FOREWORD

This research was performed in the Laboratory of Aviation Psychology, The Ohio State University, Columbus, Ohio, under Contract No. AF 33(657)-10763 during the period of 15 March 1963 to 31 January 1964. This research was prepared under the provisions of Project No. 7184, "Human Performance in Advanced Systems," and Task No. 718403, "Man-Machine Systems Research." Dr. George E. Briggs was the principal investigator representing The Ohio State University. The contract was monitored by Captain Karl L. Wiegand, Chief, Systems Research Branch, Engineering Psychology Division, Behavioral Sciences Laboratory.

The authors are pleased to acknowledge the invaluable assistance rendered by several members of the laboratory staff. Messrs. Edwin R. Lassettre and Lonnie D. Whitehead assumed a major share of the responsibility for designing and preparing the large amount of computer programming required in this research project. These two individuals must be accorded a primary share of the credit for the enormous programming effort which was necessary in order to generate the complex stimulus environment described in the report. Miss Carolyn Black assisted in the collection and analysis of data. Miss Janis Frye and Miss Barbara Lindig provided invaluable editorial assistance.

ABSTRACT

The first of a series of experiments investigating the value of automated hypothesis-evaluation aids in multiman-machine systems devoted to assessing or diagnosing threat is described. In the experiment, an eight-man team evaluated threats posed by a hypothetical Aggressor. The team made these evaluations on the basis of intelligence information gathered during simulated reconnaissance overflights of Aggressor's territory. IBM 1401 and 7090 computer facilities generated the highly complex, real-time stimulus environment (data base) which is described in detail. The primary output of the threat-evaluation team was the commanding officer's posterior probabilities estimates as to Aggressor's most likely hostile strategies. During half of the experimental trials, the commander had access to computer-produced posterior probabilities based upon a modification of the Bayes Theorem. The major experimental issue was whether or not these would aid the commander in his hypothesis evaluation. Also investigated was the effect of data-processing load upon system operation. Although some improvement in the posterior probabilities estimates resulted from the commander's having access to the hypothesis-evaluation aid and this improvement became more pronounced as system load increased, the main-order effect of access to the aid was not found to be statistically significant. Throughout the entire experiment, solutions of posterior probabilities based upon the Modified Bayes Theorem were compared with the commander's estimates. The Bayesian solutions were significantly superior to the commander's.

PUBLICATION REVIEW

This technical documentary report is approved.

Walter F. Grether

WALTER F. GRETHER
Technical Director
Behavioral Sciences Laboratory

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

I. HYPOTHESIS SELECTION AND THREAT EVALUATION

There are many situations in which human beings are called upon to select or formulate hypotheses in explanation for the occurrence of a certain set of observed data. These hypotheses are tentative statements susceptible to revision in the light of new data. Diagnoses made by a physician provide a convenient example. Upon careful scrutiny of a set of physical indications, the physician selects or formulates one or more hypotheses which best explain the occurrence of these indications and which may imply both the etiology and eventual remedy for the disturbance underlying these observed indications. Within the context of military affairs, individuals are similarly required to formulate hypotheses or make "diagnoses" which are to account for the occurrence of certain critical events taking place in some hostile environment. An intelligent judgment of the intent or threat implied by these events is required before effective counteraction can be recommended. This "threat-evaluation" activity is to be expected in any military organization in which the significance of intelligence data (obtained by any type of sensor) is assessed. The command and control facilities of North American Air Defense Command (NORAD) provide a specific example. The enormous difficulty of these threat-evaluation tasks is a product of at least the following factors: (a) the range of possible hypotheses which account for environmental events may be large; (b) the volume of input data to be evaluated may also be very large; (c) the input data on any given occasion are expected to be fallible to an unknown degree and often will be only fragmentary in nature; (d) the environmental contingencies which relate these data to the alternative hypotheses may be obscure or unknown.

Edwards (refs. 6, 7) has recently taken the lead in illustrating how Bayesian statistics might be applied profitably in the design of systems in which hypothesis-selection or "diagnostic" tasks are performed. Evidence is accumulating which suggests that human hypothesis selection in the form of posterior probabilities estimation [estimation of the probability of alternative hypotheses in the light of observed data, $P(H|D)$] is inferior to posterior probabilities estimation generated by computer facilities using Bayes Theorem (refs. 6, 7, 9). The general conclusions of these studies seem to be that humans typically do not extract all of the certainty from input data and that they are generally unwilling to estimate extreme probabilities. If these conclusions are valid, automated hypothesis selection or posterior probabilities estimation may provide a partial solution to the system design problems implied by the various perplexities mentioned in the preceding paragraph. Edwards' contention is that a system in which probabilistic information is processed within a Bayesian paradigm will have these advantages:

1. Bayes Theorem is optimal in a mathematical sense for extracting as much certainty as possible from available information.

2. Fallible or degraded information may be used to better advantage in a Bayesian system than in a deterministic system.

3. The Bayesian paradigm may allow a more efficient allocation of probabilistic inference tasks between men and machines (ref. 7).

In actual application Bayesian information-processing systems will operate upon a large number of complicated environmental events. In addition, information-processing and decision-making tasks may be required at several levels within such a system. Therefore, reasonable feasibility tests of a Bayesian system seem to require the facility for generation of a large and complex stimulus environment and the facility for information retrieval, information display, and for operator intercommunication. The multiman-machine systems simulator at the Ohio State University Laboratory of Aviation Psychology was easily adapted to fulfill these requirements. The details of the stimulus environment and of the system operating upon it are fully developed in later sections of this report. Briefly, however, the entire operation represents the simulation of a threat-evaluation system which functions within the Bayesian paradigm. The system operates upon a simulated hostile environment from which 25 sets of independent and controlled input data may be extracted. Twenty hypotheses comprise the mutually exclusive and exhaustive set of response alternatives available to a hypothetical "Aggressor" whose military forces occupy the hostile environment. In addition, there exists a set of contingency relationships between input data and hypothesis categories. These characteristics match the requirements of the Bayesian paradigm.

A team of eight subject-operators produces threat evaluations on the basis of events observed in the simulated hostile environment. Although information-processing and decision-making tasks are performed at several levels within the team, the highest level decisions (the estimations of posterior probabilities) are made by one operator who is designated the "commanding officer" (CO) of the threat-evaluation team.

The experiment described below is the first of a series devoted to estimating the feasibility of a Bayesian information-processing paradigm which might be applied whenever hypothesis selection or diagnosis is required in a systems context. The specific concern in this introductory experiment is an evaluation of the usefulness of a computer-implemented Bayesian solution of posterior probabilities as an "aid" to a primary decision maker (the CO) charged with the responsibility of making threat evaluations in a complex environment. A reasonable evaluation of the complex system paradigm which Edwards proposes cannot be made in a single experiment or perhaps even in several experiments. As a starting point, however, it was hypothesized that Bayesian posterior probability estimates could serve as aids to a simulated high-level decision maker. In all probability there will be many problems associated with implementing automated hypothesis-selection procedures. Hopefully, some of these problems will at least be identified as a result of the current experiment. Emphasis upon reasonable performance measures useful in threat-evaluation systems will be a feature of this introductory experiment as well as in those which will follow. The reader is cautioned that the results of this introductory experiment will be more heuristic than definitive in character. In large-scale system research, of which the present experiment is an example, variables have a perverse way of confounding themselves and lack of easy replication presents some rather excruciating problems.

Because this is the first in a series of experiments, a rather detailed explanation is presented of certain methodological features which are expected to apply throughout the entire series. Let us begin by describing a modification of Bayes Theorem which was used to generate posterior probabilities in the current experiment.

II. THE MODIFIED BAYES THEOREM (MBT)

Bayesian probability theory has received considerable attention in recent years by scientists interested in evaluating probabilistically oriented information. Although the basic concepts were originally proposed by Rev. Thomas Bayes in 1763, only in recent times has any appreciable research been performed which explicitly utilizes his concepts. Essentially, the Bayesian paradigm deals with a finite hypothesis or outcome set wherein each hypothesis is assigned some a priori probability of being true at a particular instant in time. Some event (A) is then observed for which conditional dependencies or probabilities have been established with each of the hypotheses in the finite set. As a result of the occurrence of A, the Bayesian theorem generates a new or an a posteriori hypothesis probability which is, in fact, the normalized product of the a priori and the conditional probabilities. A simple example can be made from a weather forecasting problem. At this particular time of year, the likelihood ratio of fair weather to foul is 4 to 1, i.e., the P(fair weather) = .80 and P (foul weather) = .20. These are two hypotheses (exhaustive within our categorization) for which a priori probabilities have been specified. We now observe the barometer and note that it is falling at the rate of 10 mm. of mercury/hour. Our past experience has shown us that the conditional probability ratio of foul weather to fair for this rate of pressure decrease is 9 to 1. These probabilities could be represented as the P (air pressure decrease of 10 mm./hour given fair weather) = .05 and P (air pressure decrease of 10 mm./hour given foul weather) = .45. Applying Bayes Theorem to this weather prediction problem, we can compute the probability of either fair or foul weather by normalizing the product of the a priori probability (.80 or .20) and the conditional probability (.05 or .45). The resultant products are .0400 and .0900; normalized, they become the a posteriori probabilities of .31 for fair weather and .69 for foul weather. Aside from illustrating the Bayesian probability process, this example also represents the shift of the "most likely hypothesis" from fair weather to foul weather as the result of observation of an event. Some other event or observed datum could just as well have increased the a posteriori confidence in the fair weather hypothesis.

The example just given is admittedly simple, but it does serve to demonstrate the straightforward nature of the Bayesian approach. However, certain modifications seem to be desirable when the theorem is to be applied to more elaborate information-processing problems which occur, for example, in the complex environment about to be described. Dodson (ref. 5) has modified Bayes Theorem in order to deal with:

1. Multiple event or data categories; this requires classification of all possible states in an event set so that all states of an

event given any hypothesis category can be assigned a conditional probability. Each event or data set then is treated as conditionally independent from every other event set and evaluated, order independent, with the a posteriori probabilities derived from one event set serving as the a priori probabilities for the next event set.

2. Uncertainty concerning which state of a particular event or data set has been observed; this extension of Bayesian theory would not be necessary if each and every possible observed event state could be recognized with absolute certainty. However, when an observation is made, correct categorization of the resulting datum within a data or event set may be uncertain and it appears desirable to be able to reflect this uncertainty by allowing a probability distribution to be assigned across all possible states within an event or data set. "All-or-none" judgments regarding data states are thus rendered unnecessary. Dodson's extension allows for this observational uncertainty and therefore appears to have great practical significance.

Dodson's formulations have been termed the Modified Bayes Theorem (MBT) and appear as follows with notational simplifications:

$$P(H_i|D) = \sum_{k=1}^{\mu} P(D_k) \left[\frac{P(H_i)\ P(D_k|H_i)}{\sum_{i=1}^{n} P(H_i)\ P(D_k|H_i)} \right] \qquad (1)$$

This equation applies to the evaluation of an observation from one event or data set. The bracketed expression is Bayes Theorem with $P(H_i)$ being the a priori probability of hypothesis $\underline{i}$ and $P(D_k|H_i)$ being the conditional probability that an event will occur in state $\underline{k}$ if hypothesis $\underline{i}$ is true. The denominator is merely a normalizing constant which assures that the a posteriori probabilities sum to 1.0 across the H_i set. $P(D_k)$ is the probability that an event has been observed in state or level $\underline{k}$ (see Dodson's second modification). Note that $\sum_{k=1}^{\mu} P(D_k) = 1.00$. For each event or data class there are μ states where μ varies according to the event or data class. $P(H_i|D)$ is the a posteriori probability that hypothesis $\underline{i}$ is true given the probabilistic estimates that any of the $\underline{k}$ states in the data class have been observed. There may be observations in many different data or event classes to be evaluated by means of the MBT. The $P(H_i|D)$ calculated by means of the observations and conditional probabilities for one data class become the a priori probabilities $P(H_i)$ used in the calculation of $P(H_i|D)$ for the next event or data class and so on until all data observations have been evaluated.

In terms of the present experiment the H_i set refers to 20 possible response alternatives available to the hypothetical Aggressor. There are 25 sets of events or data which are to be observed in the stimulus environment. Each set has several possible levels or states. The $P(D_k|H_i)$ or conditional probability values are the environmental contingency rules which relate every state of each of the 25 data classes to each of the 20 alternatives (H_i). The $P(H_i|D)$ or posterior probabilities are to be estimated by a primary decision maker on the basis of observed data in the various classes and knowledge of the environmental contingencies $[P(D_k|H_i)]$. For the purposes of the present experiment, the input data used in these posterior estimations were developed by several members of the system operator team. The correct environmental rules were provided for the decision maker and were used in MBT solutions of the posterior probabilities.

III. EXPERIMENTAL VARIABLES AND MEASURES

Two independent variables were manipulated in the experiment. The first variable concerned permitted access to the MBT hypothesis-selection aid. As we will discuss in greater detail in Section V, System Task Description, the system's "commanding officer" (CO) was required to produce interim and final threat evaluations (regarding certain critical events in Aggressor territory) which were based upon data observed in the simulated hostile environment. These evaluations were in the form of posterior probabilities estimates. On half of the experimental sessions (according to the design described below) the CO was permitted access once every 10 minutes (real time) to computer-generated MBT posterior probabilities estimations in order to augment his own evaluations. This condition was called the "aid" condition. On the other half of the experimental sessions (the "no-aid" condition) the CO was required to make his interim and final posterior estimations without benefit of the MBT aid. In either case, however, the experimenter was provided with the MBT posterior estimates which were to be compared with the correct posterior probabilities and the CO's estimates of these probabilities. Under both the aid and no-aid conditions the MBT solutions provided for the experimenter were generated incorporating the <u>same</u> data used by the CO in providing his personal estimates. This was done in order to compare the MBT solutions to the human solutions throughout the course of the entire experiment.

The second variable described the data-processing <u>load</u> on the system (three levels). Load was defined as the number of Aggressor developmental groupings terminating in an experimental sessions. For a precise description of a "developmental grouping" refer to page 9. For the moment, however, we shall simply say that the number of developmental groupings terminating per experimental session determined both the amount of data processing imposed upon the system operators and the number of posterior probability estimations required of the CO. There were either one, two, or four developmental groupings terminating each session. In the entire experiment there were 56 developmental groupings seen by the operator team. This variable was included in order to investigate possible interactive effects of load and aid-usage as well as to determine the operating characteristics of the entire system.

The experiment was conducted over a period of 6 weeks. There were four sessions per week and the individual sessions were of 4-hour duration. The order in which the experimental conditions were presented is as follows:

	Week (4 sessions per week)					
	1	2	3	4	5	6
MBT Aid	Aid	No Aid	Aid	No Aid	Aid	No Aid
Load Condition	1	2	4	4	2	1

Several performance measures were taken in order to indicate the accuracy both of the primary system output (the CO's interim and final posterior probability estimations) and the data upon which this output was based. For each specific buildup of Aggressor forces (called a "developmental grouping") the CO produced hourly (real-time) estimates of the posterior probabilities of the various hypotheses which could account for the observed data. In addition, when it appeared to the operator team that a particular grouping was in its terminal stages of development, a final estimate of the posterior probabilities was made. Therefore, we have a time-dependent sequence of posterior estimations for each grouping plus a final (presumably best) estimate of the posterior probabilities. These interim and final estimates were scored for both the human- and MBT-generated posterior probabilities. Scoring the accuracy of the data (developed by other members of the operator team) upon which the CO's posterior estimates were based presented several problems. A large number of the classes of data did not become available for observation until the later stages of buildup of a grouping. This condition was due to the time-dependent nature of the stimulus environment. Therefore, only the final data estimates generated by the operator team were scored. The interim data estimates, which were obtained for the incomplete set of data, were recorded by the operators but not scored by the experimenter.

IV. THE STIMULUS ENVIRONMENT

Members of the experimental system operator team served, essentially, as observers and evaluators of events taking place within a fictitious geographical area which comprised the homeland of an "Aggressor." Let it be emphasized that the judgments and evaluations generated by the system operator team produced no subsequent effect upon further events in this fictitious geographical area. The operators, playing the role of passive observers and evaluators, attempted to exploit information gained within the fictitious geographical area for an understanding of the events. The term "stimulus environment" or "data base" will refer to the characteristics of the fictitious geographical area and to the characteristics of all events taking place within this area. Specifically, the following dimensions describe the stimulus environment: (a) Aggressor homeland characteristics, (b) Aggressor surface and air forces description, (c) developmental grouping or maneuver characteristics of these surface and air forces, (d) alternative strategies to be used by Aggressor in deploying his surface and air forces,

(e) data classes which describe attributes of these developmental groupings and which allow inferences to be made about the probable Aggressor strategy or intent signified by each developmental grouping, and (f) environmental rules or contingency statements which relate the attribute data classes to the strategies or alternative courses of action available to Aggressor. In their preliminary description of our man-machine systems simulator, Feallock and Briggs described some of the more general features of the stimulus environment (ref. 8, pp. 49-52). In addition, they described the method by which the real-time characteristics of the environment were achieved through the use of computer facilities. The reader will be asked to refer to the Feallock and Briggs report from time to time for certain descriptions which, in the interests of economy, will not be repeated in the present report. Several changes have been made in the basic paradigm described by Feallock and Briggs; these changes will be noted in the present report whenever necessary.

Generation of a realistic complex stimulus environment for the sake of realism alone achieves very little. Our primary concern in the generation of the stimulus environment was to achieve environmental realism tempered by the facility for experimental control. The astute tactician may well recognize certain inadequacies in the portrayal of the deployment of Aggressor's military forces. We would ask such an individual to recognize the fact that certain trade-offs affecting degree of realism were necessary in order to develop a reasonably realistic environment in which rather basic human judgmental processes could be studied in a controlled manner.

A. Aggressor Homeland Characteristics

Figure 1 portrays the fictitious geographical area comprising Aggressor's homeland with its fixed installations and major transportation facilities. Each of the four sides are 1024 miles in length and each represents a border of contention. Within this homeland area critical events in the form of concentrations of vehicles, mobile weapons, and aircraft were to be observed by the system operator team.

