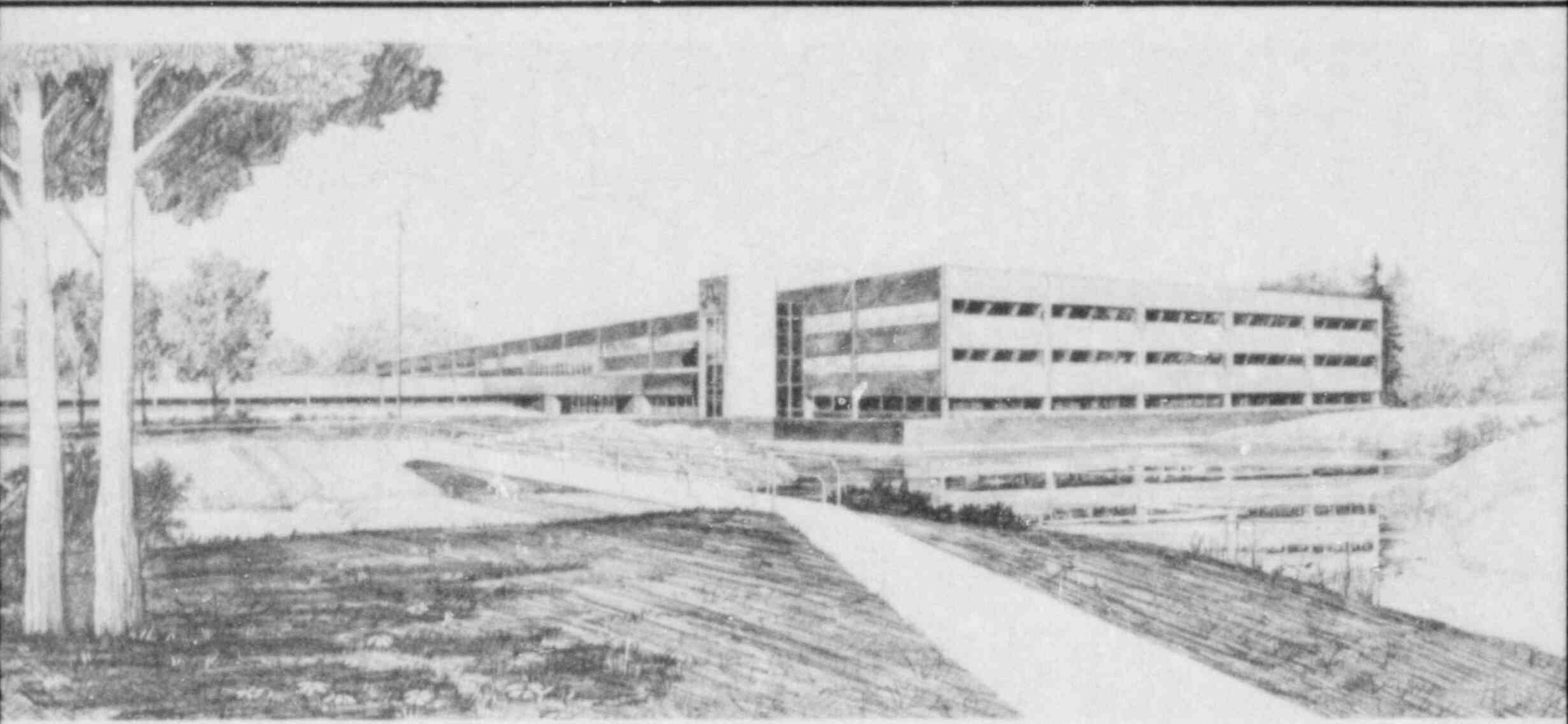
NUREG/CR-2916
EGG-2214



Idaho National Engineering Laboratory
Operated by the U.S. Department of Energy

An Empirical Examination of Evaluation Methods for Computer Generated Displays: Psychophysics

Rohn J. Petersen
Robert L. Smith
William W. Banks
David I. Gertman

September 1982

Prepared for the
U.S. Nuclear Regulatory Commission
Under DOE Contract No. DE-AC07-76IDO1570

EG&G Idaho

Available from

GPO Sales Program
Division of Technical Information and Document Control
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555

and

National Technical Information Service
Springfield, Virginia 22161

NUREG/CR-2916
EGG-2214
Distribution Category: RX

AN EMPIRICAL EXAMINATION OF EVALUATION METHODS FOR COMPUTER GENERATED DISPLAYS: PSYCHOPHYSICS

Rohn J. Petersen
Robert L. Smith
William W. Banks
David I. Gertman

Published September 1982

EG&G Idaho, Inc.
Idaho Falls, Idaho 83415

Prepared for the
U.S. Nuclear Regulatory Commission
Washington, D.C. 20555
Under DOE Contract No. DE-AC07-76ID01570
FIN No. A6308

ABSTRACT

An investigation was performed to evaluate the perceptual aspects of safety parameter display systems (SPDSs) in nuclear power plant control rooms. Three SPDS configurations (star, bars, and meters) were evaluated in a series of four experiments. Subjects for the investigation were qualified nuclear plant operators and engineers at EG&G Idaho, Inc., at the Idaho National Engineering Laboratory. The techniques reported herein were demonstrated to be sensitive to differences in human performance which result from using different display formats to present the same safety parameter information.

FIN No. A6308—CRT—Display Design and Evaluation

SUMMARY

Modern computer technology is currently being applied to the control of nuclear power plants. The United States Nuclear Regulatory Commission is tasked with regulating the nuclear power industry and, therefore, is concerned with the infusion of computer technology in nuclear control rooms. In addition to the hardware aspects of applying computer technology to the nuclear application, the human factors aspect of introducing this technology is also a major topic for consideration in formulating acceptance criteria. In formulating acceptance criteria, a definitive set of minimum standards for the human factors aspects of computer systems must be developed and tested. These minimum standards are not completely extractable from the existing literature. Much of the data needed is not readily available and, hence, must be derived from research conducted specifically to solve specific standards issues.

The research reported herein is a part of an extensive methods development program which focuses initially upon developing evaluation methods for safety parameter display systems (SPDSs). Several methods have been identified including, psychophysical methods, multivariate rating scales, checklists, and decision analysis using plant simulations. The overall goal of this program is to examine the interrelationships of these methods to develop an overall, cost effective technique for evaluating computerized displays. This report describes the psychophysical methods which can be used to objectively evaluate the effect of display format on the perceptual performance of nuclear power plant operations. The determination of acceptance criteria for display formats is one of the critical questions related to establishing minimum standards for computer displays.

The perceptual aspects of SPDS design were evaluated in four experiments. Experiments 1 and 2 address an SPDS function of alerting the operator to the occurrence of abnormal plant parameters. The experiments were conducted using an apparatus (tachistoscope) which precisely controls the exposure duration of the display. The display was initially exposed to the subjects for 5 ms. Following each exposure, the subject was asked if any of the parameters were abnormal and the response was recorded. The exposure duration was then increased 10 ms, and the process was repeated. This continued until the subject responded correctly on 81 consecutive trials. This procedure was repeated with three different display formats, a circular profile (star), deviation bars, and clustered meters. The data were analyzed using a signal detection procedure, to get an unbiased view of the subject's perceptual sensitivity, and analysis of variance (ANOVA), to test the experimental hypotheses. The major conclusion from these experiments was that the experimental paradigm is sensitive to performance differences imposed by the various display formats, even though information content was held constant.

Experiment 3 was conducted primarily as a training exercise for the subjects prior to Experiment 4. The design of Experiment 3 was such that some useful information regarding evaluation methodologies could also be gleaned. The basic design of the experiment was the same as for Experiment 2. However, the task of the subject was changed from a simple detection task to a more complicated localization task, where the subject must report the location of all abnormal parameters. The data from this experiment are interesting because they are almost inversely related to those of the first two experiments. In Experiments 1 and 2, the star display facilitated performance better than the bar and meter displays. In Experiment 2, however, performance was better on the meter display than on the other two. This result suggests that the task or situation in which the display is to be used is a critical consideration in developing evaluation methodologies

The fourth and last experiment conducted during this study investigated the subjects ability to identify the status of individual parameters as a function of different display formats. The experimental paradigm used in Experiment 4 was quite different than those used in the previous experiments. The subject's task was to listen to a digitized recording of a parameter name, visually locate that parameter in the display, and report the status of the parameter. In addition to the

dependent measures used in the other experiments, a chronometric measure was used. The analysis of the data from this experiment did not yield significant differences. This result was different from the data derived from a pilot study which found a significant difference. In comparing the pilot study and the experiment, it was noted that the subject's familiarity with the display formats changed radically from the pilot study (low familiarity) to the experiment (high familiarity).

In summary, the main conclusions from this research are:

1. The methods examined in this study are sensitive to changes in performance elicited by changes in display format
2. In developing methods for display evaluation, the effect of task (context) was found to be critical and must be controlled in future studies
3. Operator familiarity with the different display formats may be an important variable which must be specifically controlled when evaluating display formats.

ACKNOWLEDGMENTS

We would like to acknowledge the support and assistance provided by Harold Blackman, Mike Bray, Mike Clark, and Randy Shaw. They were instrumental in portions of the design of the stimuli, training of the subjects, collecting data, and statistical analysis. We would also like to thank the LOFT Program and the EG&G Operator Licensing Branch for allowing reactor operators to participate in the experiments.

CONTENTS

AN EMPIRICAL EXAMINATION OF EVALUATION METHODS FOR COMPUTER GENERATED DISPLAYS: PSYCHOPHYSICS

INTRODUCTION

Computer generated displays are potentially powerful and flexible tools for presenting information to nuclear power plant operators. Such displays may be used in "advanced" control rooms and in Safety Parameter Display Systems (SPDSs). The purpose of the research discussed herein is to develop and describe objective methods for evaluating the effect of introducing graphic displays in nuclear power plant control rooms. The data from this work can provide a basis for the Nuclear Regulatory Commission (NRC) to objectively evaluate specific licensee developed display designs. The work will also provide a base from which the NRC can judge licensee display design and implementation processes. The NRC is tasked with regulating the commercial power reactor industry. It is, therefore, of paramount importance to anchor prescribed regulations to empirically derived data.

Background

Recently, there has been an effort on the part of the utilities and commercial vendors to develop real-time computer based information systems. These systems are oriented toward improving plant operation and control. Presently, there are few human engineering design standards or test criteria which can be directly applied to assess the effectiveness of graphic displays as information carriers. By adopting a narrow view of the regulatory world, one can see the NRC as primarily concerned with setting minimum standards for the information offered the reactor operator. Merely presenting information is not sufficient grounds to assure that the information is incorporated in the operator's decision process, witness the Three Mile Island incident. We must assure not only that the information is available, but also, that it is presented in a manner that encourages the operator to use it in making operating decisions. The long-range focus of this research project is to provide the NRC with information which will assist in establishing criteria for evaluating automated information systems.

It is clear that data and reliable methodologies are needed to support the guidelines and evaluation efforts. Banks, Gertman, and Petersen (1981)[1] set forth some initial design criteria; however, the data to date cannot completely satisfy all of the requirements for quantitatively supported design guidelines. The data are simply not available in many areas which are germane to the total specification of design criteria. Hence, research projects such as this one will be conducted to fill in specific gaps in the data base.

There are a number of different levels of consideration that must be addressed when detailing guidelines for information display, including sensory/physiological, perceptual, and cognitive/decisionmaking. Each of these areas represent a broad spectrum of concerns and issues. Sensory/physiological variables represent issues relating the physics of light and sound as they interact with an operator's sensory system. This is an area where a great deal of data are available, and the application is fairly independent of the situation. Banks et al., focused heavily on this area in their initial work. The perceptual aspects of information display are not as well understood, although there are several laboratory methodologies available for use as comparative metrics. The perceptual aspects are the focus of the research reported here, while the cognitive/decisionmaking aspects will be addressed in a later phase of the research project.

Objectives

The general orientation of this research is to address applied behavioral questions in terms of objective measures of human performance. That is, we are asking questions concerning the nature of human-computer interaction in a manner that leads directly to an examination of man-machine communication under ridgidly controlled conditions. Control is the key to successful conduct of this type of empirical research; therefore, in this report, a strong emphasis is put on the factors related to rigorous experimental control. This report presents a series of experimental methods sufficient for addressing the initial levels (perceptual) of human information processing as they are impacted by graphic display design. This research is directed toward developing methods for evaluating the adequacy (from a safety perspective) of graphic SPDSs which are currently being developed for nuclear power plant control rooms.

For this research, it was assumed that the safety parameter display (SPD) would be a CRT display which provides the operator with a single focus for monitoring key plant parameters. In addition, there are at least three basic SPD functions which:

1. *Alert* the operator to the occurrence of abnormal plant conditions

2. Aid the operator in *identifying* specific *abnormal parameters*

3. Assist the operator in hypothesizing *diagnosis of plant conditions* based upon the relative values of the parameters.

The choice of a particular experimental paradigm is usually dependent on the type of question being asked. The following questions relating to the perceptual aspects of SPD design (Functions 1 and 2) are addressed here:

1. Is the detection of abnormal plant parameters differentially affected by SPD configuration?

2. Is the ability to locate abnormal plant parameters differentially affected by SPD configuration?

3. Is the recognition of individual abnormal plant parameters differentially affected by SPD configuration?

Deviation bar graphs, clustered meters, and a circular profile or star display similar to a Westinghouse design, were selected as the SPDS formats for this investigation. Ultimately, we would like to identify those characteristics of visual displays which most profoundly influence information extraction; therefore, psychophysical and information processing methodologies were employed. In essence, we are adopting an information processing view of information extraction and focusing upon the early stages of the process. Although the perceptual elements of information extraction were of primary interest, the overall goal of relating SPD information to safe and efficient nuclear power plant operation was also recognized. An assumption that pervades the experiments described below is that the information content of the displays can be controlled so that the configuration of the display can be manipulated independently.

In addition, it must be kept in mind that the orientation of this project is to examine and demonstrate the sensitivity and applicability of a number of different evaluation methods. In this evaluation of static displays, two methods were examined: a classical psychophysical paradigm using a signal detection analysis technique, and a chronometric (reaction time) method. Another important issue is the overall goal of evaluating the displays from several different perspectives. This was accomplished by performing four experiments which focused upon three different aspects of our perceptual analysis. Experiments 1 and 2 address the question of detecting abnormal SPDS conditions, Experiment 3 relates to the localization question, and Experiment 4 focuses upon the parameter recognition question.

