UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

DTIC ELECTE DEC 20 1988 D Cb

AD-A202 462

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

DTIC FILE COPY

1b. RESTRICTIVE MARKINGS	- -
3. DISTRIBUTION/AVAILABILITY OF REPORT	Approved for public release; distribution is unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	Technical Report No. CPL 86-8
5. MONITORING ORGANIZATION REPORT NUMBER(S)	ARI Research Note 88-91
6a. NAME OF PERFORMING ORGANIZATION	Department of Psychology, University of Illinois
6b. OFFICE SYMBOL (If applicable)	n/a
7a. NAME OF MONITORING ORGANIZATION	U.S. Army Research Institute for the Behavioral and Social Sciences
6c. ADDRESS (City, State, and ZIP Code)	Champaign, Illinois 61820
7b. ADDRESS (City, State, and ZIP Code)	5001 Eisenhower Avenue, Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION	see 7a.
8b. OFFICE SYMBOL (If applicable)	PERI-BR
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	MDA903-83-K-0255
8c. ADDRESS (City, State, and ZIP Code)	see 7b.

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
6.11.02	2Q2611 02B74F	n/a	n/a

11. TITLE (Include Security Classification)

Non-Optimality in the Diagnosis of Dynamic System States

12. PERSONAL AUTHOR(S)

Barbara Barnett and Christopher D. Wickens

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim Report	FROM 85-10 TO 86-10	1988, October	42

16. SUPPLEMENTARY NOTATION

Michael Drillings, contracting officer's representative

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Dual Task Performance, Decision Making, Anchoring, Information Processing, Heuristics, Optimality, Information Load, Judgement, Salience, (OVER)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research note discusses various sources of non-optimality in the diagnosis of a dynamic system's state, looking at them within the context of a military flight scenario. Subjects examined integrated cues which varied in their informational worth under different conditions of information load and clue salience. Actual responses were correlated with an optimal response function, as well as with seven non-optimal response functions, modeled on the basis of filtering, heuristics, and salience biases. Sequential updating strategies were also analyzed.

Results from the two studies indicated that the optimal response function provided the best fit to the data. The imposition of time stress produced a slight bias in favor of processing more salient display locations. A significant performance decrement occured in secondary task conditions, manifest in a trend toward conservatism in judgement, but no biases in display sampling. Analysis of sequential updating strategies also suggested that hypothesis updating was somewhat conservative. Keywords:

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS
21. ABSTRACT SECURITY CLASSIFICATION	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	Michael Drillings
22b. TELEPHONE (Include Area Code)	202/274-8722
22c. OFFICE SYMBOL	PERI-BR

DD Form 1473, JUN 86 — Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

88 12 19 106

ARI RESEARCH NOTE 88-91

18. Subject Terms (continued)

Displays,
Stress. (SDRD)

DTIC
COPY
INSPECTED
6

A-1

ARI Research Note 88-91

Non-Optimality in the Diagnosis of Dynamic System States

Barbara Barnett and Christopher D. Wickens
University of Illinois

for

Contracting Officer's Representative
Michael Drillings

Basic Research Laboratory
Michael Kaplan, Director

November 1988

United States Army
Research Institute for the Behavioral and Social Sciences

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

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

EDGAR M. JOHNSON Technical Director	JON W. BLADES COL, IN Commanding

Research accomplished under contract for the Department of the Army

University of Illinois

Technical review by

Dan Ragland

Table of Contents

Abstract 1

Introduction 2

Overview 2
Information Processing Approach to Decision Making 3
Information Load 4
Filtering Strategies 6
Anchoring and Adjustment 8
Limitations of the Static Approach 10
Focus of Present Study 10

Experiment 1 12

Method 12
Results 18
Discussion 19

Experiment 2 20

Method 20
Results 21
Discussion 29

Conclusion 32

References 34

Appendix 38

Non-optimality in the Diagnosis of Dynamic System States

Barbara J. Barnett and Christopher D. Wickens

Department of Psychology
University of Illinois at Urbana-Champaign

Abstract

Various sources of non-optimality in the diagnosis of a dynamic system's state were investigated within the context of a military flight scenario. Subjects integrated cues of varying information worth under different conditions of information load (manipulated by number of information sources, time stress, and the addition of secondary task processing) and cue salience (manipulated by varying display formats). Actual responses were correlated with an optimal response function, as well as seven non-optimal response functions, modeled on the basis of filtering, heuristics and salience biases. Sequential updating strategies were also analyzed. Results from two studies indicated that the optimal response function provided the best fit to the data. The imposition of time stress produced a slight bias for processing more salient display locations. A significant performance decrement occurred in secondary task conditions, manifest in a trend toward conservatism in judgment, but no biases in display sampling. Analysis of sequential updating strategies also suggested that hypothesis updating was somewhat conservative.

Introduction

Overview

Increasing emphasis on automation in many modern complex systems such as the nuclear power plant or the jet aircraft has forced the role of the human operator to evolve from one of an active controller to that of a systems monitor. One major impact of the human operator's changing role in the dynamic system is that the former emphasis on skills necessary to control the manual system are of much less significance when supervising the automated system. The new role calls for a unique set of skills, requiring the human operator to scan and monitor a number of displays, to update and integrate new information with previous system information, to predict future states of the system, and to make decisions and control actions based on those predictions (Moray, 1986; Wickens, 1984). In order to effectively carry out this complex role, it has been speculated that the human operator must form some type of mental image or mental model of the system under control (Bainbridge, 1974; Moray, 1986; Rasmussen, 1981).

Few researchers have investigated the extent to which humans fail in the ability to effectively integrate information in the assessment of the dynamic system state. This omission is somewhat surprising, given the number of similarities between the dynamic integration and assessment task, and the overwhelming body of literature dealing with decisions under risk or uncertainty (see Einhorn & Hogarth, 1981; Kahneman, Slovic & Tversky, 1982; and Pitz & Sachs, 1984 for recent reviews).

Wickens (1985) summarizes three specific similarities that exist between the two domains. First, both rely on a number of information sources that must be integrated in order to gain the most possible knowledge about the true state of the world. Second, this information to be integrated may be presented either sequentially or simultaneously. Finally, such information may not always be of perfect reliability; thus the impact of each source varies accordingly.

Given the high degree of correlation between these two realms of interest, a logical extension of existing research both in cognitive decision theory and in performance assessment of control processes, is an investigation of the extent to which sources of systematic bias and non-optimality in human judgment, demonstrated in cognitive decision theory, interfere with diagnosis of the complex dynamic system state. The present study seeks to examine the extent to which such biases and heuristics operate within this dynamic realm, focusing specifically on integration of periodically updated cues of varying information value.

Before giving details of the present study, a discussion of relevent literature from both research domains is considered. An information processing approach to decision making in complex·environments is

discussed within the context of a model suggested by Wickens and Flach (in press). Secondly, the relevant literature in cognitive decision theory is reviewed, focusing specifically on judgmental heuristics and biases. This review is followed by a brief discussion of some of the limitations in the static approach to studying decision performance. Finally, a brief outline of the present experiment, including the paradigm employed and experimental hypotheses, is presented.

Information Processing Approach to Decision Making

Much reaserch has been devoted to describing the information processing strategies used by humans when making decisions (Bettman, 1979; Knapp & Tolbert, 1985; Svensen, 1979; Wallsten, 1980). Initially, the information processing approach to decision making focused on consideration of all relevant information (Payne, Bettman & Johnson, 1986). Examples of these decision models include normative decision models such as Additive Utility Theory, Bayes' Theorem, and Multiattribute Utility Theory. Such theories imply an exhaustive utilization of all information sources, weighted according to subjective or objective criteria.

The problem with such normative decision theories is that prescribed decision processes are often incompatible with our current knowledge of human cognitive processes (Brehmer, 1981; Einhorn & Hogarth, 1981; Slovic, Fischoff & Lichtenstein, 1977; Wallsten, 1980, Wickens, 1984). Furthermore, such theories do not account for research demonstrating a wide variety of decision-making strategies contingent upon contextual parameters of the decision-making task (Johnson & Payne, 1985; Tversky & Kahneman, 1981; Payne, 1976, Payne, 1982).

With increasing skepticism of the ability of normative decision theories to describe human decision performance, a vast body of research on judgmental heuristics and biases has emerged as one popular descriptive tool (Tversky & Kahneman, 1974; 1981; Kahneman, Slovic & Tversky, 1982; Slovic, Fischoff & Lichtenstein, 1977).

In one model that points out a number of problems with describing the human information processor in normative terms and emphasizes the role of these heuristics, Wickens and Flach (in press) discuss the decision-making task within the information-processing framework. Their model takes into account various phases of the judgment task in which sources of non-optimality are brought about by limitations in human cognition.

According to the model, the decision-making task may be divided into three separate information-processing components. Initially, environmental cues are sampled, leading to an overall situation assessment and diagnosis. The focus of the present study will be on this particular phase of the decision process. The second phase is to generate alternative courses of action. Finally, the response selection and

execution phase takes place. Wickens and Flach present a discussion of how human cognitive limitations can lead to non-optimal decision performance at each phase of this decision process.

While judgmental heuristics can often simplify the cognitive complexity of the decision-making process, these "rules of thumb" may result in rather systematic deviations from optimal performance (Wickens, 1984). In recent years, a number of such biases in decision-making and judgment have been demonstrated in a wide variety of choice and decision tasks (see Kahneman, Slovic & Tversky, 1982 for recent review).

The discussion to follow, however, will be restricted to those biases hypothesized to influence the dynamic information integration task pertinent to the present study. Therefore, the review will include biases related to information load (manipulated by the number of information sources, presentation of an additional secondary processing task, and by imposition of time stress), filtering strategies (the "as-if" heuristic and the salience bias), and the anchoring-and-adjustment heuristic.

Information Load

Information load has been defined as the amount of data to be processed per unit of time (Wright, 1974). Given this definition, information load may be increased by increasing the number of information sources to be processed in the decision-making task, or by reducing the amount of time available to process a given amount of information. While these sources of load are intrinsic to the decision task, a related manipulation of load involves diverting resources away from the decision task, by requiring concurrent performance of a secondary task. A number of studies provide support for the notion that this increased load leads to a decline in the optimality of performance (Knapp & Tolbert, 1985).