B. Aggressor Surface and Air Forces

Reference to the Feallock and Briggs report (ref. 8, p. 50) will indicate that units of the surface and air forces of the hypothetical Aggressor were patterned after those of the maneuver enemy of the United States Army. In the development of this feature of the stimulus environment only battalion- and regimental-size units appeared in Aggressor's surface forces and only squadron-size units appeared in his air forces. There were 33 different types of battalion-size units and 14 different types of regimental-size units. In general, infantry, armoured, artillery, rocket, missile, and reconnaissance units comprised the surface forces. There were 31 different types of squadron-size units in Aggressor's air forces. For the purposes of the present experiment, only tactical air support, reconnaissance, and air transport units comprised Aggressor's air forces.

The various surface and air units were never directly observed by the system operator team. In point of fact, it was the task of part of the operator team to establish the existence of these units by inference on the basis of observed concentrations of road and rail vehicles, mobile weapons, and aircraft. The existence of a certain unit could be inferred by an operator only by relating observed collections of these vehicles, weapons,

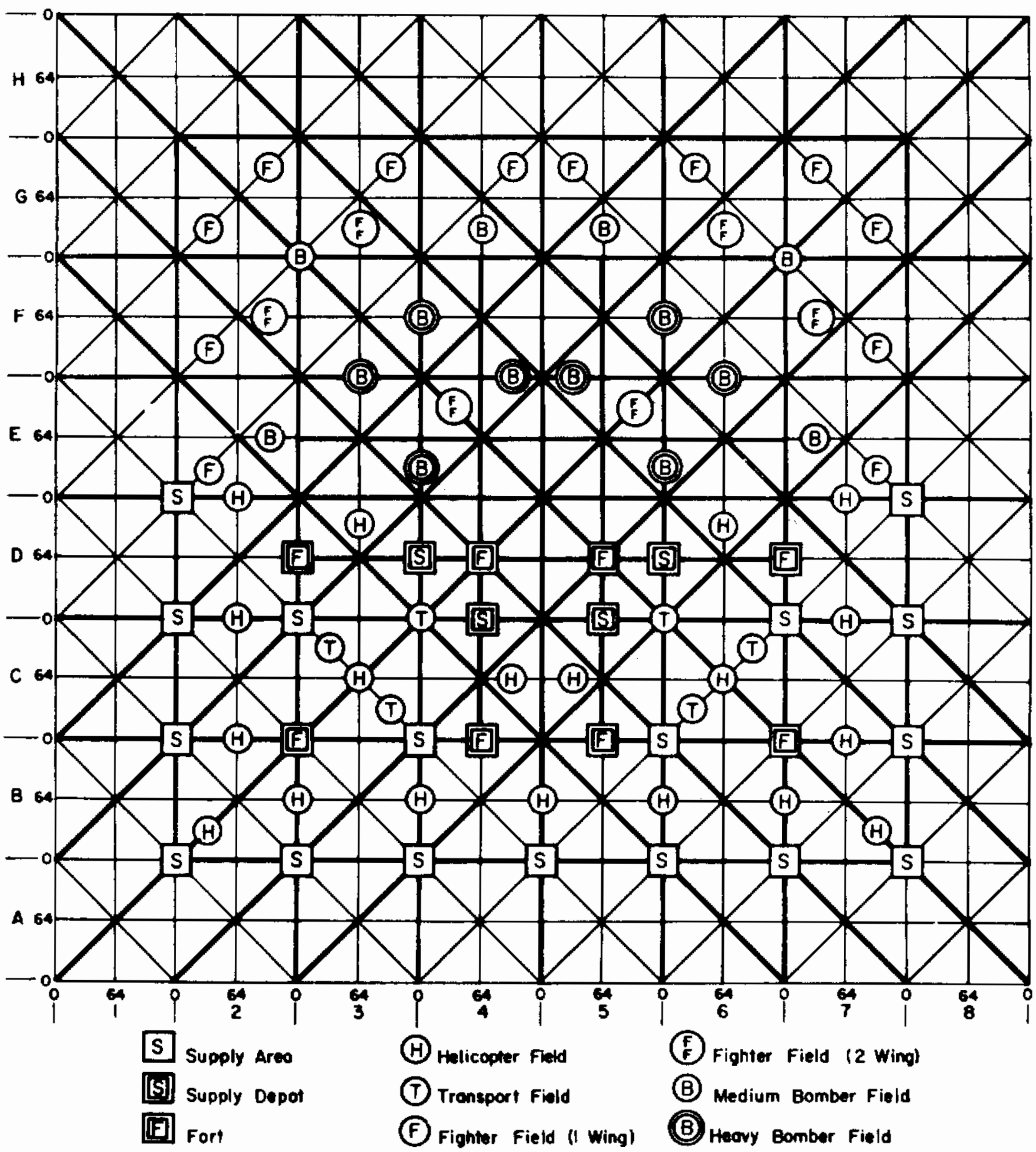


Figure 1. Map of Aggressor Homeland Showing Railroads (Double Lines), a Representative Number of Principal Roads (Single Lines), and Representative Locations of Permanent Installations. Roads actually occur at 2-mile intervals; installations shown in the upper half are found in similar locations in the lower half, and vice versa.

and aircraft (collectively termed "elements") to tabled reference sources which described in detail the composition and characteristics of the various units (see Appendix III). There were 67 different types of mobile weapons and vehicles and 30 different types of aircraft which the operator team observed. In a later section of the report, as we discuss the training given to the operator team, we will describe in some detail how the existence of a unit could be inferred from these reference tables.

C. Developmental Groupings of Aggressor Forces

The movements or maneuvers of Aggressor's forces were not aimless or haphazard events. Indeed, Aggressor's forces were deployed in an orderly and controlled manner. To an observer of Aggressor activity, it appeared that the various units departed from the center of Aggressor territory (usually from one or more forts) and moved directly or indirectly toward one of the borders where groups of these units assumed terminal positions along one of the borders. An orderly buildup of a certain number of Aggressor units culminating in a discrete or separate concentration of units along a border was termed a "developmental grouping." The time-dependent events in a particular developmental grouping could unfold over a period of from 2 to 6 days. That is, some groupings developed very rapidly (2 days) and others developed more slowly (up to 6 days). In the present study there were often as many as 16 separate groupings in various stages of development appearing at the same time in the environment. The number of developmental groupings terminating (at peripheral locations) on any given day was carefully controlled by the experimenter. In fact, as noted earlier, system load was defined as the number of developmental groupings terminating on a given day. By now the reader's curiosity is undoubtedly aroused about the disposition of a particular developmental grouping <u>after</u> it assumed a position along one of the borders of contention. After remaining in their terminal positions for a short period of time, the units involved in a developmental grouping simply disappeared from the environment. Aggressor units were never shown crossing a border (mounting an attack into "friendly" territory), nor were they shown returning to their original starting locations or to some other location. By way of explanation for this somewhat curious state of affairs, let it be recognized that the basic task given to the system operator team was the evaluation of the <u>threat</u> posed by each of these developmental groupings as they unfolded <u>within</u> the confines of Aggressor territory. Through the use of this particular scheme it was possible to present to the operator team a large number of simultaneous stimulus events without "cluttering" the environment with units returning to starting locations. On the following day, however, the exact course of action taken by Aggressor in each of the developmental groupings on the previous day was made known to the operator team. This implies that some type of action was, in fact, taken by Aggressor although the team did not actually observe any action subsequent to the termination of a buildup of a particular developmental grouping. This scheme does no particular violence to our basic paradigm—that of measuring the threat evaluations produced by a team of passive evaluators of events taking place in a hostile environment. The operator team was not required to recommend any action to counter these threatening actions on the part of Aggressor. That an open-loop situation prevailed should now be apparent. Since no counteraction was taken or recommended by the team, the events taking place in the stimulus environment were not affected by the evaluative judgments produced by the operator team.

The existence of a certain developmental grouping could be discovered by the operator team as soon as units of this grouping began to congregate at some peripheral location. When a developmental grouping had been discovered in this manner, the operator team assigned a unique identification number to the grouping. These numbers were provided by the experimenter on what was called an "advance intelligence report." This daily report listed the developmental groupings (identified only by the unique number) beginning that day and a very gross estimate of their termination areas. The report did not mention the day on which the groupings would terminate. This scheme was necessary in order to insure agreement between experimenter and operator team about the labeling of the stimulus events. All of the various kinds of data produced by the operator team with respect to a certain developmental grouping contained this unique identification number. In this complex experimental situation data collection would have been impossible without a procedure for matching the team's responses with the correct stimuli.

The mechanics for simulating the real-time characteristics of these developmental groupings has already been described by Feallock and Briggs (ref. 8, pp. 50-52). Let us emphasize the fact that each developmental grouping was simulated by means of a prepared script of the time-dependent events which described the types of Aggressor units in a grouping and their various activities occurring from beginning to termination of the developmental grouping. These scripts were developed on the basis of the attribute information discussed below. Simulation of many such groupings in various stages of development was produced by combining these developmental grouping scripts into an "environment script." The environment script, prepared for each experimental session, described all of the time-dependent events taking place throughout the entire Aggressor territory during that session.

D. Alternative Aggressor Strategies

Although there were several subtasks involved in the threat evaluations made by the operator team, the primary task was to determine, on the basis of observed characteristics of a developmental grouping, the strategy Aggressor would employ in the utilization of the forces included in the grouping. For the purposes of the present experiment there were 20 different strategies or response alternatives available to Aggressor. There were three basic dimensions which defined these strategies. The first dimension specified whether the developmental grouping was to represent a forthcoming actual hostile action by Aggressor (a "real" attack) or whether the grouping represented only exercises or maneuvers being conducted by Aggressor for training purposes (the latter were called "rehearsals" or "practice"attacks). The second dimension specified the probable endurance of the threatened action in terms of the depth to which the forces in the grouping could penetrate into friendly territory given the available logistics support. There were four penetration depths specified for this dimension: 100-mile, 50-mile, 25-mile, and 10-mile depths of penetration (into friendly territory). At this juncture it must be noted that this second dimension was incorrectly described in the Feallock and Briggs report. They referred to the second strategy defining dimension as a "frontage" (ref. 8, p. 51, and Appendix IV, p. 137). It is true that frontage or the lateral dispersion of forces in the terminal positions of a developmental grouping was an attribute to be recognized by the operator team. Frontage, however, was not a defining dimension of the set of alternative strategies available to Aggressor. The third dimension specified the particular infantry-armoured tactic to be initiated by the ground forces represented in a developmental grouping. There were five permitted tactics: double pincer, multiple penetration, double envelopment, single envelopment, and penetration. For a precise description of these tactics the reader should refer to Army Field Manual FM 30-102 (ref. 2).

These three dimensions were chosen because of their apparent mutual independence. In Appendix I of this report are listed the 20 alternative strategies defined by various logical combinations of these three defining dimensions. The 20 Aggressor strategies comprise the mutually exclusive and exhaustive set of alternative hypotheses required in the Bayesian paradigm. These strategies are the hypotheses that account for the occurrence of observed attribute data describing the characteristics of a developmental grouping.

E. Attribute Data Classes

The major features of a developmental grouping could be described by 25 types of information which we have termed "attribute data." In general, these data described the ground and air forces constituency of the developmental groupings, the logistics support provided, and the movement and order of battle characteristics of the groupings. Each of the 25 specific data classes had several possible levels or states. For example, one such data class referred to the number of battalions of artillery with range of up to 20,000 meters. There were three possible states applicable in this data class: there could exist either 0, 1, or 2 battalions of this type of artillery in a developmental grouping. One and only one state or level in each of the 25 data classes applied to a single developmental grouping. In Appendix II are listed the 25 types of attribute data along with a short description of each attribute and its possible levels or states.

Following a procedure specified below in a discussion of the rules relating attribute data classes to the alternative strategies, laboratory assistants generated a library of developmental grouping scripts (one script described one grouping). The characteristics of a certain developmental grouping were determined by prescribing one state or level in each of the 25 attribute data classes. As we shall see, for a given script the particular configuration of the levels or states prescribed in each attribute data class determined which of the 20 strategy types the developmental grouping was to represent. From the viewpoint of the experimenter, therefore, the prescribed selection of one level of each of the 25 attribute data classes represented a "true" description of the developmental grouping generated by means of this selection. In the experiment a primary task for several members of the operator team was to determine (after a sequence of observations) which level or state existed in each of the 25 attribute data classes for each developmental grouping under surveillance. This attribute data information, after being developed by certain team members on the basis of environmental observation, was forwarded to a primary decision maker (the CO) who attempted to estimate or judge the strategy to be used by Aggressor in each developmental grouping for which he had attribute data. These estimations were in the form of posterior probabilities.

F. Environmental Rules Relating Attribute Data to the Alternative Aggressor Strategies (Data-Generation Model)

There were, in fact, two different classes of rules which described contingent relationships in the stimulus environment. We have already encountered the first class of rules in the discussion of the characteristics of Aggressor's surface and air forces. This class of rules enabled the system operator to infer the existence of Aggressor units on the basis of observed numbers of elements (vehicles, mobile weapons, and aircraft). The rules were provided in tabled reference sources (also called "inference tables"), examples of which can be seen in Appendix III. For the moment, however, our interest is in the second class of contingency rules. These rules were, in fact, conditional probabilities relating each level or state of every attribute data class to each of the 20 possible strategies (or hypotheses). These conditional probabilities, $P(D_{J_k}|H_i)$, answered the following question: for any given hypothesis or strategy (H_i), what is the probability that the k^{th} level or state of attribute data class J will occur? Consider the following example:

D_X — Number of tactical air support squadrons

State		H_1 (A real attack, 100-mile depth of penetration, double-pincer tactic)
1	0 Sqdns.	.09
2	1 Sqdn.	.05
3	2 Sqdns.	.77
4	3 Sqdns.	.09

In this example the four conditional probabilities state: given hypothesis or strategy type 1, the probability of observing 0 tactical air support squadrons is .09; 1 squadron, .05; 2 squadrons, .77; and 3 squadrons, .09.

A matrix of conditional probabilities was constructed over all levels of the 25 data classes and across all 20 hypotheses. The completed 20 × 103 matrix described the true conditional relationship between the mutually exclusive and exhaustive hypothesis set and the data classes. The probability values appearing in the completed matrix were assigned according to several criteria. First, every effort was made to assign the conditional probabilities in a manner consistent with the tactical doctrine for the maneuver enemy of the U. S. Army. Frequent reference was made to the various manuals describing the characteristics of various Aggressor maneuvers (refs. 1, 2, 3). Second, the conditional probabilities were assigned in such a manner that no unique datum-hypothesis relationships would exist. This means that no single level or state of any data class was perfectly related to any specific hypothesis or strategy (i.e., with a probability of 1.00). Third, the conditional values were assigned in a manner that would allow the subjects to make discriminations (among alternative strategies) which were implied by the way in which the strategies were defined. For example, the definition of a 100-mile depth of penetration strategy implies the observation of more logistics support than does a 10-mile depth of penetration strategy. Finally for any given hypothesis, the conditional probability values assigned to the various levels within any single data class were made to sum to 1.00 and no $P(D_{J_k}|H_i)$ entry of 0.00 was allowed in any cell of the matrix.

In order to show how discrimination could be made among the hypotheses by the operator team, and in order to illustrate how the matrix of $P(D_{J_k}|H_i)$ values entered into the preparation of the developmental grouping scripts, we have extracted a small portion of the 20×103 $P(D_{J_k}|H_i)$ matrix (see table 1).

Table 1 should help to clarify the notion that the conditional relationship between a certain hypothesis and the various levels of each of the 25 data classes provided the defining characteristics of the hypothesis or strategy. On the basis of the distributions of conditional probabilities within each of the four data classes in table 1, H_1 can be generally characterized as having, with the highest probability, at least six mechanized rifle battalions, at least two squadrons of air support, at least two units of fire (ammunition) for its infantry and armoured units, and a terminal position at least 11 miles from the border.

With the details of the environment thus specified, it is now possible to describe how the individual developmental grouping scripts were generated. Each script can be considered as one "sample" from a multidimensional population describing, in probabilistic terms, the defining characteristics of the hypothesis class from which the sample was drawn. At least four samples of each of the 20 hypothesis categories were selected by this procedure. Using the "package" of the 25 selected defining attributes for each developmental grouping, laboratory assistants prepared the individual scripts.

Table 1 can also be used to illustrate how discrimination among the 20 hypothesis categories was achieved by the CO as he estimated his posterior probabilities. The conditional probability values were assigned in such a manner that they would be consistent with the way in which the hypothesis categories were defined. Using table 1, compare the conditional probabilities between H_{10} and H_{20}; notice that H_{20} is merely a rehearsal for the same kind of actual attack defined by H_{10} (see the footnote definitions in table 1). In defining these conditional relationships, it was reasoned that H_{20} and H_{10} should appear to be very similar except perhaps that fewer supplies would be carried during a rehearsal and that, if the developmental grouping were a rehearsal, it might be expected to build up farther away from the border than an actual attack of this type. Similarly, a comparison of the conditional probabilities between H_1 and H_{10} reveals that although both are classed as actual attacks, H_1 is a very large-scale attack and H_{10} is very small. The conditional probabilities in data classes I, X, and XIII reflect

TABLE 1

AN EXTRACTED PORTION OF THE $P(D_{J_k}|H_i)$ MATRIX

D_{J_k} Attribute Data Class	State	Strategy Type (H_i)* H_1	H_{10}	H_{11}	H_{20}
I. Mechanized Rifle Battalions (Btns)	1 - 0 Btns	.01	.18	.01	.18
	2 - 1 Btn	.01	.04	.01	.04
	3 - 2 Btns	.01	.09	.01	.09
	4 - 3 Btns	.03	.37	.03	.37
	5 - 4 Btns	.08	.28	.08	.28
	6 - 6 Btns	.39	.02	.39	.02
	7 - 7 Btns	.39	.01	.39	.01
	8 - 9 Btns	.08	.01	.08	.01
X. Tactical Air Support	1 - 0 Sqdns	.09	.78	.46	.88
	2 - 1 Sqdn	.05	.19	.44	.10
	3 - 2 Sqdns	.77	.02	.04	.01
	4 - 3 Sqdns	.09	.01	.06	.01
XIII. Logistics Support (Units of fire for infantry and armoured units)	1 - 0 UOF	.01	.03	.69	.88
	2 - 1 UOF	.03	.27	.20	.10
	3 - 2 UOF	.29	.65	.10	.01
	4 - 3 UOF	.67	.05	.01	.01
XVIII. Terminal Activity Zone (Miles from border of advance units)	1 - 0-5 Miles	.01	.40	.01	.01
	2 - 6-10 Miles	.10	.48	.01	.01
	3 - 11-25 Miles	.39	.10	.10	.10
	4 - 26-50 Miles	.30	.01	.10	.68
	5 - 51-150 Miles	.20	.01	.78	.20

* H_1 = Actual, 100-mile depth of penetration, double pincer.
H_{10} = Actual, 10-mile depth of penetration, penetration.
H_{11} = Rehearsal, 100-mile depth of penetration, double pincer.
H_{20} = Rehearsal, 10-mile depth of penetration, penetration.

this fact. In addition, as data class XVIII shows, a very large buildup might occur farther away from a border due, in part, to the number of units involved. The reader is reminded that the discrimination task for the primary decision maker in the experiment was considerably more difficult than that illustrated in the present example owing to the greater number of hypothesis and data classes. Also, since there were no unique datum-hypothesis relationships, a developmental grouping could never be categorized within one of the hypothesis classes with infinite certainty (i.e., with a probability of 1.00).