EXPERIMENT 1: DETECTION

Classical psychophysical methods (Fechner, 1869;[2] Cattell, 1893;[3] Jastrow, 1888;[4] Urban, 1910[5]) were used in Experiments 1 and 2 albeit in a modified form. In classical psychophysics inferences about ease of information extraction are derived from the display exposure time required for the subject to make a response. If the task and information available remain constant and only the display configuration is changed, it can be argued that any differences in the exposure requirements reflect the subjects ability to extract information from the display. Assume we have two different display configurations (A and B), each portraying the same anomalous reactor condition. Suppose we conduct an experiment and find that Configuration A requires a longer exposure duration before the operator can detect the abnormal condition. The implication is that Configuration A is inferior relative to Configuration B, that is, subjects can extract information more efficiently from Configuration B.

Unfortunately, the conclusion presented above ignores a number of psychological elements which are important when making inferences based upon human performance. Using the classical psychological methodology, we inferred a direct and unmediated relationship between exposure duration and ease of information extraction. Put in terms of an information processing model, we assumed that lengthening the exposure duration provided the observer with additional information. The requirement of additional time to extract information was then assumed to be a measure of the difficulty in extracting information from the display configuration. This seems to be a reasonable assumption, until we consider that there are other mental events which intervene between extraction of information (perception) and making a response. One such event is the decision about whether or not to make a response based on the information extracted. Another factor which mediates the observer's responses is the amount of information and the context in which the information is presented. The observer's response is based not only upon the information presented in this display, but also on a number of other factors. A factor such as the consequences of the response, referring to the payoff that the observer associates with the response, directly affects the observers performance. For instance, responses which have major ramifications, such as overriding a safety system, will require much more information to elicit than a less consequential response, such as adjusting power from 92 to 93%.

Information is seen as the driving force behind the decision process, and the confidence that the decision maker has in his decision is related directly to the confidence he has in his ability to extract information pertinent to the decision. For our purposes in considering experimental methodologies, we were concerned with the observer's ability to extract information from a display, that is, the ability of the subject to distinguish information (signal) from the background of the display (noise). Neural activity can be thought of as modulated by information laden signals from the environment. The normal activity of the neurons can also be thought of as noise, just as the normal background energy of the environment is considered noise. Figure 1 represents the concept of a normal probability density function for this noise. When a signal is presented, it modulates the normal "noise" activity. The observer's task is to decide if the resulting neural activity is sufficiently different from the normal "noise."

Figure 2 represents the concept of a probability function of signal plus noise. The observer's mental activities then include both encoding information from the environment and making a decision that determines the type of response to be made.

With regard to our initial efforts in evaluating displays, we are not necessarily interested in the response and decision aspects of the operator's mental processes, although they will be important

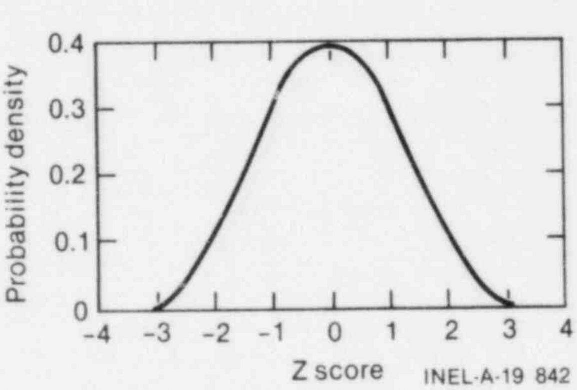
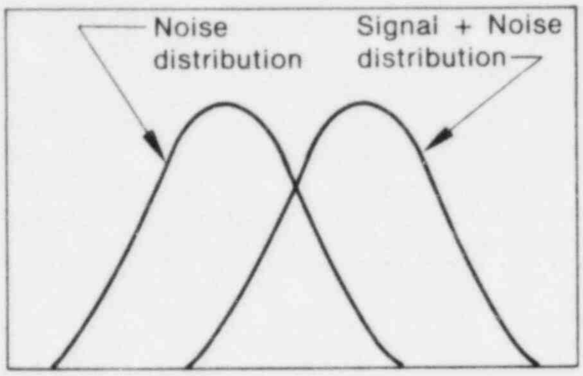


Figure 1. A normal probability distribution curve.



INEL 2 1851

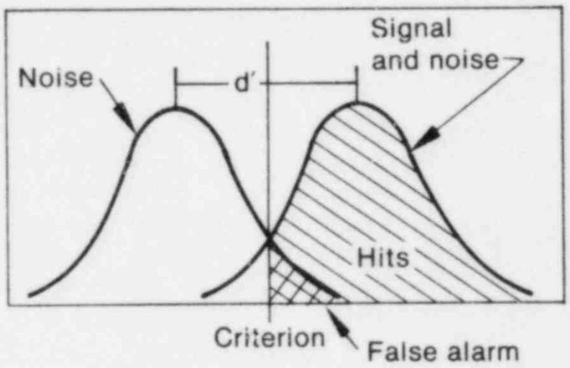
Figure 2. Noise and signal plus noise distributions.

later in our investigations. We are currently very interested in obtaining a clear view of the operator's sensitivity to the information contained in the display. The important point is that an unbiased view of sensitivity cannot be obtained without separating sensitivity from the observer's rationale for making decisions (response bias).

The theory of signal detection (TSD) offers a general methodology for obtaining independent and quantitative estimates of both sensitivity and response bias (Tanner and Swets, 1954;[6] and Green and Swets, 1966[7]). Basically, TSD assumes two normal distributions of equal variance: a noise distribution and a signal plus noise distribution, just as depicted in Figure 2. In addition, the theory assumes that the observer sets a criterion (decision point) and responds positively to any stimulus values falling above the criterion and negatively to any falling below the criterion. By making these assumptions, independent quantitative values can be calculated for both the observers' sensitivity to the signal and the response bias of the observers. Determining the sensitivity for each of the three display configurations (bars, star, and meters) will allow us to meaningfully compare the three displays, as to their relative ease of extracting information.

By conducting an experiment in which the experimenter controls the presentation (that is, signal and noise) and nonpresentation of the signal (that is, noise alone), we can record the subject's responses to these two situations. From these data, we can calculate the probability of the observer responding positively when the "signal" is present (a hit) and the probability of making the same response when only "noise" is presented (a false alarm). See Figure 3 for a graphical representation. Since TSD assumed normal distributions, we can calculate the relative differences (in z scores) in the means of these distributions and the distance (in z scores) from the mean of the "noise" distribution to the criterion. The differences in the means of the distributions is a direct measure of the sensitivity of the operator to the display. The distance from the "noise" mean to the criterion is the measure of the response bias. The interested or confused reader is referred to Appendix A where Massaro's, 1975,[8] detailed treatment of TSD is presented.

Given that the experiments being reported here were conducted in order to demonstrate extremely reliable differences in performance, the alpha level for determining the significance of the results of the experiments was set at the 0.01 level.



INEL 2 1841

Figure 3. A graphical representation of the relationship between "hits" and "false alarms."

Method

This section presents the method for Experiment 1. It describes the subjects, apparatus, instructions to the subjects, stimuli, subject training procedure, test procedure, and design for Experiment 1.

Subjects. Ten adult volunteers were used as subjects in this investigation. Their ages ranged from 23 to 49 years and all had vision correctable to 20/20. Five of the subjects were nuclear plant operators and five were engineers. The first group of subjects were qualified reactor operators from the Loss-of-Fluid Test (LOFT) reactor. They had a mean of 8.3 years of reactor operating experience. Each operator received his initial reactor training in the U.S. Navy.

The second group of subjects were EG&G Idaho, Inc., engineers. These engineers were not trained in the details of the LOFT plant or the significance of the parameters displayed on the SPDS formats. This was not considered a limitation because the detection task only required identification of normal or abnormal display states based on color or shape changes.

Apparatus. A dual channel tachistoscope (Gerbrands Model G1180) equipped with an automatic slide changer (Model G1180) and adaptation field logic interface (Model G1159) was used for stimulus presentation. This device was equipped with a four-channel timer (Model 300-4T), two shutters, one beam splitter, and associated shutter drive console (Gerbrands Corp., Arlington, Massachusetts). All testing was conducted in a room 4.57 x 6.10 m with 1.52-m-high partitions placed around the subject's position.

Illumination levels were recorded using a Gossen cadmium-sulfide cell light meter. A hemispherical diffuser was used to measure ambient room illumination levels from the subject's test position. Spot attachments of 15 and 7.5 degrees were used as necessary to reduce the meter's angle of acceptance when measuring illumination levels on specific areas of the rear projection screen.

On the simulated CRT display, the red and green information was at an illumination of 700 LUX with an average screen illumination of 525 LUX. Average ambient room illumination throughout all presentations was 1.75 LUX.

Instructions to Subjects. Prior to the testing, instructions to subjects were generally as follows:

> This is a visual recognition experiment in which we are attempting to determine the value of various display configurations. The type of displays we are currently interested in are Safety Parameter Displays (SPD) for nuclear power plants.
>
> During the experiment, you will be asked to observe the screen and report when you detect an abnormal parameter on the SPD. You will be in control of the display's appearance so you can merely identify the state the display represents, that is, all normal parameters or some abnormal parameters. There will be three different configurations for SPD used in this experiment. Figure 4, Examples A and B, show a typical bar graph display in both normal and abnormal states. Note that the abnormal states are represented by red bars and by red numerical reading which indicates the actual state of the parameters. These two forms of recognizing abnormal displays will be found on all display configurations. Figure 4, Examples C and D, are normal and abnormal meter configurations. Meter needle positions and colored numerical readings indicate normal and abnormal conditions. Normal and abnormal star configurations are shown in Figure 4, Examples E and F. Star shape and colored numerical readings indicate normal and abnormal conditions.
>
> The displays will be shown to you for only a brief period of time. If you cannot determine the state of the display, abnormal/normal, make your best guess. The display will then be shown to you for a slightly longer period of time. This will continue until you are consistently making the correct response. That is, correctly identifying normal displays as "normal" and abnormal displays as "abnormal."

Stimuli. The stimuli used in the experiments were 35 mm duplicate slide transparencies of photographs of reactor transient data displays on a cathode ray tube. The photographs were taken with a Contax Model RTS camera using a Zeiss Planar f2.8, 66 mm, macro lens. The CRT image was displayed through a Dunn Instrument camera 631 system. Ektachrome 200 color slide film was used. The stimuli are described in three parts: content, parameter format, and display configuration.

Content—Stimuli content refers to the actual reactor transient data which was pictured on the test slides. The data came from recordings of plant instrument readings before and during experiments in the LOFT reactor.

The LOFT reactor is a 50 MW(t) pressurized water reactor at the Idaho National Engineering Laboratory. EG&G Idaho, Inc., operates the LOFT facility to conduct reactor transient testing for the NRC. This testing has included small and large break loss-of-coolant experiments and other operational transient experiments. The slides used in display evaluation pictured normal conditions before and abnormal conditions during the following LOFT experiments:

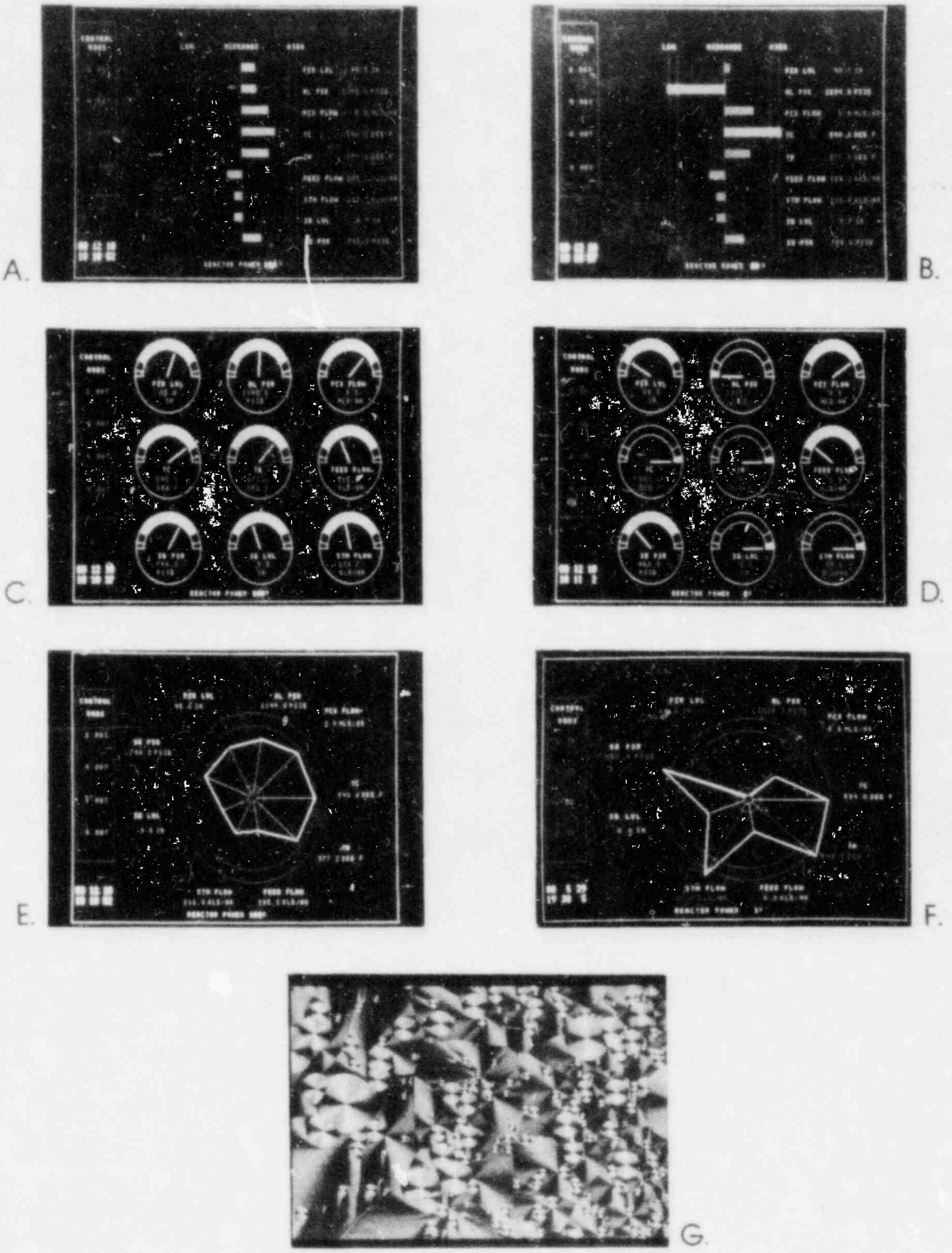


Figure 4. Examples of the stimulus material used in the detection experiment.

1. Experiment L6-1, a steam load decrease—this transient was conducted on October 8, 1980. It was initiated from 75% power [37 MW(t)] by closing the main steam control valve. Power initially decreased, steam flow decreased and primary pressure increased. The reactor shut down (scrammed) on high primary system pressure at 22 s after experiment initiation.

