THE EFFECT OF WORK LOAD HISTORY ON OPERATIONAL ERRORS IN AIR TRAFFIC CONTROL SIMULATION: THE HYSTERESIS EFFECT--EXPECTANCY PERSEVERANCE OR SHORT-TERM MEMORY OVERLOAD?

by

MARK WAYNE SMOLENSKY, B.A.

A DISSERTATION

IN

PSYCHOLOGY

Submitted to the Graduate Faculty
of Texas Tech University in
Partial Fulfillment of
the Requirements for
the Degree of

DOCTOR OF PHILOSOPHY

Approved

December, 1990

ACKNOWLEDGEMENTS

First, I want to thank my mentor and committee chairperson, Dr. Charles Halcomb, for his guidance and support throughout my graduate career at Texas Tech. Next, I want to thank my committee members, Dr. Vernon Perez, Dr. Clay George, Dr. Matthew Lambert, and Dr. Jerry Ramsey for their contributions toward my graduate training as, first and foremost, a Psychologist and second, as an Engineering Research Psychologist. I want to thank all of the people whom I have come to call friends at Texas Tech. These are the people who stood by me in the worst and best of times and I am quite positive that without them I would never have completed the curriculum. I particularly want to thank my closest of friends, Jennifer Willis-Eaglin, Gary Chong, Jerry Sparkman, Rob Weiss, Jeff Gofter, Terri Belloff, and Meghan Carmody. I want to thank my new supervisor at the FAA Technical Center, Dr. Lloyd Hitchcock, for serving in the role of surrogate chairman to my committee during my final leg of the dissertation work. I also want to thank my co-workers at the FAA Technical Center, especially Mr. Norm Hitchner and Dr. Earl Stein. Finally, I want to thank my Mom and Dad, Sister, Grandmother and Grandfather, who instilled in me the desire to always strive for my personal best and to never settle for second. Thanks to all!

TABLE OF CONTENTS

ABSTRACT

Most laboratory investigations of simple vigilance tasks and complex monitoring tasks have assessed performance with the underlying premise of an invariable work load. That is, signal probabilities within sessions have traditionally been held constant. This, however, limits the generalizability of the results since real-life vigilance and complex monitoring tasks require operators to work under widely varying levels of work load. Air Traffic Control Specialists (ATCSs), for example, must monitor their radarscopes during varying levels of air traffic activity and complexity within a work shift and consequently their work load varies accordingly.

Among the handful of laboratory investigations that have assessed monitoring performance within the context of intra-session work load variability, the most dramatic behavioral patterns noted have been either a performance decrement in signal detection efficiency under moderate or low work load immediately subsequent to a period of high work load (Colquhoun and Baddeley, 1964, 1967; Krulewitz, Warm and Wohl, 1975) or a disproportionate recovery of performance to reduction in work load (Cumming and Croft, 1973; Goldberg and Stewart, 1980). These studies emphasized the importance of *work load history* in assessing monitoring performance for a given time period. Evidence for these performance decrements in operational settings is best represented by its documented occurrence in air traffic control (ATC) where operational errors (OEs) are reported to occur most often under low to moderate levels of work load (Aviation Safety Institute, 1974; Allnutt, 1976) and particularly when immediately preceded by a period of high work load (Biggs, 1979).

Two predominant explanations have been offered for the occurrence of these performance decrements in the laboratory. One explanation is that the operator establishes certain expectancies

concerning signal:event ratios during a monitoring session that persevere even after signal probabilities shift downward (Cumming and Croft, 1973). An alternate explanation is that there exists a temporal disparity between short-term memory (STM) information processing and transmission that causes the STM buffer to become overloaded (Goldberg and Stewart, 1980). The STM buffer continues to be overloaded for a time even after signal probabilities shift downward. The purpose of this dissertation is to investigate, through the process of controlled simulation, the role of expectancy perseverance and STM overload in the performance decrements that are collectively referred to as the *hysteresis effect*.

Forty-five psychology students served as ATCSs for two one-hour sessions. Students were selected on the basis of their scores on a former Federal Aviation Administration (FAA) ATCS entrance exam. Those students passing the exam were instructed in basic ATC techniques and in the use of the PC-based ATC simulation package, TRACON. This software served as the radarscope in a full-scale, low-to-moderate fidelity mockup of a Terminal Radar and Approach Control (TRACON) workstation. An ATC session began with low task demand, rose to either a high or moderate task demand approximately 20 minutes in the session, remained at that level for 20 minutes and then decreased sharply to low task demand and remained at the low task demand for the remainder of the session. To test the explanation that expectancy perseverance is the cause of the hysteresis effect, each participant experienced a cue treatment indicating either a forthcoming downward shift in task demand from the current demands of the task or a continuance of the current task demand, **and** a no cue treatment. For the cue treatment, both a visual cue and a verbal cue were given simultaneously immediately prior to the downward shift in task demand. To test the explanation that STM overload is the cause of the hysteresis effect, participants experienced either a high to low, moderate to low, or high to high (no shift) task demand session. Sessions were recorded on videotape. OEs were recorded, transcribed and coded from the videotapes. OEs were categorized using a taxonomy of Handoff, Keying, Navigational, Pilot and Memory errors.