V. SYSTEM TASK DESCRIPTION

The mission of the experimental system was to produce periodic probabilistic evaluations of the capabilities and intentions of certain developments of hostile forces in a hypothetical environment. The complexity of the stimulus environment and the difficulty of evaluating the threat posed by the hostile forces required information-processing and decision-making tasks at several different levels within the system and a high degree of coordination between and within these task levels. In order to discuss how each system operator contributed to the overall system mission, it is necessary to summarize the flow of major classes of information in the system. Figure 2 illustrates the general task levels in the system and describes the major classes of information emanating from each task level.

Two degrees of explanation will be provided with respect to the tasks illustrated in figure 2. First, a brief summarized explanation of these tasks will be presented along with the description of the flow of major classes of information. This level of description is intended for the reader who is less interested in a detailed account of the affairs conducted at each system operator position. A more complete description of the various system operator tasks, however, begins on page 17.

IBM 1401 and 7090 computer facilities simulated the activities of the many individuals who gather and process photographic, radar, and infra-red sensor data obtained on hypothetical overflights of Aggressor territory. In figure 2 these activities are represented by the primary input data-processing level. On demand from any of the intelligence staff officers (ISOs) this primary level produced verbal descriptions of Aggressor elements (mobile weapons, vehicles, and aircraft), element activity, and locus of element activity. Arrow no. 1 refers to the output from the primary input data-processing level. Acting upon the output from this primary level and making use of tabled contingency data, system operators at the ISO level identified and tracked concentrations of weapons and vehicles and produced periodic probabilistic estimates of the state or level of each of the inferred attributes of specific Aggressor activities or developmental patterns. These estimates were then passed forward to the command level (arrow no. 2). As noted earlier, for each specific developmental pattern of Aggressor surface and air forces there were 25 different attributes or defining characteristics of the pattern which the ISOs attempted to determine. Each of these attributes had several possible states or levels and, in general, they referred to such things as the amount and types of surface or air units, the amount of logistics support, methods of transporation, and unit order of battle. Each of the ISOs (labeled A through D in figure 2) was responsible for a subset of these attribute data. In a sense, each served as a "content expert." One ISO was a logistics "expert," one was an order of battle "expert," and so on. The ISO labeled "OPNS" in figure 2 served as a coordinator and planner of the hypothetical reconnaissance overflights. Upon receiving these periodic statements of the inferred attributes of certain Aggressor activity or developmental patterns, the commanding officer was responsible for providing a probabilistic judgment [P(H|D) about the strategy to be used by Aggressor in each of the developmental patterns being observed. There were 20 alternative strategies which could be used by Aggressor either

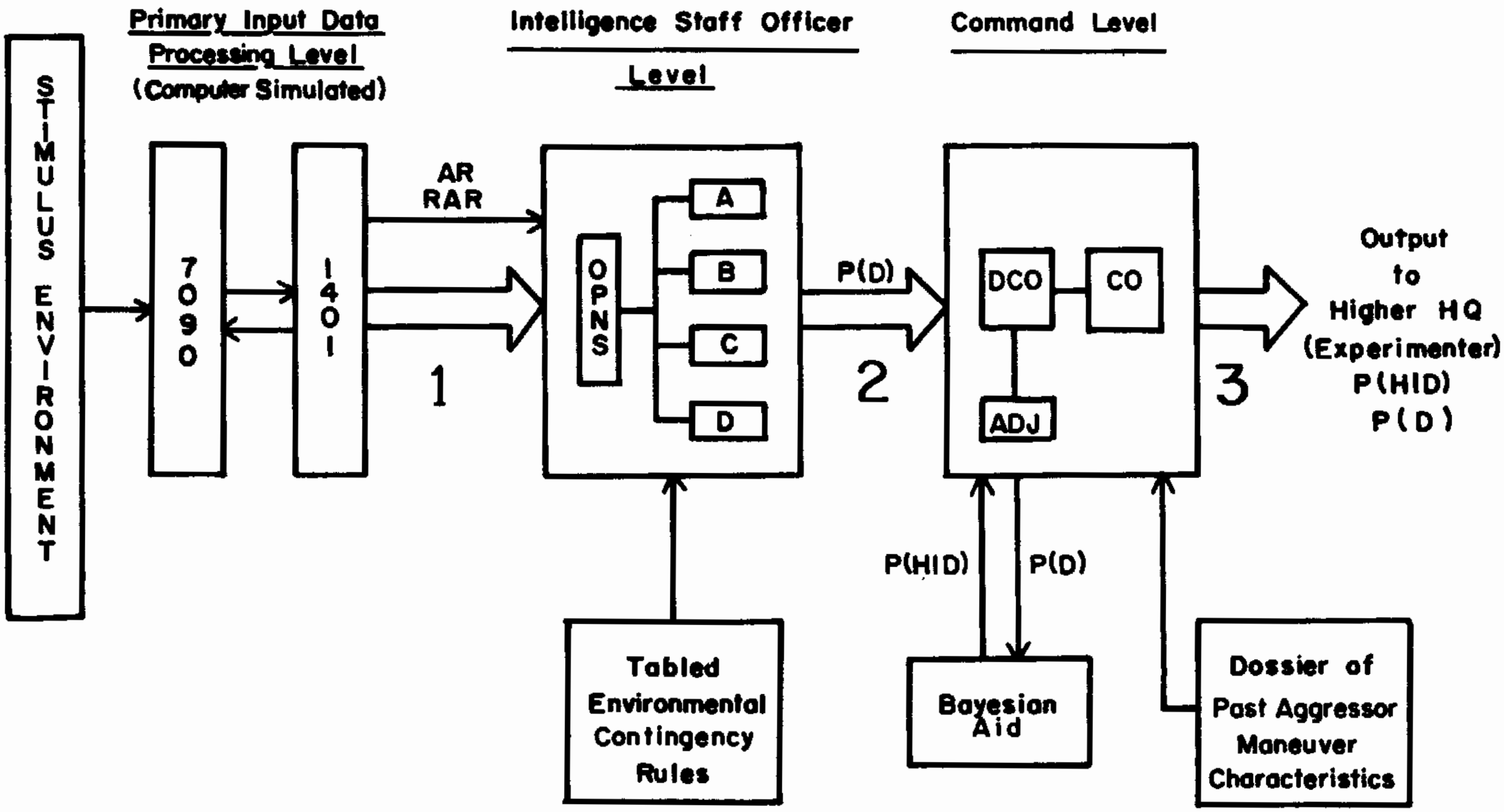


Figure 2. System Task Level Description (see text for explanation).

in mounting actual attacks or in conducting maneuvers in his homeland area. In order to make judgments of probable strategy, the commanding officer made use of an additional source of information in the form of a dossier of known previous Aggressor actions. This dossier contained a verbal summarization of the characteristics of each of the 20 alternative actions which Aggressor could take. This dossier was given to the CO as an "historical record" of the characteristics of previous Aggressor activity. Of greater importance than the dossier, however, was the computer-implemented Bayesian hypothesis-selection aid. Upon entering into the 1401 computer the current probabilistic estimates of the various possible levels or states of any or all of the 25 attributes or predictors for any single recognized developmental grouping, the CO could obtain a Bayesian estimate of the posterior probability of each of the 20 possible hypotheses (Aggressor strategies) which could account for the observed attribute data for that specific developmental grouping. These probabilistic attribute data are symbolized by P(D) in figure 2.

Each hour during each experimental period the CO produced interim probabilistic judgments [P(H|D)] of the type of strategy to be used by Aggressor in each of the developmental groupings which happened to be under surveillance at that time. When it appeared to the system operators that a given grouping was in its final stages of development, a final estimate of the type of strategy and the interim estimates were forwarded to the experimenter. Similarly, the ISOs forwarded their interim and final probabilistic estimates [P(D)] of the state or condition of each of the 25 attributes for each specific grouping. Thus, the primary data which the experimenter collected from the experimental system consisted of a "package" of interim and final P(H|D) and P(D) estimates for each specific Aggressor developmental grouping.

Before discussing in detail the various individual tasks within the experimental team of operators, it should be mentioned that the team was not modelled after any particular existing intelligence information-processing branch within the Air Force. Instead, the task format incorporates features which are common to many systems involved in the evaluation of intelligence information. A detailed explanation of the task at each system operator position follows.

A. Intelligence Staff Officer Level Tasks

As figure 2 indicates, there were five system operator positions at the intelligence staff officer (ISO) level. One of these operators was called an "operations liaison officer" (OPNS in figure 2). The other four were staff officers, each being responsible for developing a particular subset of the attribute data describing each specific Aggressor developmental grouping. The specific tasks delegated to each of these officers are indicated below.

1. Operations Liaison Officer: Two facts about the stimulus environment must become apparent to the reader before we can discuss the tasks given to the operations liaison officer. First, the system operator team was not given unlimited access to information about events in Aggressor territory. Since unlimited access to complete intelligence information would be highly unrealistic, the amount of information available for use

by the system operator team was controlled by the experimenter. This control was achieved by regulating the amount of area in Aggressor territory about which the team could request information. In each 5-minute (real-time) game cycle the team could request information about events in an area not to exceed 1/32 of the total area in Aggressor territory. The location of this specified area was not controlled, i.e., the team could request information about any specific area in Aggressor territory with the restriction that in any 5-minute, real-time period the area requested did not exceed 1/32 of the total 1024 × 1024-mile area. Second, because events in the stimulus environment were time-dependent, information about these events became obsolete rapidly, particularly since the game cycles were run in double time.

These two facts about the characteristics of the stimulus environment implied two critical functions for the system operator team. First, an efficient allocation of the area in which to "reconnoiter" was manditory if the team was to observe most of the critical events in Aggressor territory. Second, updated or current information in "critical" areas was always required since the characteristics of the patterns of Aggressor activity changed rapidly over time. It was the task of the operations liaison officer to decide which area of the map to reconnoiter each 5-minute (real-time) period during the game. His objective was to allocate the restricted amount of team information-gathering capability to areas which appeared to contain the greatest amount of Aggressor activity. Efficient allocation of area could be made only if the operations liaison officer worked in close cooperation with the other ISOs whose primary tasks involved identification and tracking of critical events in Aggressor territory. By using his knowledge of locations where critical events would most likely take place (e.g., at supply depots, airfields, railroads, etc.), and by following carefully the unfolding developmental patterns of Aggressor activity (which were displayed symbolically on a large map of Aggressor territory), the operations liaison officer selected areas of Aggressor territory in which intelligence information was to be gathered.

The actual gathering of this intelligence information by reconnaissance overflights as well as the interpretation and collation of this information was, as we have previously stated, simulated by our computer facilities. In a sense the operations liaison officer acted as a reconnaissance overflight "mission planner" since his decisions regarding the area to be reconnoitered determined what events in Aggressor territory would be observed by the operator team. He thus acted as a link between the ISOs who evaluated the intelligence data, and the simulated "operational" units that collected the intelligence data and the simulated imagery analysts who processed the data. Specifically, he always initiated the first phase of a two-phase information retrieval sequence. The first phase, called an "acquisition request" (AR), was used to specify an exact area in Aggressor territory over which photo, radar, and infra-red sensor records were to be obtained. The AR was prepared by the operations liaison officer on a standard form and was punched on IBM cards for entry into the 1401 computer. The AR instructed the computer to extract all processed reconnaissance data obtained over the specified area in the specified time interval from a tape which held in storage all events (in proper time sequence) taking place in the entire Aggressor territory. After these data were extracted from the environment tape, they were placed in storage on the 1401 RAMAC file and were made available for use by the ISOs. All of the reconnaissance data extracted from the environment tape were stored in the RAMAC file and subsets of these data could be withdrawn on demand by the ISOs. Reconnaissance information was held in RAMAC storage only during the specific experimental day in which it was extracted from the environment tape, i.e., information was not stored on the RAMAC file from day to day. Therefore, the oldest reconnaissance information which the ISOs could request from RAMAC storage was information extracted from the environment tape after the start of the current experimental day.

After an AR had been processed by the computer the operations liaison officer was notified on a digital display console by the appearance of an AR log number which was to be used by the ISOs whenever they requested information about the area at the time specified on that particular AR. In order to present a summary display of the status of all ARs entered on a certain day, the operations liaison officer maintained a map of Aggressor territory upon which was plotted the area covered by the various ARs that had been entered. Each AR plotted on the map could be identified by its log number and time of entry into the computer. By noting the time of an AR, an ISO could gauge how current the information would be if extracted by means of this AR.

The second phase of the sequence necessary to extract information from computer storage was initiated exclusively by the four ISOs who were assigned the task of developing the 25 types of attribute or predictory information for each Aggressor developmental grouping. This second phase was called a "reconnaissance analysis request" (RAR). The procedure for initiating an RAR was the same for each ISO. Therefore, before describing the tasks unique to the four remaining ISOs, it might be profitable to explain the RAR procedure and describe exactly what it accomplished for the ISOs. The RAR was, in essence, an interrogation made of the library of reconnaissance information stored on the 1401 RAMAC file. An RAR was always specific to the information on one or more ARs. This meant that an RAR report would only yield intelligence data if its area was concurrent with the area previously specified on one or more ARs. Before entering an RAR, the ISO checked the AR status board to see if there was a relatively current AR covering the area of interest. If there were none, he would request the operations liaison officer to enter an appropriate AR. The RAR was written on a standard form by an ISO and was carried by messenger to the IBM 1401 data-processing center. On the RAR the ISO specified an AR log number, the types of sensor records to be examined (photo, radar, infra-red), the geographic area of concern, and level of description desired (i.e., which quality levels of sensor description). Most important, however, it allowed the ISO to specify classes of events in Aggressor territory to be described. In general, the ISOs were interested in the following types of information relative to events in the stimulus environment.

a. Element (vehicle, mobile weapons, or aircraft) description and enumeration.

b. Element activity description, for example, "on-loading," "off-loading," "dispersing," "landing," or "taking-off."

c. Operations facility, refers to the locus of element activity in terms of certain permanent features in Aggressor territory, e.g., on roads, at rail sidings, on runways, or at supply depots.

d. Composition class, refers to a description of a collection of elements in terms of the most predominantly occurring element in the collection. For example, the convoy composition class of a medium tank battalion was "tracked vehicle" (TRV) since tanks were the most frequently occurring vehicle in this type of convoy.

e. Speed of vehicles, mobile weapons, or aircraft.

f. Movement direction of vehicles or mobile weapons.

g. Exact geographic coordinates of concentrations of elements.

h. Dispersal distance between elements when in bivouac or attack position.

i. Road convoy length to the nearest 10 yards.

j. Dispersion area (in acres) occupied by units in bivouac or in attack position.

The RAR enabled the ISOs to make an amazing number of different interrogations of the library of reconnaissance information. These interrogations could range from extremely general queries to highly specific queries about environmental events. The generality or specificity of the request depended primarily upon the instructions given to the computer with regard to the first four types of events listed above. The indications which the ISOs made on the RAR with respect to these first four event categories provided the computer with a set of filtering instructions for each of these four event categories. In each event category the computer could be instructed to "filter" on a specific event (e.g., "1/4-ton utility truck" in the element category, or "on-loading" in the activity category), or it could be instructed to pass all events in a category (e.g., all events, any activity, etc.). This latter instruction was called an "open filter." The following series of examples should make clear how the generality or specificity of the request varied with the filtering instructions provided by the ISO on the RAR.

Example 1 (a case of extreme report specificity): Suppose an ISO wanted to determine (for a specific area) whether or not there were any 3/4-ton cargo trucks (element) part of a wheeled vehicle convoy (composition class) which were onloading (activity) at a rail siding (facility). The ISO would instruct the computer to make the appropriate filtering operation in each of these four event categories, i.e., Event Category on RAR

Element	Activity	Facility	Composition Class
02T	Y21	Z31	X5U

The three-character symbols in each category above are the filtering instruction codes for 3/4-ton cargo truck, onloading, rail siding, and wheeled vehicle convoy, respectively. This is an example of an extremely specific report request. The ISO has instructed the computer to report the existence (in the area specified on the RAR) only of those collections of elements having 3/4-ton cargo trucks onloading at rail sidings provided that the 3/4-ton trucks are part of wheeled vehicle convoys.

Example 2 (a case of extreme report generality): Suppose the ISO was interested in whatever events were occurring in a specified area. In this case he would use an open filter in each of the four event categories, i.e.,

Event Category on RAR

Element	Activity	Facility	Composition Class
(Black)	(Blank)	(Blank)	(Blank)

A blank space in any category on the RAR was an open-filter instruction. In the present example, the computer would report the existence of any element part of any composition class engaged in any activity at any facility within the area specified on the RAR.