Number of information sources. Wickens (1984) discusses the impact of the number of information sources on the human's ability to make decisions. Limitations in attention and working memory restrict the decision maker from processing more than a few sources of information. Wright (1974) has shown that the difficulty in processing a number of information sources is severely confounded under conditions of time stress. Performance deteriorated as information load was increased, either by increasing the amount of data with which an individual must cope, or by decreasing the time available for processing. Payne (1976), furthermore, found an increase in the use of decision heuristics as the number of decision alternatives increased.

Given the bias effects that emerge as a result of too much information, it seems somewhat ironic that decision-makers still seek more information than they can efficiently integrate, in an attempt to get "all the facts" (Wickens, 1984).

Dual task performance. Information load may be increased further

through the addition of a secondary task to be performed concurrently with the primary decision task. When two tasks are time-shared, attention must be divided between both tasks. Realizing that individuals are limited in the cognitve resources that may be allocated at any given time (Wickens, 1984), one may infer that under certain conditions of information load, in this instance adding a secondary task, the resource limit may be reached, resulting in a performance decrement. For example, most decision tasks are dependent upon substantial resources of working memory. Therefore, the decision making task should be sensitive to concurrent tasks of similar cognitve demands (e.g. running memory task, mental addition task).

Time Pressure. Under conditions of time pressure, subjects must often resort to some heuristic in order to deal with the information overload (Simon, 1981). Ben Zur and Breznitz (1981) propose three ways in which the decision-maker may simplify the decision task under such conditions: acceleration, filtration, and strategy adjustment. Each will be discussed in turn below.

Acceleration refers to a "speeding up" in the information integration process under conditions of time pressure. Such a strategy is of limited usefulnes, however, and would only be beneficial under conditions where the decision maker was not processing information at a maximum rate initially, and thus has more resources available to devote to the processing task (Wickens, 1984).

A filtering strategy restricts the total amount of information that one must process. Such a notion has received considerable attention within the visual performance literature (Hockey, 1970; Sheridan, 1981; Wright, 1974; Wright & Weitz, 1977). Sheridan, for example, refers to the narrowing of attention under conditions of time stress as "cognitive tunnel vision." Specific types of filtering strategies employed by decision-makers in a variety of tasks will be discussed further below.

A shift in processing strategies occurs when the decision maker consciously or unconsciously switches the means by which information is processed under conditions of time stress. Some support for this notion lies in the predictions of Payne's (1982) contingent decision theory, which argues that humans' judgment behavior is contingent or dependent upon the type of decision problem presented. Similarly, Tversky and Kahneman (1981) discuss the importance of "framing effects" in how an individual perceives a problem.

A number of studies have contrasted alternative-based processing with attribute-based processing (Bettman, 1979; Payne, Bettman, & Johnson, 1986; Svenson, 1979). With alternative-based processing strategies, one processes a single alternative completely, weighting and combining each relevant attribute to arrive at an overall evaluation, before attempting to evaluate a second alternative. Attribute-based processing, on the other hand, focuses on the evaluation of a single attribute across all possible alternatives. Under conditions of time

pressure, individuals often switch from alternative-based processing strategies to more attribute-based strategies.

Evidence for all three types of simplification strategies exist within the literature. Aside from those studies mentioned previously, some research shows that under stressful conditions, subjects may use a combination of two or more of these simplifying techniques. Ben Zur and Breznitz (1981), for example, observed subjects using a combination of filtration and limited acceleration strategies under time stress. Payne, Bettman and Johnson (1986) observed a combination of all three types of strategies in subjects under time pressure. Furthermore, they noted that individuals adapt to time pressure in an ordered fashion, accelerating processing initially, then focusing (selectively) on necessary information, and finally changing processing strategies from an alternative-based to an attribute-based strategy if the first two simplification techniques are insufficient.

A study by Payne, Bettman and Johnson (1986), furthermore found that under time stress, several of the attribute-based simplifying heuristics were clearly superior to the more normative (alternative-based) strategies. This finding lends support to the notion that simplifying heuristics are often quite adaptive to difficult circumstances, and in that respect are not necessarily "non-optimal" processing strategies.

Filtering Strategies

It has been established previously that subjects are often forced to filter information under conditions of information overload (Ben Zur & Breznitz, 1981, Tversky & Kahneman, 1974; Wright, 1974). This filtering assumption helps to account for the finding that more information does not necessarily improve decision-making performance. Not only may the filtering task itself compete for valuable processing time, but the filtering strategies used may be non-optimal, thereby eliminating valuable sources of information (Wickens, 1984).

Several criteria for filtering excess information have been identified. Once more, the discussion to follow will be limited to those filtering strategies hypothesized to influence information integration within the dynamic environment. These strategies include filtering by display salience (Payne, 1980; Wallsten, 1980; Wickens, 1984), discounting cue reliability (Kahneman & Tversky, 1973), and the "as-if" heuristic (Johnson, Cavanagh, Spooner & Samet, 1973). Each will be dealt with in greater detail below.

Filtering by salience. Filtering by salience describes a situation in which more perceptually salient information is processed first, and thus has a greater influence on the decision-maker's final judgment (Payne, 1980; Wickens, 1984). It is argued that systematic filtering effects will surface as information load increases. Deviance from optimal performance should vary as a function of whether the salient cues

are of high or low information worth. Specifically, the greater the information worth of salient cues, the smaller will be the non-optimal effects of this narrowing based upon the physical properties of stimuli.

Evidence supports this notion in several recent studies. Payne (1980) discusses a number of influences in decision performance attributed to information salience. Specifically, he states that the nature of problem representation is dependent upon wording of the problem within the text, ordering of sentences within text, whether or not the information was explicitly displayed as part of the stimulus object. Wallsten and Barton (1982) found that stimulus dimensions were processed sequentially from most to least salient. When equating the likelihood of these dimensions, the more salient dimensions carried greater weight in the final judgment. Furthermore, the effects of salience were strengthened under time pressure. Jones and Wickens (1986) also found that the integration of probabilistic information from different regions in space was differentially influenced by the physical locations of the information sources. Russo (1977) found a salience bias in unit price information displayed to consumers. Unit pricing, suggested as a remedy to the problem of calculating the most economical product, was used by consumers only when presented in a simple, organized list, rather than when displayed under each separate brand.

Given that the physical position of information sources on a display panel is a highly salient feature (Van Cott & Kinkade, 1963; Wickens, 1984), the location of these sources used in the monitoring task should influence processing order, and thus the relative weight each has in the integration task (Wallsten & Barton, 1982). Particularly, the more salient locations on the display panel should be at the center (recommended position for crucial warning and frequently-used indicators), and the upper left positions (the point of origin when scanning the entire panel; Van Cott & Kinkade, 1963). Thus, when all other factors are equated, manipulating the location of highly-informative cues should produce different judgments if the observer is filtering by salience.

It should also be noted that there may be other stimulus features, besides spatial position, that may define salience. For example, loud, bright, or colored cues may be more salient. Furthermore, certain semantic or psychological categories of cues may be more salient than others. For example, operators may "trust" visually displayed messages more than verbal messages, or cues bearing on a decision that come from one source, rather than another. These factors may be referred to as psychological salience (Wickens, 1984).

The "as if" heuristic. The "as if" heuristic, a term coined by Johnson, Cavanagh, Spooner & Samet (1973), refers to the tendency of operators to discount differences in cue reliability when formulating a diagnostic hypothesis. The decision-maker treats all cues "as if" they were equally reliable and diagnostic. Optimally, diagnostic weight should be a function of the variable's correlation with the decision criterion.

To the extent that the "as if" heuristic is employed, however, the obtained weighting appears to be more of an "all-or-none" step function (Wickens, 1984). While filtering by salience is more of a perceptual filtering strategy, the "as if" heuristic is more of a cognitive filtering phenomenon.

Decision-makers using the "as if" heuristic, therefore, would weigh most cues as if they were of equally high information worth. If this heuristic is in operation, deviations from optimality would be greatest when several variables of low diagnostic weight indicate a different state from only a few variables of high weight. An optimal decision should emphasize the few cues of high information worth, while the human operator using such a strategy is predicted to to place more emphasis on the large number of cues supporting the alternate hypothesis. Payne, Bettman & Johnson (1986) postulated an "equalweight" bias, similar to the "as if" heuristic, in which subjects summed only values within alternatives, ignoring information about the relative importance of alternatives.

Support for such a bias comes from a number of other researchers (Schum, 1975; Snapper & Fryback, 1971; Trope, 1982; Wickens, et al., 1986), within several different contexts. Schum noted that in criminal trials, jurrors often fail to discount testimony of unreliable witnesses. Snapper and Fryback discussed failures of subjects to optimally discount unreliable reports. Trope found that subjects did not discount confidence in judgment for effects of their own unreliable memory. Wickens et al. found that subjcts' performance did not vary with respect to the informativeness of a predictor cue. Performance was invariant across high, moderate, and uninformative stimuli.

Anchoring and Adjustment

In cases where information is sequentially updated and a number of separate judgments are called for over time, evidence supports the notion of an "anchoring-and-adjustment" updating strategy (Tversky & Kahneman, 1974). According to this heuristic, the subject's initial response serves as the "anchor" from which all further responses are updated on the basis on new information. Several researchers have postulated similar updating strategies (Einhorn & Hogarth, 1985; Lopes, 1981; Wallsten & Barton, 1982). Lopes (1981), and Einhorn and Hogarth (1985) have described general core formulations of sequential information processing, whose parameters can be adjusted to produce differing degrees of anchoring (primacy) or, its counterpart of recency.

One consequence of such an adjustment strategy is that individuals often err on the side of conservatism when updating their initial response. Several studies show that individuals do not respond optimally at extreme conditions, rather they often demonstrate a "central tendency effect" (Lee, 1971), in which low probabilities are overestimated, while high probabilities are underestimated. Wickens (1984) discusses a

similar phenomenon, the sluggish beta effect. Manifest in signal detection tasks, this principle states in essence that individuals are often less risky in judgments than is optimal if the ideal response criterion is low, and less conservative than is optimal when the ideal criterion is high. In conclusion, it appears that human judges often favor less extreme responses, a bias that could lead to suboptimal performance if an extreme response is warranted.

