actually mean by using more specific dimensions. This process is aided by probing the judge on his/her suggestions. It is important to determine if (a) the list of value dimensions is exhaustive and (b) if some of the value dimensions are redundant. Ideally, the relationship between the lower level dimensions and higher level dimensions are hierarchical. Edwards (1982) suggests that the relationship between the dimensions should meet the basic criteria of being an exhaustive and nonredundant set of explanatory value dimensions, even though this ideal is unlikely to be met. As an example, consider the adaptation of Edwards' (1982) value tree generated for the job selection problem above. Here the primary dimensions on the left side of Figure A-1 have been disaggregated into lower layers of attributes.

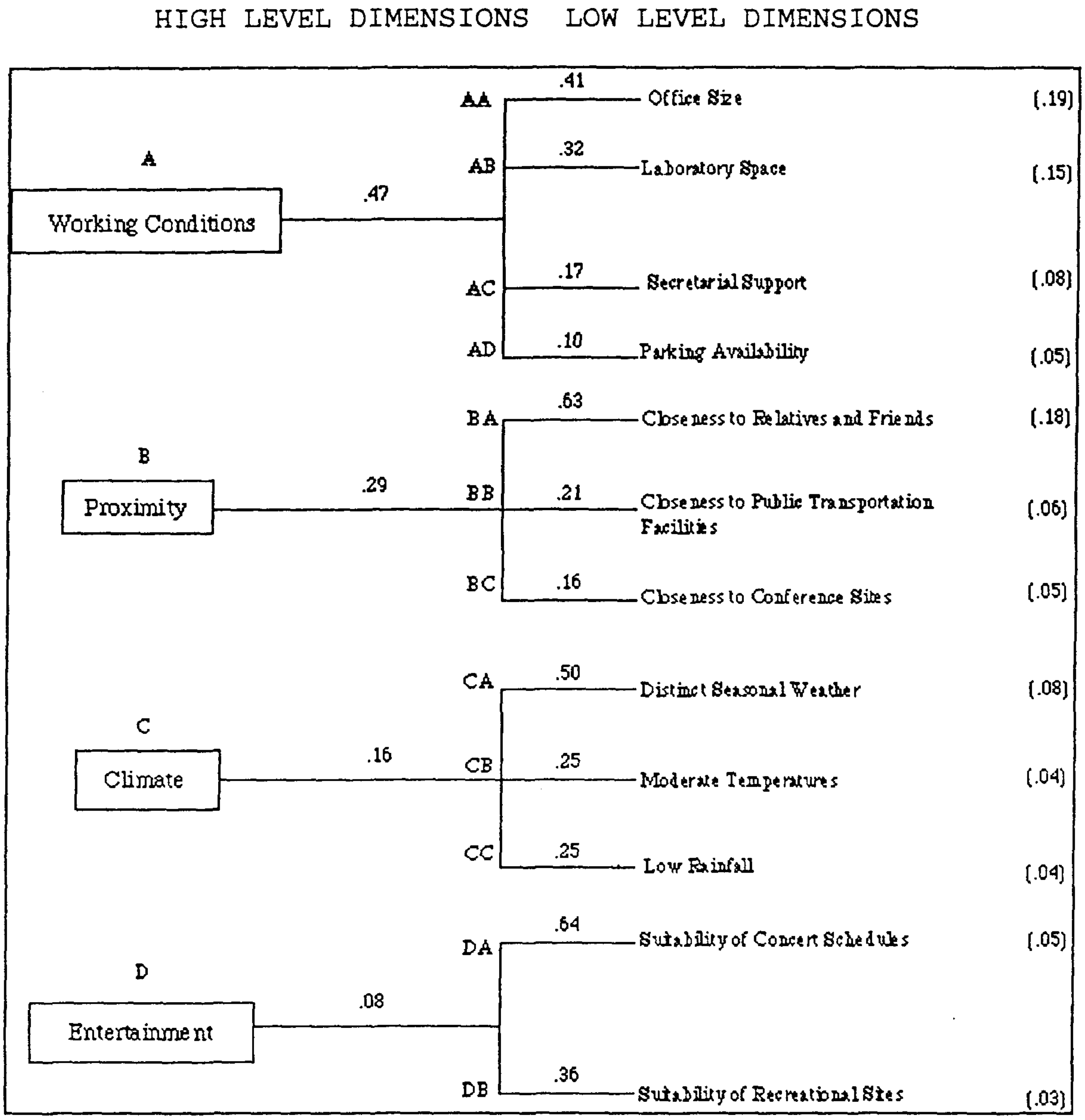


Figure A-1. Value tree for job preference.

The point at which the analyst should stop disaggregating is usually when the dimensions are at the lowest possible level of measurement, or judging. This is the point at which operationalizing the dimensions becomes exceedingly difficult. Sometimes the quantification process is easy, such as when the dimension is a monetary one where cost can readily be specifiable. Often times, further disagreggating high level dimensions does not lead to quantification.