2. Experiment L6-2, a loss of primary flow—this transient was conducted on October 7, 1980. It was initiated from ~75% power by tripping power to the primary coolant pumps. At ~2 s after initiation, the reactor scrammed due to low primary system flow. Following the scram, primary system pressures and temperatures decreased.

3. Experiment L6-3, a steam load increase—this transient was conducted on October 9, 1980. It was initiated from ~75% power by opening the main steam control valve to increase secondary system steam flow and, therefore, increase reactor power. The power increased, primary pressure decreased, and the reactor scrammed at ~16 s after initiation due to low primary system pressure. Reactor power at the time of scram was 84%.

4. Experiment L6-5, a loss of feedwater—this transient was conducted on May 29, 1980. It was initiated from ~75% power by tripping power to the secondary system feedwater pump. Feed flow decreased, steam generator liquid level decreased, and the reactor was manually scrammed at ~23 s, when steam generator liquid level corresponded to a low level trip setpoint in commercial pressurized water reactors.

5. Experiment L3-5, a small break loss-of-coolant accident (pumps off)—this transient was conducted on September 29, 1980. It was initiated from ~100% power by opening a leak path in the primary system cold leg downstream of the primary coolant pumps. The primary coolant pumps were tripped at ~1 s. The reactor was manually scrammed at ~5 s after experiment initiation. The experiment proceeded until ~40% of the primary mass inventory had been lost, and then the system was recovered.

6. Experiment L3-6, a small break loss-of-coolant accident (pumps on)—this transient was conducted on December 10, 1980. It was initiated from ~100% power and was identical to Experiment L3-5 except that the primary coolant pumps were left running until 40 min after experiment initiation. Because the pumps were kept running, much more mass left the primary system during Experiment L3-6 than during L3-5. At the conclusion of Experiment L3-6, the reactor core was briefly uncovered.

The data shown on the test slides were recorded on the LOFT data acquisition system during the above experiments. The data were collected from 380 channels at 1000 samples per second to produce a sample rate of ~3 samples per channel per second (1000/380). Four seconds of data were collected in a buffer and averaged by the data system. The averages of each channel reading were transmitted to the display generation computer for recording and display. Thus, the displayed instrument readings on the test slides represent the numerical average of 12 readings of the same instrument over 4 s (3 samples per channel per second x 4 s = 12 samples per channel).

Two other types of data appeared on the test slides: the first was control rod positions as indicated by rod bottom lights. These data were binary (the rod is IN or if not, it indicates OUT) and were transmitted as a current status, with no averaging from the data acquisition system, to the display computer every 2 s. The second type of data were date/time which were generated by the data acquisition system and carried as a flag by the sets of data received by the display computer.

Parameter Format—Test slides were made of the three different safety parameter display formats (star, bars, and meters). These formats displayed data which provided an overview of LOFT plant conditions. Each format displayed exactly the same plant parameters. The normal (green), caution (yellow), and alert (red) parameter limits were identical for each format.

Plant parameters in the display were selected by reviewing LOFT plant operating manual emergency procedures. Those parameters listed as symptoms of plant transients were selected for the displays. The details associated with parameters on the test slides are shown in Table 1.

Parameter normal values and ranges used on the test slides are shown in Table 1. The design assumption inherent in range selection was that the normal value and range were for steady state operations. Thus, a "normal" operation at LOFT, such as a slow power ascension, may cause one or more parameters to leave the prescribed normal range temporarily.

All of the displays represented the normal value from Table 1 as a central value in the display, with the range bracketing that normal value. For some parameters, the normal value and range were fixed, and for others, the normal value and range were a direct function of reactor power (that is, the normal value of feed flow increased with increasing reactor power).

All test displays used green to represent the central 85% of the range, yellow for the 10% adjoining the green, and red for the outer 5% of the range. The green was for normal, yellow for caution, and red for alert in the color standard used for these displays.

Display Configuration—Three safety parameter display formats were represented among the test slides. Each display showed control rod status in a box to the left, date/time in the lower left, and reactor power at the bottom. The only difference in the displays was the method used to show normal values, ranges, and interrelationships between Parameters 2 through 10 of Table 1. The display formats are described in the following paragraphs.

Table 1. Normal values and ranges for the parameters used in the display stimuli

	Parameter	Display Symbol	Instrument	Display Value	Instrument Range	Normal Value and Display Range
1.	Reactor power	Reactor power	RE-T-77-1A2 RE-T-77-2A2 RE-T-77-3A2	Highest reading channel	1 to 125%	Value displayed and used to normalize other parameters
2.	Pressurizer liquid level	PZR LVL	LT-P139-006 LT-P139-007 LT-P139-008	Medium reading channel	5 to 70 in.	Normal = 44 in. Range = 37 to 51 in.
3.	Primary hot leg pressure	IHL PSR	PT-P139-002 PT-P139-003 PT-P139-004	Medium reading channel		Normal = 2140 psig Range = 2125 to 2155 psig
4.	Primary flow	PCS flow	FT-P139-27-1 FT-P139-27-2 FT-P139-27-3	Medium reading channel	0.5 to 5.6 Mlbm/h	Normal = 3.8 Mlbm/h Range = 3.7 to 3.9 Mlbm/h
5.	Primary cold leg temperature	TC	TE-P139-28-2	Reading	500 to 650°F	Normal = 532.5°F + [3.75 (% power)]/100 Range = 525 to 540°F + [7.5 (% power)]/100
6.	Primary hot leg temperature	TH	TT-P139-032 TT-P139-033 TT-P139-034	Highest reading channel	500 to 650°F	Normal = 533.4°F + 0.385 (% power) Range = 525.42°F + 0.353 (% power) 541.37°F + 0.417 (% power)
7.	Secondary feedwater flow	Feed flow	FT-P004-72-2	Reading	0 to 300 klbm/h	
8.	Secondary steam flow	STM flow	FT-P004-12	Reading	0 to 300 klbm/h	
9.	Steam generator liquid level	SG LVL	LT-P004-008B	Reading	-145 to 57.5 in.	Normal = [0.1 (% power)] in. Range = Normal ± 2 in.
10.	Steam generator pressure	SG PSR	PT-P004-101A	Reading	0 to 1200 psig	Normal = [893.05 - 1.2055 (% power)] psig Range = [834.91 - 1.394 (% power)] psig
11.	Control rod positions	Control Rods 2, 4, 6, and 8	LS-CRDM2 LS-CRDM4 LS-CRDM6 LS-CRDM8	In (red) Out (green)	In—Out	Not applicable

Deviation bars (Figure 4, Examples A and B)—this display used a central vertical line to indicate the normal value. Parameter deviations from this value were shown as bars to the left or right of normal. High- and low-range values were shown as vertical lines. Parameter descriptions and digital values were on the right of the display. As parameter values reached the 85% (green-yellow) and 95% (yellow-red) barriers, the bar indicator and digital values on the display changed to the appropriate color. On this display, primary coolant system parameters were grouped at the top with secondary system parameters grouped at the ground.

Meter display (Figure 4, Examples C and D)—this display represented parameter values as needle positions on nine meters drawn on a cathode ray tube. The green, yellow, and red ranges were shown on the meters with only the color corresponding to the current parameter value lit. Digital values (color coded) and parameter descriptions were inside each meter.

Circular plot, star (Figure 4, Examples E and F)—this display represented parameter values as positions on the spokes of a circle. A small inner circle represented range minimums with an outer circle representing maximums. Current value spoke positions were tied together to form a nine-sided polygon. Digital values and parameter descriptions were shown around the outside of the maximum-range ring. A background ring showed the 85% range values, and digital parameter indications changed color corresponding to 85 and 95% values.

Intertrial mask (Figure 4, Example G)—an intertrial mask display was presented after each trial for the duration of the intertrial interval. The mask consisted of an enlarged photograph which consisted of a pseudo-random color pattern.

The simulated CRT display subtended a horizontal visual angle of 13.4 degrees and a vertical angle of 11.4 degrees from the subject's test position.

Subject Training Procedure. Both nuclear plant operators and engineers were used as subjects in these experiments. Operator subjects were given more extensive training than engineer subjects. This training was to prepare them for the more complex display testing done in the parameter recognition (Experiment 4) part of the investigation. Each of the operators was briefed on the three SPDS formats and on the normalization schemes for displayed parameters. An engineering simulation of LOFT was used to drive each display so that, in real time, each operator subject observed the same simulated plant evolutions on each display format. The evolutions were:

1. Power ascension from 50 to 75% power accompanied by charging and draining to reduce primary boron concentration
2. From 75% power, an excessive steam load increase was simulated allowing the operators to observe a reactor scram caused by low pressure
3. Power descent from 100 to 75% power accompanied by charging and draining to increase primary boron concentration.

Following simulation training for operator subjects, each subject was required to correctly sketch each display format and explain the parameter normalizations.

The second group of subjects were EG&G Idaho, Inc., engineers. These engineers were not trained in the details of the LOFT plant or the significance of the parameter displayed, as were the reactor operators. This was not considered a limitation because the detection task (Experiments 1 and 2) only required identification of normal or abnormal display states based on color or shape changes.

Training of the engineer subjects was limited to familiarization with each display because of the nature of the detection task. The subjects were shown each display in normal and abnormal states to ensure that they knew how these states were represented.

Test Procedure. Three types of SPD configurations (bar graphs, meters, and star) were used as separate conditions in this experiment. Each subject (S) was presented with three blocks of trials for each condition. Each block contained 9 normal displays, and 18 abnormal displays. The order and sequence of the trials were randomized. Each subject was familiarized with the displays as per the instructions and then given a series of 30 warmup trials before actual testing was initiated. In addition, they were given detailed instructions before the session began. Following instructions and the warmup, testing began with a test display presented for 5.0 ms. The exposure duration was then increased by 10 ms per block until the subject made no errors during three successive blocks. The subjects (Ss) responses were recorded at each intensity level. Between every presentation, the masking slide was displayed to eliminate the possibility of establishing latent images. Each block of trials consisted of a single display type (for example, meters). The order of presentation of the test blocks was balance across subjects and type of display configuration. The subjects were given a 15-min rest between display configuration changes.

Design. A "within subjects" nested design was used in this experiment. Four independent variables were manipulated: Three fixed variables, display configuration, exposure duration, and type of subject (operators versus nonoperators), with the random variable subjects being nested within type of subjects. Three dependent variables were examined in this experiment: perceptual sensitivity (d'), response criterion (β), and response accuracy (percent correct). In addition, the following orthogonal planned comparisons were conducted. The meter scores were compared to the bar and star scores, and then the bar scores were compared to the star scores.

Results

The data from Experiment 1 are shown in Figure 5. An analysis of variance of these data was conducted in three separate parts: The first part analyzed the perceptual sensitivity (d') of the subjects as a function of display type and exposure duration, as shown in Figure 5, Plots A and D. The analysis revealed that both display type [$F(2,16) = 10.88$, $p < 0.01$] and exposure duration [$F(7,56) = 21.17$, $p < 0.01$] are significant main effects, that is, sensitivity changes with display type and exposure duration. In addition, a significant interaction was shown for the display type and exposure duration [$F(14,112) = 2.15$, $p < 0.01$], that is, sensitivity changes as a function of display type and exposure duration.

Since one of the objectives of this experiment was to evaluate the three display formats, orthogonal planned comparisons of the data were conducted. The first comparison was meters versus bars and star. This comparison revealed a significant difference [$t(16) = 4.02$, $p < 0.01$], that is, bar and star formats were better for detection than meters. The second comparison, bars versus star, showed no significant difference for detections.

The second part of the analysis considered the accuracy of the subject's responses in terms of percent correct. These data are plotted in Figure 5, Plots B and E. As with sensitivity, two main effects were found: display type [$F(2,16) = 12.94$, $p < 0.01$] and exposure duration [$F(7,56) = 24.13$, $p < 0.01$]. Again, significant interaction was found: display type by exposure duration [$F(4,112) = 5.15$, $p < 0.01$]. Orthogonal planned comparisons of the data on the type of display format showed the same pattern of results as the first part of the analysis, that is, the difference between meters versus star and bars was significant [$t(16) = 4.90$, $p < 0.01$]. The second comparison, bars versus star, showed no significant difference.

The third part of the analysis examined the data in terms of the subject's response criterion (β). These data are plotted in Figure 5, Plots C and F. The only significant main effects was exposure duration [$F(7,56) = 9.58$, $p < 0.01$]. No interactions were found to be significant and no comparisons were conducted on these data.

Discussion

The data from Experiment 1 indicate that star and bar formats may be better display configurations than the meter format for detection tasks. Although star and bar formats cannot be distinguished on a strictly statistical basis, visual examination of Figure 5, Plots A and B leads one to infer that the star format may hold some advantage over the bar format for this task. In addition, examination of Plots D and E, the data representing the significant interaction of display format and exposure duration for both perceptual sensitivity (d') and response accuracy (percent correct), reveals that the star format promoted consistently better performance on this detection task. Therefore, the star format apparently transmitted information concerning parameter conditions better than the meter format, which required that the viewer have longer exposures (more information) to accurately assess the condition of the display. Interestingly, the rate of change in the subject's ability to extract information seemed greater with the meter format, perhaps due to a ceiling on the subject's responses using the star format. Given that the order of presentation for exposure duration was fixed (5 to 75 ms), the subjects may have been engaging in more perceptual-pattern learning from the meter display than from the other two formats.