When anchoring and adjustment biases are applied to the scenario of changing evidence in favor of one of two hypotheses, with a "neutral" or indifference point in between, then suboptimal performance may be modeled in one of two ways: the "Elastic" model or the "Viscocity" model.

The Elastic Model. Performance described by this model may be described as an elastic band or spring fastened at the neutral point. Slight updates away from this neutral point (i.e. confirming an already favored hypothesis) are more or less optimal, while such updates are much more conservative at the extremes, as the "force" from the elastic resists further travel in the direction away from neutral. Conversely, this force dictates that updates towards the neutral point are much more liberal at the extremes, and more or less optimal the closer one's belief gets to the neutral point. Support for this model stems from work by Einhorn and Hogarth (1985), described by their surprise/contrast model, in which extreme positions are aided little by confirming information, but decreased a lot by disconfirming evidence.

The Viscocity Model. This model assumes that an individual's updates are always rather conservative (Lee, 1971). In other words, regardless of the current position (whether neutral or extreme), new evidence is given less credibility than it should, rather like a constant force applied through a highly viscous medium. This model in essence describes the "sluggish beta" phenomenon described previously (Wickens, 1984).

There are instances, however, in which differing degrees of anchoring and adjustment might be more or less optimal, depending upon the circumstances under which the updating is taking place. For example, when the human operator takes successive readings from a single source, it is assumed that the second reading is more current (and therefore more representative of the current state of the system) than the first. In this instance, conservative updating strategies would lead to less optimal performance. On the other hand, if the second piece of data is viewed after the first simply because the human operator got to it second (e.g. the monitor is checking a series of printouts in the order they were stacked on the desk, not in the order they were actually sampled and printed), then an anchoring and adjustment strategy is relatively more optimal than a recency strategy.

Limitations of the Static Approach

Given the extent to which heuristics have been found to enter into the judgment and decision-making process, one might conclude that the human operator within the dynamic system is merely a "bundle of biases" on the verge of disaster (Wickens & Flach, in press). The high safety record of both aviation and the process control industry, however, suggests that such a conclusion might be in error. What then accounts for the human's ability to perform reasonably well under difficult circumstances? The answer may lie in inherent differences between real world decision making, and decision making within the laboratory setting.

Originally, the focus of investigations concerning heuristics and biases had been on problems caused by humans' limitations and errors in decision-making within a static environment (single-outcome gambles, consumer product selection, etc.). More recently, however, several researchers have questioned the ability of such static models and experiments to capture the essence of complex decision-making behavior (Einhorn & Hogarth, 1981; Klein, 1983; Payne, Bettman & Johnson, 1986).

As evidenced by performance records, humans are often quite adaptive in judgmental processes used to cope with many complex environments (Payne, 1982; Einhorn & Hogarth, 1981). In fact, evidence shows that several biases identified in discrete situations may result from heuristics that are functional in a more continuous environment (Einhorn & Hogarth, 1981; Klein, 1983; Payne, Bettman, & Johnson, 1986). Theories of judgment and choice that do not consider the continuous perspective exclude the adaptive learning element, an important determinent of human behavior.

The present study, therefore, will investigate a number of these judgmental heuristics and biases within a dynamic environment, focusing on the extent to which these biases are manifested at different levels of processing load. Furthermore, it will attempt to identify the extent to which such behavior may be efficient and adaptive, given the information load and time constraints of the system.

Focus of the Present Study

Two experiments were performed in order to investigate cognitive biases within the supervisory control context. Subjects were asked to make repeated diagnoses of the state of a dynamic system under conditions of varying information load (via time stress and addition of a secondary task). The paradigm employed was a simulated military flight mission, in which subjects were instructed to imagine that they were pilots monitoring a number of cues, and deciding, based upon overall status information, whether the current and predicted state was such that it was most optimal to continue or abort the mission. In Experiment 1, under conditions of slight or moderate time stress, subjects integrated information from five or eight diagnostic cues of varying information

worth in order to arrive at a diagnosis of the present state of the mission. In Experiment 2, under a condition of moderate time stress, subjects integrated five or eight information sources, again with varying information worth, performing with and without an additional secondary running memory task.

In both experiments, the system under control was dynamic in the sense that subjects were asked to make repeated updates of cues that were correlated over time, or over the course of the same "mission." As a result of the cognitive demands of such a task, it was hypothesized that a number of the simplifying heuristics or biases described above would be manifested in subjects' responses, and furthermore that such manifestations would increase as a function of increasing information load. By comparing the subjects' judgments with the optimal decision model, systematic deviations from optimality should be observed. Various non-optimal information-extraction strategies were modeled. The extent to which the subject's deviations are correlated with non-optimal heuristic functions established for each type of filtering strategy should determine the decision-making and integration strategies used. These experiments also attempted to identify which, if any, of these heuristic strategies produce adaptive responses in the judgment task, given the system constraints and demands of high information load.

The first study was essentially a pilot to the second experiment, and thus discussion of the results from Experiment 1 will be brief. However, given the similarities between the two experimental paradigms of both studies, and the analysis techniques employed, a thorough description of this paradigm will be presented within the method section of Experiment 1, while only modifications of the paradigm for the second study will be described within the method of Experiment 2.

Experiment 1

Method: Experiment 1

The Flight Paradigm. Subjects were told to envision themselves as military aircraft pilots, midway through a flight mission. They were instructed to monitor a number of cues from an intelligent on-board computer providing diagnostic information on the state of the aircraft, as well as information about weather conditions, enemy strength, navigational equipment status and pilot fatigue. Based on their integration of information from periodic cue updates, subjects judged repeatedly whether to carry out the mission to its final destination, or to abort it and return to the point of origin. In this sense, they were not merely diagnosing the current state of a system as in many previous examinations of internal models, but were predicting its future state given current information.

Information load was manipulated between subjects by presenting either five or eight information sources that were periodically updated during the course of the mission. Each cue was dichotomous in that it provided one of two possible types of information: support for completing (+) or aborting (-) the mission. These discrete representations were meant to mimic the output of an intelligent on-board computer that would condense a cue's continuous analog value into a discrete value, based upon current mission status (airspeed, altitude, etc), as well as predictions into the future (distance to final destination, weather conditions ahead, etc.).

To create the illusion of a slowly-changing integrated system, each mission (block of 20 trials) consisted of cues that were autocorrelated. These autocorrelations were based on temporal update functions specifically generated for each cue. In order to evaluate a subject's ability to integrate of information over time, it was crucial to generate a series of realistic cue values within the dynamic system. Therefore the cue update functions were created so as to represent a logical progression of that cue value over a period of time. For example, during the course of a mission, the weather cue would not be expected to randomly fluctuate between + and -, but rather, would stay positive (indicating no immediate weather threats) for a number of trials, then take on a negative value (indicating a potential weather hazard) for several trials as the aircraft flies through a storm front, but return to positive once the craft has passed through the source of danger. It is important to note that since the intelligent computer is predicting the future state of each cue's discrete value given current analog information, it is not unrealistic that a steadily-declining analog cue may be converted to a discrete cue capable of switching from a negative to a positive value. For instance, a negative cue representing fuel availability might switch to a positive value if a reduction in airspeed or change in headwinds resulted in a reduced rate of fuel consumption.

Each cue varied as to the total amount of information, or information worth, it carried. Information worth was defined as the weighted combination of a cue's reliability and diagnosticity. These dimensions were defined as follows:

Reliability: Refers to the accuracy with which a cue conveys the true state of the world. For example, in a highly-reliable cue, the displayed value almost always indicates the true state of the world (bearing in mind that no indicator is of perfect reliability). An unreliable cue, on the other hand, may reflect an inaccurate or incorrect value.

Diagnosticity: Refers to how informative each cue is in evaluating potential success or failure of the mission. For example, fuel is much more diagnostic of potential mission failure than is pilot fatigue. That is, a fatiqued pilot will be more likely to complete a mission successfully than will be an aircraft with an empty fuel tank.

Specific reliability and diagnosticity values were not presented, but rather subjects were given the combined information worth of each cue they would be monitoring. These information worths were given arbitrary numerical values, with a maximum weight of eight, a minimum of 1.

The eight information sources utilized in the flight scenario, along with the information worth of each cue, are listed in Table 1. Cues were grouped functionally into one of three categories: external factors, system factors, and a pilot factor. The display of information within each of these three categories was portrayed by a differently-shaped symbol. The table also contains the maximum range of value reversals ("+ to -" or "- to +" that occurred during a 20-trial block. This range then estimates the "bandwidth" or time constant of the cue values.

Subjects were instructed to weight each cue's value according to its information worth and sum the products across cues to arrive at a judgment about the potential success or failure of the mission. Their weighted judgments were recorded on a continuous scale ranging from -10 (maximum amount of information supports aborting the mission) to +10 (maximum amount of information supports continuing the mission. The alogrithm for combining the cues optimally was calculated as follows:

$$\text{Opt. Resp.} = \left[\sum (+XC) - \sum (-XC)\right] \times k$$

where $\sum XC$ = weighted cue value and
k = constant, converts function to +10/-10 scale

Cue Formats. Three cue formats were varied within subjects for each of two cue number conditions. In the center-informative condition, the cues of high information worth were centrally located in the display. In

Table 1

Cues used in the Experimental Paradigm, along with Information Worth, Functional Group, and Range of each

CUE	INFO WORTH	FUNCTIONAL GROUP	RANGE To ⟶ From
*Fuel	8	System	+ - + -
*Engine Temperature	7	System	+ -
Hydraulic Pressure	6	System	+ -
*Enemy Intent	5	External	- + - +
Weather	4	External	+ - + -
*Navigational Aids	3	External	+ - +
Headwinds	2	External	+ - + -
*Pilot Fatigue	1	Pilot Fatigue	+ - + -

* Cues used in the five-cue condition.

the left-informative and right-informative conditions, the cues of high information worth were placed at the left and right of the display, respectively. It is hypothesized that if subjects used a strategy of filtering by salience, the optimality of their judgments should vary as a function of the location (and inferred salience) of these highly-reliable cues. For example, if a subject was placing more emphasis on cues at the left of the display (and therefore partially filtering the righthand cues), task performance should be closer to optimal for the information-left condition than for the information-center or information-right conditions. Figure 1 illustrates the three types of display formats for both five-cue and eight-cue conditions.