Figure A-1 is an example of a job preference value tree that expresses the value relevant objectives and attributes for comparing alternate job locations. The numbers that are assigned to the high level dimensions (branches) and the low level dimensions (twigs), are weights that reflect the relative importance of the attributes for the overall evaluation. In this particular case, the weights were elicited in a two stage process. The judge was first asked to consider the lower level dimensions. Each set of lower level dimensions were rank-ordered in terms of their importance in reflecting the overall primary dimensions. For example, office size (AA) was viewed by the judge as being the most important attribute in the "working conditions" dimension. Next, the judge distributes 100 points across the lower level dimensions. For example, office size is given a rating of 41, laboratory space, 32, secretarial support, 17 and parking availability, 5. These numbers are then normalized to add to 1.0. The same procedure is followed for the attributes under the remaining branches. Finally, the judge is asked to follow the same procedure for ranking and weighing the primary dimensions.

Given the numerical assignments in Figure A-1, weights on the attribute levels are easily found by multiplying through the tree. For example, the weight for twig AA (office size) is determined by multiplying the normalized weight for A (.47) by the normalized weight for AA (.41) to yield .19. This is conducted through the value tree for all attributes.

Knowing the structure and value of the dimensions that define job preference, it is now possible to generate utility values by combining ratings of particular cities across the attributes with the attribute weights derived from the value tree. The ratings are generated using the direct rating approach discussed above. Each city is located on the relevant attribute continuum on the basis "how much" the city possesses the attribute in question. Typically, judgmental attributes range between 0 and 100. Once the cities are all rated across all the attributes, aggregate values are generated using a simple weighted average model. This model defines the overall value of city location X $(X = 1, 2, \ldots\ldots 5)$ as $V(X) = \Sigma W_i V_i(X_i)$ where $V_i(X_i)$ is the value of city X on the ith attribute, W_i the importance weight of the ith attribute, and V the value of X.

The results of the job preference example is summarized in Table A-1. The direct ratings across attributes are listed under each city being judged. The direct ratings are followed by the aggregate value which is the weighted sum given in equation 1.0 above. Cost of living estimated from Median housing prices is used to make the final comparison.

Table A-1

Values and Median Housing Costs For Five Cities

	Job location				
Attribute List	Atlanta	Ann Arbor	Chicago	Boston	Los Angeles
AA	90	60	20	10	30
AB	50	90	35	25	40
AC	25	30	25	25	10
AD	60	75	40	10	40
BA	90	70	70	30	10
BB	95	65	95	30	85
BC	95	25	90	40	80
CA	40	70	80	80	50
CB	90	30	10	30	95
CC	60	40	40	55	90
DA	95	80	95	95	95
DB	90	100	70	60	90
Values	73	64	51	34	44
Costs ($)	137K	110K	141K	179K	247K

The final comparison is made by plotting housing cost against aggregate value. Figure A-2 shows the results of the plot. Benefits are plotted along the abscissa (horizontal axis); more is better than less. Cost is plotted along the ordinate (vertical axis); less is better than more. Points that are low and to the right are preferred over those that are high and to the left.

Appendix B

Information Display Study

The manner in which information is displayed is an important link in the design of the user interface. The following study serves to examine the utility of several information displays in supporting multidimensional judgments of training performance. A guiding objective for this study is that any reasonable display design will support a rapid form of information extraction and integration, which is viewed as an essential element of the after action review process.

Background

The present report, in part, identifies a rapidly growing research literature addressing the problem of choosing the optimal information representation to use for various judgment and decision-making tasks. Although, graphic displays that induce holistic processing have been shown to be superior in some military command and control settings, as well as integration of diagnosticity and reliability information in intelligence reports (see, Nawrocki, 1973, Wickens & Scott, 1983), the efficiency of these types of displays has yet to be fully determined. When different versions of object (configural) displays have been compared to more traditional displays such as bar graph and tabular formats, results have been conflicting. For example, although Carswell and Wickens (1987) found superior task performance with object oriented configural displays over bar graph displays (so called separable displays) in complex multi-information tasks, others have found that separable displays could support task performance at levels equal to that of configural displays (see Sanderson et al, 1989). While, some of the results of these and other studies reflect differences in performance measures, the studies also serve to demonstrate a fundamental lack of knowledge on how information presented in different graphic forms is used by people in the process of integrating the various components of a display in task performance. The research proposed here examines the relative merits of a three display types in a task requiring the execution of complex judgment.