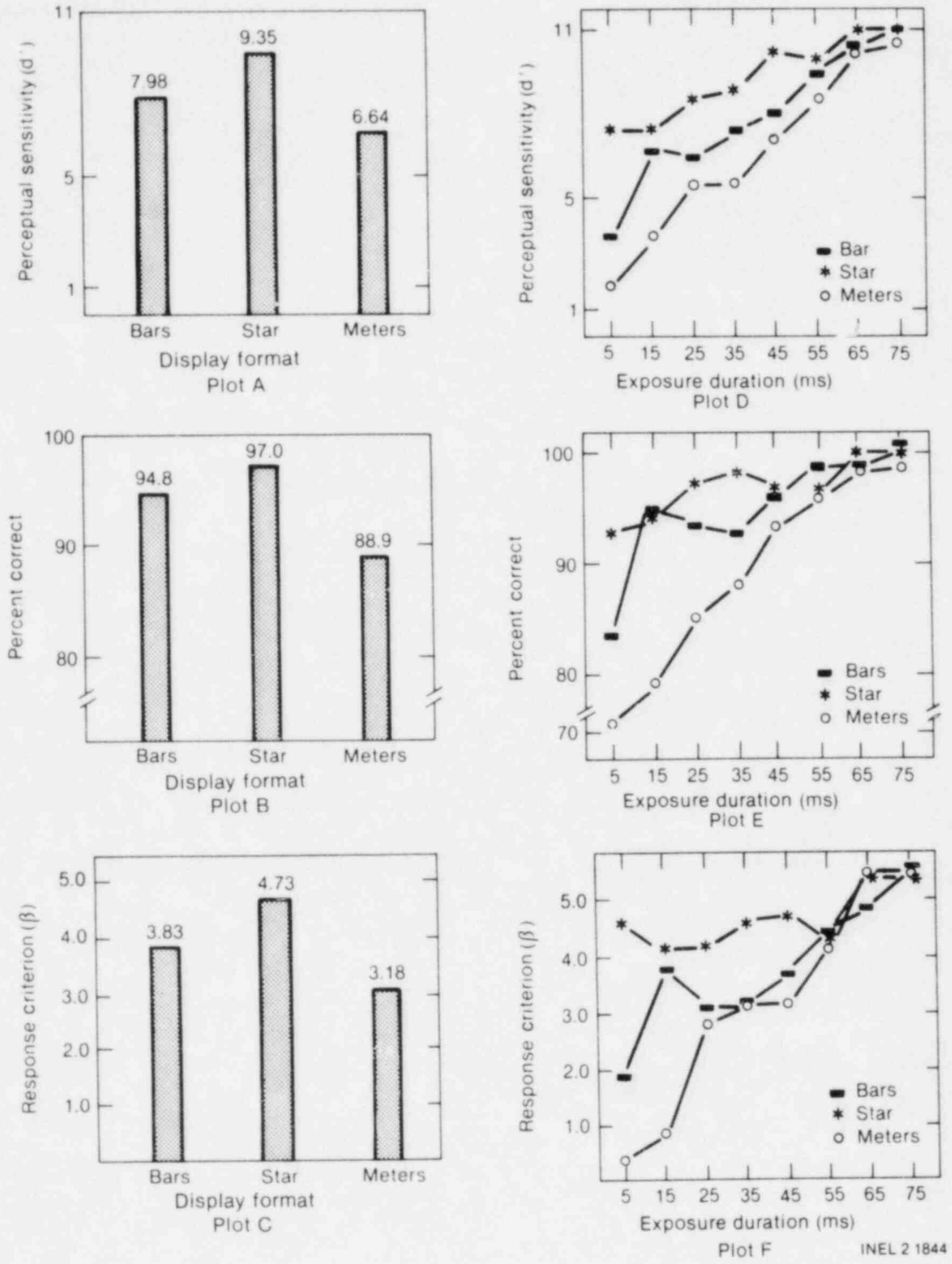


Figure 5. Plots of the data from Experiment 1.

Longer exposure times increased the measured sensitivities (d') of the subjects and produced more correct responses. This is not surprising, since the amount of information available for making a decision would usually increase with a longer exposure, and the more information available, the better the decision would be.

The background and experience of the subjects were not significant variables in this task. Operators could not be distinguished from nonoperators on the basis of their performance. Therefore, we can conclude that the detection task is purely perceptual in nature and is not influenced by the differences in training and experience between these two groups.

Finally, there was no differential effect of display format on the response criterion used by the test subjects. Therefore, the differences shown in the accuracy measure are not due to shifts in the response strategy of the subjects as a function of the type of display being viewed.

EXPERIMENT 2: DETECTION

Experiment 2 represented an extension of Experiment 1, but focused upon the operator population. Thirteen current and former LOFT operators were added to the original five operators used for Experiment 1. This was done to satisfy the requirements of the project's later phases (Experiment 4) when operators were the only valid test subjects. Utilizing a large sample of operators in Experiment 2 increased the probability of being able to get a reasonable sample of these people for the later experiments. The correlational techniques used in the overall multimethods approach demanded that the same subjects be used in all experiments.

The basic experimental methodology remained the same; however, the data were collected using an Apple computer to record the experimenter's responses to the subjects determination of the state of the display.

Method

The methodology for Experiment 2 was the same as that for Experiment 1, except for subjects and design described as follows:

1. Eighteen current and former LOFT operators were used for Experiment 2. These operators had the same characteristics as those described in Experiment 1, except that the age range became 26 to 44 and the mean years of experience changed to 9.5.

2. A "within subjects" design was used for Experiment 2. Two independent variables were manipulated: display type (configuration) and exposure duration. The same dependent variables examined in Experiment 1 were examined here. The same planned comparisons used in Experiment 1 were used here.

Results

The data from Experiment 2 are shown in Figure 6. The analysis was conducted for the data collected during the first five exposure durations. It appears that a lack of variance precluded formal analysis of the data from exposure durations beyond 45 ms. A mode error, probably indicating division by zero, was encountered at all durations longer than 45 ms. A MANOVA of the first five exposure durations yielded significant results for:

1. The display type variable, Hotelling's [$F(6,504) = 27.58$, $p < 0.01$]

2. Exposure duration, Hotelling's [$F(12,755) = 11.80$, $p < 0.01$]

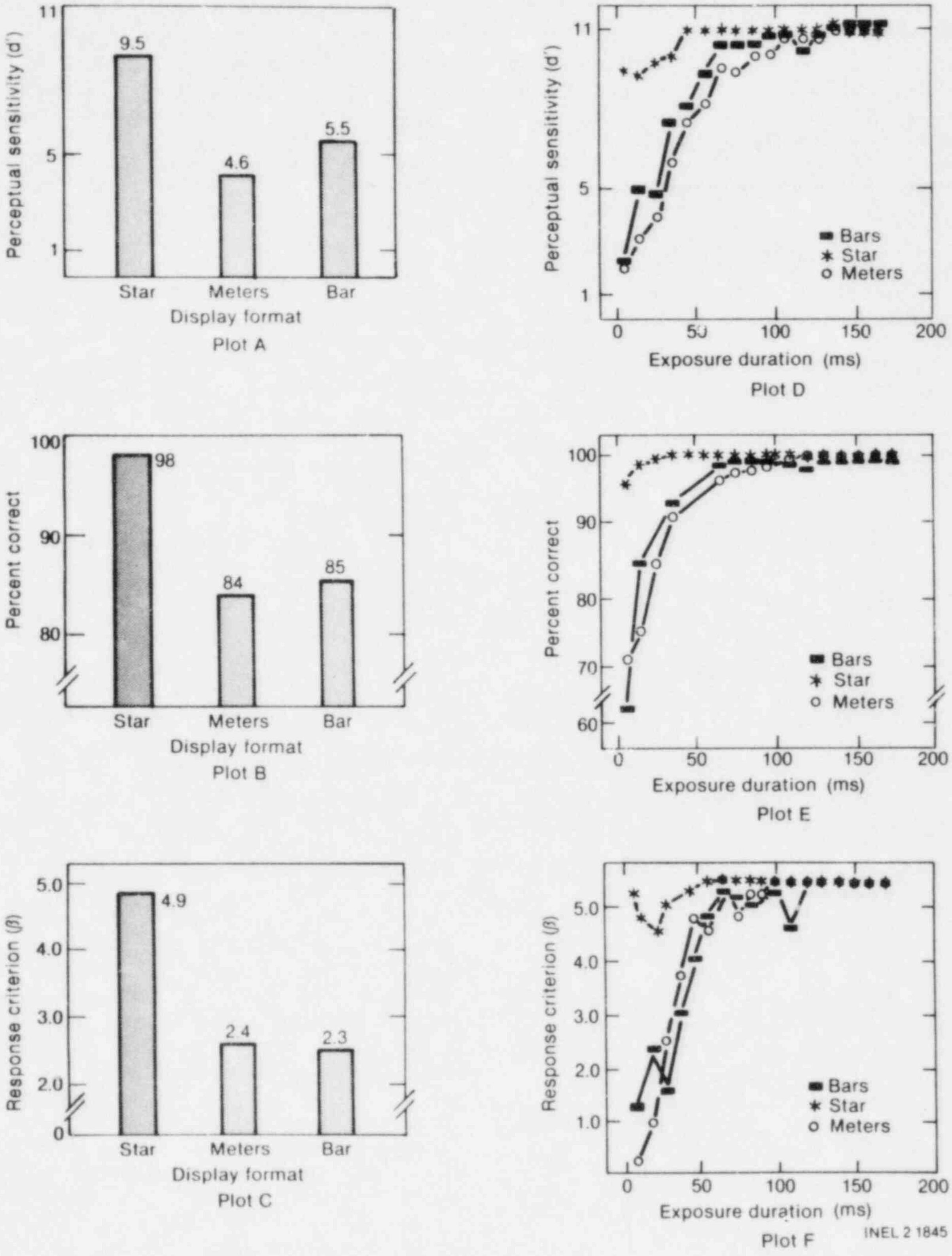
3. The interaction of display type and exposure duration, Hotelling's [$F(24,755) = 3.76$, $p < 0.01$].

The MANOVA tables for these data are given in Appendix B.

The univariate analysis of variance (ANOVA) revealed that for the perceptual sensitivity (see Figure 6, Plots A and D), both main effects were significant and the interaction was almost significant: display type [$F(2,255) = 72.65$, $p < 0.01$, $MS_e = 8.55$], exposure duration [$F(4,255) = 23.19$, $p < 0.01$, $MS_e = 8.55$], and display type by exposure duration [$F(8,255) = 2.35$, $p < 0.01$, $MS_e = 8.55$]. The orthogonal planned comparisons found meters to be significantly different [$t(255) = 5.93$, $p < 0.01$] than bars and star for the detection task. Bars were also shown to be significantly different [$t(255) = 7.07$, $p < 0.01$] than star.

A second ANOVA, this time using the percent correct data, analyzed the data shown in Figure 6, Plots B and E. In this analysis both main effects, display type [$F(2,255) = 57.20$, $p < 0.01$, $MS_e = 97.96$] and exposure duration [$F(4,255) = 30.89$, $p < 0.01$, $MS_e = 97.96$], and the interaction of display type and exposure duration [$F(8,255) = 6.66$, $p < 0.01$, $MS_e = 97.96$] were found to be significant. The orthogonal planned comparisons found meters to be significantly different [$t(255) = 4.54$, $p < 0.01$] than bars and star for the percent correct data. In addition, star displays were shown to be significantly different [$t(255) = 6.82$, $p < 0.01$] than bar displays.

The third ANOVA examined the response bias on criterion variable (β) from Figure 6, Plots C and F. This ANOVA showed both main effects


Plot A

Plot B

Plot C

Plot D

Plot E

Plot F

Figure 6. Plots of the data from Experiment 2.

and the interaction to be significant. The main effect of display type [$F(2,255) = 38.96$, $p < 0.01$, $MS_e = 5.15$] is shown in Figure 6. The exposure duration main effect [$F(4,255) = 9.90$, $p < 0.01$, $MS_e = 5.15$] is not depicted graphically. The interaction of display type and exposure duration [$F(8,255) = 3.25$, $p < 0.01$, $MS_e = 5.15$] is shown in Figure 6. The orthogonal planned comparisons of meters versus bars and star was shown to be significant [$t(255) = 3.22$, $p < 0.01$], as was the comparison of star versus bars [$t(255) = 6.91$, $p < 0.01$].

Discussion

The data from Experiment 2 generally support the conclusions drawn from Experiment 1, including some of the extrapolations made from trends in the data. There were some slight differences in the data, primarily, regarding the response criterion measure which exhibited significant differences in Experiment 2, but which were not significant in Experiment 1. These results indicate that the technique analyzed in these experiments was very sensitive to changes in response criteria and perceptual sensitivity for the detection task and was quite reliable as a performance measure.

It must be recognized that detection represented only the first stage of cognitive processing. Later stages assumed a larger role in determining the overall effectiveness of modern decision-aiding techniques such as the SPDS. The reader is, therefore, cautioned not to generalize the results of Experiments 1 and 2 beyond the context of visual detection. This work was the first step in a systematic examination of the cognitive elements involved in utilizing decision aiding in the nuclear power plant control rooms. The next step in our evaluation work was to examine the recognition attributes of various SPD formats. In the next two parts of our examination (spatial localization, Experiment 3, and parameter recognition, Experiment 4) we extended the demonstration of measurement techniques, first to a new task and then to a measurement technique.

EXPERIMENT 3: SPATIAL LOCALIZATION

The subject's ability to determine the locus of information on an SPD was examined in Experiment 3. The question being investigated concerned the ability of subjects to locate particular information in the three SPDS configurations. The subject's task was to record the location of abnormal parameters when a display was presented using the tachistoscope.

Experiment 3 was conducted as the training module for the parameter recognition part of our examination performed in Experiment 4. In Experiment 4, subjects were asked to locate specific parameters on the displays and respond appropriately, depending on the status of the parameter. The goal of the training was to have the subjects attain the ability to locate specific parameters automatically without searching the display for the parameter in question. Therefore, in Experiment 3 the subjects were only shown the display for a brief, fixed period of time. The task of the subjects was twofold:

1. To learn the names and locations of the parameters in the display
2. To locate all abnormal parameters presented on the display.

The subjects recorded the location of the abnormal parameters and were encouraged to articulate the names of these abnormal parameters. The subject's ability to name the parameters was measured in a criterion test at the conclusion of Experiment 3, where they were given a blank data sheet for each display and asked to write out the names of all parameters. All subjects attained a perfect score on the criterion test before proceeding to Experiment 4.

Given that some familiarization training was necessary for Experiment 4, Experiment 3 was designed to provide that training and produce data relating to the subject's ability to specify the location of the abnormal parameters presented in the displays. For this task, the subjects were asked to use a graphic tablet stylus to point to the locations of the abnormal parameters on an overlay which was placed on the tablet. The overlay was a stylized representation of the SPD being used, that is, bars, star, or meters. Details of the methodology and results and conclusions are presented in the following sections.

Method

Experiment 3 utilized essentially the same paradigm described for the preceding experiments with the following exceptions:

1. The subject's task was changed from a detection task to a localization task. Now, instead of merely reporting the detection of an abnormal condition, the subjects must mark the location of all abnormal parameters on the graphic tablet.

2. The duration of display presentation was set and held constant at 750 ms. This duration was derived during a short pilot study. The number of display presentations was limited to one warmup block and three test blocks of 27 slides each.

3. In addition to marking the data sheets with the location of the abnormal parameters, the subjects were asked to learn the names and locations of the parameters by articulating the parameter name as he marked the parameter. The goal of this task was to have the subject be able to correctly label each parameter by the end of the experiment. In this way, the experiment served as the training for Experiment 4.

4. The intertrial mask was not used in this experiment. Instead, a dark screen was used.

5. An Apple II plus computer and Apple graphics tablet were used to record the data from this experiment.

6. The design of this experiment was essentially the same as Experiment 1, with three test blocks being substituted for exposure duration.