Information Load. In addition to manipulating the number of cues presented between groups of subjects, information load was varied within subjects by imposing two different levels of time stress. Under conditions of slight time stress, subjects had eight seconds in which to view and integrate information from the five or eight cues. In the moderate time stress condition, however, subjects had only four seconds of viewing time. The time stress variable was manipulated within subjects, therefore within each group, every subject performed four missions in the moderate and four missions in the slight time stress conditions for each of the three display formats.

Procedure. Subjects were seated 70 cm in front of a cathode ray tube (CRT) screen. Information cues and the confidence judgment scale were displayed on the screen. The subject's right hand rested on a joystick mounted with a response trigger. This control was used to move the cursor on the judgment scale; the trigger to record the final judgment. The experimental session was controlled and data collected by a PEARL II: Portable Laboratory Computer System interfaced with an Imlac graphics processor (Heffley, Foote, Mui & Donchin, 1985).

A typical trial sequence consisted of simultaneous presentation of the five or eight cues for either four or eight seconds (depending on the information load condition). An auditory warning tone sounded after a brief period (three seconds for the moderate time stress condition, and seven for the slight time stress condition), warning subjects that only 1 second of information viewing time remained. Following this presentation, the cues disappeared from the screen and the subject then had 10 seconds in which to make a response on the judgment scale. At the beginning of every mission, the response arrow on the judgment scale initiated from the zero or neutral position. On all following trials, the arrow initiated from the point of the previous response. A second marker remained at this previous response point until the updated response was made. This allowed subjects to make adjustments as cues were udpated over the course of the mission.

Each subject participated in a two-hour practice session, followed by three experimental sessions. The practice session was designed to familiarize the subject with the flight scenario and integration instructions, and then to provide practice blocks using each of the three

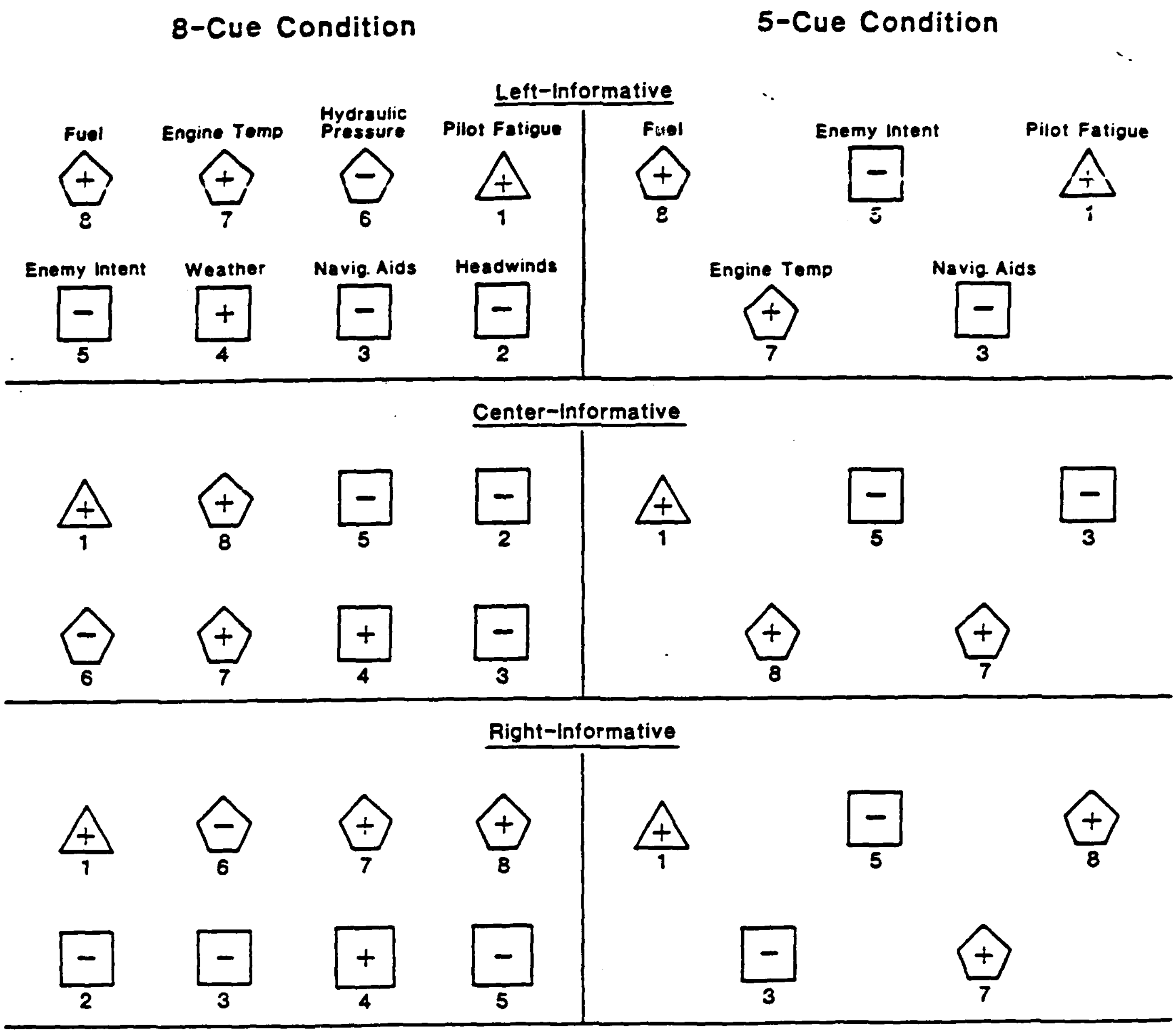

Figure 1. Three display formats for 8-cue and 5-cue conditions. Information worths were displayed only during the practice session.

cue formats, under both conditions of time stress. During practice blocks only, information worth values were displayed beneath each cue and subjects received feedback as to the optimal response following every integration response. These information displays enabled subjects to become familiar with each cue's worth, and to internalize how a cue's value impacted the optimal judgment. By the end of the practice session, each subject's response closely tracked the optimal response.

Each of the three experimental sessions consisted of eight missions (four blocks each of the 4-second and 8-second conditions), using a different cue format for every session. Cue worths were not presented during these sessions, although the verbal cue labels were always presented, and the location of each remained constant during a session. Subjects were randomly assigned to one of two cue number conditions, so that each subject consistently made decisions based upon either five or eight cues. The display format for particular cues varied from experimental session to session. The order of format presentation was counter-balanced across subjects in each condition.

The blocks within each condition differed in that each contained a unique random combination of potential functions for each individual cue. Therefore, no subject saw the exact same combination of cue values within an experimental session. Trials were individually generated off-line, adhering to the temporal functions for each cue. Specific manipulations of cue values provided instances to test for the utilization of simplifying heuristics in the diagnostician's judgments.

Subjects. Three male and three female students from the University of Illinois served as paid volunteers in this study. By random assignment, two males and one female served in the eight cue condition, and two females and one male served in the five cue condition. Two types of performance incentives were offered to subjects. The first bonus was analyzed on a trial-by-trial basis. Subjects received a 1-cent bonus for every "correct" response (decision to optimally abort or continue the mission) and were penalized one-half cent for every incorrect decision. This payoff schedule therefore rewarded subjects equally for both "hits" and "correct rejections," and penalized equally for "misses" and "false alarms." The equal emphasis on both courses of action therefore should have eliminated any unnecessary criterion shifts (e.g. a subject who was biased towards carrying out the mission, despite overwhelming evidence to abort).

The second performance incentive was a $5 bonus paid to the subject in each cue number condition, whose overall responses correlated most closely with optimal responses. Subjects therefore were rewarded not only for selection of the appropriate action, but also for the accuracy with which their weighted responses tracked optimal responses.

Analysis. For each block of trials, different non-optimal information-extraction strategies were modeled (including the "as-if" strategy and strategies relying on cue salience rather than cue information worth). Product moment correlations were then computed between each of the

modeled response functions and the subjects' actual responses. These correlations served as the basis for further analyses. A high positive correlation between any one function and a subject's response implies that the subject employed the strategy in question. It was assumed that the function with the highest correlation with the subjects' responses is the best description of the subjects' integration behavior.

Seven non-optimal integration strategies were modeled. These included two "as-if" functions (one in which all cues were weighted equally, and one in which the proportion of trials where subjects' responses agreed with an as-if strategy rather than an optimal strategy was calculated), and five functions where more emphasis was placed on cues in specific locations on the screen. These models, simulating filtering by salience strategies, included center, top, top-left, left and right filtering strategies. Actual weights used to model each strategy are presented in the appendix.

Results: Experiment 1

Since Experiment 1 was a pilot study with few subjects (n=3 per cue condition), and therefore low statistical power, formal statistics were not performed. Rather, the purpose of this study was to establish that subjects could perform, with some level of proficiency, the task at hand; to assess in general terms the impact of stress manipulations; and to establish the viability of the data analysis procedures. An examination of these correlations showed that overall, the optimal response function gives a better fit to the subjects' actual responses than do any of the nonoptimal models. Furthermore, speed of presentation influences to some extent the degree of optimality. The responses in the slight time stress condition had a higher correlation with the optimal response (.977) than did responses in the moderate stress condition (.968). A second influence of speed is a higher correlation within the non-optimal top and top-left response functions that results from the faster presentation rate. Therefore, it appears that under some time stress, subjects respond less optimally by placing more emphasis than is optimal on cues in the upper-lefthand portions of the display screen.

Alternative non-optimal models of processing were found to "fit" optimally with the varying display formats. For example, when the highly informative cues were placed on the right, the right response function correlated more closely with the subject's response than did other models of processing. This finding again reflects the overall optimal level of performace of subjects in this integration task. Within each location-specific model, the correlation dropped slightly as the number of cues increased. This finding further supports the notion of a decline in performance as information load increases. In general, however, across most conditions a high degree of optimality was obtained.

Discussion: Experiment 1

The results from this initial experiment revealed that subjects were able to perform the task at hand quite optimally, given the level of information load that was employed. Subjects' responses were modeled well by the optimal response function. Some evidence provided support for the hypothesis that increasing information load leads to a decline in optimality. This decline appeared to be in the form of a filtering by salience (top-left display position), an observatrion that reinforces the results obtained by Wallsten and Barton (1982).