<u>Overview of Research Design</u>. The task to be used in this study is designed to require subjects to integrate multiple items of information, having differing importance, into a series of judgments for which true values (criterion scores) exist. Specifically, the task requires subjects to make judgments regarding the Move, Shoot, and Communicate dimensions described in this report. A large set of training events will be presented to each subject in this study. Each event will be described by an array of information cues (subtasks) that serve to reflect performance on the above three dimensions (primary tasks). These subtasks will be those identified by SIMNET instructors in the

multiattribute utility analysis described in the present report and listed in Figure 2. Each subject will make four judgments for each of the many training events to be judged. These judgments will call for assessing (a) the quality of a platoon's movement, (b) the quality of engagement characteristics (i.e., Shoot), (c) the quality the platoon's communication, and (d) an overall judgment of total performance quality of the training event. The cues (subtasks) will be presented in one of three display formats: configural triangular display, bar graph display and a numeric (tabular) display. Subtask values will be generated using examples taken from UPAS time-line data-sets and thus will be realistic of the information in SIMNET events.

General Hypotheses. The hypotheses in the proposed study center around a group of related key issues. Because the AAR task process is (a) multidimensional in nature, (b) uncertain in the sense that AAR task dimensions do not perfectly predict the quality of training performance, and (c) requires a relatively rapid form of information processing by trainers in order to ensure the timely feedback of training performance, several specific and directional hypotheses regarding the utility of displays can be made. First, the complexity of the AAR process reflected in the above task dimensions means that the optimal mode of information processing should be intuitive in nature. A holistic scheme for extracting and integrating the relevant multidimensional information, as opposed to a scheme that favors the analysis of independent task dimensions, should be the preferred approach in quality assessments by trainers. Therefore, a configural display that is designed to support integrated judgments should be superior to its separable counterpart. Similarly, analytical assessments of the AAR information from a separable display format will lead to poorer judgments of training quality.

The criteria for quality of judgments will be made using lens model indices described below. In addition, a primary index of intuitive processing is that it is a very fast form of cognition. When trainers are proceeding in an intuitive fashion, one should see more judgments per unit time. Conversely, when analytical processing is taking place, the rate of judgments should be significantly reduced. Thus, intuitive judgments should result in more accurate as well as more rapid assessments of training performance quality.

Configural-Separable Continuum. The displays described below can be scaled (monotonically) according to their tendency to induce intuitive processing. For example, the triangle display is configural, and thus should support intuitive processing. The tabular display is separable and should induce an analytical approach to judgment. Finally, the barograph display can be considered an intermediate display in the sense that both intuitive and analytical processing may be supported. That is, while the independent vertical bars induce an analytical process, the contour associated with the top of the bars may support a configural type of judgment. The triangle display should be an optimal way of presenting AAR information while the tabular display should be the least optimal. The performance characteristics on the bar graph display should be found laying between the triangle and tabular displays.

Cue-Criterion Characteristics. Four mathematical rules will define the four criterion variables to be judged in this study. Three of the rules will define the relationship among the three primary tasks and their respective subtask arrays. The fourth rule will specify the mathematical relationship among the total performance criterion and the three primary task variables. These rules will serve to produce the true criterion scores that subjects will attempt to judge. The rules producing the true scores will be linear.

The following rules incorporate both the independent and conditional properties of the value tree illustrated in Figure 4. These rules define the relationships among the primary tasks and the subtask arrays using coefficients derived from the MAU analysis of training experts' judgments in the present study. Movement is defined by the subtasks expressed in the value tree using the weights generated by instructors' ratings of each subtasks diagnostic value of the primary task:

Movement = b_1(Navigation) + b_2(Speed) + b_3(Timing) + b_4(Security/Density)

where the weights represent the standardized versions of the validity coefficients described in the value tree of Figure 2. Here, the subtasks- Security and Density, will be combined for the purpose of yielding a more discriminable information source (an information source that, in theory, can produce identifiable changes in Movement criterion true scores).

The primary tasks- Shoot and Communicate are defined in the same manner- by their respective subtasks weighted by standardized validity coefficients.

Shoot = b_1(Fire Distribution) + b_2(Trigger Line) + b_3(Kill Zone) + b_4(Concealment)

Communicate = b_1(Timely) + b_2(Accurate) +b_3 (Comprehensive) + b4 (Succinct)

Finally, the total performance true scores is a linear function of the primary tasks and their standardized coefficients from the value tree in Figure 2.

Total Performance = b_1(Movement) + b_2(Shoot) + b_3(Communicate)

Judgment Process. The subtasks described in Table 2 will represent the information arrays for judgments of training performance. Each judgment of Move, Shoot, and Communicate will themselves be integrated to produce the final judgment of quality of training performance (total performance). Thus, a two stage judgment process will be necessary: (a) making judgments of the primary task values on the basis of subtask information, and (b) using the these judgments as input into the final judgment on performance quality.

It is possible, given the task relationships presented in Figure 2, to make the total criterion judgments on the basis of the subtasks alone (i.e., integrating subtasks across primary tasks). However, this would seemingly be more difficult because the diagnostic value of subtasks for predicting their respective primary tasks is much different than the diagnostic value of the subtasks for predicting the total performance. In theory, the most efficient approach for judgments of total performance would be to use the judgments of primary tasks as input to the final quality assessment, rather than organizing the subtasks in two very different ways; one for primary tasks judgments and one for total performance judgments. However, this can be verified empirically by a detailed examination of the judgment data.