Results

The data from Experiment 3 are shown in Figure 7. A multivariate analysis of variance (MANOVA) revealed a significant effect for display type, Hotelling's [$F(6,300) = 85.43$, $p < 0.01$]. Test blocks and the interaction terms were not

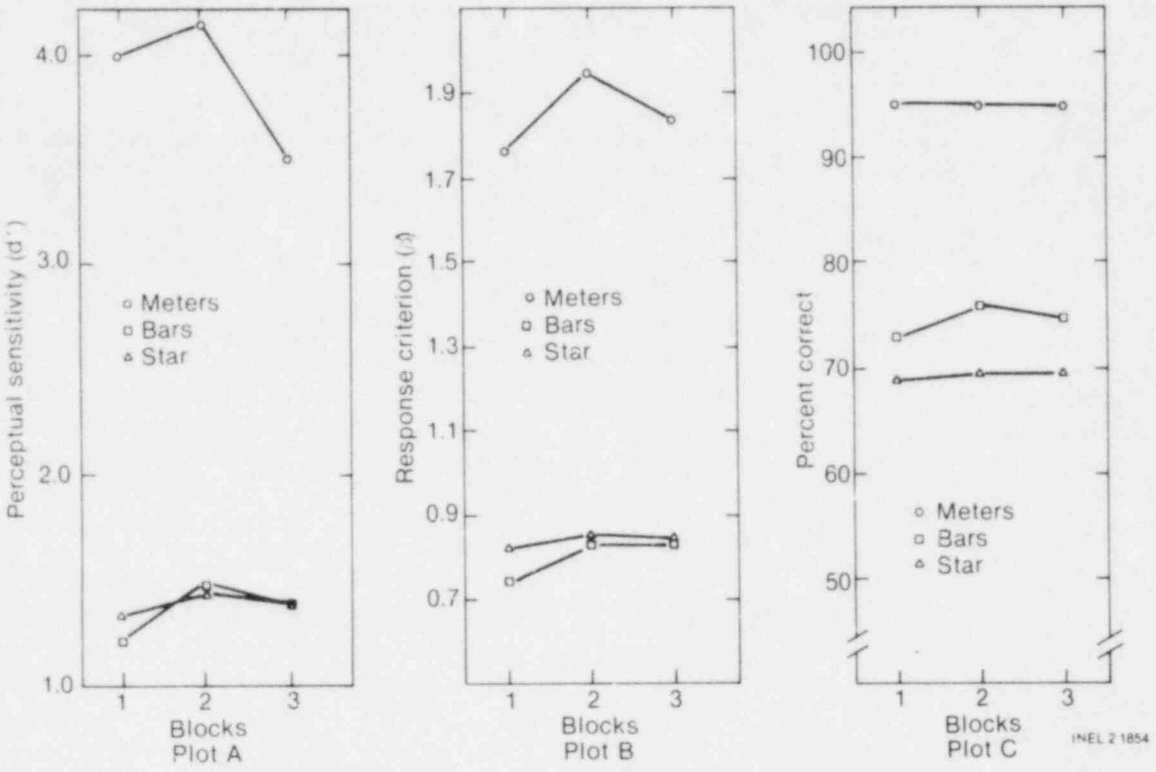


Figure 7. Plots of the data from Experiment 3.

significant. The univariate ANOVA showed that the display type variable was significant across all three dependent variables, see Figure 7: for the sensitivity measure (d'), Plot A, [$F(2,153) = 38.05$, $p < 0.01$, $MS_e = 3.00$]; for percent correct, Plot C, [$F(2,153) = 186.74$, $p < 0.01$, $MS_e = 53.58$]; and for the response bias (β), Plot B, [$F(2,153) = 30.89$, $p < 0.01$, $MS_e = 0.61$]. Orthogonal planned comparisons of these data reveals that for (d'), meters versus star and bars were significantly different [$t(153) = 12.02$, $p < 0.01$], while bar displays were not significantly different than star displays. For response bias (β), the meters versus star and bars comparison was significant [$t(153) = 11.14$, $p < 0.01$] and, again, the other comparison was not significant. For percent correct, the meter versus star and bars comparison was significant [$t(153) = 18.94$, $p < 0.01$], as was the comparison of bars versus star [$t(153) = 3.98$, $p < 0.01$].

Discussion

The results from Experiment 3 are interesting from the standpoint of demonstrating the influence that the task (context) exerts on the way in which information is processed. The data from this experiment show a very different pattern than those from Experiments 1 and 2. In this experiment, the meters format facilitated the subject's responses to meters compared to the bar and star formats, while in Experiments 1 and 2 the subject's responses to meters were inhibited relative to the other formats. The important point of this demonstration is that a single experiment does not provide a comprehensive evaluation of a format or set of formats. Format evaluation is a task/situation dependent process. The next experiment examines the parameter recognition task using an additional dependent measure, reaction time.

EXPERIMENT 4: PARAMETER RECOGNITION

The parameter recognition question was addressed in Experiment 4. That is, is the subject's ability to recognize the status of individual parameters affected by the format of the SPDS display? There are several methodologies that could be useful in addressing this question and, in fact, a method similar to that used in Experiments 1 and 2 could be used. However, there are two points that should be made regarding the use of that paradigm in this context. The first is conceptual and regards the purpose of the project which is to test a variety of methodologies for assessing SPDSs. Since we have demonstrated that a psychophysical method, such as the one used in the previous experiments, is very sensitive and can differentiate the display formats on the basis of subject performance, the argument can be made that a different paradigm should be tested.

The second consideration is methodological. If the paradigm of Experiments 1 and 2 were utilized, the subject's task would be to report all abnormal parameters. The exposure duration would be set at a short interval, and increased until the subject was consistently reporting all of the abnormal parameters. This procedure was tested in a short pilot study where it was observed that the subjects very consistently reported no more than four or five abnormal parameters, even though seven or eight may have been presented. This observation appeared to be independent of the display format, indicating that the problem was at the level of information storage or retrieval (memory) and not at the information input (perception) stage. This is a situation similar to that reported by Miller, 1956,[9] and Sperling, 1960[10] and 1963.[11] Their conclusions were that short-term memory was imposing a constraint on the amount of information the subject could remember for reporting. Sperling solved this problem by devising an experimental technique, called the partial report technique, in which, the subject was exposed to a large number of items but only asked to report a portion of them. The indication of which items to report was made just prior to or simultaneous with the presentation of the display. Sperling demonstrated that more information can be taken in from a display than can be held in memory and reported on.

Given the data from the pilot study and the orientation of the project to investigate a number of evaluation methods, Experiment 4 was designed to utilize a paradigm similar to Sperlings partial report technique and to demonstrate another type of experimental methodology. Given control over the information content of the displays, there are several methodologies which have been developed to address the preliminary stages of cognitive processing (perception). One prominent method, reaction time (Donders, 1868;[12] Estes, 1975;[13] Posner, 1975;[14] Pachella, 1974;[15] and others), requires the subject to view a display, make some sort of decision regarding the display, and make a response as quickly as possible.

Information extraction is assumed to involve a complicated set of mental operations each of which requires a finite amount of time. It is reasoned that when more processing is required to make a response, more time will also be required. The main premise is that more processing time will be required when information is more difficult to extract from a particular display configuration. Therefore, one way of assessing ease of information extraction would be to measure the time required to respond to the various displays. Of course, more complicated decisions or responses will also increase reaction time and must be controlled.

Both the decision and the response related components of reaction time can be assumed to remain constant, if both the task and the response remain constant across the experimental conditions. Therefore, any change in response time can be attributed to a change in the amount of perceptual processing required to extract information from the different display configurations. A detailed description of the methodology used in conducting Experiment 4 is presented below, followed by the results and a discussion of the experiment.

Method

The same paradigm described for the preceding experiments was used for Experiment 4, with the following exceptions:

1. Additional apparatus used for Experiment 4 consisted of an Apple II plus computer with 64K memory, supertalker (a voice digitizer), real-time clock (Apple

clock), and a modified input/output (I/O) board to control display presentation and monitor subjects' button responses. The software included Apple DOS 3.3, and a control program written in BASIC. A movable, button box 35.6 cm wide by 21.5 cm deep by 8.5 cm high placed on the table in front of the subject. Two buttons were mounted on the inclined portion of the top of the box. The buttons protruded 2 mm above the box surface. Each button was 34 mm in diameter with a throw of 7 mm. The button mounted on the right side of the box was green, and the one on the left was red.

2. After the subject had been trained for the particular display type (from Experiment 3), they were informed that they could participate in a reaction time experiment. It was explained that they were to listen to the computer voice in a particular parameter and find that parameter in the display. The subject's task was to determine if the parameter was normal or abnormal. If the parameter was abnormal, the subject pressed the left (red) button. If the parameter was normal, the subject pressed the right (black) button. Feedback was supplied on a CRT located to the left of the subject. The subject's reaction time and the correctness of the response were shown on the screen. In addition, the subject initiated each trial by depressing a foot switch.
3. The test procedure was modified as follows:
   a. The display was presented for a constant 2000 ms.
   b. The accuracy and speed of the subject's responses were recorded. Any responses longer than 2500 ms were not recorded.
   c. No mask was presented during the intertrial interval.
4. A response time dependent measure was added to the experiment design.

Results

The data for Experiment 4 are shown in Figure 8. A multivariate analysis of variance (MANOVA) revealed no significant effects on interactions. See Appendix B for the MANOVA tables.

Discussion

Although the results of Experiment 4 do not appear interesting, they are potentially informative. In fact, they are interesting to the extent that they represent yet another task where the relationships between the display formats change as a function of the task the subjects engage in. The informative aspect of these data occurs when one considers the lack of differences shown here, compared to the results of a pilot study using this same paradigm where a significant difference was demonstrated. The results of the pilot study are shown in Figure 9. The analysis of the data from the pilot study found a significant difference between the display formats, indicating that the chronometric technique would be another useful technique with which to evaluate displays. The problem developed when the formal experiment was conducted, and the data were not consistent with the pilot study.

There may be, however, a serendipitous side to this problem. The pilot study was conducted using a random set of 12 subjects (not necessarily operators), while the 18 subjects (all operators) who participated in the experiment contributed 8 h and were subjected to the training session and the series of experiments (in order). The experimental subjects viewed the three formats much more often than the pilot subjects. Additionally, the training the pilot subjects received consisted of one block of 27 trials for the localization experiment (Experiment 3), while the experimental subjects ran 4 blocks of 27 trials for that experiment. All of this leads to the hypothesis that familiarity with the displays resulted in a change in the performance of the subjects (perceptual learning beyond accuracy toward automaticity, LaBerge, 1976;[16] LaBerge and Samuels, 1974;[17] Gibson and Levin, 1981[18]). Familiarity with the displays apparently overcame the benefit observed in Experiment 3 and the pilot study for spatially distributing the information across the display surface. Familiarity may have reduced the need to search for the specific parameter. When the parameter was announced, the experienced subjects knew exactly where to look, while the inexperienced subjects from the pilot study had to

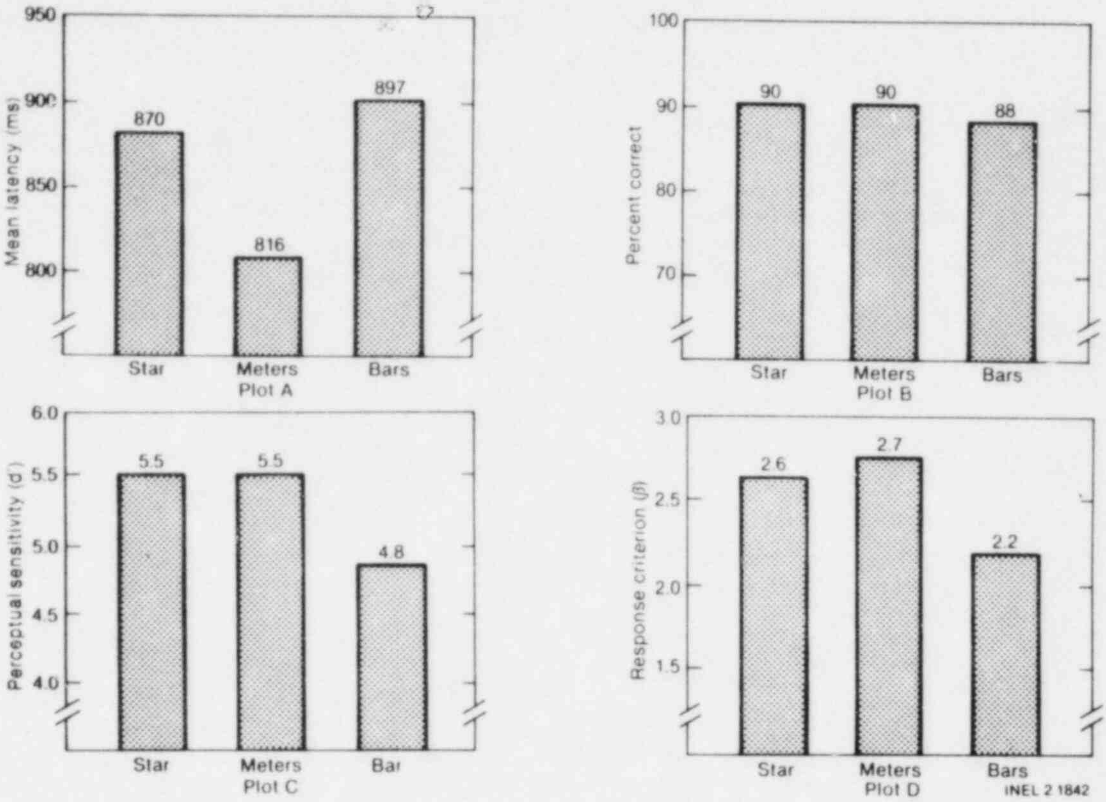


Figure 8. Plots of the data from Experiment 4.