The data suggest, therefore, that while some decline in performance may be attributed to increased information load, the cognitive demands of the current task did not impose an information load great enough to create severe departures from optimality. Based upon the results of the pilot work, Experiment 2 was designed with the purpose of increasing the cognitve demands of the human operator without changing the general paradigm employed in Experiment 1. The relevant changes made to the paradigm will be discussed in the method section for the following experiment.

Experiment 2

Method: Experiment 2

Subjects. Nine male and nine female university students served as paid volunteers in Experiment 2. Four males and five females served in the eight cue condition, while four females and five males served in the five cue condition. The paradigm employed and the procedure used in this study was identical to that of Experiment 1, with the following exceptions:

Information Load. Rather than using two different levels of time stress, all subjects performed the task under moderate time stress (4-second presentation), and information load was manipulated by the addition of a secondary running memory task during half of the simulated flights. In the secondary task condition, subjects listened to word lists (composed of 1 and 2-syllable common words) presented over headphones at the rate of one word every three seconds. The subject's task was to monitor the word lists, and to verbally "shadow" or repeat the list, one word back in the sequence. For example, if the subject heard "cat" as the first word, no response was called for. If "boy" was the second word, the subject then responded with "cat," the previous word in the sequence. When presented with "eat" as the third word, the subject responded with "boy," again the previous word in the sequence. The running memory task was performed continuously throughout the mission in the secondary task conditions.

Addition of the secondary task increased demands on the working memory and information-processing capacities of the monitor, while also representing a task similar in nature to one that might be required of a pilot in an actual flight task (for example, monitoring radio messages and responding verbally).

Context-Free Judgment Task. After completing the three experimental sessions using the standard task, subjects were asked to perform the same general weighted judgment task without the contextual flight cues. The purpose of this additional condition was to assess an individual's ability to perform the necessary mental calculations without the benefit of contextual cues. If subjects are able to perform the mental calculations, yet demonstrate biases within the flight judgment task, then this source of bias may be attributed to simplifying heuristics employed, rather than to difficulties in performing the required calculations. Conversely, no difference in performance between the two paradigms may imply that subjects performed both tasks identically.

In the context-free version, subjects were presented with five or eight signed numerical values (+7, -4, etc.), and asked once more to sum the products across cues to arrive at a weighted judgment, recorded on the same -10/+10 scale. No cue labels or shapes (as appeared with the cues in the flight scenario) were presented in this context-free version. Subjects again performed the context-free task with and without the secondary running memory task.

Immediately following the eighth mission of the third experimental session, subjects performed an additional eight blocks of the context-free integration task. The general format of the context-free blocks was always the same as the cue format presented to a subject earlier that session for the flight scenario blocks. For instance, if the fuel cue (with an information worth of 8) was always presented in the upper-lefthand portion of the screen during the final session of the contextual task, then the subject would always see a +8 or -8 in the upper-left during the context-free blocks. Figure 2 illustrates how the information display changed from the flight scenario task to the context-free task for the center-informative, 8-cue format.

Memory for Information Worth. Following the completion of each mission, subjects were quizzed on the information worths of each of the information cues. If a subject responded incorrectly to any value, a brief review of the cues and their associated information worths was provided. This control ensured that any potential response bias in judgment was not due to a memory failure of each cue's worth, but rather to some inappropriate integration strategy.

Results: Experiment 2

The discussion below will focus initially on the product moment correlations of subjects' responses with the optimal function, and then with the non-optimal ppmodeled functions.

Correlations with Optimal Function. Separate correlations for each block of trials were computed, and these data showed that overall, the optimal response function gives a better fit to the subjects' actual responses than do any of the non-optimal models. Figure 3 plots mean correlations of subjects' actual responses with the optimal function. The figure is partitioned by cue number, with the eight-cue condition plotted on top; the five-cue condition on the bottom. Both secondary-task and primary task conditions are plotted for each of the two cue number conditions. Correlations are plotted as a function of varying display formats for each of these individual conditions. As can be observed in Figure 3, all such correlations are extremely high (all mean correlations are > .890). Thus, subjects generally performed the task in an optimal manner.

In order to evaluate possible task, cue and format effects, the correlation values were normalized utilizing a Fischer's Z transformation. These values were then analyzed by a 2 x 9 x 2 x 3 x 4 (cue number x subject x task x format x repetition) analysis of variance (ANOVA). Results of the analysis for the optimal function revealed that subjects' performance of the primary integration task was significantly worse when concurrently performing the secondary task, than in conditions of the primary task alone

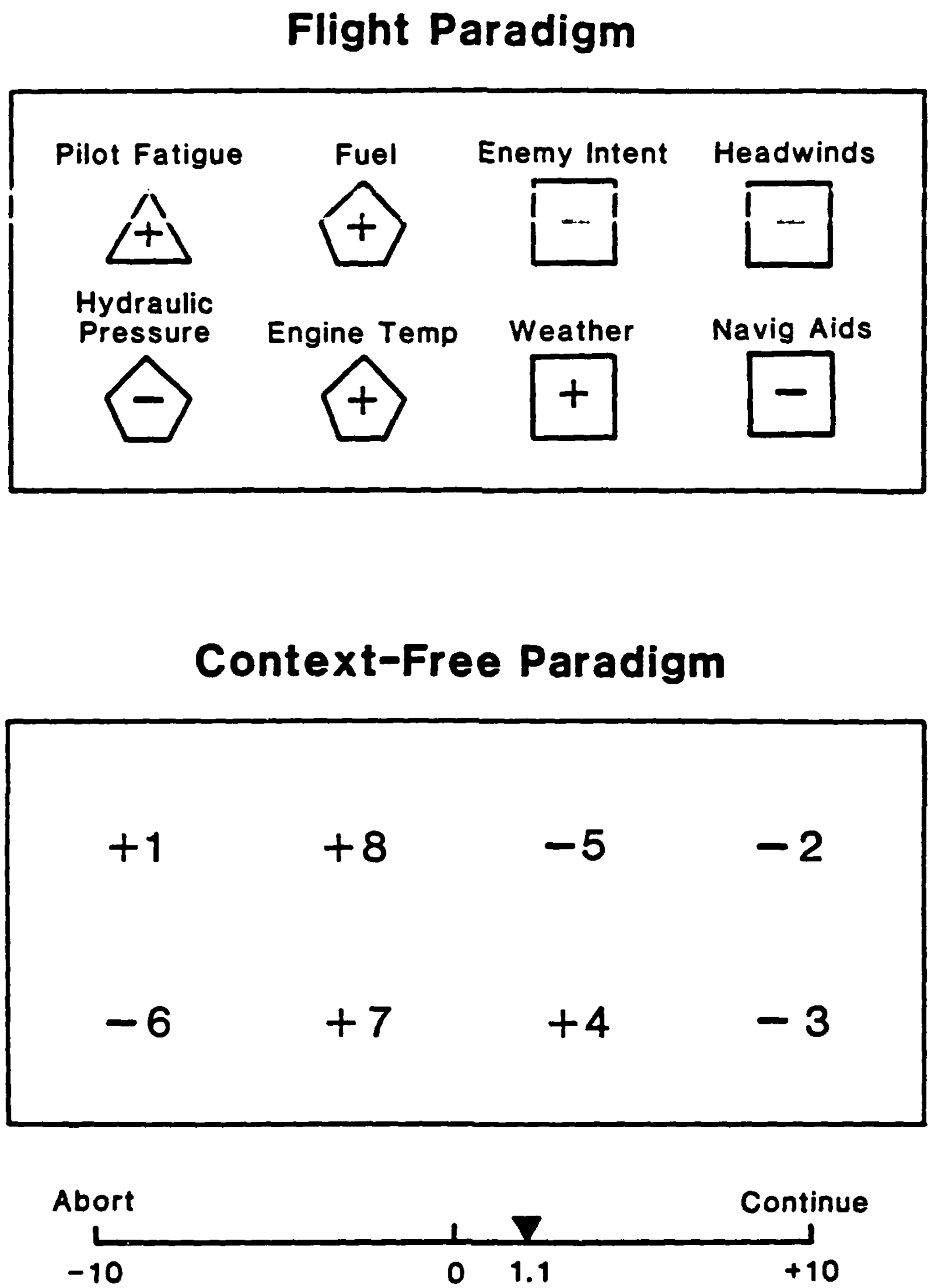

Figure 2. Illustration of contextual and context-free display formats for the 8-cue condition.