Information Displays

The subtasks will be presented according to three different display formats. These displays were chosen on the basis that they contain display elements that vary on their configural properties from high configurality to low configurality. In addition, these displays represent a wide range of formatting properties that are being used today in display systems (see Wickens & Andre, 1990 for review).

Triangular Display. The triangular display is intended to represent the data in two ways. First, at each apex of the triangle is a scaled polygon which presents the subtasks for a primary dimension on four orthogonal radials. Each radial is labeled at its endpoint with its associated subtask. The resulting points, when connected, produce a four sided polygon object. The shape of the polygon indicates the values on each subtask. Integrating subtask values in the judgment of primary tasks is, in theory, accomplished by attending to the area of the polygon. If all four subtasks are at their highest value (i.e. performance on each subtask is highest), the area of the square polygon will be at a maximum value. This would in turn indicate the highest score in the primary task criterion.

A second feature of the display is the reticle symbol that is present inside the triangle itself. The placement of the reticle supports the integration of the primary tasks in judgments of total performance. When the reticle is in the center of the triangle, this indicates that all subtasks for all three primary tasks are at their highest value. In this configuration the total criterion score would also be at its highest. Subtask polygons that depart from their highest values attract the reticle. Thus, the more a polygon deviates from the largest square, the closer the reticle moves toward the polygon and the lower the total performance value becomes.

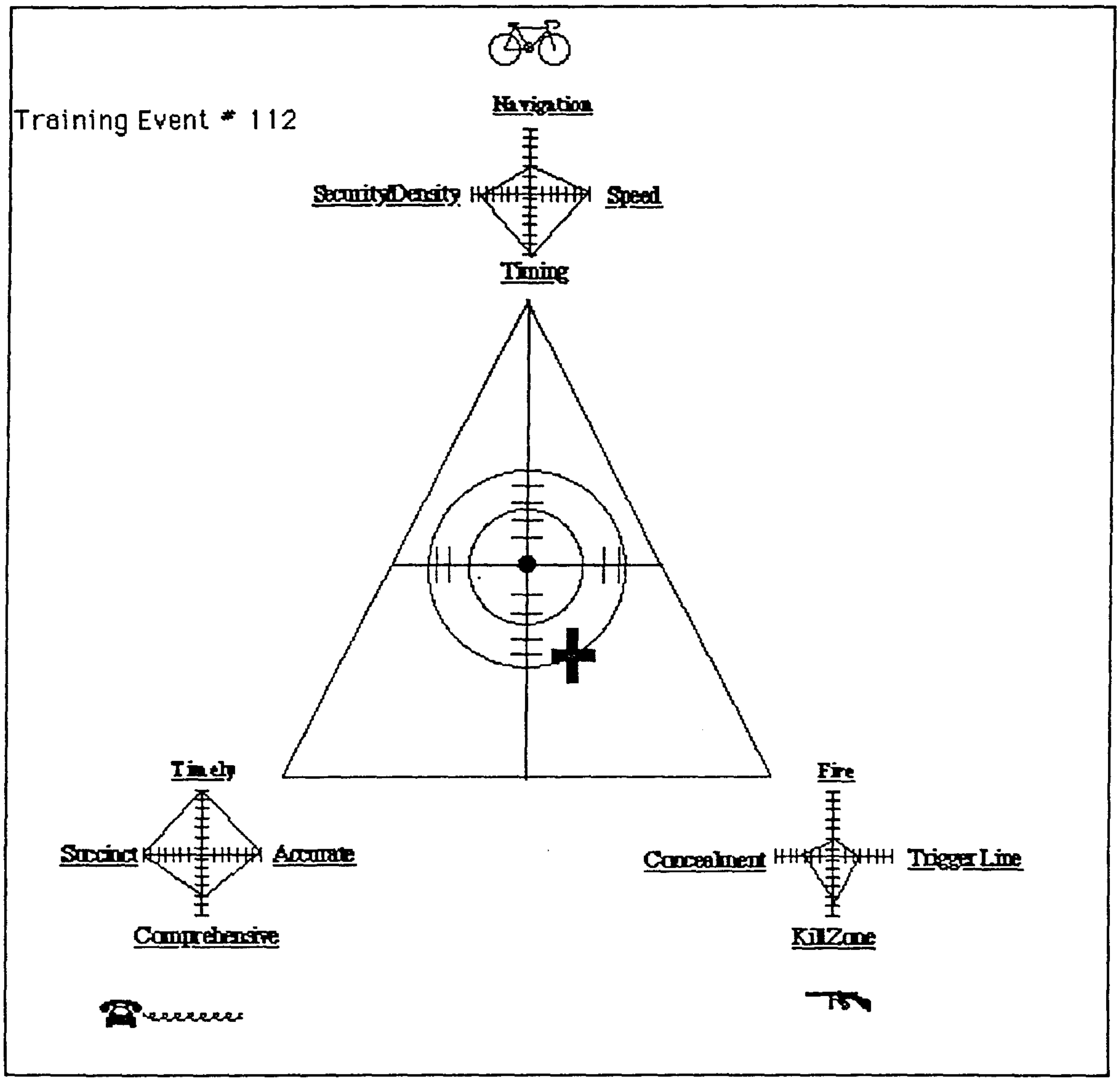


Figure B-1. The triangular representation of the primary tasks as a set of objects- bicycle (Move), gun (Shoot), telephone (Communicate), along with their subtask polygon information arrays.