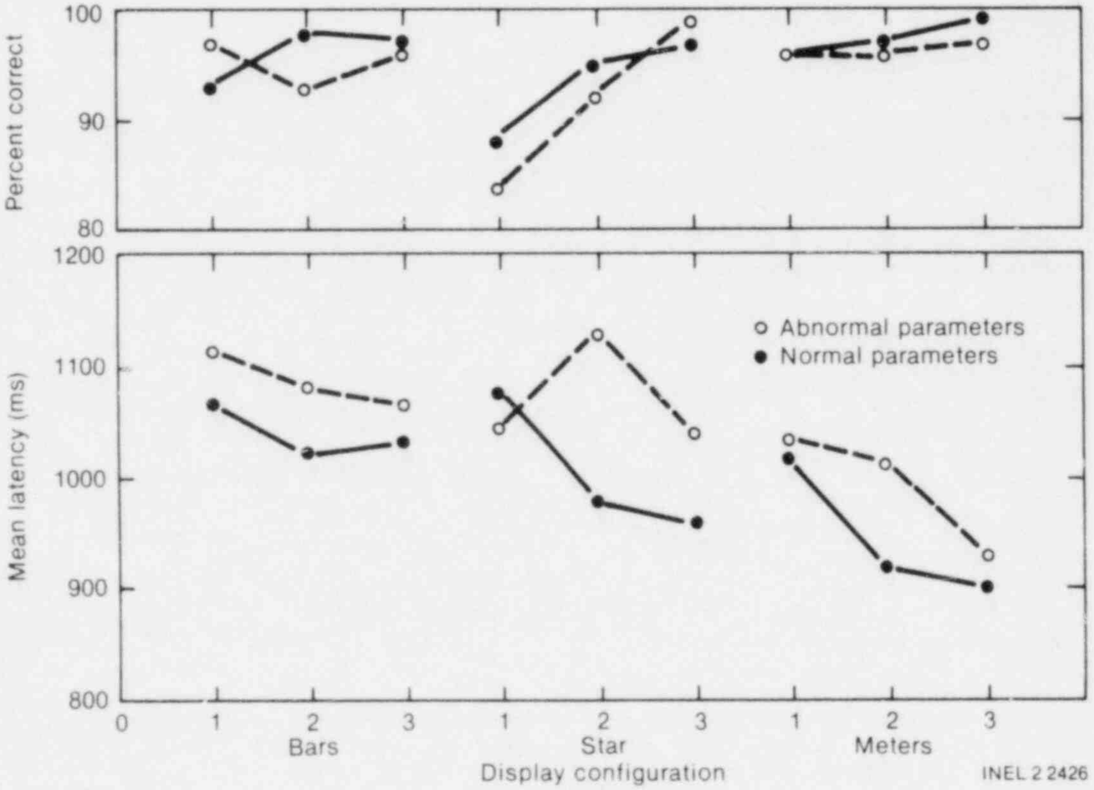


Figure 9. Reaction time and percent correct data from the pilot study for Experiment 4.

engage in some form of search behavior. This result is serendipitous to the extent that it points out an aspect to utilizing displays which interacts with aspects which are considered more perceptual. Both task context and format familiarity are areas where more data are needed to determine the boundaries of the phenomenon before minimum standards can be articulated. There are, no doubt, other areas besides task context and familiarity which markedly influence the evaluation of display formats. Areas such as mental workload and stress, schema congruency, and the implementation paradigm, used to introduce and develop the display, are all candidates (along with others) for investigation. It must also be emphasized that these considerations are not confined to the perceptual aspects of display evaluation, they will be manifest in all of the areas investigated.

CONCLUSIONS

This investigation represents the first leg in developing a set of display evaluation techniques. These techniques will be analyzed using a multiple regression technique to arrive at an optimized methodology for generating data which will support the development of minimum standards for the inclusion of computer generated displays in nuclear power plants. The techniques reported here are used to assess the perceptual aspects display evaluation. User preference, decisionmaking, and sensory aspects of display evaluation will be investigated in future analyses. These diverse measurement techniques will provide a more complete composite for display evaluation than any of the techniques alone.

The techniques reported herein were demonstrated to be sensitive to differences in human performance which result from using different formats to present the same information. The formats used in these experiments were very similar to one another, so any differences that were observed were indicative of very sensitive instruments. The fact that the techniques yielded quite different sets of results indicates not only that the test instruments were very sensitive to changes in performance but also, that a satisfactory evaluation methodology must take into account the task(s) which the display is intended to serve. This conclusion also is supported by other investigators (Christ, 1977;[19] Crawford, 1977;[20] and Tullis, 1981[21]). Again, it should be emphasized that format evaluation is task/situation dependent.

Additionally, the data from Experiment 4, and the pilot study for that experiment, brought out another important aspect to display evaluation. It appears that the subjects' performances were influenced by their familiarity with the displays. As we saw in Experiment 4, differences that appear at a low level of familiarity may disappear as the displays become more familiar. The reverse may also be true, displays that work well for the unfamiliar viewer may be unsatisfactory for the more familiar user.

As was stated above, this research is but the initial step toward developing an evaluation methodology which takes into account the multitude of variables and considerations which impact upon the assessment of graphic/alphanumeric displays. Still to come are the development of methodologies which go beyond the perceptual aspects of display evaluation, including experiments which emphasize the impact of display format or other cognitive processes. Decisionmaking during real-time, controlled transients is an area that will receive top priority during the simulation phase of this work. Other areas, such as, sensory/physiological, individual preference, social, and organizational issues, will be examined during checklist evaluations and multivariate rating scales.

REFERENCES

1. W. W. Banks, D. I. Gertman, R. J. Petersen, *Human Engineering Design Considerations for Cathode Ray Tube-Generated Displays*, NUREG/CR-2496, EGG-2161, April 1982.

2. G. T. Fechner, *Elemente der Psychophysik*. Leigepi: Breitkopf und Hartel, 1869.

3. J. M. Cattell, "On Errors of Observation," *American Journal of Psychology, 5*, 1893, pp. 285-293.

4. J. A. Jastrow, "A Critique of Psychophysic Methods," *American Journal of Psychology, 1*, 1888, pp. 271-309.

5. R. F. Urban, *Psychological Review*, 1910, in E. G. Boring, "The Stimulus-Error," *American Journal of Psychology, 32*, 1921, pp. 449-471.

6. W. P. Tanner and J. A. Swets, "A Decision Making Theory of Visual Detection," *Psychological Review, 61*, 1954, pp. 401-409.

7. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, New York: Wiley, 1966.

8. D. W. Massaro, *Experimental Psychology and Information Processing*, Chicago: Rand McNally Publishing Co., 1975.

9. G. A. Miller, "The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information," *Psychological Review, 63*, 1956, pp. 81-97.

10. G. Sperling, "The Information Available in Brief Presentations," *Psychological Monographs, 74*, (11, Whole No. 498) 1960.

11. G. Sperling, "A Model for Visual Memory Tasks," *Human Factors, 5*, 1963, pp. 19-31.

12. F. C. Donders, "Over De Snelheid Van Psychische Processen," Onderzoekingen gedaan in bet Physioligisch Lo boratorium der Utrechtsche Hoogeschool. 1868-1869, Tweede reeks, II, 92-120. Translated by W. G. Koster in W. G. Koster (Ed.), *Attention and Performance II, Acta Psychologica, 30*, 1969, pp. 412-431.

13. W. K. Estes, "The Locus of Inferential and Perceptual Processes in Letter Identification," *Journal of Experimental Psychology: General, 104*, 1975, pp. 122-145.

14. M. I. Posner, "Psychobiology of Attention," in M. S. Gazzaniga and C. Blakemore (eds.), *Handbook of Psychology*, New York: Academic Press, 1975.

15. R. G. Pachella, "The Information of Reaction Time in Information-Processing Research," in B. H. Kantowitz (ed.), *Human Information Processing; Tutorials in Performances and Cognition*, Hillsdale New Jersey: Lawrence Erlbaum Associates, 1974.

16. D. LaBerge, "Attention and Automatic Information Processing," in P. M. A. Rabbit (ed.), *Attention and Performance, V.*, London: Academic Press, 1976.

17. D. LaBerge and S. J. Samuels, "Toward a Theory of Automatic Information Processing in Reading," *Cognitive Psychology, 6*, 1974, pp. 293-323.

18. E. J. Gibson and H. Levin, *The Psychology of Reading*, Cambridge: MIT Press, 1981.

19. R. E. Christ, "Four Years of Color Research for Visual Displays," *Proceedings of the Human Factors Society 21st Annual Meeting, San Francisco, California, October 1977.*

20. A. M. Crawford, "Interactive Computing Graphics for Simulation Displays," *Proceedings of the Human Factors Society 21st Annual Meeting, San Francisco, California, October 1977.*

21. T. S. Tullis, "An Evaluation of Alphanumeric, Graphic and Color Information Displays," *Human Factors, 23*, 5, 1981, pp. 541-550.

APPENDIX A

A PRIMER ON THE THEORY OF SIGNAL DETECTION

APPENDIX A

A PRIMER ON THE THEORY OF SIGNAL DETECTION

The theory of signal detection (TSD) offers a unique view of information extraction that is valuable here because it avoids many of the problems which burden other methodologies, while offering independent/quantitative estimates of sensitivity and response criterions. In order to give the reader a theoretical background sufficient to understand TSD, the following paragraphs are presented from Massaro, 1975,$^{A-1}$ which are a very understandable overview of the signal detection paradigm.

"There are only two possible states of the world in the psychophysical task: SN (signal present) and N (no signal). Accordingly, it is only natural to assume that there are only two possible outputs of the sensory system: s (sensation) and n (no sensation). However, according to a multistate theory—the theory of signal detectability (Green and Swets, 1966)—many sensory states are possible. The central assumption of this theory is that no threshold or barrier exists that must be overcome for a sensation. Rather, there is always some background noise in the sensory system, which always produces some positive sensation value. Therefore, even though there are only two stimulus trials, many possible outputs of the sensory system can occur when it processes either of these two stimulus inputs. Although the subject could be presented with a constant stimulus from trial to trial, he actually knows a different amount on each trial, a difference that can extend over a wide range. For a given stimulus event, the sensory system can output any of a number of sensation values corresponding to the magnitude of the sensation.

"The transition matrix corresponding to the sensory system of multistate theory, therefore, consists of two stimulus states and m sensory states.

$$\begin{array}{ll} & \text{Sensory State} \\ & \begin{array}{ccccccc} S_1 & S_2 & S_3 & \cdots\cdots & S_i & \cdots\cdots & S_m \end{array} \\ \text{Stimulus} \begin{array}{c} \text{SN} \\ \text{S} \end{array} & \left[\begin{array}{ccccccc} p_1 & p_2 & p_3 & \cdots\cdots & p_i & \cdots\cdots & p_m \\ q_1 & q_2 & q_3 & \cdots\cdots & q_i & \cdots\cdots & q_m \end{array}\right] \end{array}$$

$$\sum_{i=1}^{m} p_i = 1, \qquad \sum_{i=1}^{m} q_i = 1$$

"As can be seen in the transition matrix, a signal gives rise to sensory state s_i with probability p_i and a no-signal trial gives rise to this state with probability q_i. It is assumed that the sensory states are ordered in magnitude along with some dimension, for example, the magnitude of sensation. In this case, the magnitude of sensation given by sensory state s_i is less than that given by sensory state s_{i+1} (s_i s_{i+1}).

"If s_1 is the smallest magnitude for A sensory state, then we would expect q_1 to be larger than p_1. In other words, it should be more likely for a no-signal trial to elicit sensory state s_1 than for a signal trail to do so. In contrast, if sensory state s_m is the largest magnitude of sensation, we would expect that p_m q_m. Somewhere along the continuum of sensory states, then, the relative values of p_1 and q_1 are larger at the extreme values of s_i than in the middle range of values.

"According to multistate theory the decision system is faced with any of a whole range of sensation values X, given the sensory system. The task facing the decision system in the multistate theory, therefore, differs significantly from its task in the threshold models. In a two-alternative task, however, the decision system can divide the range of sensory output into two classes: those that are smaller than some sensation value s_i and those that are larger than this sensation value. In multistate theory, these two sets of values are treated differently by the decision system. It is assumed that the decision system responds 'no' for sensation values that are below the cutoff sensation value and 'yes' for sensation values that are larger than this cutoff value.

"In this model, then, it is assumed that the decision system chooses a cutoff or criterion value C such that if the sensory value s_i from the sensory system is equal to or exceeds this value, a 'yes' response is executed; otherwise, the observer says 'no.' This decision rule can be represented by the following transition matrix:

$$\begin{array}{ll} & \text{Response} \\ & \begin{array}{cc} \text{yes} & \text{no} \end{array} \\ \begin{array}{ll} \text{Sensory} & s_i \geq C \\ \text{state} & s_i < C \end{array} & \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \end{array}$$

"As can be seen in the transition matrix representing the decision system, the decision rule is assumed to be deterministic rather than probabilistic. Given a total value from the sensory system, the response is determined with probability 1. This decision rule contrasts with the probabilistic decision rule of the general two-state model.

Likelihood Ratio

"In the mathematical formulation of the theory of signal detectability, the decision rule of the subject is assumed to be analogous to one in statistical decision theory. It is assumed that the decision system assigns conditional probabilities to the output of the sensory system. The decision system computes the conditional probability that the output X from the sensory system arose from the SN trial, P (X|SN), and the probability that it arose from an N trial, P (X|SN). The likelihood ratio L is a ratio of these two probabilities:

$$L = \frac{P(X|SN)}{P(X|N)}$$

"The decision system has a criterion value, so that if the likelihood ratio exceeds this value it responds 'yes'; otherwise, it responds 'no.' The unbiased observer would have a criterion value set at 1, the point at which it is equally likely that X came from an SN or an N trial.

"The exact predictions of multistate theory are not easily derived from the transition matrices of the sensory and decision systems because of the large number of sensory states. The theory, however, can be formalized and actually simulated or modeled in a straightforward manner. First, we shall describe how the multistate theory can be simulated, using a dice game analog. For this analogy we imagine an experimenter with 3 dice, including one stimulus or signal die. The signal die is imprinted on three of its sides with the value 3, and on the remaining three sides with the value zero. The other two dice in the game are normal dice with the values 1 through 6 on their six respective sides. The experimenter rolls the three dice and announces the sum of the values on their faces. A subject must decide, on the basis of the total of all the values, whether the value showing on the signal die was a 3 or a zero.

"The dice in this game are analogous to the stimulus trial in a simple detection experiment; the total dice value presented to the subject by the experimenter represents the sensation—the information from the sensory system. The subject in the dice game performs the task of the decision system, judging whether the signal die is 3 or zero on the basis of the total value. We use a dice game analogy because, in a real psychological task, we cannot observe the value X that is produced by the sensory system. We do not know what sensation a stimulus produces in a subject; we only know the response he chooses to make. If we do not know the nature of the sensation, we do not know the exact decision problem he faces in choosing that response.

"In a dice game model of a particular sensory system, on the other hand, we have direct control over the possible outcomes of the roll of dice, and therefore, over the values that can be transmitted from the sensory system to the decision system. Thus we can change our dice game to represent different experiments in which variables are manipulated to affect the operations of the particular model of the sensory system clarifies the operation of the sensory system and the task that the decision system faces, given the output of the sensory system.

"The multistate theory challenges the assumption that there is a threshold. Although the experimenter presents either a signal or no signal trial, the sensory system always gives a certain positive sensation value. This theory can best be understood by simulating it with our dice game. The multistate theory assumes that the output of the sensory system is analogous to the total number of points obtained from the three dice. One die is our special signal die with three 3s and three 0s; a 3 corresponds to those trials on which a signal is presented and zero corresponds to no signal or blank trials in the psychophysical task. The other two dice are normal dice whose six sides each have a number from 1 to 6; these dice are analogous to the noise that multistate assumes to exist in the sensory system. We roll all three dice; the decision system, given only the combined total, must say 'yes' or 'no' whether the stimulus die is showing a 3 (signal) or a zero (non-signal), respectively.