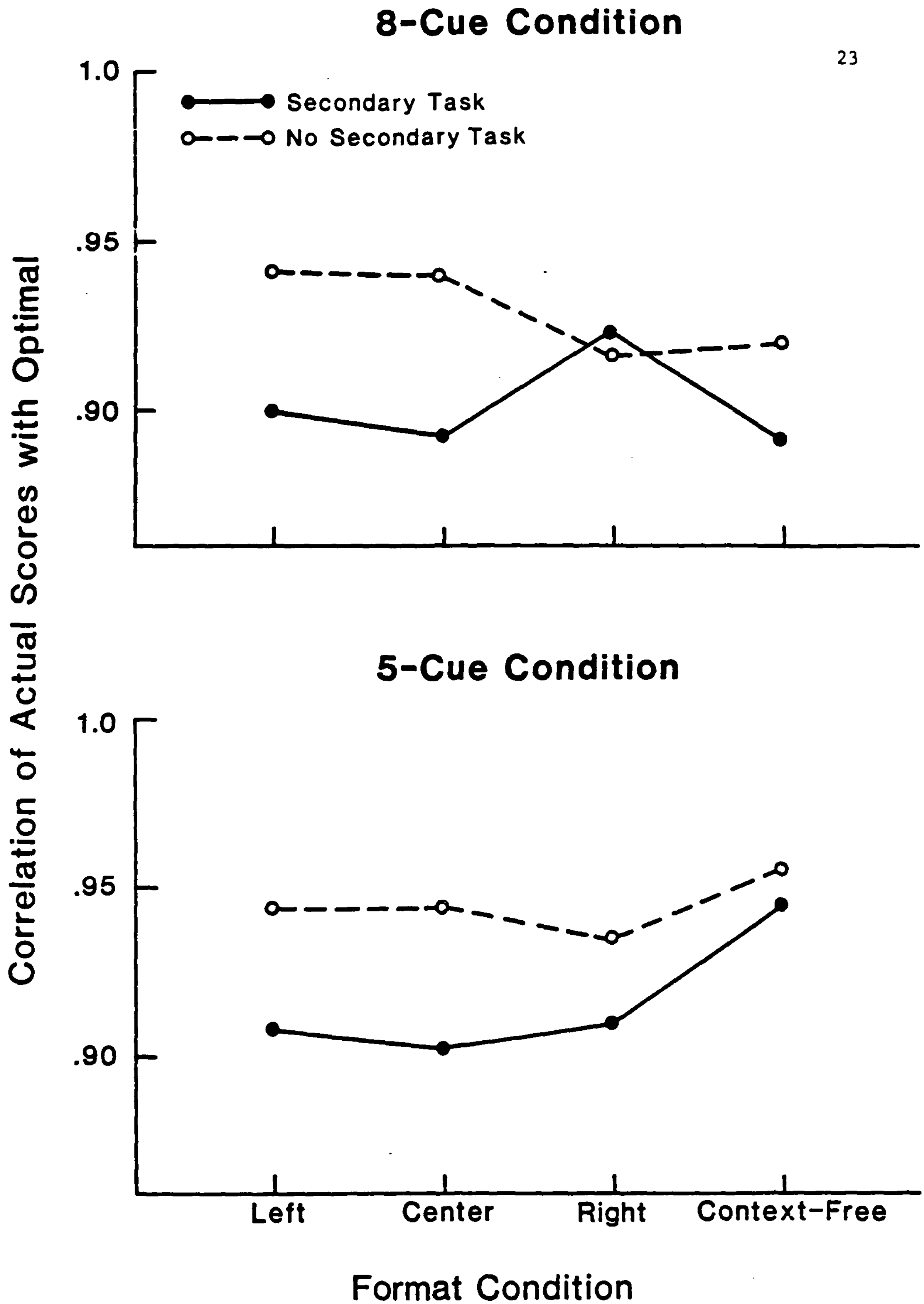

Figure 3. Mean correlations of actual with optimal response as a function of display format for 8-cue and 5-cue conditions, with and without the secondary task.

($F(1,16) = 25.43$, $p=.0001$). This finding is consistent with the hypothesis that, under conditions of greater information load, the optimality of a subject's performance will decline.

There was no main effect, however, of cue number ($F(1,16)=.79$, $p=.389$), nor of display format ($F(2,32)=.37$, $p=.692$). Furthermore, none of the interactions between any of the three primary variables (cue number, task, format) were significant. These findings suggest that additional information sources in the eight cue condition did not increase information load enough to cause a decline in optimality. Optimal performance is further emphasized by the non-significant format effects. Subjects were able to weigh each cue's value according to information worth, regardless of the cue's location in the display.

Furthermore, no significant differences were found between the contextual and context-free paradigms $t(35)=.42$, $p=.676$. This difference was consistently non-significant across cue number and secondary task conditions. This result further substantiates the claim that subjects were able to perform the weighted integration task optimally, with or without contextual cues. Finally, memory for information worths was very accurate (error rates were less than 1% for both groups in all conditions).

Correlation with Non-Optimal Functions. Just as significant effects of the optimal correlation suggest the degree to which the independent variables induce departures from or conformity with optimality, so the ANOVA applied to the correlation of subjects' responses with different non-optimal functions can be interpreted as showing how certain independent variables may have induced different non-optimal forms of behavior.

Because the non-optimal functions were fairly highly correlated (across scenarios) with the optimal functions, it was necessary first to partial out the variance attributable to the latter. Once this correlation was partialed out, the overall correlations of subjects' presponses with each non-optimal function (a "residual" of nonoptimal behavior) were very low. This finding once more emphasizes the optimal nature in which subjects performed the integration task. An ANOVA similar to that described for the optimal function was performed for each of the modeled non-optimal functions. Results of these analyses failed to produce significant influences of cue number, secondary task condition or format for any of the functions. This lack of significance, however, is not surprising, given the high degree of optimality across subjects and conditions.

Further analyses of the data from Experiment 2 consisted of a multiple regression analysis to quantify the weighting strategies of subjects, an investigation of the extent to which subjects "anchored" responses from one trial to the next, and finally, an analysis of secondary task performance across varying conditions. Each of these will be discussed in turn below.

Multiple Regression Analysis. The multiple regression analysis (regressing subject response on the five or eight cue value predictors) quantified the weighting strategies of the subjects, to determine the extent

to which each information source was actually weighted optimally. This analysis represents a more detailed description of subject performance than does the more global correlation analysis. It was hypothesized that systematic biases described previously ("as-if" strategy, filtering by salience, etc), would be manifested by differential weighting of the cues under varying conditions of information load and cue location.

Beta weights were calculated for each subject, and averaged across subjects in like conditions. The resulting average beta weights are portrayed in Figures 4 and 5, which depict the weights as a function of cue format and secondary task condition respectively, for both 5-cue and 8-cue conditions. The optimal weighting function is also represented by the straight solid line. The one salient feature that emerges from the two figures is the extent to which subjects' actual weightings closely resemble those of the optimal function. This linear pattern within the subjects' weighting schemes is to be expected, given the significantly high correlation with the optimal response function discussed previously.

One interesting trend that emerges from Figures 4 and 5 is the tendency for subjects to be biased (although not significantly) towards a "central tendency effect" (Lee, 1971). Under all conditions of format and secondary task, subjects tend to underweight the highly-informative cues, while over-weighting the less-informative information sources. This tendency which varied little among differing display formats, appeared to be slightly exaggerated with secondary task loading (Figure 5). An analysis of the standard errors between conditions revealed that this effect was not statistically reliable. However the consistency of this trend across both graphs, both task conditions and all of the high weight cues suggests that it is real, and in fact only low statistical power prevented significance from being obtained. Therefore, although subjects were generally optimal in performing the integration task, there is still some evidence to support the claim that humans are generally conservative in judgment.

<u>"Anchoring-and-Adjustment" Analysis.</u> This analysis determind the extent to which subjects tended to "anchor-and-adjust" their sequential responses. As discussed previously, such a tendency is manifested when subjects fail to adjust current judgment as much as is optimal. Such an analysis consists of observing the degree to which responses are adjusted relative to the changes prescribed by optimal responses. The measure used to quantify updating strategies was the ratio of actual to optimal adjustments:

$$\frac{\text{Actual}(n) - \text{Actual}(n-1)}{\text{Optimal}(n) - \text{Optimal}(n-1)}$$

Optimal adjustment performance yields a ratio of one. Conservative adjustment performance (under-adjusting) yields a ratio less than one, while risky performance (over-adjusting) yields a ratio greater than one.

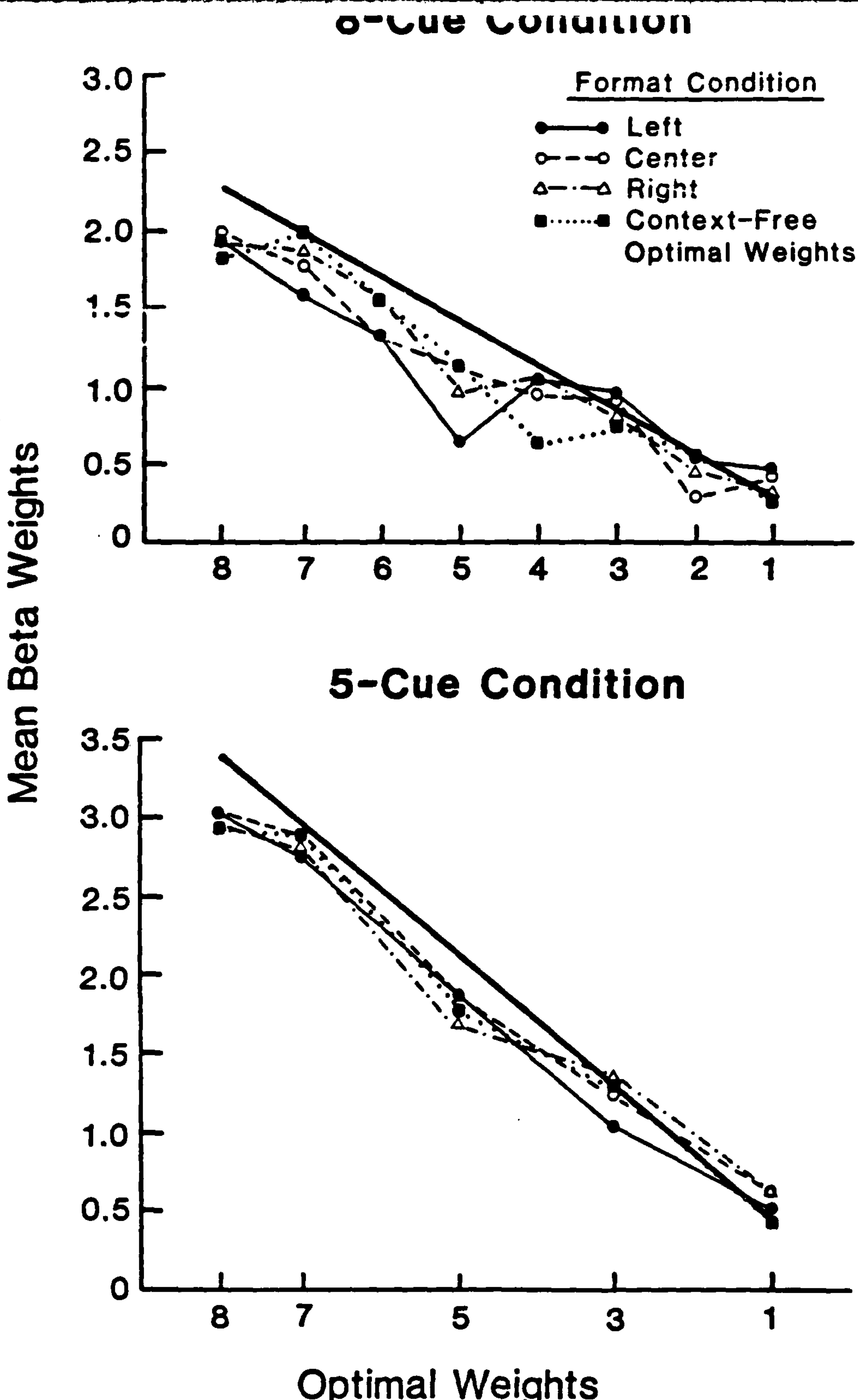

Figure 4. Average beta weights as a function of optimal weights for 8-cue and 5-cue conditions, for varying display formats.