Bar Graph Display. The bar graph display represents each subtask value according to the height of a shaded bar. Each bar is labeled with the name of the subtask it represents. The bar height indicates the value on the standardized performance dimension (see below) for a given subtask. These values can range from 0 to 100.

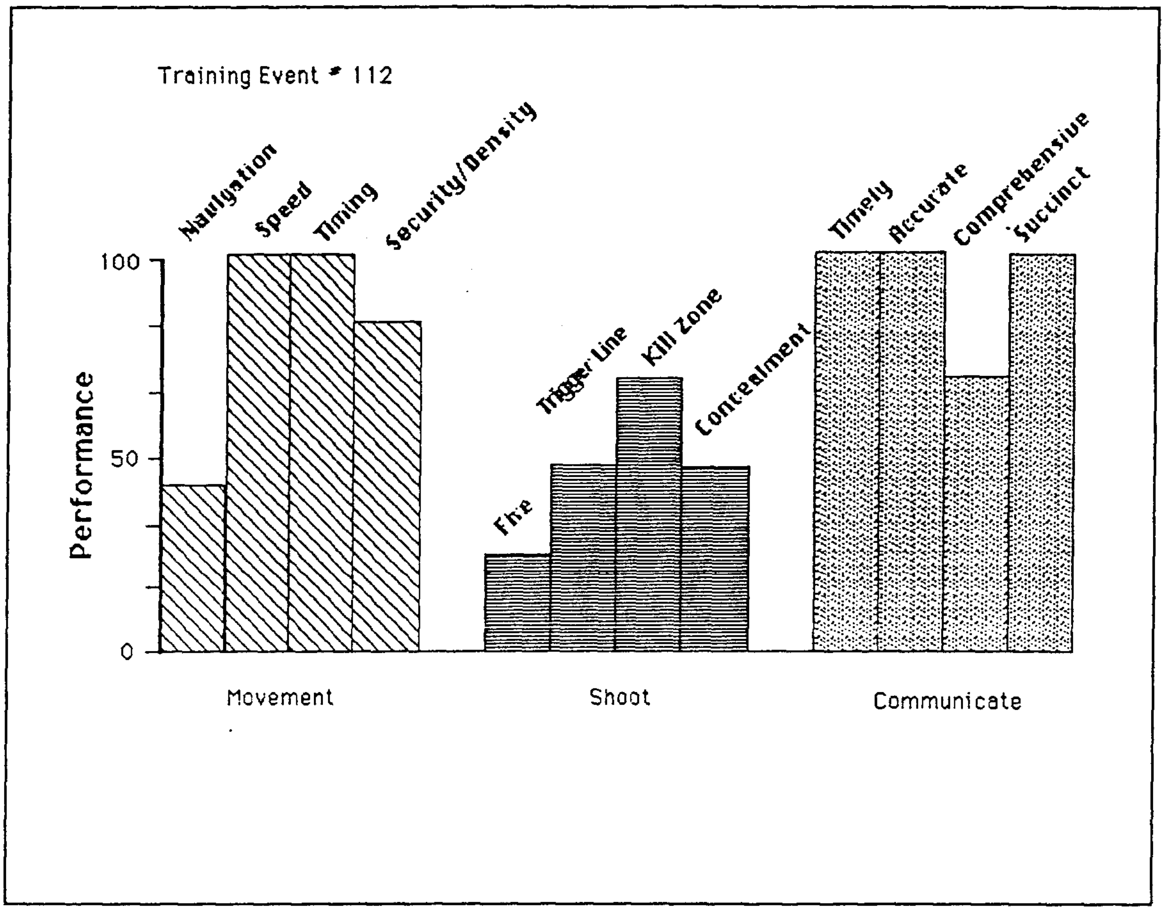
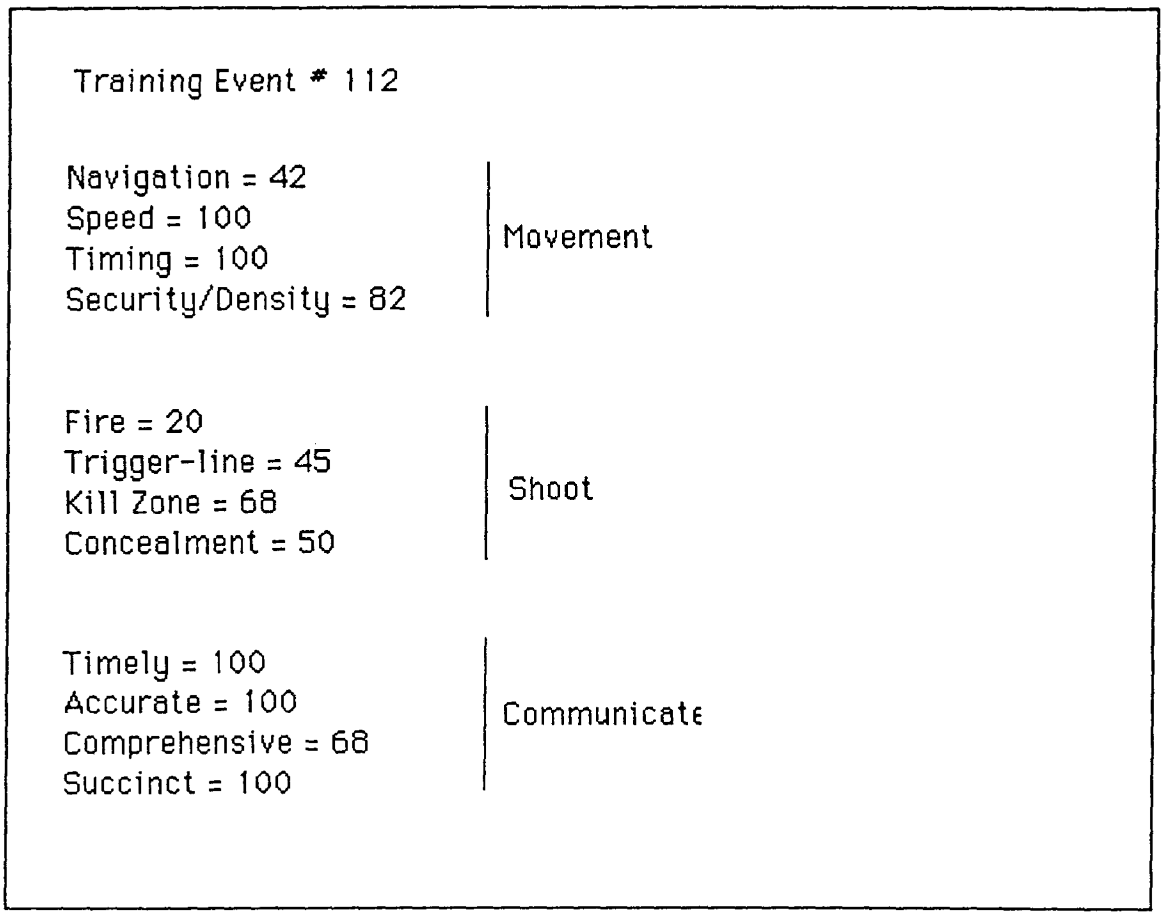