"In this case we have a range of possible combined totals from 2 (0 + 1 + 1) to 15 (3 + 6 + 6). Given the total value of 2, the decision system can say 'no' with absolute confidence. Similarly, it can say 'yes' to the value of 15 with 100 percent accuracy. In fact, the combined values up to 4 could not occur if the signal die is 3; and the combined values of 13 and over could not occur if the special die is zero. So up to 4, and over 13, the decision system can be 100 percent correct. For all values in between, the decision system cannot be absolutely certain of its decision. The possible outcomes in this dice game given that a 3 or zero is showing on the signal die are presented in Table 1. The total value of 6 can occur in seven ways; five ways with a zero on the signal die and two ways given a 3 on the signal die. Therefore, given the total value of 6, the probability of a signal being present is 2 out of 7 or 2/7.

"This particular dice game can be analyzed in terms of a transition matrix of the sensory process. Since the dice totals are in the range of 2 to 15, and these correspond to the possible number of sensory states, we have 14 possible sensory states. Of course, there are only two possible stimulus trials, signal (3) or no signal (0). Therefore, the sensory process can be represented by the following transition matrix:

Sensory State

$$\begin{array}{ll} & \begin{array}{cccccccccccccc} 2&3&4&5&6&7&8&9&10&11&12&13&14&15 \end{array} \\ \text{Stimulus} \begin{array}{c} 3 \\ 0 \end{array} & \begin{bmatrix} 0&0&0&1&2&3&4&5&6&5&4&3&2&1 \\ 1&2&3&4&5&6&5&4&3&2&1&0&0&0 \end{bmatrix} \end{array}$$

Each entry in the matrix is divided by 36.

For example, if the odd die is showing a 3, the sensory system will output a total value of 10 points is 6/36. In this case, $p_{10} = 6/36 = 1/6\ 0.17$.

"The multistate theory assumes that the detection system transmits a wide range of values as in the totals in the dice game. The decision maker receives one of these values, computes—in some sense—the likelihood of a signal being present given this value, and makes his response according to some decision rule. This theory does not require the concept of a threshold—a barrier that must be overcome before a sensation is possible. On the contrary, it assumes that the detection stage always has a sensation value, and that the sensation can vary across a wide range. In our dice game simulation, the output of the detection system is very informative when the value is 2 or 15, but essentially uninformative when the value is 8 or 9. The optimal decision procedure in this situation is to establish a cutoff point, say the total 8-1/2, at which the decision maker would cease to respond 'no' and begin to respond 'yes.'

"In the dice game model of this theory, the decision maker is faced with one of the many possible values from 2 through 15. His task is to translate the value he received into one of only two possible responses. For a few of the combined totals he might receive, this could be done with confidence. Combined totals of 2, 3, or 4 mean that the signal die could only be showing zero, and, therefore, the correct response could only be 'no.' For combined totals above 12, on the other hand, the signal die would have to be showing 3. Given a combined total of 13, 14, or 15, therefore, the decision process could choose a 'yes' response with absolute confidence. Values between 5 and 12, however, represent varying probabilities of either condition.

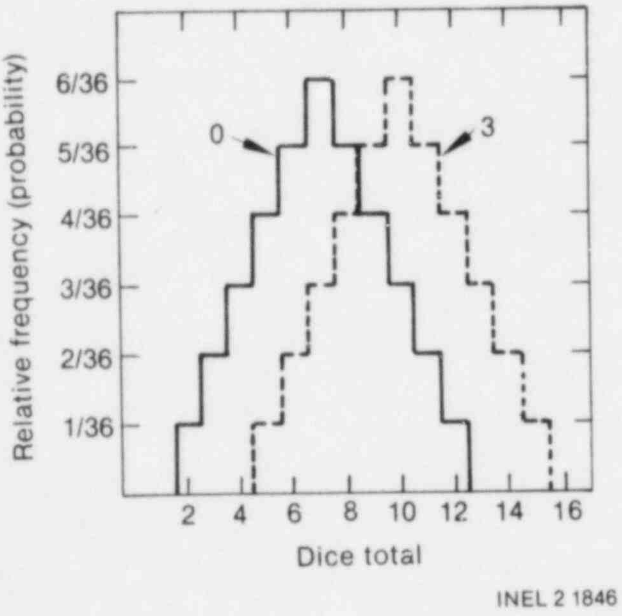
"Another way of seeing the relationship between the sensory and decision systems, is to plot the distribution of total scores given in Table 1. The expected distribution of totals given

"Table 1. Frequency of possible totals given the signal die showing 3 or 0 for a block of 72 trials

	Frequency given signal die is	
Dice Total	3	0
2	0	1
3	0	2
4	0	3
5	1	4
6	2	5
7	3	6
8	4	5
9	5	4
10	6	3
11	5	2
12	4	1
13	3	0
14	2	0
15	1	0
Total combinations	36	36

that the signal die contains a zero or a 3 are shown in Figure 1. The total number of points represents the output of the sensory system, the values given to the decision system. These values, then make up a decision continuum, with some total values giving 100 percent probability of a correct 'yes' at the extreme right, and those giving 100 percent certainty of 'no' at the extreme left. As the values given to the decision process increase from left to right, uncertainty grows; at 8 and 9, it is very unsure of the answer. Thereafter the probability of 'yes' over 'no' increases, until at 13 and beyond the decision maker again knows 100 percent of the information needed to choose a response correctly. The subject in our dice game arrives at his decision choosing some criterion value C, such that if the total is equal to or exceeds that value, he says 'yes'; otherwise he says 'no.' The criterion value that the subject chooses is influenced by experimental variables assumed to affect the decision system.

"As in the two-state models of detection, two processes in the multistate model contribute independently to the final result. The output of the sensory system is affected by the nature of the stimulus and the state of the sensory system. The decision rule is affected by the subject's knowledge of the probability of a signal trial, the payoffs in the experimental situation, and the attitude of the observer.



"Figure 1. The expected distribution of totals as a function of whether the signal die showed 3 or 0.

"Returning to the dice game model of multistate theory, let us change the nature of the stimulus by changing the number representing an SN trial on the signal die from 3 to 6. We now have, as before, three dice: two of them are normally marked with the values 1 through 6, each imprinted on one of the six sides. The signal die is now marked 6 on three of its sides and zero on the remaining three. Again, the decision maker is told only the combined value showing on all dice, and his task is to say 'yes' or 'no,' given this combined value, whether the number showing on the signal die is 6 or zero.

"The range of possible totals will now be from 8 to 18 when 6 is on the signal die, and from 2 to 12 when zero is on the signal die. Therefore, changing the signal level from 3 to 6 changed the possible outputs of the sensory system, as represented by the total number of points showing on the three dice. The decision maker in this case can be certain that 6 is not present given 2, 3, 4, 5, 6, and 7, and that it is present given 13 through 18. The number of trials on which he can be absolutely certain has increased significantly and his performance should be correspondingly better. Table 2 presents the possible outcomes for each total given that 6 or zero is showing on the signal die.

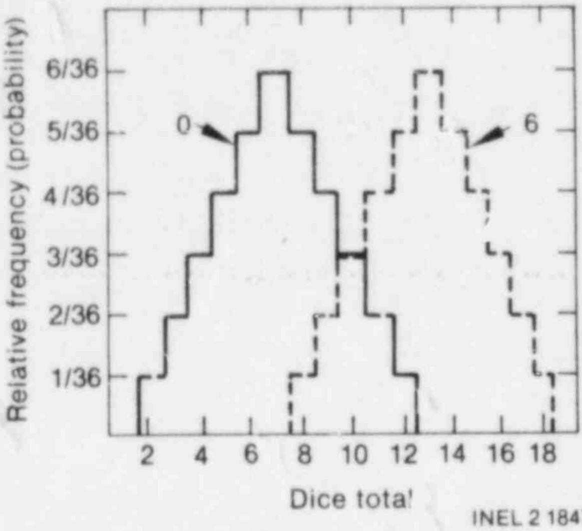
"The following transition matrix of the sensory system presents the relationship between the two stimulus trials and the sensory state outputted by the sensory system.

Sensory State

$$\begin{array}{r|ccccccccccccccccc|} & 2&3&4&5&6&7&8&9&10&11&12&13&14&15&16&17&18\\ \text{Stimulus } 3 & 0&0&0&0&0&0&1&2&3&4&5&6&5&4&3&2&1\\ 0 & 1&2&3&4&5&6&5&4&3&2&1&0&0&0&0&0&0 \end{array}$$

Each entry in the matrix is divided by 36.

"As can be seen in the transition matrix, there is a much larger number of sensory states in which the decision system can be confident about its decision. Figure 2 presents the distribution of total values from the sensory system given the values 6 and 0 showing on the Table 2 signal die. It should be noted that the distance between the two distributions corresponding to the 6 and 0 on the signal die is twice the distance between the 3 and 0 distributions shown in Figure 1.



INEL 2 1847

"Figure 2. The expected distribution of totals as a function of whether the signal die showed 6 or 0.

"Table 2. Frequency of possible totals given the signal die showing 6 or 0 for a block of 72 trials

	Frequency given signal die is	
Dice Total	3	0
2	0	1
3	0	2
4	0	3
5	0	4
6	0	5
7	0	6
8	1	5
9	2	4
10	3	3
11	4	2
12	5	1
13	6	0
14	5	0
15	4	0
16	3	0
17	2	0
18	1	0
Total combinations	36	36

"Correspondingly, we can change the signal die in the opposite direction and choose two very similar numbers, zero and one, representing to signal and signal trends, respectively. Now the decision system is faced with a much more difficult task: of the 72 possible combinations of the three dice, on only two of them will the probability of a correct response be one. As an exercise, work out the possible outcomes for each total given that a one or zero is shown on the signal die, and represent the distribution of these outcomes along a decision contained as in Figures 1 and 2.

"These variations of the signal die can be thought of as reflecting ways in which the efficiency of the sensory system can be made to vary. Analogously, we could change the nature of the noise (the numbers on our normal dice), which would also affect the efficiency of the sensory system. If stimuli are very different from each other, they will be easier to distinguish than if they are very similar. We are very much more sensitive to the difference between a red light and a green one, than to that between a red and red-orange light. The absolute values on the signal die—the stimulus variables in the experimental situation—determine the values that the sensory system transmits to the decision system; the signal die values do not influence the operations of the decision system since the same decision rule or algorithm can be maintained. In both cases, the decision system will want to choose a criterion value that lies halfway between the means of the two distributions. This value will optimize the percentage of correct responses in both cases.

"It should be noted that the decision system must have the information about the range of values transmitted by the sensory system in order to apply its decision rule reliably. If the subject did not know the range of values from SN and N trials, he would not know where to set his criterion and could not respond appropriately. In our dice game, the subject is told the range of values in advance, so that he has the necessary information. In a real signal detection experiment, we practice the subject in the task before the experiment so that he can learn the range of sensation values in the experiment. After each trial, feedback must be given about which stimulus event was presented, allowing the subject to learn the range of sensation values resulting from SN and N trials.

"Other variables in the task affect the operations of the decision system. To illustrate, take up our dice game again with a 3 representing signal trials and a zero representing no signal trials. However, instead of a 3 appearing on three of the six sides, it now appears on five. This manipulation of the a priori probability changes the likelihood of a signal trial from 50 percent to 5/6 ≈ 83 percent, but does not change the possible total values presented to the decision system. The a priori probability refers to the likelihood that the signal die will contain a 3 before the dice total is known. On a given role of the dice, the sensory system possesses no more or no less knowledge about the stimulus than it did when the 3 occurred on only three sides. The values transmitted by the sensory system, or the range of totals given by the three dice, does not change with changes in the a priori probability. All that has changed is the likelihood or the probability that any of the ambiguous values given by the sensory system represents the presence of a 3 on the odd die. The expected frequency of occurrence of 3 and 0 given the possible totals when the a priori probability should accordingly change the rule of the decision system. The decision system should be biased to say 'yes, a 3 was presented,' much more often than in the previous example where zero and 3 trials were equally likely.

"To determine the relative frequency of possible totals with an a priori probability of a signal trial (3) being 5/6, we can simply weigh the frequencies given in Table 1. Before the total is determined, we know that on the average, a 3 will be presented five out of six times. Accordingly, for each of the possible totals from the two normal dice, on 5/6ths of the trials, a signal (3) will be added to this total and on 1/6th of the trials, no signal (0) will be added to the total. For example, given the total value five from the three dice, the relative frequency of a signal (3) will be five times the relative frequency of no signal. With equally likely signals and no signal trials, the total 5 arises from 3 one out of five times, and from 0, four out of five times. If the a priori probability is 5/6ths, the total five will result from a signal trial five times as often, relative to a no signal trial. Therefore, multiplying the relative frequencies 1 and 4 for the total 5 by five and one, respectively, gives the relative frequencies of 5 and 4 when the a priori probability is 5/6.

"The frequencies in Table 3 are converted into relative frequencies or probabilities in the following transition matrix of the sensory system:

"Table 3. Frequency of possible totals given the signal die showing 3 or 0 when a priori possibility of 3 is 5/6 for a block of 216 trials

Dice Total	Frequency given signal die is 3	Frequency given signal die is 0
2	0	1
3	0	2
4	0	3
5	5	4
6	10	5
7	15	6
8	20	5
9	25	4
10	30	3
11	25	2
12	20	1
13	15	0
14	10	0
15	5	0
Total combinations	180	36

Sensory State

$$\begin{array}{ll} & \begin{array}{cccccccccccccc} 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 \end{array} \\ \text{Stimulus} \begin{array}{c} 3 \\ 0 \end{array} & \left[\begin{array}{cccccccccccccc} 0 & 0 & 0 & 5 & 10 & 15 & 20 & 25 & 30 & 25 & 20 & 15 & 10 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 \end{array} \right] \end{array}$$

The entries in the first and second rows are divided by 180 and 36, respectively.