Example 3: The ISOs could filter using any of the logically possible combinations of these four event categories. A frequently used combination occurred when the ISO was merely interested in searching for a certain type of element in a specific area. Suppose he was interested in the number and location of units having medium tanks in an area. His filtering instructions would be as follows:

Event Category on RAR

Element	Activity	Facility	Composition Class
19/	(Blank)	(Blank)	(Blank)

The three-character symbol (19/) is the appropriate filter instruction for "medium tank." The computer would report the existence of all collections of elements having medium tanks regardless of their activity, facility, or composition class in the area specified by the ISO.

The extremely large range of interrogations available to the ISOs can be made apparent by considering the number of filtering instructions possible in each of the four event categories. Including the open-filter instruction in each case, there were 180 element filters, 12 activity filters, 10 facility filters, and 12 composition class filters. This allowed an ISO to make over 259,000 different interrogations of the library of reconnaissance data. When an area had been specified by an AR, the ISO was free to enter as many RARs on the area as he wished. In fact, use of the RAR was restricted only by the volume of information that each ISO could handle in the 4-hour, real-time game period.

There were two methods of reporting to an ISO the results of his interrogations on an RAR: a condensed or abstracted version of the report presented on a digital display console and a complete machine-printed report (hard copy) which was available a short time after the abstracted version appeared on the digital display. When the RAR has been processed, the officer who initiated the request was notified of the report log number over his digital display. The abstract report could be viewed by an ISO on his digital display when he entered the appropriate instructions in the control register of his digital display console. The RAR abstract on the digital display consisted of a grid coordinate map representation of Aggressor territory upon which were overlayed coded symbols representing elements or groups of elements detected in the area specified on the RAR. Each symbol was coded with a footnote number which corresponded to a condensed one-line, alpha-numeric description of the event represented by the symbol. This condensed description appeared at the top of the display and included the total number of elements, the composition class, speed, direction, activity, facility, and exact geographical coordinates involved in the event and the time the data were extracted from the environment.

The hard-copy report of an RAR was a more complete account of the information requested by an ISO. All 10 classes of events described above could be presented on the hard-copy report of an RAR. An additional feature of the hard-copy report is worthy of note. Suppose an ISO instructed the computer to "filter" on "medium tanks" in a specified area. On the hard-copy report of this request the computer would print out a description of all other elements that accompanied the medium tanks in a concentration of elements. Such a report might include, for example:

1 2-1/2 Ton Cargo Truck
6 1/4 Ton Utility Trucks
32 Medium Tanks
1 Tank Retriever
1 5 Ton Wrecker
2 Amphibious Armoured Carriers

Finally, let us emphasize that the verbal descriptions of events recorded on hard copy varied according to the quality level of the sensor records which the ISO wished to examine. The above listing of elements assumes that the ISO had requested level 1 (top quality) photo records. Suppose he had requested level 2 photo; the description would have read:

6 Small Wheeled Vehicles
4 Medium Wheeled Vehicles
33 Medium Tracked Vehicles

Under level 3 photo the description would have been:

43 Self-Propelled Vehicles

Having described the method by which environmental information could be obtained by the ISOs, we can now direct our attention to their specific tasks.

2. Staff Officer for Main Attack Units: This officer was one of four assigned the task of developing the 25 types of probabilistic attribute information [P(D)] about each specific developmental grouping. These 25 types of attribute information were grouped into four classes, each class being assigned to one ISO. The four classes were: (a) main attack forces description, (b) combat support forces description, (c) logistics support description, and (d) order of battle description. The main attack forces in any developmental grouping consisted of infantry and armoured units. Therefore, the officer responsible for describing the main attack force in each grouping developed information about the predictors or attributes which described the infantry and armoured constituents of such groupings. Three of the 25 attributes described the infantry and armoured characteristics of a grouping. They are listed as follows with the number of possible states or conditions in each:

I. Mechanized Rifle Battalions. This attribute had eight possible states: in any single Aggressor developmental grouping there could be either 0, 1, 2, 3, 4, 6, 7, or 9 mechanized rifle battalions.

II. Medium Tank Battalions. Seven states were possible: there could be either 0, 1, 2, 3, 4, 5, or 6 medium tank battalions in any single grouping.

III. Heavy Tank Battalions. Six states were possible: either 0, 1, 2, 3, 4, or 5 heavy tank battalions could exist in any single grouping.

In order to make probabilistic judgments about the state or condition of each of the above three attribute data classes for each grouping, the ISO for main attack units made interrogations (by means of the AR-RAR sequence) of computer storage about events such as element description and enumeration, composition classes, road convoy length, etc., which were known to imply the existence of the three types of Aggressor infantry and armoured units. After receiving information about these events either by digital display or hard-copy presentation, the ISO related these observed data to his tabled contingency rules which enabled him to infer the existence of these three types of units. Most frequently he made use of contingency tables 1 through 5 as listed in Appendix III. On many occasions he extracted information of significant value to one of the other ISOs. Such information was immediately shared with the interested ISO.

During the course of a 4-hour experimental period, the ISO for main attack units made numerous extractions of information from the library of stored reconnaissance data by means of the AR-RAR sequence. For each specific Aggressor developmental grouping he made interim and final probabilistic judgments about the level or condition of each of the three attributes which reflected the infantry and armoured constituents of the grouping. These judgments were forwarded to the command level and were used (in addition to other attribute information developed by the other ISOs) by the CO in his decision as to which of the alternative hypotheses (or strategies) best accounted for the occurrence of the observed attribute data.

3. Staff Officer for Combat Support Units: There were several different kinds of units in Aggressor's surface and air forces which were said to support the main attack units. Artillery, missile, rocket, reconnaissance, and tactical air units comprised these supporting forces. The officer responsible for developing information about combat support forces made probabilistic judgments about the level or state of each of the nine types of attribute data which reflected the combat support configuration for each developmental grouping. The nine attributes were:

IV. Artillery Battalions (with maximum range of 10,000 meters). Three possible states existed: either 0, 1, or 2 battalions per grouping.

V. Artillery Battalions (with maximum range of 20,000 meters). Three states were possible: either 0, 1, or 2 battalions per grouping.

VI. Artillery Battalions (with maximum range of 30,000 meters). Three states were possible: either 0, 1, or 2 battalions per grouping.

VII. Rocket Battalions. Three possible states: either 0, 1, or 2 battalions per grouping.

VIII. Intermediate Range Ballistic Missile Battalions. Three possible states existed: either 0, 1, or 2 battalions per grouping.

IX. Ground Reconnaissance Battalions. Two possible states existed: either 0 or 1 battalion per grouping.

X. Tactical Air Support Squadrons. Four states were possible: either 0, 1, 2, or 3 squadrons per grouping.

XI. Aerial Reconnaissance Squadrons. Four states were possible: either 0, 1, 2, or 3 squadrons per grouping.

XII. Surface-to-Air Missile Battalions. Four states were possible: either 0, 1, 2, or 3 battalions per grouping.

In order to make probabilistic judgments about the state or condition of each of the above nine attribute classes for each grouping, the ISO for combat support units also made interrogations (by means of the AR-RAR sequence) of computer storage about events which would imply the existence of these various units. After receiving such information over digital display or hard copy, the ISO related the observed data to his tabled contingency rules to infer the existence of these units. Contingency tables 1, 2, 3, 4, 5, and 11 (Appendix III) were used most frequently. After numerous extractions of information from the library of reconnaissance information, this ISO also produced interim and final probabilistic judgments [P(D)] about the state or condition of the nine attributes for each developmental grouping. These judgments were also passed forward to the command level.

4. Staff Officer for Logistics Support: Five attributes described the logistics support in any Aggressor developmental grouping.

XIII. Units of Fire for Main Attack Forces (four possible states).

XIV. Units of Fire for Combat Support Forces (four possible states).

XV. Terminal Activity Logistics Dispersion (four possible states).

XVI. Supply Timing for Main Attack Forces (three possible states).

XVII. Supply Timing for Combat Support Forces (three possible states).

The officer responsible for developing this logistics attribute information for each developmental grouping made frequent use of the AR-RAR information-retrieval sequence already described. Contingency tables 8 and 9 (in Appendix III) were used by the officer in the formulation of his probabilistic judgments of the level or condition of predictors XIII and XIV above. With respect to predictors XV, XVI, and XVII, however, probabilistic judgments arose as a result of a different process. The logistics ISO's judgments regarding the level or condition existing in predictors XV, XVI, and XVII came, not from relating observed events to contingency rules tables, but from a careful observation of the spatial and temporal relationships existing in the patterns of logistics activity in each unfolding developmental grouping. A judgment regarding predictor XV, for example, came from a careful measurement of the geographical separation (in miles) of the supply units after these units had reached their terminal locations. Judgments regarding predictors XVI and XVII arose as a result of careful observation of the temporal order of appearance at the terminal location of supply units, main attack units, and combat support units. From such observation the ISO could determine whether the supply units had moved into the terminal location of a grouping before, after, or concurrently with the appearance of the main attack or combat support units.

Both interim and final probabilistic estimates of the condition of each logistics attribute were formulated for each specific grouping and were forwarded to the command level.

5. Staff Officer for Order of Battle Description: The final eight predictors described not only the spatial and temporal attributes of a developmental grouping but also the methods of transportation employed by Aggressor during the buildup of these groupings. These eight attributes were as follows:

XVIII. Terminal Activity Zone (distance in miles from a border of contention of the forward unit in a developmental grouping). There were five possible states.

XIX. Terminal Activity Developmental Pattern (configuration or placement of forces laterally along a border of contention). There were four possible states.

XX. Attack Position Lateral Dispersion (dispersal distance in miles along a border of contention of an entire developmental grouping). There were five possible states.

XXI. Attack Position Depth (distance in miles involved in the placement of forces perpendicular to a border, i.e., the distance in miles from the most forward unit in a grouping to the unit farthest in the rear). There were four possible states.

XXII. Attack Buildup Timing (temporal order of appearance at terminal positions along a border of main attack and combat support forces). There were three possible states.

XXIII. Transportation Methods (particular combination of road, rail, and air transport facilities being used to transport the units in Aggressor's forces in any developmental grouping). There were three possible states.

XXIV. Group Transportation Speed Class (road and rail convoy speed during buildup of a developmental grouping). There were five possible states.

XXV. Developmental Period (length of time in days from the beginning to the termination of the buildup of a developmental grouping). There were six possible states.

Predictors XVIII through XXI described the geographical or spatial dimensions of a developmental grouping. In order to make probabilistic judgments about the level or state of each of these predictors, the officer worked in close cooperation with the other three ISOs who identified and tracked the various Aggressor units involved in the buildup of a particular grouping. When a grouping appeared to be in its final stages of development and when all or most of the Aggressor units involved in the development had been identified, the order of battle officer could make probabilistic judgments relative to these four predictors by a careful observation of the pattern of dispersion of the units and careful measurement of this dispersion. This process required coordination with the other ISOs who were in possession of the geographical coordinates of the various units as a result of the AR-RAR sequence. A large overlay map facilitated the process of keeping track of the location of units as well as the patterns of their dispersal. Predictors XXII and XXV described some of the temporal characteristics of the buildups of a grouping. Judgments about the level or state of these two predictors could be made only if the order of battle officer carefully followed the buildup of each grouping and the temporal order of appearance of the various classes of units. Predictors XXIII and XXIV described characteristics of the transportation facilities involved in the buildup of a grouping. With respect to these two predictors, the order of battle officer made use of the AR-RAR sequence to identify the various road, rail, and air vehicles used to transport Aggressor's units. Contingency tables 2, 3, 6, and 7 (Appendix III) were helpful in making judgments regarding predictor XXIII, and table 10 was helpful in making judgments regarding predictor XXIV.

From the preceding discussion it should be apparent that a large amount of the information used by the order of battle officer became available only at the terminal stages of the buildup of an Aggressor developmental grouping. This is particularly the case with predictors XVIII, XIX, XX, XXI, and XXV. A premium was therefore placed upon rapid and accurate work by the order of battle officer during the later stages of the buildup of a grouping in addition to a careful following of events during the early stages of the buildup. The time-dependent nature of these predictors is an extremely important characteristic which will be noted further in connection with the results which we obtained.

In summary, with respect to the tasks of the four ISOs who developed the attribute or predictor information, we have seen that each ISO performed identification and tracking functions with respect to concentrations of various Aggressor elements such as road, rail, and airborne vehicles. In addition, coordination among the ISOs was necessary in order to determine the spatial and temporal characteristics of the maneuvers of the various Aggressor units during the buildup of specific developmental groupings. Information extracted from the library of reconnaissance data was related to tabled contingency rules in order to make probabilistic judgments about the level or state of each of the attribute data classes. For other predictors (XV through XXI and XXV), however, these probabilistic judgments were made by careful observation of spatial arrangements in the environment and careful following of time-dependent events. The criticality of mutual cooperation among ISO team members was emphasized.

B. Command Level Tasks

There were three command level positions in the experimental system: a commanding officer (CO), a deputy commanding officer (DCO), and an adjutant (ADJ). The specific responsibilities of each of these individuals are indicated below.

1. Commanding Officer: The primary responsibility of the CO was to provide interim and final probabilistic judgments of the type of strategy to be used by Aggressor in each specific developmental grouping which appeared in the simulated environment. The judgments or decisions which the CO made with respect to each grouping represented the final stage in a series of information-processing and decision-making tasks performed at several levels within the system. After the ISOs had developed their interim or final probabilistic estimates of the state or level of each of the 25 attributes or predictors for each recognized developmental grouping, the CO utilized this attribute information to generate his interim or final probabilistic estimates [$P(H|D)$] regarding the strategy to be used by Aggressor in these groupings. These $P(H|D)$ judgments answered the following question: which of the 20 possible alternative Aggressor strategies accounts for the occurrence of the attribute data observed regarding a particular grouping? The question could also be phrased: given these observed attribute data about a developmental grouping, to what extent do any of the alternative strategies account for the occurrence of the data? In a sense, the $P(H|D)$ estimations reflected the CO's degree of certainty that any one of the 20 possible Aggressor strategies could explain the occurrence of the attribute data observed relative to a particular developmental grouping. The magnitude of the judgmental or decision task imposed upon the CO should be apparent upon consideration of the number of different types of attribute information developed, the range of alternative actions available to Aggressor, and the amount of activity occurring in the simulated environment at any given time. It was our specific intention to place the primary decision maker (the CO) in a situation in which many types of information would be available about critical events and in which there could exist many alternative explanations for the occurrence of these events.

In the discussion of the Modified Bayes Theorem (MBT) it was mentioned that the "true" environmental rules or contingent relations between attribute data and Aggressor strategies [$P(D|H)$] were used in the current study in the MBT solutions of posterior probabilities [$P(H|D)$]. It was also true in the current study that the CO had access to these true conditional probabilities relating attribute data to alternative strategies. The $P(D|H)$ data were available in two different formats: (a) an actual matrix of the $P(D|H)$ values and (b) a verbal abstraction of the $P(D|H)$ values. The size of the $P(D|H)$ matrix was 20×103 (20 hypotheses and 103 possible data states). This large matrix soon proved to be unwieldy for rapid use by the CO. The verbal abstractions of these $P(D|H)$ values were incorporated into what we termed a "dossier" of Aggressor strategy. A section of the dossier was devoted to each of the 20 possible alternative strategies Aggressor could use. In each section (i.e., for each strategy) we described verbally the relation between each possible data state and the hypothesis or strategy. For example, the true contingency relationship between attribute type VII (rocket battalions) and strategy type 1 (actual attack, 100-mile double pincer) was as follows:

Attribute VI (Rocket Battalions)

Possible Levels or States	Strategy Type 1
0 Battalions	.01
1 Battalion	.10
2 Battalions	.89

The verbal abstraction of this contingency relation or $P(D|H)$ in the dossier might read as follows: "The occurrence of strategy type 1 is associated most frequently with two rocket battalions, occasionally with one rocket battalion, and very infrequently with zero rocket battalions."

Upon reading the above abstracted version of the sample contingency relation, the reader may well wonder why this verbal type of presentation was to be preferred over the large matrix of exact $P(D|H)$ values. In a pilot study completed before the experiment being reported was begun, we found that the utilization of large matrices of probabilistic information was a most excruciating task for our subjects. Our major reason for introducing the verbal abstractions of the contingency rules was simply to describe known characteristics of Aggressor activity in the form which we felt would be most familiar to our primary decision maker.

Upon receiving a set of attribute data for a particular developmental grouping, the CO related these data to his dossier of environmental rules in order to judge which strategy or strategies best accounted for the occurrence of the data. Of critical importance was his proficiency in determining which attributes or predictors were most useful in making discriminations among the alternative strategies. This task was complicated by the fact that any level of any predictor was not unique to any one of the 20 alternative strategies, i.e., knowledge of any particular datum did not allow a positive identification of an Aggressor strategy. Lack of uniqueness in the datum-hypothesis relationship forced him to consider the entire range of attribute data although some predictors were better than others for certain discriminations. It was this very lack of uniqueness in the datum-hypothesis relationship which forced his judgments to be probabilistic in nature. Theoretically, the CO could never be infinitely certain that only one strategy would account for the occurrence of a set of attribute data.

On half of the trials in the experiment the CO had access to the Bayesian hypothesis-selection aid (see page 6). The CO was free to request MBT aid for any developmental grouping about which he had attribute data. In addition, there was no restriction on the number of times he could request MBT aid for the same developmental grouping. The exact procedure he followed in requesting the MBT solutions and the conditions under which these solutions were obtained are worthy of note. In order to obtain an MBT solution of $P(H|D)$ for a particular developmental grouping, the CO relayed to the data-processing center the current estimates of the level or state of the various attributes describing the developmental grouping under consideration. Frequently, the set of attribute data relayed to the data-processing center was incomplete, i.e., there were data estimates on fewer than the 25 possible types of attributes. Incomplete attribute sets occurred most frequently in the early stages of the buildup of a developmental grouping and were due to the time-dependent nature of the appearance of certain types of attribute data. For example, information about attribute types XVIII, XIX, XX, XXI, and XXV (see page 26) was available only in the terminal stages of the buildup of an Aggressor grouping. It was also true that the CO monitored the progress of the ISOs as they developed the attribute data for each developmental grouping. Frequently, he found it necessary to revise the probabilistic attribute data estimates which he received from the ISOs. It is important to note, therefore, that the attribute data which the ISOs generated was subject to revision by the CO before he used the data in making his decision or before he entered the data into the computer for a Bayesian solution.