Figure B-2. Bar graph display presenting values of subtasks as the height of individual bars.

In the bar graph, the ordinate has indexing marks with the general scale label being performance. The bars are spatially arranged so that three groups of bars define their respective primary tasks.

Tabular Display. The tabular display presents training event information by listing the numeric values associated with each subtask on a standardized performance dimension which again ranges from 0 to 100. One column of subtask values will be displayed for each training event, and it will define each of the primary tasks.



Training Event # 112

Navigation = 42
Speed = 100
Timing = 100
Security/Density = 82

Movement

Fire = 20
Trigger-line = 45
Kill Zone = 68
Concealment = 50

Shoot

Timely = 100
Accurate = 100
Comprehensive = 68
Succinct = 100

Communicate

Figure B-3. Tabular display of subtask information represented numerically.

In all of the above displays the subtask information values represent scales having differing metrics. Therefore, the raw subtask values will be first converted to standard scores before being displayed. Thus, the actual representation of the subtask values will always be in terms of the distributional properties of the subtask rather than in terms of the subtask's unique metric.

Subjects. Fifteen SIMNET instructors will be used for the study. They will be randomly assigned to one of three display groups generating five subjects per group.

Study Characteristics. There are two general approaches typically used in conducting these kinds of judgment studies. One approach is to treat the study as a judgment skill acquisition experiment. Here the idea is to examine the process whereby subjects discover a policy for weighing the subtasks by providing them with outcome feedback in the form of the true criterion values after each judgment is made. Learning to discover an efficient way to integrate the information (produce the judgment policy) is achieved by providing outcome feedback in the form of the true criterion scores after every criterion judgment. After many judgments and outcome feedback occur, the subjects develop a weighing system that reflects the diagnostic value of each subtask in predicting the primary task scores (a policy that best matches the mathematical rules generating the true scores). Similarly, the subjects learn the weights associated with the primary tasks in predicting the total performance criterion. Subjects will not be told these weights. The focus of such an experiment is to examine the process by which multi-task probability learning emerges.