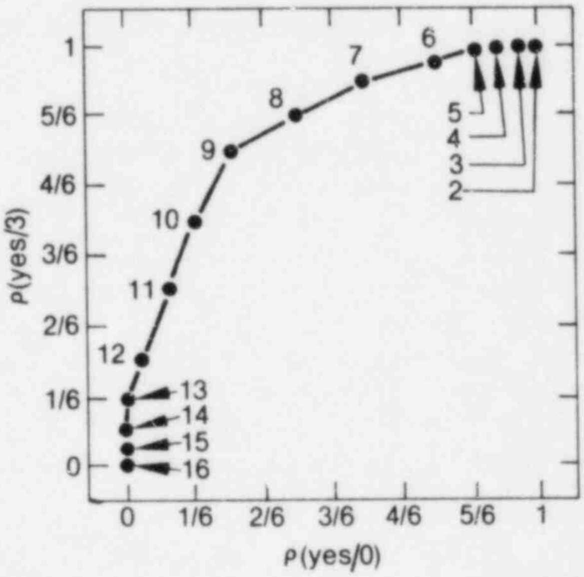
The transition matrix is identical to the matrix generated when the a priori probability of a 3 was 3/6. The number and the values of the sensory states are the same and the transition probabilities are identical in the two cases. The sensory system only makes contact with the stimulus events on a trial. Therefore, manipulations in the a priori probability cannot affect the operations and hence the outputs of the sensory system. Accordingly, our measure of the Table 3 sensitivity of the sensory system should not change with changes in a priori probability. Since the transition matrix for the sensory system does not change with changes in a prior probability, neither does the distribution of sensory states as plotted in Figure 1. Changes in a priori probability do not affect the relative frequency of occurrence of each total given the value showing on the signal die. Therefore, a change in a priori probability does not affect the shape of the distributions or the distance between the means of the two distributions. This contrasts with the manipulation of signal intensity (6 or 3) which directly affects the distance between the two means of the distributions.

"According to this analysis, our measure of the sensitivity of the sensory system should not change with changes in a priori probability, and with the placement of the criterion, since these changes do not affect the operations of the sensory system. We can see that the percentage of the correct responses in the task cannot be used as a measure of the sensitivity of the sensory system. Assume that the observer knows that the a priori probability of a 3 is 3/6 in one case and 5/6 in another. If he simply responded on the basis of this information, he could achieve a performance of only 50 percent correct in the former case but could maintain a performance of 83 percent correct in the latter. He would simply respond 'yes' on every trial regardless of the total value. Then if 'percent correct' were used as a measure of the sensory system, we would conclude that the system was more sensitive with increases in a priori probability—an incorrect conclusion since the actual operations of the sensory system did not change.

"Note that the decision system must have access to the a priori probabilities in the experimental situation to use an optimal decision rule. The experimenter might either tell the observer the a priori probabilities or give the observer feedback from trial to trial so that the latter could discover this fact for himself. As noted, the decision rule can also be affected by the payoff in the situation. If the observer is rewarded for saying 'yes' correctly by a greater increment than he is punished for saying 'yes' incorrectly, it will pay off for him to say 'yes' more often even when the probability is relatively low that 'yes' is the correct response.

"According to multistate theory, the a priori probability of a signal trial (3) affects the decision system's choice of a criterion value. We can actually determine the receiver-operating characteristic predicted by multistate theory which changes in the criterion value. Assume first, for example, that the observer has a strong bias to say 'no'—the a priori probability of a signal trial is very low. The observer may respond 'no' all the time. Therefore, P(yes|3) = 0 and P(yes|0) = 0. As the observer becomes more willing to respond 3, the P(yes|3) increases but so does P(yes|0) as demonstrated in Figure 3. Figure 3 presents the P(yes|0) and P(yes|3) at each possible criterion value on the dice game task when the signal die contains a 3 as the signal trial and a 0 for a no signal or blank trial.

"Now it is necessary to provide a measure of the sensitivity of the sensory system that is invariant with changes in the criterion value. This is given by the distance between the means of the two distributions. How can this distance be calculated in a signal detection task? In this task, we are given the results represented by a confusion matrix. From our hit and false alarm probabilities, we can determine the criterion value of the subject with respect to the means of each of the two distributions and, therefore, compute the distance between the means of the distribution. Assume that in our experiment, an observer responded P(yes|signal) = 0.9 and P(yes|no signal) = 0.3. Therefore, the distribution for the signal trial



INEL 2 1840

"Figure 3. An ROC curve which represents P (yes|3) and P (yes|0) for each possible criterion value in the dice game task. The subject responds 'yes'—a 3 was presented— if the total is larger than or equal to the criterion value. Otherwise, he responds 'no.'

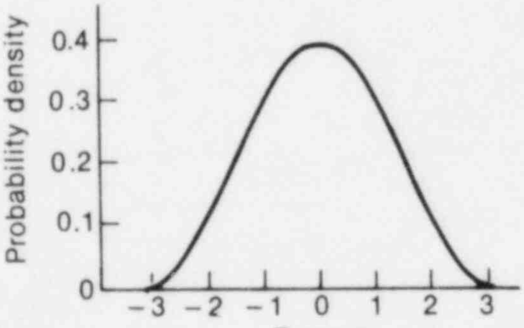
would be drawn so that 90 percent of it lies to the right of the criterion value. Similarly, the distribution for the no signal trail would be drawn so that 30 percent of it lies to the right of the criterion value.

"Since the experimenter does not know the number or values of the sensory states in an actual experiment, multistate theory assumes that the sensory states are distributed normally in a bell-shaped curve with variance equal to 1. This distribution is called the normal distribution and is shown in Figure 4. The normal distribution is similar to the distributions generated by the dice game but is drawn as a smooth curve, since it can take on all positive and negative values, not just the integral values given by the dice game. The normal curve has a strict relationship between the area represented under the curve and the distance along the horizontal axis. Since the curve is symmetrical, 50 percent of the area lies to the right of the mean. The distance from the mean is given by z scores. Since we know the shape and the variance of the curve, we can compute a z score for each percentage of the area to the right or the left of the mean. Similarly, we can derive the percentage of the area between a point along the horizontal axis and the mean. Table 4 gives the distance between the mean and the criterion as a function of the percentage of the distribution contained between these points.

"Table 4. The z score distance corresponding to the percentage of the distribution between the mean and the criterion value. (If the criterion lies to the left of the mean, the z scores are negative.)

Percentage	z score	Percentage	z score
0	0	26	0.707
1	0.025	27	0.739
2	0.050	28	0.772
3	0.075	29	0.806
4	0.100	30	0.842
5	0.125	31	0.878
6	0.150	32	0.915
7	0.176	33	0.954
8	0.202	34	0.995
9	0.223	35	1.037
10	0.253	36	1.080
11	0.280	37	1.126
12	0.306	38	1.175
13	0.332	39	1.226
14	0.358	40	1.282
15	0.385	41	1.340
16	0.403	42	1.405
17	0.440	43	1.476
18	0.468	44	1.555
19	0.496	45	1.645
20	0.524	46	1.751
21	0.553	47	1.882
22	0.584	48	2.054
23	0.610	49	2.327
24	0.643	49.5	2.575
25	0.675	49.9	3.085

"Accordingly, representing the distributions of sensory states by normal curves, we can compute the distance between the means of the two distributions. We use this distance called d prime and written as d' as an index of performance in the detection task, since it remains fixed with changes in the criterion value. To the extent that the two distributions of sensory states differ from each other, d' will be large. To the extent that the two distributions are similar, d' will be small. In our hypothetical experiment we said that P(yes|signal) = 0.9 and P(yes|no signal) = 0.3. The hit rate of 0.9 means that the criterion was set so that 90 percent of the distribution of signal trials was to the right of the criterion. This is shown graphically in Figure 5.

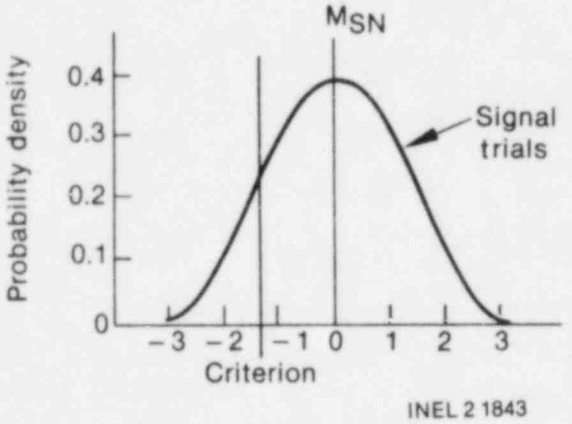


INEL 2 1865

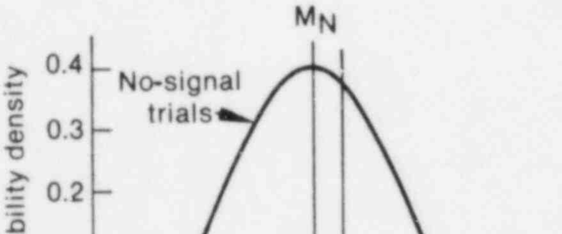
"Figure 4. The normal distribution with a mean of 0 and a variance of 1. The values along the X abscissa are called z scores.

Figure 6 is a graphic representation of the normal curve of the no signal distribution, with

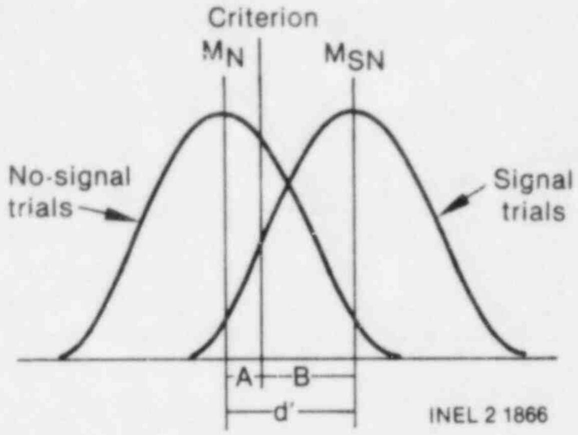
30 percent of the curve to the right of the criterion. The relationship between the two curves is given in Figure 7. Accordingly, the distance between the means is given by the sum of the absolute distance of A and B. To determine the distance B between the criterion point and the mean of the signal distribution, we want the value of the z score of a point that lies 90 percent to the left of the distribution or 40 percent to the left of the mean of the distribution, since the curve is symmetrical around the mean. Table 4 shows, that the z score which represents this distance along the X abscissa is—1.28. (The value is minus since the criterion lies to the left of the mean.) Analogously, to find the distance of the same criterion point from the mean of the no signal distribution, we want the distance between the mean of this distribution and a point that lies 20 percent to the right of the mean. This distance as shown in Table 4 corresponds to a z score of 0.52. Given that the criterion lies between the means of the two distributions, the distance between the means of the two distributions is given by the sum of the positive value of A and B. Therefore, the distance d' is given by the sum of 1.28 and 0.52, which is 1.80. Our measure of sensitivity in this task (1.8) provides an index of the sensitivity of the sensory system that is contaminated by the exact placement of the criterion value. Given a fixed distance between the means, the calculation of this distance using z scores obtains the same distance value regardless of the placement of the criterion.



"Figure 5. The criterion value is drawn so that 90 percent of the possible values of a signal trial lie above the criterion.



"Figure 6. The distribution representing no-signal trials is placed so that 30 percent of the no-signal trials lie above the criterion.



"Figure 7. The relationship between the signal (SN) and no-signal (N) curves shown in Figures 5 and 6.

"The distance between the means provides a measure of sensitivity of the sensory system. Sometimes it is also informative to have a measure of the bias in the decision system. The most straightforward measure of this bias is simply the overall or marginal probability of a 'yes' response. This probability lies between zero and one and reflects the willingness of the observer to say 'yes.' To calculate the marginal probability of a 'yes' response, we take the average of the conditional probabilities of a 'yes' response given signal and no signal trials, weighted by the probability of occurrence of the signal and no signal trials.

$$P(\text{yes}) = P(SN) \cdot P(\text{yes} \mid SN) + P(N) \cdot P(\text{yes} \mid N)$$

This measure of P(yes), then, indexes the willingness of the observer to say 'yes,' independent of sensitivity of the sensory system.

"The multistate theory provides a method of data analysis in order to provide a measure of the sensitivity of the sensory system that is independent of changes in the decision system. This method of analysis conditions even though the variables affecting the decision rule differ in the different situations. This method achieves Fechner's original goal: to relate the sensation of the observer to the stimulus situation although the response of the subject does not appear to be directly related to the stimulus situation."

Reference

A-1. D. W. Massaro, *Experimental Psychology and Information Processing*, Chicago: Rand McNally Publishing Co., 1975.

APPENDIX B

MANOVA TABLES

APPENDIX B

MANOVA TABLES

MANOVA tables for Experiments 2, 3, and 4 are presented in Tables B-1, B-2, and B-3, respectively.

Table B-1. MANOVA for Experiment 2

	df	MS	MS error	F	p
Display type	6	—	—	27.59	0.00001
d'	2	621.40	8.55	72.65	0.00001
β	2	200.63	5.15	39.00	0.00001
%	2	5603.68	97.96	57.20	0.00001
Duration	12	—	—	11.8	0.00001
d'	4	198.38	8.55	23.19	0.00001
β	4	50.98	5.15	9.9	0.00001
%	4	3025.87	97.96	31.90	0.00001
Display duration	24	—	—	3.76	0.00001
d'	8	20.13	8.55	2.35	0.0185
β	8	16.73	5.15	3.25	0.0015
%	8	652.87	97.96	6.67	0.00001

Table B-2. MANOVA for Experiment 3

	df	MS	MS error	F	p
Display type	6	—	—	3.42	0.00001
d'	2	114.07	2.99	38.04	0.00001
β	2	18.96	0.61	30.89	0.00001
%	2	10,005.67	53.58	186.74	0.00001
Test blocks	6	—	—	0.84	0.54 NS
d'	2	0.85	2.99	0.28	0.75 NS
β	2	0.15	0.61	0.24	0.79 NS
%	2	26.54	53.58	0.49	0.61 NS
Display duration	14	—	—	0.84	0.60 NS
d'	4	0.78	2.99	0.26	0.90 NS
β	4	0.04	0.613	0.06	0.99 NS
%	3	10.83	53.58	0.20	0.94 NS

Table B-3. MANOVA for Experiment 4

	df	MS	MS error	F	p
Display type	8	—	—	1.06	0.39 NS
d'	2	9.61	10.26	0.94	0.39 NS
β	2	3.98	4.05	0.98	0.38 NS
%	2	79.52	55.69	1.43	0.25 NS
RT	2	91,156.96	32,018.49	2.85	0.06 NS
Test blocks	8	—	—	1.14	0.33 NS
d'	2	2.48	10.26	0.24	0.79 NS
β	2	2.41	4.06	0.59	0.55 NS
%	2	26.77	55.69	0.48	0.62 NS
RT	2	40,188.17	32,018.49	1.26	0.29 NS
Display duration	16	—	—	0.53	0.95 NS
d'	4	3.17	10.26	0.31	0.87 NS
β	4	5.76	4.05	1.41	0.23 NS
%	4	37.95	55.70	0.68	0.61 NS
RT	2	2,824.50	32,018.49	0.09	0.99 NS

120555078877 1 ANRX
US NRC
ADM DIV OF TIDC
POLICY & PUBLICATIONS MGT BR
PDR NUREG COPY
LA 212
WASHINGTON DC 20555

EG&G Idaho, Inc.
P.O. Box 1625
Idaho Falls, Idaho 83415