In addition to the judgmental or decision functions which the CO exercised in making his $P(H|D)$ estimates, there were certain managerial functions which he performed. First, he was responsible for insuring that the system produced the required outputs at the specified time. These outputs were, as we mentioned, the interim and final $P(D)$ and $P(H|D)$ estimates for each developmental grouping. Second, the CO monitored and directed the internal functioning of the system operator team. The CO could direct that an AR be entered on any particular area which he judged to be of importance (provided that the area did not exceed the limits already mentioned). In addition, he could enter RARs himself if it happened to be necessary. Therefore, in addition to being the primary evaluator of the intelligence data collected on each developmental grouping, the CO exerted control over the various information-processing tasks performed at the ISO level. Finally, he was responsible for maintaining strict adherance to the game rules specified by the experimenter. Communications made by the experimenter to the system operators regarding rule infringement or other performance adjustments were always made through the CO.

2. Deputy Commanding Officer (DCO): The DCO assumed responsibility for a large share of the managerial functions required at the command level. In particular, he monitored the activity of the ISOs by reviewing and summarizing their attribute data estimates before presentation of these data to the CO. The DCO also monitored the unfolding patterns of developmental groupings which were displayed on the large map of Aggressor territory. He did so in order to verify that all Aggressor developmental groupings reported to the CO on the "advance intelligence report" (see page 10) were, in fact, being recognized correctly by the ISOs. Finally, he was responsible for preparing the requests for MBT aid which the CO wished to enter.

3. Adjutant: Maintenance of the summary or status map (upon which were plotted Aggressor units and their locations) and the relaying of messages between the system operators and the data-processing center were the primary responsibilities of the adjutant. In addition, he was responsible for presenting the completed data packages to the experimenter when they were required.

VI. SYSTEM OPERATOR SELECTION AND TRAINING

From a pool of 30 male applicants 13 were initially selected to begin training as experimental subject-operators. They were selected on the basis of the following criteria: (a) cumulative academic grade point average at least "C"; (b) class level—only undergraduate juniors or seniors, graduate students, or professional school students were considered; (c) long-term availability—a minimum of three consecutive quarters (about 35 weeks) retainability was required; (d) apparent maturity as judged by two senior staff members who interviewed all applicants. The ages of the 13 applicants selected ranged from 19 years to 28 years; the average age of the group was 23 years.

The training received by the system operators consisted of approximately 114 hours of lecture sessions, demonstrations, problem-solving sessions, and on-the-job training. This training program was divided into three phases. The first two phases were conducted in a classroom setting and the final phase was conducted in the systems research vehicle itself. Before the training program began, the trainees were advised that there were to be eight system operator positions available in the actual experimental work on the research vehicle and a number of ancillary positions of lesser importance such as runners and monitors. The trainees were made to understand that the system operator positions would command higher rates of pay and would be superior in terms of "prestige" to the supporting positions. In addition, the hierarchical structure within the experimental team of system operators was explained to the trainees. It was established that the rate of pay for a system operator would be commensurate with the degree of responsibility he assumed in the experimental team.

Following an explanation of the various available positions, the trainees were informed that initial assignment to the eight desirable system operator positions would be contingent upon their performance in the first two (classroom) phases of the training program. Trainee performance was assessed periodically during these first two phases of training by objective-type tests which included both standard objective test items and problem solution of a sort which would be expected in the actual experimental system operation. Thus, the trainees were made aware of the fact that they were competing for the most desirable positions in the experimental work which was to follow.

A detailed account of the various training phases seems desirable for several reasons. Knowledge of the materials and techniques presented to our experimental subjects will perhaps make the balance of the report more intelligible to the reader. In addition, since the same general training procedures are anticipated for additional or replacement system operators in future studies in the series, the interests of economy in future reports can be served by a comprehensive statement of training procedures in this initial research report.

A. Training Phase I (40 hours)

After the trainees were made aware of the purpose of the research program and the types of Air Force operations which were to be simulated, they were introduced to basic information regarding the characteristics and contingencies of the stimulus environment. This information can be conveniently grouped into three classes: (a) descriptive information referring to environmental constants such as fixed installations and transportation facilities in Aggressor homeland, Aggressor equipment and weapons (elements), the structure or organization of Aggressor surface and air forces, and finally, the activities which these forces were permitted; (b) inferential intelligence information referring to those items of information such as supply requirements, vehicle or weapon allocation, and movement characteristics which would be useful to the system operator in establishing the existence of any particular Aggressor unit; (c) sensor data characteristics referring to the type and quality of description given by the various simulated airborne sensors. A more detailed explanation of these information classes follows:

1. Descriptive Information:

a. Aggressor Homeland Characteristics: The trainees were required to know thoroughly the types and locations of fixed installations, the types and locations of transportation facilities, and the map coordinate system by which any points in the homeland area could be located.

b. Surface Force Description: In the present study the developmental patterns of Aggressor ground forces which the system operator teams would observe consisted of controlled numbers of battalion- and regimental-size units. There were 33 different types of battalion-size units and 14 different types of regimental-size units in Aggressor's ground forces. The trainees were required to be thoroughly familiar with these different classes of units. In addition, a thorough knowledge of the 67 types of Aggressor's major mobile weapons and vehicles (collectively called "elements") was demanded of the trainees.

c. Air Forces Description: There were 31 different types of squadron-size units in Aggressor's air forces which the trainees were to be aware of in addition to some 30 different types of airborne vehicles (also called "elements"). Again, as with Aggressor surface forces, the trainees were required to demonstrate knowledge of the types of air elements and their permitted activities.

2. Inferential Intelligence Information: Let us review briefly the task to be performed by several of the system operators. Four of the operators, who were called "intelligence staff officers" (ISO), had the task of retrieving and analyzing information in the form of verbal descriptions of types and numbers of elements (mobile weapons, vehicles, and aircraft), element activities, and locus of element activity. On the basis of such information the ISOs were responsible for identifying Aggressor units and following the activities of these units as the developmental patterns of Aggressor unit maneuvers unfolded in the hypothetical environment. In order to infer the existence of a particular unit at a particular location,

given fragments of the above types of input information, the ISOs made use of additional sources of intelligence information in the form of tabled contingency rules which related certain environmental situations to the existence of certain types of Aggressor units. Such tables were called "inference tables." As an example of the use of these tables, consider the following situation. Suppose one of the ISOs had received information that, at a certain set of geographical coordinates, there was an Aggressor unit consisting of 36 elements (vehicles and major mobile weapons) in bivouac dispersed over 4.8 acres with a space of 25 yards between each of the elements. There is an inference table which relates bivouac area and element dispersal distance to Aggressor units. The ISO, given the above information, would consult such an inference table and determine which unit or units satisfy the given environmental situation.

In Appendix III to this report we list the types of inference tables most frequently used by the ISOs, showing in each case the contingency relationships involved between environmental situation and Aggressor unit. In many cases, these tables do not provide a unique unit identification. That is, a given environmental situation might be related to more than one type of Aggressor unit. Therefore, in order to make a positive identification of a unit it was often necessary for an ISO to request additional information which might provide a unique identification.

3. Sensor Data Characteristics: The information utilized by the team of system operators consisted of verbal descriptions of the type, number, and activity of Aggressor weapons and vehicles which were prepared by imagery analysts or interpreters from photo, radar, or infra-red sensor data obtained from reconnaissance overflights of Aggressor territory. The overflights and the activities of the imagery analysts were simulated by our computer facilities. In order to simulate the fact that there might exist several graded quality levels of sensor data (reflecting either inherent degree of sensor resolution or degraded sensor data such as that obtained, for example, from a radar scope photo taken with too much bias or video gain) or that there might have been different amounts of time taken by the imagery analysts to interpret these sensor records, different levels of description were made available for each of the elements to be seen by the system operators. It was important, therefore, that the trainees have knowledge of (a) the types of information available from the different classes of sensor records (photo, radar, or infra-red), and (b) the amount of discriminability between Aggressor elements that would result from use of the various levels of sensor data.

The prospective system operators were well-drilled in the types of information provided by each class of sensory data. In addition, the trainees were provided with several tables which listed the descriptions of all Aggressor elements under the various levels of sensor resolution or quality. Using these tables, the operator could determine how fine a discrimination could be made between elements when a given level of a sensor was used. For example, using a level 1 photo (highest level), the description given to an operator of a 2-1/2-ton cargo truck was "2-1/2-ton cargo truck," and unequivocal description. Under a level 2 photo, however, the description would read "medium size wheeled vehicle." In this case, the operator could not distinguish a 2-1/2-ton cargo truck from a 5-ton truck, an amphibious armoured carrier, or a 140-mm rocket launcher since they also were described as "medium size wheeled vehicles" under level 2 photo. Using level 3 photo (poorest), the description of a 2-1/2-ton cargo truck would read "self-propelled vehicle." The discriminability situation is even poorer in this case since there were many types of mobile weapons and vehicles which were described as "self-propelled vehicle" under a level 3 photo.

A fairly large amount of descriptive material and contingency data was thus presented to the trainees in phase I. In fact, 44 separate tables, charts, and diagrams describing various facets of the stimulus environment were given to the trainees. Memorization of data by the trainees was minimized and was required only for certain critical information that would be frequently used in the system tasks. Major emphasis was placed on knowledge of the types of tabled information available and upon speed and accuracy in utilizing this information. Each trainee was given a handbook containing all the information presented in phase I. After a lecture was presented covering a certain block of information, the trainees were given sufficient time in which to make a careful study of the material presented in the lecture. There were two objective examinations given during this phase.

B. Training Phase II (24 hours)

The training in phase II began with the presentation of a series of problems designed to acquaint the trainees with certain information-processing tasks they would encounter as system operators. These problems often involved a sequential usage of tabled reference sources when one reference source alone would not provide a unique solution to the problem. After the trainees had solved a number of these problems, they were given an examination consisting of similar problems to be worked within a time limit. Both speed and accuracy had been stressed during the practice problem sessions.

The next block of training in phase II consisted of instruction and practice on several of the individual operational procedures which were to be required of the system operators. The most important of these operational procedures concerned the methods for retrieving information from computer storage. It has already been mentioned that our computer facilities simulated the activities of many hundreds of individuals who gather and process intelligence data obtained by aerial reconnaissance overflights. In order for a system operator to retrieve some desired portion of this reconnaissance data, there were two necessary operations. The first operation, called an acquisition request (AR), specified a geographical area for which sensor records were to be obtained. The second operation, a reconnaissance analysis request (RAR), was a request for a specific subset of the information made available by the AR. A detailed account of this two-stage retrieval process begins on page 18 of this report. The trainees in the present study were required to demonstrate proficiency in initiating both of these types of requests.

As we have already mentioned, eight of the 13 trainees were selected at the end of phase II to be trained "on the job" in phase II as system operators. The five other individuals were cross-trained in several different extra-system jobs such as in monitoring tasks and preparation of stimulus materials. Two of the five were designated as system operator alternates and were informed that they would take the place of any of the eight originally chosen should any unusual circumstance develop.

C. Training Phase III (50 hours)

On-the-job training in the Comcon research vehicle began immediately for the eight individuals selected at the end of phase II. They were assigned to various levels of responsibility within the system on the basis of their performance in the earlier phases of training. After a thorough briefing on the duties and responsibilities of each individual system operator position and on the techniques and rules for coordination and communication between team members, the team began the task of observing and evaluating the events in a stimulus environment whose characteristics were quite similar to those present in the actual experiment. Having read the System Task Description section of this report, the reader will no doubt agree that the task given to the experimental team was indeed complex. Successful accomplishment of the mission given to the experimental team required that individual tasks be performed with accuracy and that effective team member coordination be established. Accordingly, in phase III individual tasks such as initiation of the information-retrieval sequence and use of the tabled reference material were practiced at length. In addition, techniques for intrateam coordination were practiced using the various intercommunication equipment provided for the system.

In the early stages of phase III, the problem cycles were run in real time, i.e., 4 hours of real time were equal to 4 hours of time in the hypothetical Aggressor environment. As the training progressed, double-time problem cycles were introduced, i.e., in 4 hours of real time the team observed 8 hours of events in Aggressor "world." The actual experiment was performed under these latter conditions. The stimulus environment presented to the team in phase III consisted of the essential features of the environment used in the experiment. Time-contingent Aggressor unit maneuvers or developmental patterns were observed by the team and individual units and groupings of these units were identified and tracked. The information relevant to these activities which the system operators retrieved from storage was presented to them in the form of computer hard copy only since the digital displays were not yet available for use at this time.

With the completion of 50 hours of practice in the system under the above conditions, the formal training program for the system operator team was ended. There remains only a description of the activities of the system operator team in the time interval between completion of the training program and the start of the actual experiment. Since the experiment being reported was to be the first in a series of studies, a preliminary series of experimental cycles was necessary in order to establish workable procedures for collection of the large amount of data which the experimental system would provide. Such preliminary investigation depended, of course, on the utilization of subjects whom we could presume had reached a fairly high degree of task efficiency. For a period of approximately 100 hours of experiment time, the Comcon system was operated under several different load and procedural configurations. We were particularly interested in determining how many simultaneous Aggressor developmental patterns the system could handle and how much input information could be processed by the operator team in a 4-hour game period.

Near the end of this preliminary or "shakedown" phase, the digital displays became available and were incorporated into the system for use by the operators. The system operators became thoroughly proficient in the use of these displays before the actual experiment began.

In summary, the individuals serving as subjects in this experiment were exposed to a large amount of information and procedures which had to be mastered before the experiment (and later experiments in the series) could begin. A quick tally reveals that they received some 114 hours of formal training plus additional practice in a preliminary phase of the experiment. Additional subjects have been trained since the completion of this initial experiment. We have found that the training time per subject can be reduced considerably. Through the use of a remote monitoring digital display and the closed-circuit television facilities, a trainee can observe nearly all the events taking place in the experimental operation. This drastically reduces the number of hours spent in explanation of necessary tasks and procedures. For these newer subjects, we have found that a desired entry level of proficiency can be attained after 50 to 60 hours of formal training.

VII. RESULTS

Two classes of performance measures are to be considered. One class of measures reflects the accuracy of human- and MBT-generated posterior probabilities estimations; the other class reflects the accuracy of the attribute data produced by the ISOs. The major methodological peculiarity of this introductory experiment involved collection of the most important data (posterior probabilities estimations) from a single subject over a considerable period of time. There appears to be no elegant way in which data obtained under these circumstances can be analyzed statistically. The various nonparametric statistical methods which we did use in the analysis of these data effectively precluded any statistical treatment of possible important interactions between the MBT-aid usage and load variables.

A. Posterior Probabilities Estimation

The CO was allowed to distribute his posterior estimates across the hypothesis set in any way he desired provided, of course, that on any one occasion these estimates summed to 1.0 across the hypothesis set. The MBT aid was programmed so that it would provide similar responses. These responses were treated in two different ways. First, a verified certainty measure was taken in which the human- and MBT-generated a posteriori probabilities compared with the correct hypothesis for each of the 56 developmental groupings. Verified certainty was simply the value of the $P(H|D)$ estimate placed in the correct hypothesis category for each developmental grouping. The subject's probabilistic estimate in a certain hypothesis category was assumed to represent his degree of certainty that the hypothesis would explain the occurrence of the observed data. This certainty estimate was "verified" by comparing it with the "true" state of nature or hypothesis for that developmental grouping. For example, suppose that strategy type 14 was the "correct" Aggressor response alternative for a developmental grouping and the CO's $P(H_{14}|D)$ was .50. His score for this developmental grouping would be, very simply, .50. Second, a more gross but nonetheless important measure was taken. The highest posterior estimate (or the "first choice" hypothesis category) made by human or MBT on each trial was scored as being correct or incorrect. We termed these dichotomous scores. As we shall see, there are certain advantages for each of these two performance indications. Also, remember that both interim and final $P(H|D)$ estimates were obtained for each developmental grouping. Results 1 through 5 below all refer to final estimates.

1. Human Judgment of $P(H|D)$ with and without MBT Aid: The solid and dashed lines of figure 3 indicate the accuracy of final human judgment of posterior probabilities under the MBT aid and no-aid conditions at each of the three levels of system load. Verified certainty scores were the measures. Although performance under the aid condition was superior to that under the no-aid condition at all levels of load, the difference between the aid and no-aid conditions was not large enough to meet the conventional requirements for statistical significance. Using the Sign Test (ref. 10), the hypothesis of no differences could not be rejected ($p = .1635$, one-tailed test). Observe, however, that the effects of the aid condition are most pronounced at the two higher load levels.

2. Overall Human $P(H|D)$ Estimation as Indicated by Dichotomous Scores: The verified certainty scores do not show the precise number of occasions on which the human's highest $P(H|D)$ estimates (for each of the 56 trials) were correct or incorrect. For each of the 56 final estimates of $P(H|D)$, the highest value (or first choice) of $P(H|D)$ was scored correct or incorrect. On eight of the 56 trials, the human's first choice among hypotheses could not be indicated unequivocally. This could occur, for example, whenever the CO assigned equal highest probabilities to two or more hypotheses. Under these conditions dichotomization was not possible. In the 56 trials, however, he made 48 unequivocal first choices. Another way of saying this is that on 48 of 56 trials the human "committed himself" to one hypothesis which he judged (with varying degrees of certainty) as best in accounting for the observed data. Now, out of these 48 unequivocal commitments, the human was correct on 14 occasions.

3. Overall Accuracy of MBT-Generated $P(H|D)$ Values (Dichotomous Scores): The dotted line in figure 3 indicates the average verified certainty scores of the MBT-generated final $P(H|D)$ values obtained using the same input data as used by the human in his final judgments. Dichotomizing the MBT solutions in the same manner as the human solutions of $P(H|D)$, it was found that the MBT generated 25 correct commitments out of 56 trials (there were no equivocal estimates produced by the MBT).