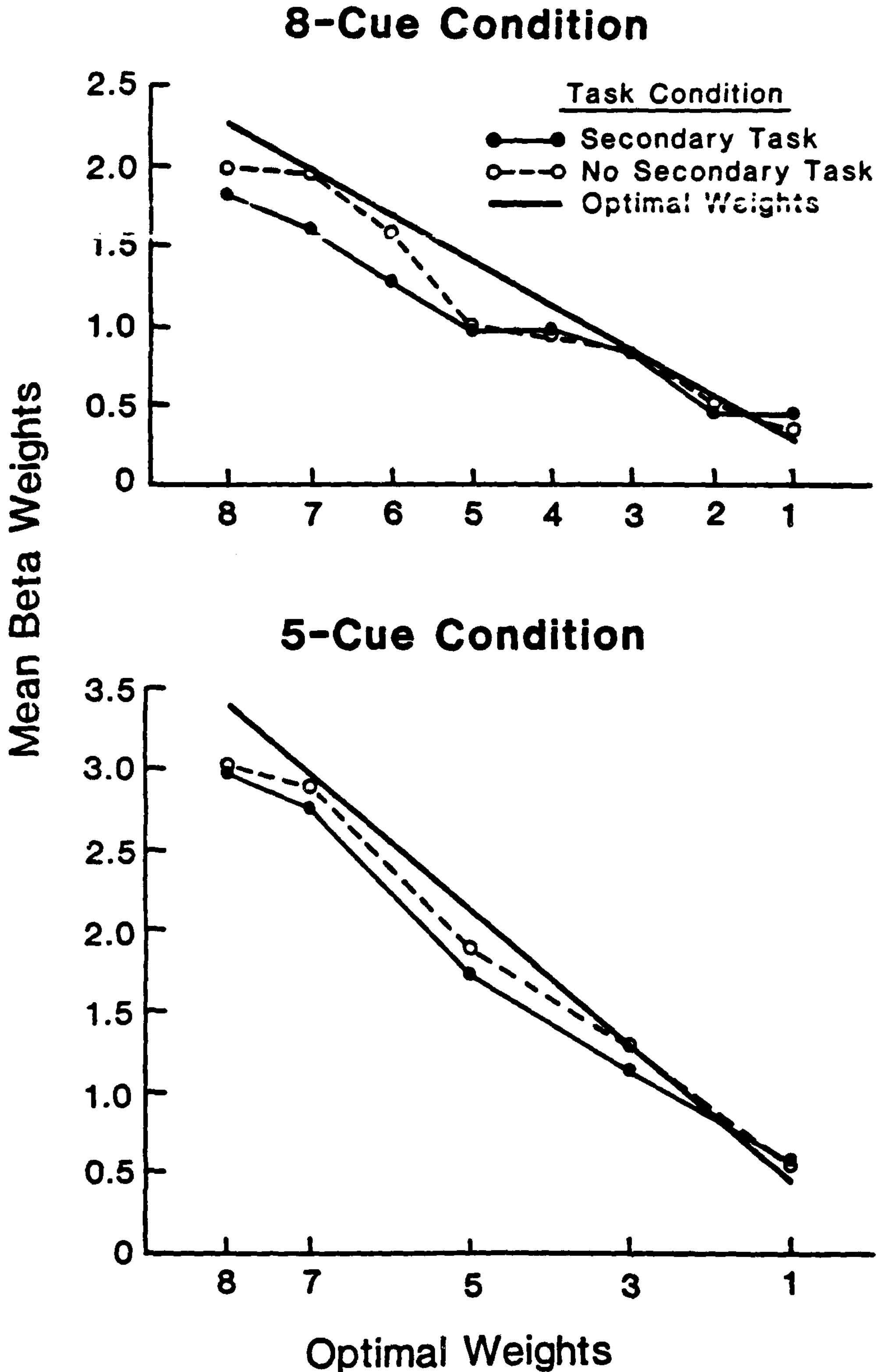

Figure 5. Average beta weights as a function of optimal weights for 8-cue and 5-cue conditions, with and without the secondary task.

The two models proposed in the introduction regarding types of updating strategies predict two differing patterns of performance. The Viscocity Model predicts that subjects would be consistently conservative in adjustment strategies, thus yielding a ratio less than one, regardless of where on the response scale the adjustment lies. The Elastic Model, on the other hand, predicts that responses will become progressively more conservative (underadjusting) at extreme values as the optimal response moves away from neutral, while becoming gradually more risky (overadjusting) at the extremes when the optimal response moves toward neutral.

Results from this analysis are summarized as follows: Overall, the ratio of actual to optimal adjustments revealed that subjects' strategies were either optimal (between .80 and 1.2) or somewhat conservative (less than .80). Only one subject deviated from this pattern, demonstrating a risky updating strategy. There were no significant differences in updating strategy as a function of cue number, format, or task conditions. Thus, limited support is given for the Vicsosity Model, suggesting that subjects are generally conservative in the achoring-and-adjustment task. Correlations of this ratio with the value of the optimal response (indicating potential relations between updating strategy and where the optimal response is on the scale) were not significant, either for adjustments towards or away from the neutral point. Therefore, it appears as if adjustment strategies did not change as a function of where the response was on the scale. Thus, the data do not support an Elastic Model as a description of updating strategies used in the present judgment task.

Analysis of Secondary Task Performance. Finally, an analysis of secondary task performance was conducted in order to observe any potential trade-offs that may have occurred between primary and secondary task performance. Some evidence of the potential for such a trade-off was discussed previously within the context of optimal performance correlations. The significant main effect of primary versus secondary task conditions suggests that primary task performance declined significantly as a function of the additional information load imposed by the secondary task.

Further analysis of secondary task performance consisted of determining secondary-task performance levels (expressed as percentage of words correctly repeated) for varying cue number and format conditions. Results of the 2 x 9 x 4 (cue number x subject x format) ANOVA revealed no significant main effect for cue number $F(1,16)=.300$, $p=.590$, or for format $F(3,48)=1.46$, $p=.236$. Furthermore, the interaction between format and cue number was not significant $F(3,48)=.60$, $p=.616$.

From this analysis, it may be concluded that the only interference caused by the secondary task was a decrease in optimality of primary task performance. When performing the two tasks simultaneously, however, no significant primary or secondary task decrements can be attributed to effects of other experimental manipulations (format and cue number).

Discussion: Experiment 2

The results of Experiment 2 agree with those of Experiment 1 in that generally, subjects were able to perform the task very well. Subjects' responses were modeled quite well by the optimal response function. (mean correlations were all > .890). In this study, however, significant decrements in performance were demonstrated when information load was increased by the addition of a secondary running memory task. Furthermore, some trends suggested that subjects were slightly more conservative in judgments than is optimal, especially under conditions of increasing information load. However, the results of both experiments failed to reveal systematic biases or deviations from optimality in subject judgments. This lack of effect contrasts with the results of some other researchers.

For example, Jones and Wickens (1986) found that subjects consistently over-adjusted responses when updating beliefs, while subjects in the present study were generally optimal or slightly conservative (displaying some "central tendency" effects) in their updating strategies. This lack of consistency between the two studies may be attributed to differences in the types of information that was presented to subjects. In Jones and Wickens' study, subjects viewed a number of cues presented either as bargraphs or as the distance from the center in a pentagon display. Each cue could take on one of fifteen possible values. In the present study, however, dichotomous cue values were presented within the labeled cue shapes. Furthermore, subjects in Jones and Wickens' study had five seconds in which to make a response, while subjects in the current study had ten seconds. Therefore, with more information of different types to integrate in the Jones and Wickens study, it may be that optimality was more difficult to obtain, and thus, given conditions of greater information load, update responses were more extreme.

Further differences lie in the differing effects of manipulations of processing load on cue sampling. The results of Experiment 1 suggested that under conditions of greater time pressure, subjects relied more on top and top-left filtering strategies, when compared to conditions of relatively little time stress. Such a trend was not observed however in Experiment 2. This difference may be attributed to the different manipulations of information load in the two experiments. In the first experiment, information load was manipulated by varying the amount of available integration time (4 seconds vs. 8 seconds). In the second experiment, however, only the 4-second condition was used, and information load was manipulated through addition of the secondary processing task (i.e., diverting cognitive resources away from integration). Thus, optimal correlations were quite high in the first study, and differences in correlations of the non-optimal top and top-left functions changed only as a result of the decrease in available processing time. In the second study, while the optimal correlation still provided the best fit to the response function, the overall correlations were lower than those of the first study. Therefore, a decline in optimality was observed generally when comparing

results of Experiment 1 with those of Experiment 2, however no further systematic effects of salience were observed within the different manipulations of information load in Experiment 2.

Other research on salience biases by Wallsten and Barton (1982) suggests that time pressure must be severe in order to find systematic salience biases. Subjects in their initial study revealed no significant salience biases, under conditions of relatively little information load and time stress. In their second study, however, salience biases were found only under conditions of severe time pressure, when payoffs were moderate and thus when performance criteria when not as high. This finding, together with the findings of the present research, seem to suggest that salience biases occur only under extreme conditions of time stress or information load, a condition that was not met sufficiently in the present study (as judged by the overall level of optimal performance maintained by subjects).

One other finding may be explained in terms of this lack of systematic bias. No significant differences were found between the contextual and context-free versions of the task. This lack of difference imples that subjects performed both tasks rather optimally, and again emphasizes the notion that within the contextual version, subjects' sampling and integration performance was not influenced by the psychological salience of varying cues.

Since results of the present study revealed no systematic trends or biases, an alternative explanation must be raised in order to account for the effects of the secondary task condition on overall optimality. Returning to the information processing model of decision making presented in the introduction (Wickens & Flach, in press), the system diagnosis phase may be divided into a number of stages. Initially, the human operator samples cues within the environment. These cues are then weighted according to some strategy (information worth, in this instance) and mentally integrated. Finally, a response is made based upon the integrated diagnosis. Both perceptual biases and cognitive biases may be present in this process. Perceptual biases are those pertaining to the cue sampling phase (such as display salience effects), while cognitive biases influence judgment at the weighting and integration phase (e.g. the "as-if" weighting scheme and conservative updating strategies).