A second approach would be to train the subjects to some level of judgment skill achievement using outcome feedback, remove this outcome, and then examine the nature of judgment decay. The former approach would support hypotheses regarding the learning process while the latter approach, decay functions (forgetting, but see Mahan, 1994). The multi-task probability learning approach seems the most reasonable given some of the logistical constraints of such a study. Here, a primary constraint is scheduling trainers for participation. Configuring the study as a learning experiment would mean that data collection is a one-time event. In contrast, using the alternate approach would mean first training the subjects to some criterion level of performance and then examining them under experimental conditions. However, either approach, in theory, should illuminate the various strengths and weaknesses of the three displays. The first approach is suited for training a naive subject on how to use information, the second approach for evaluating a trained subject (i.e., expert) performing without immediate outcome feedback as to the quality of his decisions.

Procedure. Each subject will be presented with his respective display format. Each subject will be shown many different subtask value variations of that display that characterizes many training events. For each display, the subject will be asked to make four judgments - one for each of the three primary dimension and one for total performance. The subject will be free to alter the order in which he makes the judgments at any time. That is, the subject will not be constrained to judge primary tasks first, and total event performance last. The subject can reverse that order, or execute any sequence of judgments he chooses. Once the judgments are made, the subject will receive immediate outcome feedback as to the true criterion scores for the primary tasks and total performance criterion from the mathematical rules which define the relationships among the subtasks, the primary tasks and the total criterion. The outcome feedback will theoretically teach each subject to learn a policy for integrating the display information into the judgments that best fits the mathematical rules producing the true criterion scores. The subjects will make judgments until their performance reaches an asymptote. Performance will be evaluated by correlating subject's judgments with true criterion scores.

Analysis. Both an individual differences and an average group analysis approach will be used to examine a variety of performance indices. For example, the measures of achievement, consistency, and matching can be computed for each subject on subsets of training judgments. These measures and their relationship to the judgment problem are defined in Figure 1 of the lens model.

Lens Model

The lens model displayed in Figure 1 can be mathematically characterized by defining the relationship among the components of the model and judgment task performance.

The correlational performance an individual achieves (i.e., achievement index) r_a, is a function of four distinct components: (a) the linear multiple correlation between the cue values and the criterion, R_e, (environmental predictability), which indexes the uppermost predictability of the judgment task; (b) the linear multiple correlation between the cue values and an individual's judgments of the criterion, R_s, (consistency index), which represents the ability of the subjects to control the execution of their knowledge regarding the judgment task; (c) the extent to which the linear model of the individual judge correlates with the linear model of the criterion, G, (matching index), which measures overall task knowledge; and (d) the extent to which the nonlinear residual variance in the model of the individual correlates with the nonlinear residual variance in the model of the criterion, designated C. Achievement (r_a) provides an index of overall judgment accuracy.

Over the course of the judgment process, a lens model analysis at different stages of policy formation for integrating the subtask information can be conducted. The outcome of these computations can serve as input to a mixed analysis of variance design where the within subject learning functions can be assessed for their pooled structural characteristics (i.e., linear and non linear profiles), and the between groups differences for significant display format effects. This analysis would also support evaluations on the interaction between these dimensions as represented in the J X D matrix in Figure B-4.