4. MBT-Aid $P(H|D)$ Estimation Accuracy: The MBT estimations described above were solutions incorporating the same final data which the CO used in making his posterior estimates. The MBT aid used by the CO, however, used data estimates obtained prior to the final $P(H|D)$ estimation. Most often,

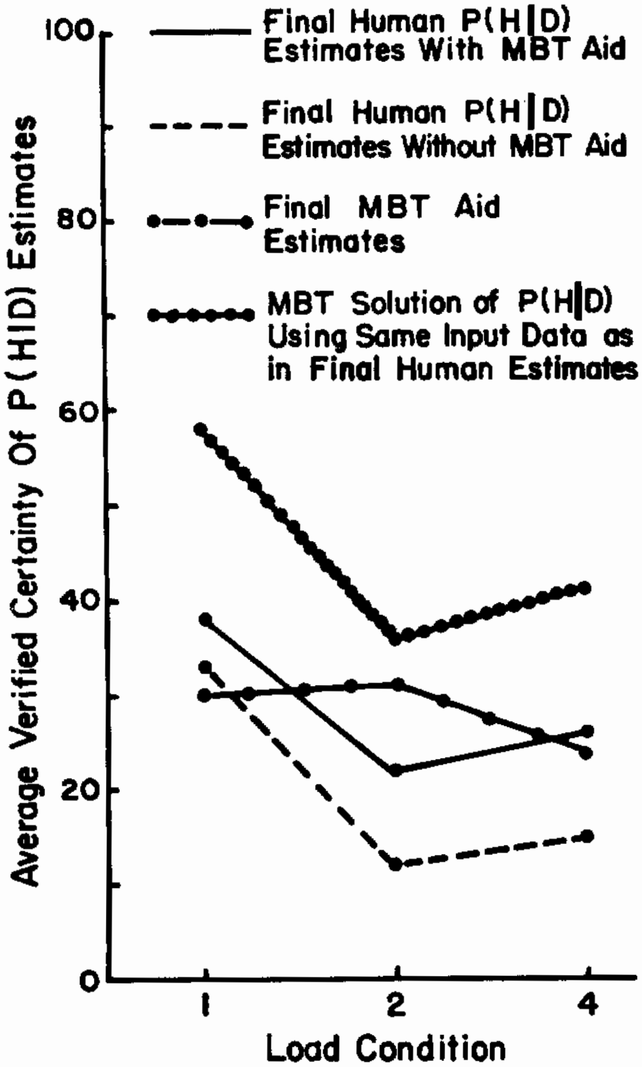


Figure 3. Human and MBT Estimates of P(H|D) Shown as a Function of System Load.

the MBT-aid solutions were based upon an incomplete set of attribute data. The MBT-aid accuracy of present concern was the last MBT-aid solution provided for the CO before he made his final P(H|D) estimations. In order to mitigate some of the possible confusion between these two MBT solutions, attributable to the time-dependent character of the stimulus environment, we illustrate in figure 4 the general location in time at which the two different types of MBT solutions were generated with respect to any particular developmental grouping.

In figure 4, T_1, T_2, and T_3 refer to MBT aids requested by the CO. Observe that these three solutions are based upon incomplete data sets. The designation T_3 represents the final MBT-aid solutions requested by the CO and is the solution of interest in this section. The designation T_4

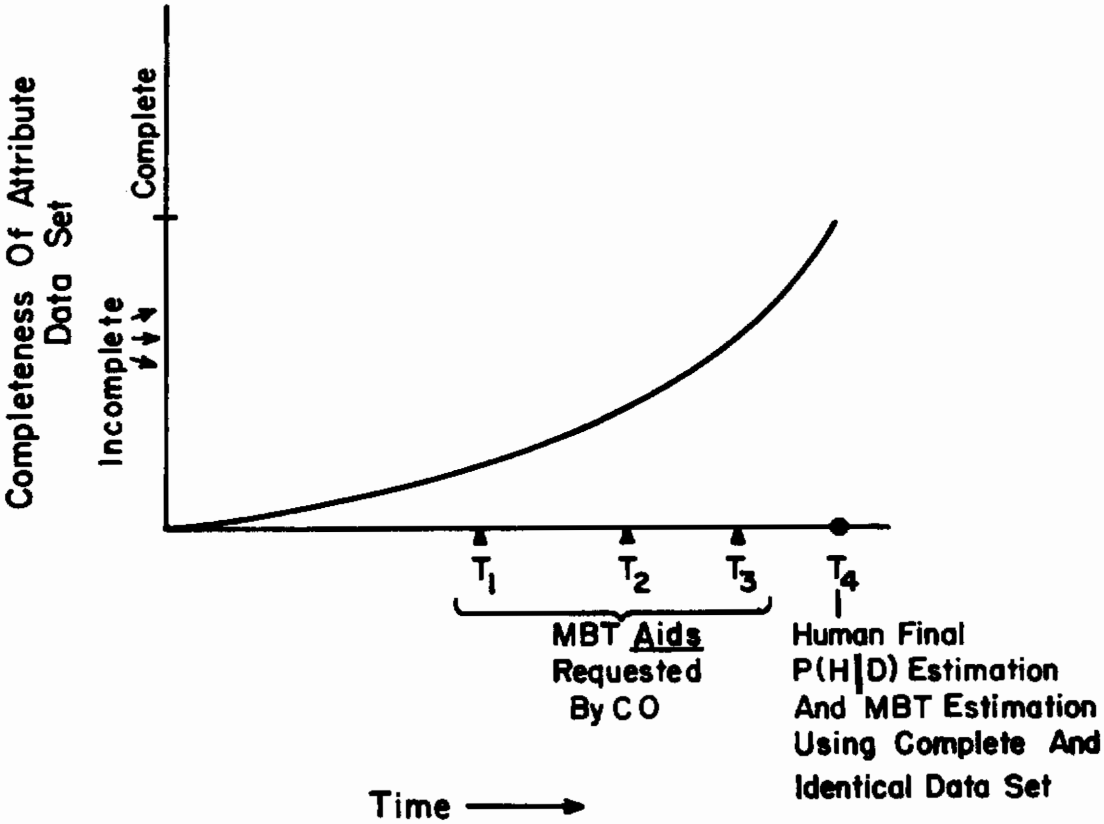


Figure 4. Location in Time of Various MBT Solutions for a Certain Developmental Grouping.

refers both to the final P(H|D) estimate made by the CO and to the P(H|D) estimate provided for the experimenter on the basis of an MBT solution which incorporated the same final data set which the CO used. (These two classes of estimates were the subject of result sections 1 through 3 above.) In terms of the dichotomous scores the final MBT aid was correct on 8 out of the 28 possible developmental groupings in the aid condition of the experiment.

The verified certainty measures of the final MBT-aid solutions requested by the CO are shown in figure 3. The input data upon which this aid was calculated varied in completeness and accuracy with the time interval (T_4 - T_3) shown in figure 4. When T_4 - T_3 was short, the aid was based upon a nearly complete set of attribute data and when long, upon only a partially complete set. The relationship of the length of this interval to input data accuracy is more complicated. As will soon be discovered, there was an important class of attribute data which became available only near T_4 in figure 4. Unfortunately, these data were often less accurate than other classes of data which were available throughout a greater portion of the time sequence. Presumably, the final MBT aid had the greatest effect on the final P(H|D) estimates made by the CO in the aid condition of the

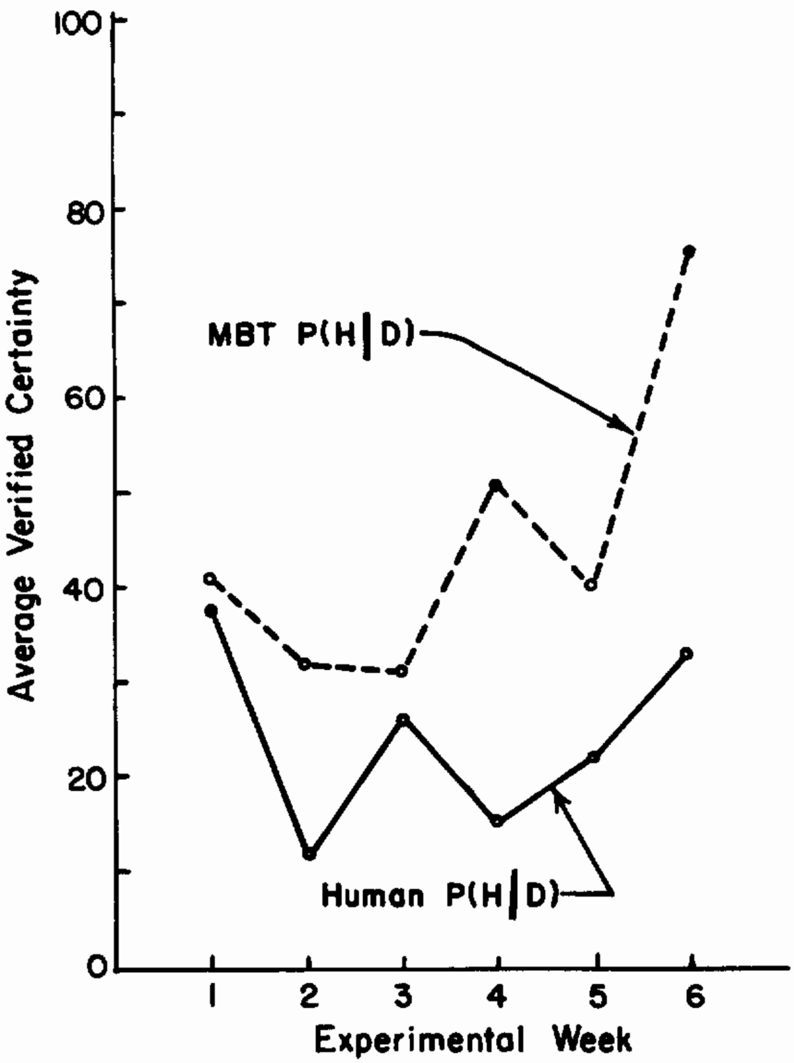


Figure 5. Comparison of Overall Human-vs. MBT-Generated Posterior Probabilities Given Identical Input Data Sets.

experiment, although there is no way to demonstrate that this was in fact the case. The overall average verified certainty scores for the final MBT aid and the human final P(H|D) estimates in the aid condition were both .28. It appears, therefore, that over all three load conditions the CO was able to take advantage of the degree of accuracy offered by the final aid although the degree to which this aid influenced his final estimates cannot be determined. Interim accuracy comparison of the MBT aids across developments of the same buildup duration was not possible because of the staggered manner in which the aids were requested by the CO.

5. <u>Comparison of Overall Human- versus MBT-Generated P(H|D) Given Identical Input Data Sets</u>: Figure 5 indicates the verified certainty scores

of the final P(H|D) estimates made by the CO and the MBT using identical input data throughout the entire experiment. The difference between these two types of estimates was statistically significant [$p < .002$, Wilcoxon Test (ref. 10)].

6. Accuracy of Interim P(H|D) Estimations Made by Human and MBT: The results listed in sections 1 through 5 all refer to the P(H|D) estimates made on the basis of the final data observed by the operator team as the developmental groupings reached their terminal stages. Yet to be considered is the accuracy of the human- and MBT-generated posterior probabilities made sequentially at intermediate stages during the buildup of the developmental groupings. It must be remembered that these interim estimates were made on the basis of incomplete and presumably less accurate data, particularly in the early stages of buildup. Also, recall that the developmental groupings terminated after differing amounts of buildup time. For the purposes of the present experiment, developmental grouping scripts were selected which depicted groupings building up over two, three, four, and five days. Appendix IV shows the accuracy of the sequence of P(H|D) judgments made by both human and MBT under the aid and no-aid conditions of the experiment. The MBT estimates shown in the various graphs refer to the solutions provided for the experimenter and calculated on the basis of the same data which the CO was using at the time of his own estimates. Inspection of these graphs reveals little in the way of systematic differences in human interim P(H|D) estimation between the aid and no-aid conditions of the experiment. Only for two-day developments does the human achieve earlier P(H|D) accuracy with the MBT aid. Notice, however, that the MBT solutions (using the same data as the CO) show an early superiority over the human estimations for those occasions on which no MBT aid was given to the CO. Also observe that at the terminal stages of the longer developmental periods (four- and five-day buildups) there is a distinct reduction in P(H|D) accuracy for both human and MBT estimations. This curious result is accounted for when we discuss the accuracy of the attribute data estimations.

B. Attribute Data Accuracy

In section V of the report was a description of the way in which four of the ISOs developed the 25 types of data upon which the CO's P(H|D) estimates were made. These data were also in the form of probabilistic estimates. For each of the 25 data classes the ISOs estimated the probability of the data class being in the various states or levels allowed in the class. This estimation was done for each developmental grouping on an interim and final basis. Verified certainty measures were taken by comparing the correct state of each data class with the probability value assigned in that state by the ISO. Only the final data estimates for each of the 56 developmental groupings were scored. The interim data estimates were not scored because they were largely incomplete, particularly at the early stages of a buildup. Since we concerned ourselves primarily with the final estimates of P(H|D), an extended analysis of the large amount of interim data was not considered of major importance. Before considering the results regarding data estimation, recall that provision was made for the CO to revise data estimates given to him by the ISOs. Therefore the accuracy of the data produced by the ISOs and the accuracy of the data actually used by the CO in making his P(H|D) judgments must be considered. The following results are of importance.

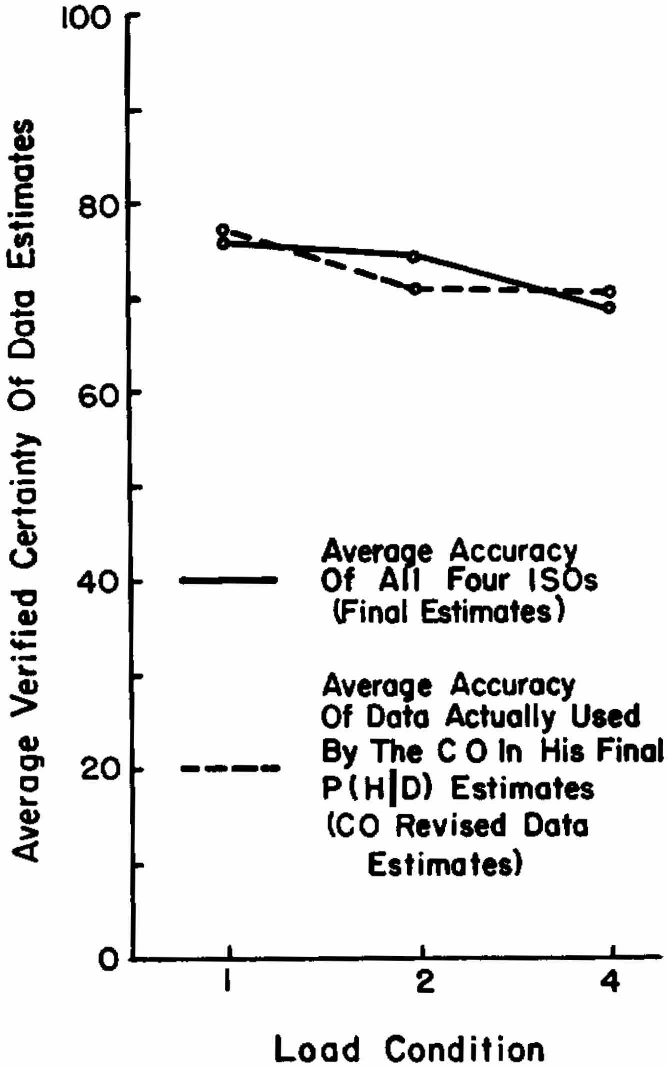


Figure 6. Overall Attribute Data Accuracy as a Function of System Load.

1. Overall Attribute Data Accuracy as a Function of System Load: In figure 6 the overall accuracy of the final data estimates produced by the ISOs and the accuracy of the CO's final revised data estimates are shown. The difference in accuracy at the three load levels for both measures is very slight. Taking the largest accuracy difference due to load, which happened to be the difference between load conditions 1 and 4 for the ISO data estimates, it was found, using the Fisher Method of Randomization for Matched Pairs (ref. 4), that the accuracy difference did not meet the conventional requirements for statistical significance ($p > .10$).