Results of the present study suggest that nonoptimality produced by concurrent task load was not manifested in the cue sampling phase, as no differential effects were observed for differing cue formats. It may be assumed that the integration task was sufficiently practiced, this aspect of performance was sufficiently automated, and thus was not harmed by diversion of resources to the secondary task (even as it might have been distorted by further time limitations). Therefore, while a significant decrement to performance was observed within the secondary task condition, such a decrement may not be attributed to sampling or salience biases. How then did the secondary task effect performance? An alternative explaination may be that the diversion of cognitive resources to the secondary task interfered with the weighting and integration phase of the diagnosis task.

Limited support for a systematic bias in the weighting phase is the trend (although not significantly so) for subjects to under-weight highly informative cues, and over-weight the less-informative cues. As may be observed in Figure 4, this trend was even more pronounced in the secondary task conditions. However, since this trend was not statistically significant, further influences must account for the decrements under secondary task conditions. Presumably, this secondary task condition did not therefore systematically bias performance as much as it simply added random noise to the weighting and integration phase of the diagnosis task.

These results are potentially of considerable importance, because a review of the literature fails to reveal studies in which decision making performance is carefully analyzed under conditions of dual task loading. It is an absence that is somewhat surprising, given the extent to which real world decisions are made in the context of concurrent cognitive activities.

Conclusion

This study examined the extent to which systematic biases may influence the quality of human operators' diagnoses of the state of a dynamic system. Generally, subjects performed the integration task optimally. Deviations from optimality were caused by the increased information load of an additional secondary running memory task. The only systematic trend that emerged from the data was a tendency towards conservatism in judgment, highlighted under conditions of increased load. This suggests that non-optimality was manifested in cognitive, rather than perceptual biases.

That individuals were significantly influenced by conditions of increased information processing load, even in a relatively simple dynamic system, is noteworthy. It can be speculated that in more complex systems, the effects of load my be magnified. Particularly, perceptual biases may result from further increase in time pressure (Wallsten & Barton, 1982), manipulations of the payoff schedule (Wallsten & Barton, 1982), presentation of continuous, rather than dichotomous, variables, or the use of less correlated, more random cues (Jones & Wickens, 1986). Furthermore, based upon results of previous studies and the present research, it would be hypothesized that manipulations of cue salience (both perceptual and psychological) may reveal differing performance effects. By investigating the extent to which systematic errors (both perceptual and cognitive) emerge within more complex dynamic systems, decision aids and automation may be developed to compensate for shortcomings in the human cognition process.

Acknowledgments

We would like to thank Lorilyn Aquino for assistance in data collection, and Mark Klein and Betty Casey for programming assistance.

References

Bainbridge, L. (1974). Analysis of verbal protocols from a process control task. In E. Edwards & F.P. Lees (Eds), The human operator in process control. London: Taylor & Francis.

Ben Zur, H. & Breznitz, S.J. (1981). The effects of time pressure on risky choice behavior. Acta Psychologica, 47, 89-104.

Bettman, J.R. (1979). An information processing theory of consumer choice. Reading, Mass. Addison-Wesley.

Brehmer, B. (1981). Models of diagnostic judgment. In J. Rasmussen & W. Rouse (Eds.), Human detection and diagnosis of system failures. New York: Plenum Press.

Einhorn, H.J. & Hogarth, R.M. (1981). Behavioral decision theory. Annual Review of Psychology, 32, 53-88.

Einhorn, H.J. & Hogarth, R.M. (1985). A contrast/surprise model for updating beliefs (Technical Report No. 13). Chicago: Center for Decision Research, University of Chicago School of Business.

Heffley, W., Foote, B., Mui, T., & Donchin, E. (1985). PEARL II: Portable laboratory computer system for psychophysiological assessment using event related brain potentials. Neurobehavioral Toxicology and Teratology, 7, 399-407.

Hockey, G.R. (1970). Signal probability and spatial location as possible bases for increased selectivity in noise. Quarterly Journal of Experimental Psychology, 22, 37-42.

Johnson, E.M., Cavanagh, R.C., Spooner, R.L., & Samet, M.G. (1973). Utilization of reliability measurements in Bayesian inference: Models and hman performance. IEEE Transactions on Reliability, 22, 176-183.

Johnson, E.J. & Payne, J.W. (1985). Effort and accuracy in choice. Management science, 31 395-414.

Jones, P.M. & Wickens, C.D. (1986). The display of multivariate information: The effects of auto- and cross-correlation, reliability, and heterogeneity. Technical Report CPL86-5) Champaign, Illinois: University of Illinois, Cognitive Psychophysiology Laboratory. ARI Research Note 87-69, AD A191-070

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

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 251-273.

Klein, N.M. (1983). Utility and decision strategies: A second look at the rational decision maker. Organizational Behavior and Human Performance, 31, 1-25.

Knapp, B.G. & Tolbert, C.A. (1985). Insights on information absorption and transmission rates in C^2I settings. (ARI Research Note 85-93). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. AD A167-331

Lee, W. (1971). Decision theory and human behavior. New York: Wiley.

Lopes, L.L. (1981). Averaging rules and adjustment processes: The role of averaging in inference. (Technical Report WHIPP 13). Madison, WI: University of Wisconsin, Department of Psychology, Wisconsin Human Information Processing Program.

Moray, N. (in press). Monitoring behavior and supervisory control. In K. Boff (Ed.) Handbook of Perception.

Payne, J.W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.

Payne, J.W. (1980). Information processing theory: Some concepts and methods apploed to decision research. In T.S. Wallsten (Ed.), Cognitive processes in choice and decision behavior. Hillsdale, NJ: Erlbaum Associates.

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

Payne, J.W., Bettman, J.R., & Johnson, E.J. (1986). Adaptive strategy selection in decision making (Technical Report 86-1). Durham, NC: Duke University School of Business Administration.

Pitz, G. & Sachs, N. (1984). Judgment and decision: Theory and application. Annual Review of Psychology, 35, 139-163.

Rasmussen, J. (1981). Models of mental strategies in process plant diagnosis. In J. Rasmussen & W. Rouse (Eds.), Human detection and diagnosis of system failures. New York: Plenum Press.

Russo, J.E. (1977). The value of unit price information. Journal of Market Research, 14, 193-201.

Schum, D. (1975). The weighing of testimony of judicial proceedings

from sources having reduced credibility. Human Factors, 17, 172-203.

Sheridan, T. (1981). Understanding human error and aiding human diagnostic behavior in nuclear power plants. In J. Rasmussen & W. Rouse (Eds.), Human detection and diagnosis of system failures. New York: Plenum Press.

Simon, H.A. (1981). The sciences of the artificial (2nd Ed.). Cambridge, Mass: MIT Press.

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

Snapper, K.J. & Fryback, D.G. (1971). Inferences based upon unreliable reports. Journal of Experimental Psychology, 3, 401-404.

Svenson, O. (1979). Process descriptions of decision making. Organiztional Behavior and Human Performance, 23, 86-112.

Trope, Y. (1982). Inferences of personal characteristics on the basis of information retrieved from one's memory. In D. Kahneman, P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.

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

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-458.

Van Cott, H.P. & Kinkade, R.G. (Eds.). (1972). Human engineering guide to equipment design. Washington, D.C.: McGraw-Hill Company.

Wallsten, T.S. (1980). Cognitive processes in choice and decision behavior. Hillsdale, NJ: Erlbaum Associates.

Wallsten, T.S. & Barton, C. (1982). Processing probabilistic multi-dimensional information for decisions. Journal of Experimental Psychology: Learning, Memory & Cognition, 8, 361-384.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus, OH: Charles E. Merrill Publishing Company.

Wickens, C.D. (1985). Information integration and the internal model. Unpublished manuscript.

Wickens, C.D., Bosco, C.M., Mane, A., Kramer, A., & Donchin, E. (1986). Information value and the event related brain potential. (ARI Technical Report CPL86-3). Champaign, Illinois: University of

Illinois, Cognitive Psychophysiology Laboratory.

Wickens, C.D. & Flach, J. (in press). Information integration. In D. Nagel & E. Wiener (Eds.). Human Factors in Modern Aviation.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. Journal of Applied Psychology, 59, 555-561.

Wright, P. & Weitz, B. (1977). Time horizon effects on product evaluation strategies. Journal of Marketing Research, 14, 429-443.

Appendix

Actual weights used to model the optimal and non-optimal functions are described separately for the five-cue and eight-cue conditions. The optimal function is derived from a linear weighting of each cue's information worth. Since the non-optimal functions do not take the information value of a cue into account, these functions are invariant across display formats. Thus, optimal values were a function of information worth, while non-optimal strategies were modeled as a function of display location. Each particular weighting scheme is designed so that the total of all weights equals ten, this value directly corresponding to the +10/-10 response scale used in the judgment task.

Optimal Functions

Optimal weights for each cue are listed beneath the information worth values given to subjects during experimental sessions. Note that both the optimal function and the information worth values are linear combinations of the cues, the former is merely the "scaled" transformation of the latter.

Eight-Cue Condition

Information Worths	8	7	6	5	4	3	2	1
Optimal Function	2.22	1.95	1.67	1.39	1.11	.83	.56	.27

Five-Cue Condition

Information Worths	8	7	5	3	1
Optimal Function	3.33	2.92	2.08	1.25	.42

Non-Optimal Functions

Weighting schemes for each non-optimal function are presented according to corresponding display locations. These functions did not vary as display formats changed across sessions.

Left Function: Emphasized cues on the left side of the display.

Eight-Cue

2	1.5	1	.5
2	1.5	1	.5

Five-Cue

3	2	1
2.5	1.5	

Right Function: Emphasized cues on the right side of the display.

Eight-Cue

.5	1	1.5	2
.5	1	1.5	2

Five-Cue

1	2	3
1.5	2.5	

Center Function: Emphasized cues at the center of the display.

Eight-Cue

1	1.5	1.5	1
1	1.5	1.5	1

Five-Cue

1.5	3	1.5
2	2	

Top Function: Emphasized cues at the top of the display.

Eight-Cue

2	2	2	2
.5	.5	.5	.5

Five-Cue

2.5	2.5	2.5
1.5	1	