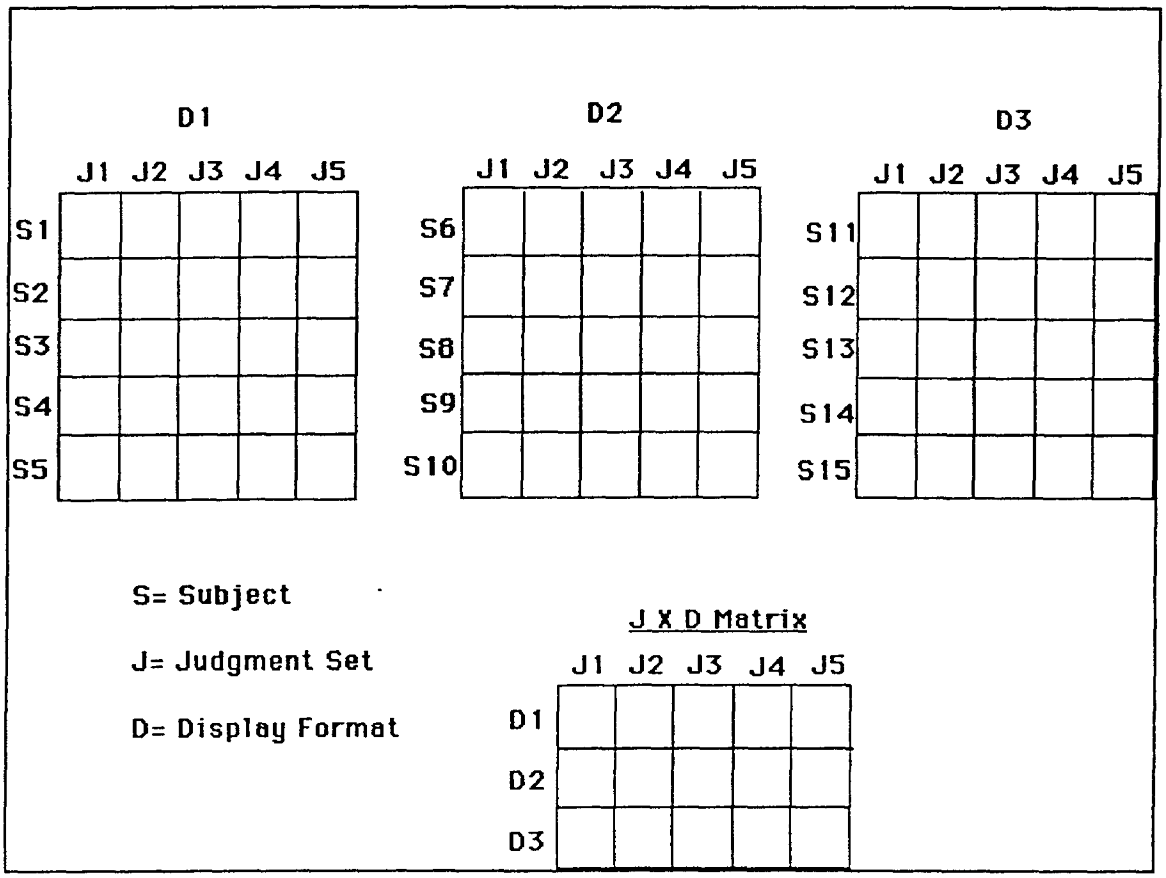


Figure B4. D X (J X S) mixed analysis of variance design matrix where D represents the between groups display factor, and J the within groups judgment set factor.

As an example, consider the design matrix presented in Figure B-4. Here, J1 through J5 are meant to represent, say, five 40 judgment sets by each subject. These measures can be evaluated for trend, within subject, by fitting single degree of freedom polynomials to the data. In addition, the design matrix shows how an analysis of the display <u>format effects</u> (D1 through D3) can proceed by collapsing across the judgment set (J) dimension. Finally, the interaction between judgments and display can be examined via the application of standard analysis of variance procedures.

In addition to the basic lens model indices, a host of other performance measures can be examined using this design strategy. For example, reaction time (time to make judgments) can be evaluated. This can be used as a component process measure of cognitive functioning. For example, a display format that has processing intense requirements (very labor intensive) should also generate slower judgments. In contrast, a display that supports easy information extraction and integration should show more judgments per unit time. In addition one can evaluate various parameters of the learning functions. For example, the examination of the slope of the learning function may prove fruitful (i.e., the steeper the slope, the more quickly the subject's learned to perform the task.

A variety of subjective measures can also shed light on the utility of the various displays in supporting the judgment process. For example, having subjects rate the diagnostic value of subtasks in predicting criterion values would allow one to create a self awareness measure of the weights used in the policy for subtask information integration. The probabilistic nature of the task means that, to a certain extent, the actual details of a subject's policy may remain implicit. Correlating these ratings with other performance indices may help reveal certain display utility patterns in the data.

Summary of Proposed Research. The overriding principle guiding the research outlined above concerns the relationship between task demands and the manner in which a task is represented. One can view the Cognitive Continuum Theory (CCT) as a possible framework useful in guiding the selection of a display that is based on the task's location on the continuum. For example, tasks requiring information integration often are best performed at an intuitive processing level. Consequently, intuitive processing should be supported by configural-like display formats because they represent integrated task elements using closed contours which have been shown to induce holistic (patterned) judgment. The linkage point between tasks, information displays, and cognition is that there must be a congruence (i.e., a direct mapping) between the task and the information processing mode of the decision maker. However, this also implies that the manner in which a task is represented and displayed must preserve the structural characteristics of the task at hand. Thus, there must be congruence between the task and the way it is displayed. CCT asserts that a complex task requiring information integration is best performed by responding to the pattern in the data and this form of processing tends to be holistic in nature. Holistic processing is often best supported by mapping the task dimensions into a closed geometric form (a configural graphical object). This should serve to preserve the congruence among the task, display, and processing characteristics. In contrast, mapping the same task into independent and orthogonal graphical representations (e.g., bar graph display), should lead to low congruence and poorer judgment performance for the integrated task.

The proposed study is aimed at generating practical information that can be useful in the design of display systems currently under development by the Army for distributed networked simulation. The practical significance of the project for the Army lies in being able to assist in the development of a display that is useful for SIMNET instructors who provide critical training feedback to soldiers during the after action review. Secondly, the proposed study is being designed to answer basic questions regarding relationships between task structure and cognitive functioning. The application of the CCT to the problem principles for the design of information displays. These principles will allow the construction of task taxonomies which will specify the most optimal display formats needed to support a wide range of tasks for a wide range of users of military systems.