2. Accuracy of the Four Major Data Classes as a Function of System Load: In order to interpret some of the P(H|D) estimation results, we needed a finer measure of data estimation accuracy than the overall measures

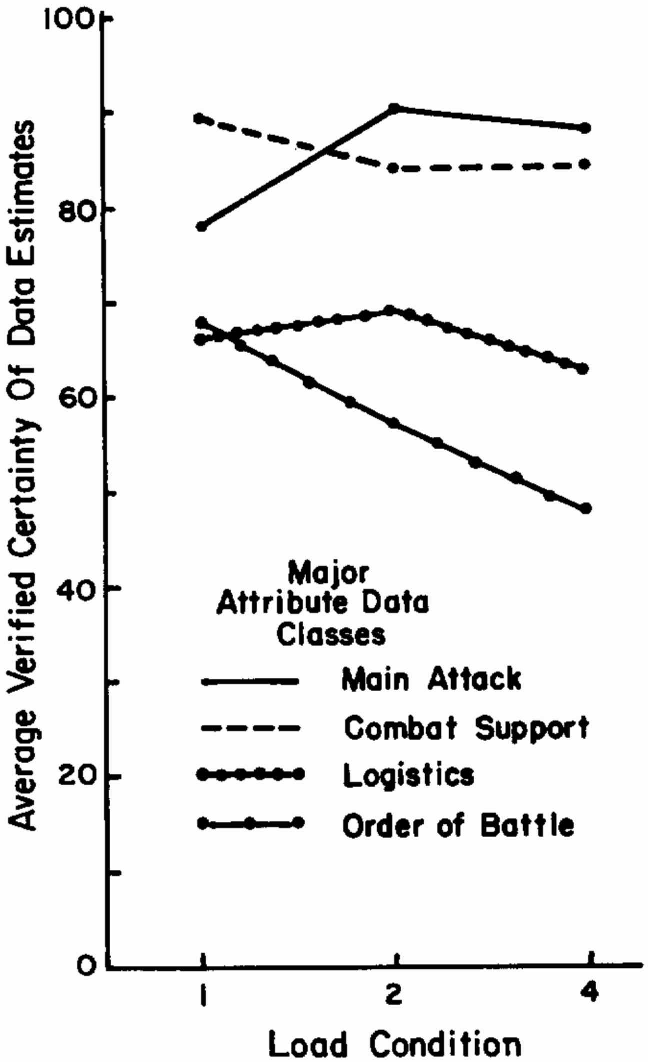


Figure 7. Accuracy of Major Attribute Classes as a Function of System Load.

described above. A finer measure of data accuracy can be obtained if we consider each of the four major data classes developed by the ISOs. The accuracy of these four classes is shown at each load level in figure 7. First, it is apparent that the different classes of data were estimated with different degrees of accuracy. This is to be expected for two reasons. First, the data in each of these four major classes were developed by a different ISO (see section V, page 23). In addition, the data estimation tasks required in these four classes were somewhat different. For example, several of the specific attributes within the logistics and order of battle data classes involved careful observation of spatial and temporal arrangements of interrelated events within a developmental grouping. The attributes in the main attack and combat support data classes, however, involved only identification and tracking of specific concentrations of weapons and vehicles. The second feature of note in figure 7 is the rather sharp decrease in order of battle data accuracy as load was increased. Due to the time-dependent character of the appearance of most of the specific attributes within the order of battle class, the ISO's estimations of these data were made very late in the buildup of each developmental grouping. At the high load levels the task of making rapid decisions as to the level of each of the order of battle attribute data sets (nearly all of which required careful spatial and temporal discriminations) was made very difficult because of the increased amount of activity in Aggressor territory. The fact that these order of battle attribute data became available at later stages of buildup and the fact that these data were considerably less accurate than those in other classes can account in large measure for the decrease

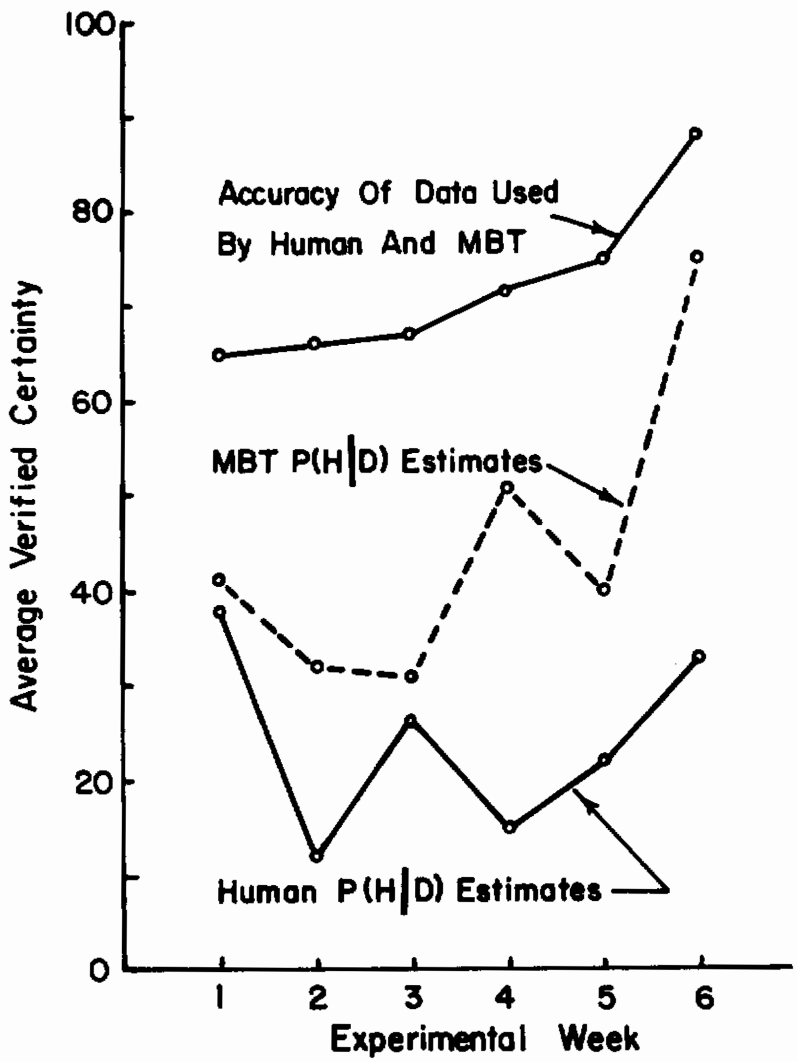


Figure 8. Final Attribute Data and P(H|D) Estimates.

in accuracy of the interim P(H|D) estimates made at the terminal stages of the buildup sequence [see P(H|D) estimation result section 6, page 41].

3. The Relationship between Final Attribute Data Estimation Accuracy and Final P(H|D) Estimation Accuracy: During the course of the experiment the overall final data estimates provided by the ISO team showed a considerable increase in accuracy. Figure 8 shows human and MBT final P(H|D) estimations in the face of this increased data accuracy. First of all, it appears that there was no systematic increase in accuracy of the human P(H|D) estimates as the data accuracy increased except perhaps at the later stages of the experiment. Second, the MBT shows an earlier and more dramatic increase in P(H|D) accuracy as the input data increased in accuracy.

VIII. DISCUSSION AND INTERPRETATION OF RESULTS

Before discussing the results of the experiment, let us summarize briefly some of the constraints and novel features in our experimental situation which will have the greatest effect upon our interpretation of these results and upon our attempt to compare these results with those obtained in previous related experiments. First of all, the real-time characteristics of the stimulus environment must be kept in mind, particularly when attempting to compare measures which might have been taken at differing relative positions in time. Second, the experimenter provided knowledge of results (for the previous day) at the start of each experimental session. The CO was told what strategy Aggressor used in each of the developmental groupings seen on the previous day. This implies a situation in which it was always possible to observe what actually happened in order to verify previous hypotheses. Third, it is obvious that sweeping generalizations cannot be made on the basis of data from a single subject. In this introductory experiment it was not possible to utilize more than one operator team. In future experiments, however, it should be possible to achieve replication of the primary decision data. Finally, some of the particular features of the Bayesian paradigm used in the experiment should be noted. Dodson's MBT paradigm (ref. 5) provided the basis for the way in which the stimulus environment was structured. This paradigm provides for the use of multiple hypothesis and data categories. In addition, it allows one to treat data observations in probabilistic terms, i.e., to allow for observational uncertainty. The true environmental contingency rules [P(D|H)] were provided for both the CO and the MBT in the generation of P(H|D) estimates. In any real-life situation, however, the true P(D|H) values would never be known, but only estimated. Human estimation of P(D|H) values in a complex environment will be a feature of later experiments in the series.

One major result of the experiment was that the MBT, used as a hypothesis-selection aid, did not induce any compelling superiority in the CO's performance over the condition in which there was no such aid. Several factors can account for this finding. First, the MBT aid itself achieved no astounding level of accuracy as figure 3 indicates. Second, it is highly conceivable that the CO viewed the aid (over which he had no control) with some suspicion, preferring to rely on his own judgment. Unfortunately, there is no way of showing the precise extent to which he relied upon the aid. We did, however, record the fact that the CO requested MBT aid several times for every developmental grouping seen in the aid condition of the experiment. From figure 3 it does appear that the CO made somewhat greater use of the aid under the higher load levels when he was required to make more rapid estimations. The data in Appendix IV does show a slightly earlier recognition of correct hypotheses on the part of the CO under the aid condition.

Let us now compare the accuracy of human- and MBT-generated posterior probabilities. First, consider the human and MBT $P(H|D)$ estimates made on the basis of the final attribute data obtained just prior to the termination of a developmental grouping. The superiority of these MBT solutions (which was statistically significant) is shown in figures 3 and 5. These data offer an increment of support for the Edward's hypothesis that computer-generated $P(H|D)$ estimations based on Bayes' theorem are superior to human estimations. One interpretation, of course, is that the MBT was able to extract more certainty from the observed data than our human decision maker. An interpretation of the results of the interim $P(H|D)$ estimates by human and MBT is more difficult. In five of the six instances cited in Appendix IV, the MBT $P(H|D)$ estimates were more correct earlier in the sequence of estimates made for developmental groupings of varying buildup duration, but MBT superiority was maintained throughout only one of these sequences. This curious result can perhaps be accounted for by considering that the MBT and human both incorporated the true $P(D|H)$ values in their respective conditions. Now recall that the order of battle data which became available at later stages in the sequence was of considerably lower accuracy than the other three major classes (figure 7). It appears that the MBT was more sensitive to the disparate condition involving correct $P(D|H)$ "rules" and these incorrect data. Under these conditions the MBT can be expected to provide incorrect $P(H|D)$ estimates. Similarly, but in the opposite direction, the MBT appeared to be more sensitive to increases in data accuracy. Figure 8 shows this relationship.

The effects of increased system load can be interpreted in several ways. First, the overall effect of increased load upon the information-processing tasks performed by the ISOs was negligible except in the case of the order of battle ISO. Increased load, as noted previously, made his tasks involving spatial and temporal discriminations more difficult and thereby greatly reduced the accuracy of his data. In turn, the reduced accuracy of these order of battle data produced the results discussed in the preceding paragraph. The effects of load upon the $P(H|D)$ estimation task of the CO is shown in figure 3. With a greater number of time-consuming $P(H|D)$ estimations required under the higher load levels, the CO was forced perhaps to rely upon the MBT aid to a greater extent.

The discussion will now be concluded by attempting (insofar as the data permit) to answer the following question: under the same conditions applicable in the present system design, what would happen if the hypothesis-selection function now performed by a human CO were automated?

1. Reasonable interim estimates of the strategy to be used by Aggressor in each developmental grouping might be obtained earlier.

2. Final automated estimates of Aggressor strategy (reported 1/2 hour before an attack is estimated to occur) may well be superior to human estimates.

In the face of the evidence gathered under the conditions of this experiment an unqualified endorsement of computer-implemented hypothesis selection can scarcely be offered. However, the problem certaintly merits further and more intensive study.

REFERENCES

1. Aggressor, The Maneuver Enemy, Field Manual FM 30-101, Department of the Army, May 1959.

2. Aggressor Military Forces, Handbook on, Field Manual FM 30-102, Department of the Army, June 1960.

3. Aggressor Order of Battle, Field Manual FM 30-103, Department of the Army, September 1960.

4. Bradley, J. V., Distribution-Free Statistical Tests, WADD Technical Report 60-661, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio, August 1960.

5. Dodson, J. D., Simulation System Design for a TEAS Simulation Research Facility, Report AFCRL 1112, PRC R-194, Planning Research Corporation, Los Angeles, Calif., 15 November 1961.

6. Edwards, W., "Dynamic Decision Theory and Probabilistic Information Processing," Human Factors, Vol 4, pp 59-72, 1962.

7. Edwards, W., Probabilistic Information Processing in Command and Control Systems, ESD Technical Documentary Report 62-345, Electronic Systems Division, L. G. Hanscom Field, Bedford, Mass., March 1963. (ASTIA Document No. AD 3780-12-T)

8. Feallock, J. B., and G. E. Briggs, A Multiman-Machine System Simulation Facility and Related Research on Information-Processing and Decision-Making Tasks, AMRL Technical Documentary Report 63-48, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, June 1963.

9. Kaplan, R. J., and R. J. Newman, A Study in Probabilistic Information Processing (PIP), Technical Memorandum 1150, System Development Corporation, Santa Monica, Calif., April 1963.

10. Siegel, S., Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

APPENDIX I
ALTERNATIVE AGGRESSOR STRATEGIES (H_i)

A. Actual Attacks

1. 100-mile penetration depth, double pincer
2. 100-mile penetration depth, multiple penetration*
3. 50-mile penetration depth, double pincer
4. 50-mile penetration depth, multiple penetration*
5. 25-mile penetration depth, double envelopment
6. 25-mile penetration depth, single envelopment
7. 25-mile penetration depth, penetration
8. 10-mile penetration depth, double envelopment
9. 10-mile penetration depth, single envelopment
10. 10-mile penetration depth, penetration

B. Rehearsals

11. 100-mile penetration depth, double pincer
12. 100-mile penetration depth, multiple penetration*
13. 50-mile penetration depth, double pincer
14. 50-mile penetration depth, multiple penetration*
15. 25-mile penetration depth, double envelopment
16. 25-mile penetration depth, single envelopment
17. 25-mile penetration depth, penetration
18. 10-mile penetration depth, double envelopment
19. 10-mile penetration depth, single envelopment
20. 10-mile penetration depth, penetration

---

* These multiple penetration strategies were incorrectly termed "multiple pincer" strategies in Appendix IV to the Feallock and Briggs report (ref. 8, p. 137).

APPENDIX II
ATTRIBUTE DATA CLASSES

Data Class	Number of Possible States	Description
I. Mechanized Rifle Battalions	8	The states or levels of data classes I through XII all refer to numbers of battalions or squadrons of the type indicated by the various data class labels indicated in column 1. The first level in every data class refers to zero battalions or squadrons.
II. Medium Tank Battalions	7	
III. Heavy Tank Battalions	6	
IV. Artillery Battalions (range up to 10,000 meters)	3	
V. Artillery Battalions (range up to 20,000 meters)	3	
VI. Artillery Battalions (range up to 30,000 meters)	3	
VII. Rocket Battalions	3	
VIII. Intermediate Range Ballistic Missile Battalions	3	
IX. Ground Reconnaissance Battalions	2	
X. Tactical Air Support Squadrons	4	
XI. Aerial Reconnaissance Squadrons	4	
XII. Surface to Air Missile Battalions	4	

Data Class	Number of Possible States	Description
XIII. Units of Fire for Infantry and Armoured Units (Main Attack Forces)	4	This data class refers to the amount of ammunition being carried by road or rail convoys which provide logistics support for infantry and armoured units.
XIV. Units of Fire for Artillery, Missile, and Rocket Units (Combat Support Forces)	4	Refers to the amount of ammunition being carried by supply convoys for these three classes of units.
XV. Dispersal Distance between Supply Units	4	Distance in miles between terminal positions of supply convoys.
XVI. Supply Timing for Main Attack Units	3	Refers to temporal order of appearance of supply units and units being supplied.
XVII. Supply Timing for Combat Support Units	3	Same as XVI.
XVIII. Terminal Activity Zone	5	Refers to the distance in miles from the border of the most forward units in a developmental grouping.
XIX. Terminal Activity Development Pattern	4	Refers to the configuration or placement of units laterally along the border of contention after these units have reached their terminal positions.

Data Class	Number of Possible States	Description
XX. Attack Position Lateral Dispersion	5	Dispersal distance in miles along a border of contention of an entire development grouping.
XXI. Attack Position Depth	4	Distance in miles involved in the placement of forces perpendicular to a border, i.e., the distance between the most forward unit in a grouping and rearmost unit.
XXII. Attack Buildup Timing	3	Refers to the temporal order of appearance at terminal positions along a border of contention of main attack units and combat support units.
XXIII. Transportation Methods	3	Refers to the combination of road, rail, and air facilities used to transport Aggressor units in any developmental grouping.
XXIV. Ground Transportation Speed Class	5	Road and rail convoy speed during the buildup of a developmental grouping.
XXV. Developmental Period	6	Length of time in days from beginning to termination of a buildup of a developmental grouping.

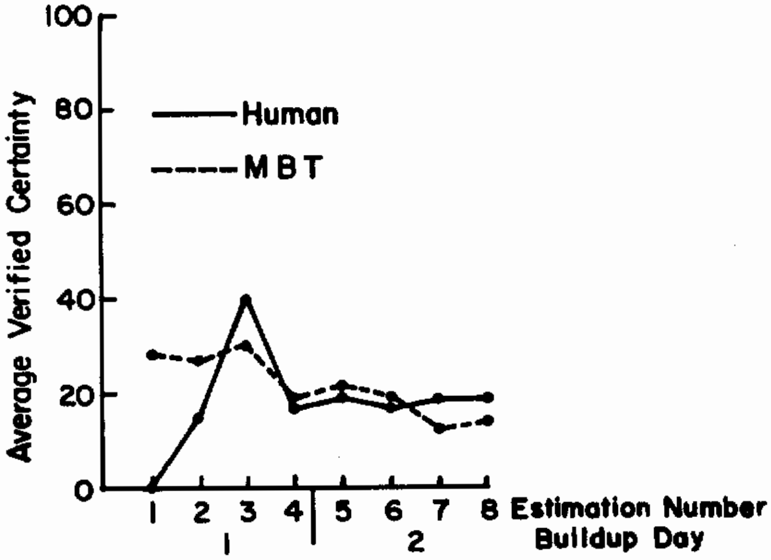
APPENDIX III
INTELLIGENCE STAFF OFFICER INFERENCE TABLES

Table	Contingency Relation
1. Ground Forces Vehicle and Mobile Weapon Allocation	Describes frequency count of each type of vehicle and/or mobile weapon in each type of Aggressor army unit.
2. Road March	Describes frequency count of each type of vehicle and/or mobile weapon and the order of march of these elements in road convoy for each type of Aggressor army unit.
3. Rail Convoy Description	Lists total number of trains used in day or night transportation of any Aggressor army unit. Also lists constituency of each type of train in terms of type and number of coaches, sleeping cars, kitchen cars, baggage cars, boxcars, and 40-, 46-, 50-1/2-, and 56-ft. flatcars.
4. Road Space for Road Convoys	Describes amount of road space (in yards) taken up by each Aggressor army unit at various road convoy speeds.
5. Bivouac Area Size	Describes amount of bivouac area taken up by each Aggressor army unit for various dispersal distances (between each element) and various lengths of time (in days) spent in bivouac.
6. Air Transportation Requirements for Surface Forces Airlift	Describes the number of airborne vehicles (transport aircraft, helicopters or ground effects machines) necessary to move various Aggressor army units.
7. Equivalent Cargo Capacity of Aggressor Transport Aircraft	Describes equivalent cargo-carrying capacities of various combinations of air transports, helicopters, and ground effects machines.

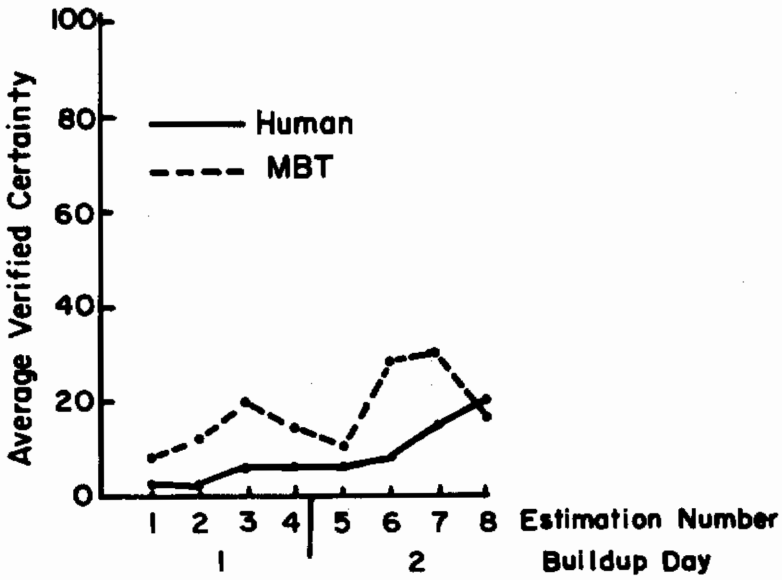
Table	Contingency Relation
8. Supply Requirements: Supplies Shipped via Road Convoy	Describes the number of various sizes of trucks, tankers, and semis necessary to provide rations and water for various amounts of time (in days) and ammunition in various quantities for each Aggressor army unit.
9. Supply Requirements: Supplies Shipped via Rail Convoy	Describes the number of boxcars, tank cars, and flatcars necessary to provide rations and water for various amounts of time (in days) and ammunition in various quantities for each Aggressor army unit.
10. Ground Forces Movement Parameters	Describes minimum, normal, and maximum rates of speed for Aggressor's wheeled, towed, and tracked vehicles on roads of poor, medium, or good quality.
11. Aircraft Allocation	Describes frequency count of each type of aircraft in each type of Aggressor air unit.

APPENDIX IV
HUMAN AND MBT INTERIM P(H|D) ACCURACY

I. Developmental Groupings of Two-Day Duration

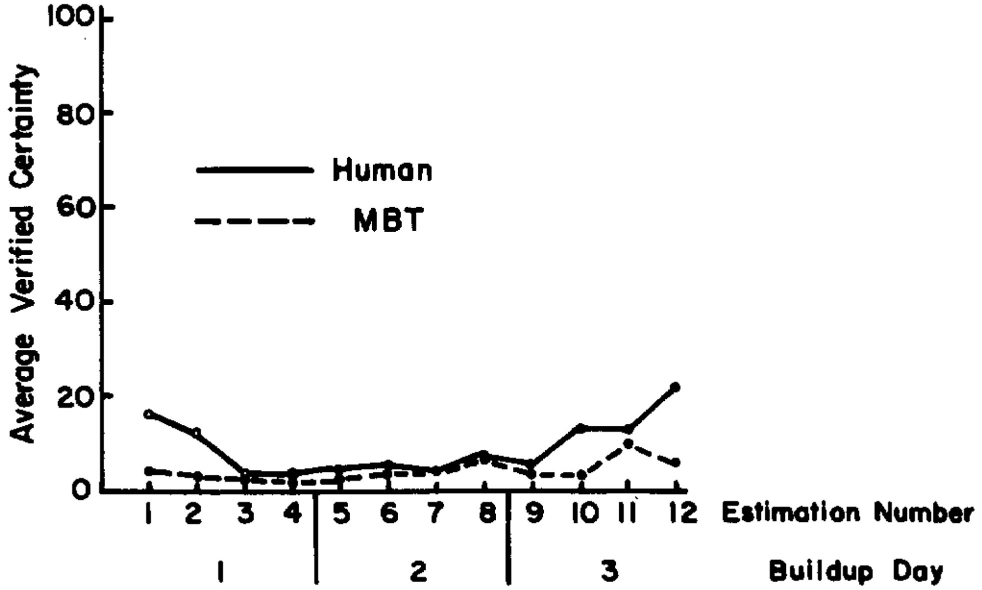


With MBT Aid

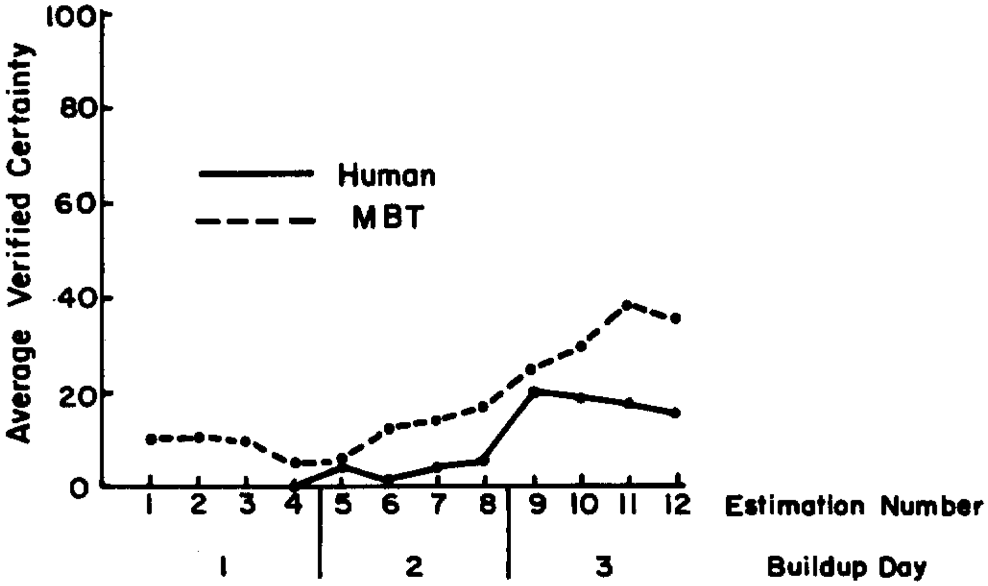


Without MBT Aid

II. Developmental Groupings of Three-Day Duration

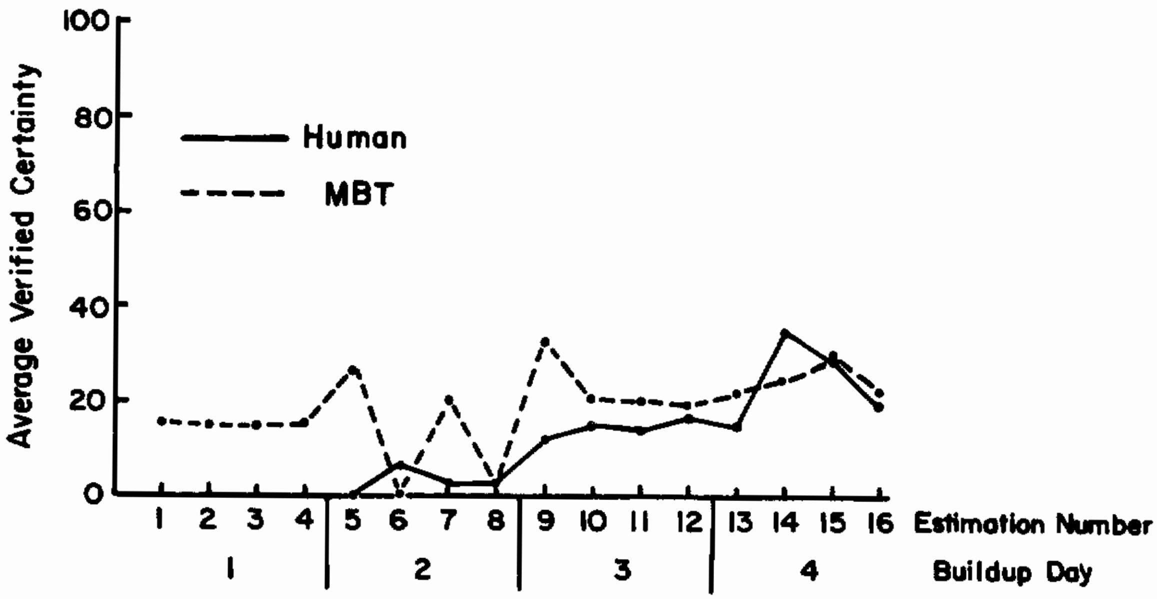


With MBT Aid

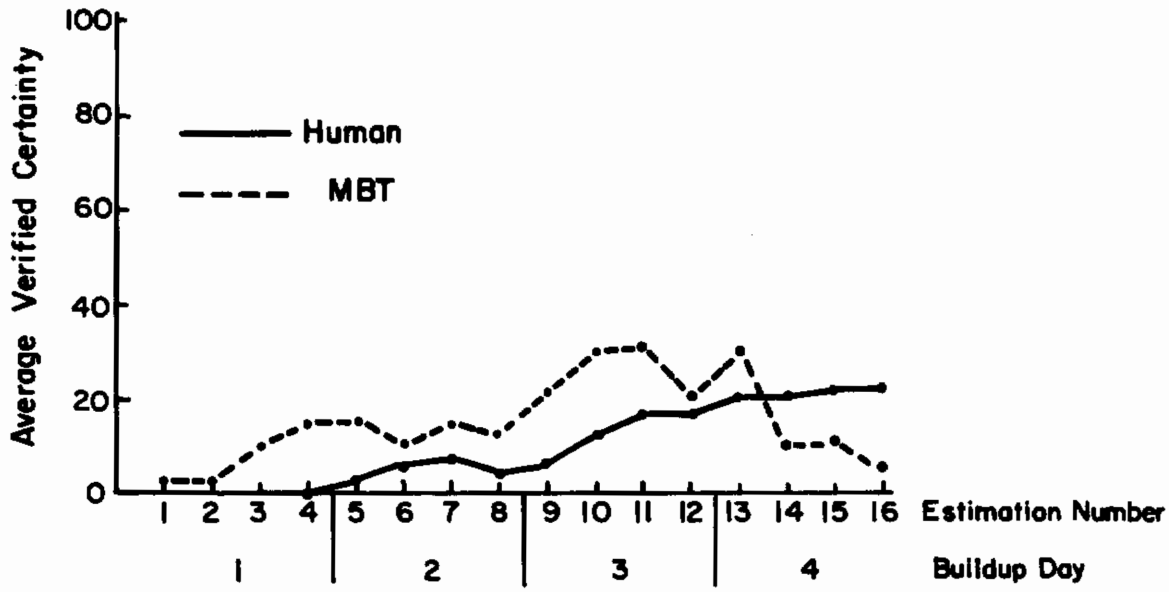


Without MBT Aid

III. Developmental Groupings of Four-Day Duration

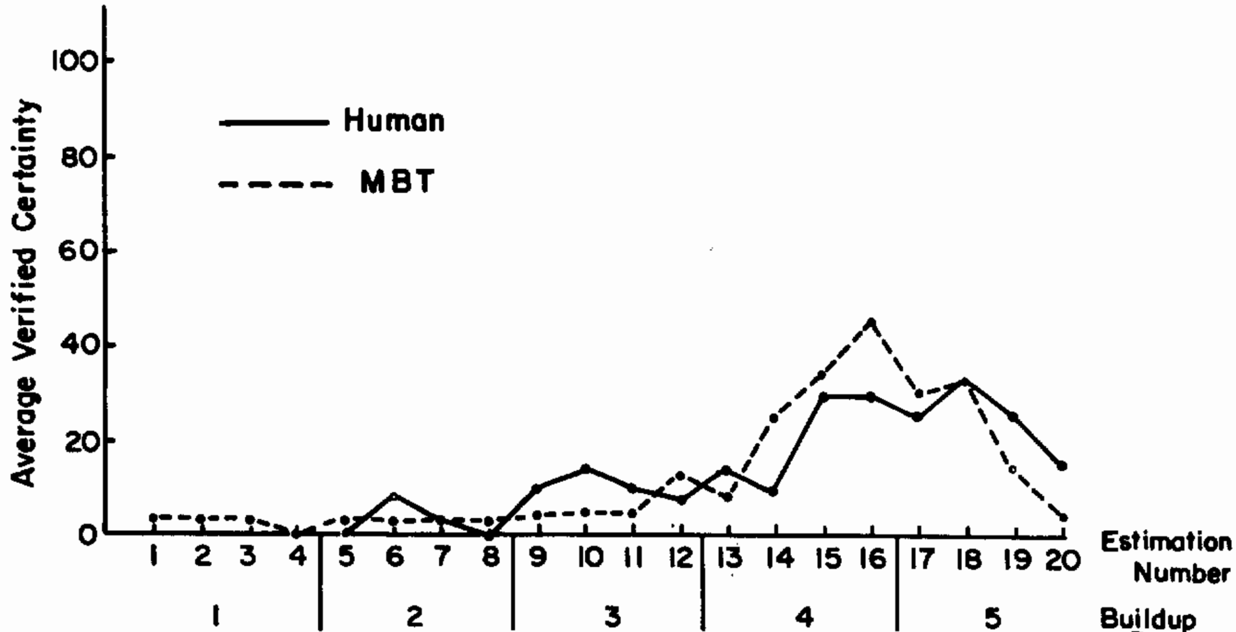


With MBT Aid

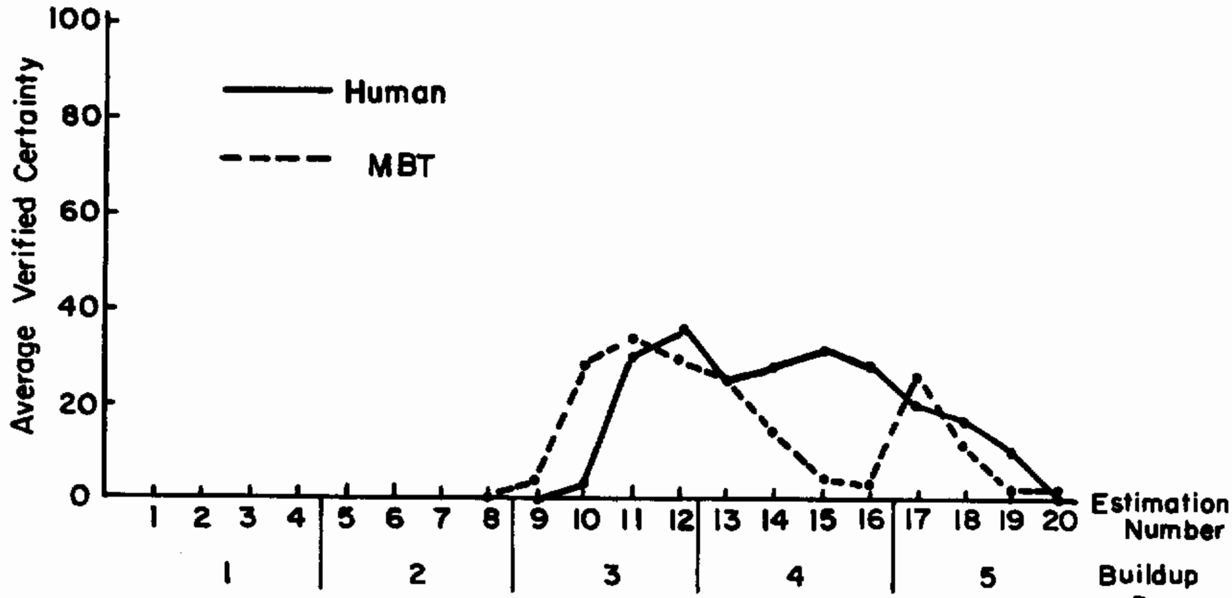


Without MBT Aid

IV. Developmental Groupings of Five-Day Duration



With MBT Aid



Without MBT Aid

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Laboratory of Aviation Psychology Ohio State University, Columbus, Ohio	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP N/A

3. REPORT TITLE

AN APPLICATION OF BAYES THEOREM AS A HYPOTHESIS-SELECTION AID IN A COMPLEX INFORMATION-PROCESSING SYSTEM

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Final report, 15 March 1963 - 31 January 1964

5. AUTHOR(S) *(Last name, first name, initial)*

Southard, Jack F.
Schum, David A.
Briggs, George E.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1964	64	10

8a. CONTRACT OR GRANT NO. AF 33(657)-10763 b. PROJECT NO. 7184 c. Task No. d. 718403	9a. ORIGINATOR'S REPORT NUMBER(S) AMRL-TDR-64-51 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* None

10. AVAILABILITY/LIMITATION NOTICES

Qualified requesters may obtain copies of this report from DDC.
Available from the Office of Technical Services, Department of Commerce, Washington, D.C. 20230. Price per copy is $1.75.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Aerospace Medical Research Laboratories Wright-Patterson Air Force Base, Ohio

13. ABSTRACT The first of a series of experiments investigating the value of automated hypothesis-evaluation aids in multiman-machine systems devoted to assessing or diagnosing threat is described. In the experiment, an eight-man team evaluated threats posed by a hypothetical Aggressor. The team made these evaluations on the basis of intelligence information gathered during simulated reconnaissance over-flights of Aggressor's territory. IBM 1401 and 7090 computer facilities generated the highly complex, real-time stimulus environment (data base) which is described in detail. The primary output of the threat-evaluation team was the commanding officer's posterior probabilities estimates as to Aggressor's most likely hostile strategies. During half of the experimental trials, the commander had access to computer-produced posterior probabilities based upon a modification of the Bayes Theorem. The major experimental issue was whether or not these would aid the commander in his hypothesis evaluation. Also investigated was the effect of data-processing load upon system operation. Although some improvement in the posterior probabilities estimates resulted from the commander's having access to the hypothesis-evaluation aid and this improvement became more pronounced as system load increased, the main-order effect of access to the aid was not found to be statistically significant. Throughout the entire experiment, solutions of posterior probabilities based upon the Modified Bayes Theorem were compared with the commander's estimates. The Bayesian solutions were significantly superior to the commander's.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Decision Making						
Simulation						
Mathematical Prediction						
Information System						
Human Engineering						
Psychology						
Bayes Theorem						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ______________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ______________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ______________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.