It was reasoned that if the hysteresis effect is due to the perseverance of expectancies, then performance should be effected by cueing since the establishment and maintenance of expectancies is primarily a matter of perception based on information regarding the recent history and present circumstances of the situation and any aids that provide information regarding forthcoming parameters of the situation should serve to assist performance. For the data to support a perseverance of expectancies theory, a performance decrement would have to occur under **both** high to low task demand and moderate to low task demand and the decrement would be ameliorated by cueing in both cases. It was reasoned that if the hysteresis effect is due to the cognitive hardware limitations of the human information processing system, namely, the STM store, then performance should not be dramatically effected by perceptual aids such as cueing. For the data to support a STM overload theory, a performance decrement would have to occur under high to low task demands and **not** under moderate to low task demands (since the STM memory buffer would never be overloaded in the latter case) and would **not** be ameliorated by cueing under the high to low task demand shift. The pattern of results do not clearly indicate support for either theory. Rather, the results for Handoff, Keying, Navigational, and Pilot OEs provide support for the perseverance of expectancies hypothesis and the results for Memory OEs provide support for the STM overload hypothesis. Furthermore, results indicate presence of the two patterns of the hysteresis effect. Non-Memory OEs display a tendency for the OE rate to lag behind a reduction in task demand while Memory OEs display a tendency to rise immediately after a sudden downward shift in task demands.

Although information concerning OEs in ATC has been gathered since 1964, there have been few studies designed to analyze the existing information. Therefore, relatively little is known of the types of errors that ATCSs make and why they make them. At least in the case of ATC, it can be concluded that there is support for the both explanations for the hysteresis effect that is dependent upon the category of OEs being considered. This information could be invaluable to the enhancement of future ATCS performance. Successful ATCS performance depends, to a great degree, on the reliable recall of relevant information in STM (e.g., aircraft, destinations, altitudes, etc). as well as prior knowledge of forthcoming changes in work load. A clear understanding of the role of memory and establishment of expectancies in the ATCS's strategic model could lead to the development of effective cueing and memory aids that can help ensure the availability of accurate essential data/information when it is needed. A memory lapse, or even a delay, during a critical point such as a downward work load shift can lead to serious consequences such as conflicting tracks and inadequate separation between aircraft, or flight into terrain obstacles. Similarly, sudden, unexpected changes in work load could disrupt the flow, or, *mental picture* that the ATCS has established to perform the job efficiently. Thus, memory and cueing aids which would efficiently facilitate recall of critical information and inform of forthcoming changes in work load would be of great value.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR: Approach Radar

ASRS: Aviation Safety Reporting System

ATC: Air Traffic Control

ATCS: Air Traffic Control Specialist

FAA: Federal Aviation Administration

FAF: Final Approach Fix

FPL: Full Performance Level

IFR: Instrument Flight Rules

ILS: Instrument Landing System

LC: Local Controller

m: Meters

MEM: Memphis TRACON

NAS: National Airspace System

nm: Nautical Miles

NTSB: National Transportation Safety Board

OE: Operational Error

R:S: Response:signal ratio

STM: Short-term memory

TC: TRACON Controller

TRACON: Terminal Radar and Approach Control

TRSA: Terminal Radar Service Area

VFR: Visual Flight Rules

PREFACE

A Falcon jet, N121GW, departed Memphis (MEM) on a training flight, instrument flight rules (IFR), remaining in the MEM traffic to practice instrument approaches. Two hours into the flight, the fatal leg began. Falcon 21GW made a go-around from Runway 17R with a clearance to climb to 2000 ft, turning right to heading 320 degrees for the downwind leg to Runway 17R. At that time, a Cessna 150, N6423K, was returning, visual flight rules (VFR), to MEM from a training flight to the west. The local control of traffic within a five nautical-mile (nm) radius of MEM was shared by two ATCSs in the tower: local controller (LC), LC 1, handled the quadrants east and west, LC 2 handled the north and south quadrants. The ATCSs could communicate directly with each other, and they shared two radarscopes in the tower. Since it was a Terminal Radar Service Area (TRSA), radar vector service was provided by two ATCSs at Memphis TRACON--ATCSs designated as approach radar (AR) AR 2 and AR 6.

Falcon 21GW was under the control of AR 6 and LC 2 (north-south) as it worked the pattern. After the go-around, LC 2 issued the clearance, inspected his radarscope, and, seeing no traffic west, handed off control to AR 6. By this time the Falcon was west of the field in LC 1's sector. AR 6 knew nothing of the inbound Cessna. Now the crucial step: when LC 2 handed off control of the Falcon to AR 6, he made no effort to coordinate with LC 1, he merely visually coordinated by looking at the radar and did not see the primary return from the Cessna. AR 6 assumed that all necessary coordination had taken place.

Meanwhile, AR 2, which had been handling the Cessna, called LC 1 and handed off control to him. LC 1 gave entry instructions and turned his attention to another aircraft in his east sector. When he accepted the Cessna, he checked the radar and saw no other traffic west. He testified later that he was unaware of the presence of the Falcon in his sector. His attention was abruptly brought back to his west sector when LC 2 asked him if he had traffic west. LC 1 looked at his radar, saw the targets 1 to 2 nm apart, and attempted to issue an instruction to Cessna 23K. A few words into his transmission, he observed a fireball to the west.

The National Transportation Safety Board (NTSB) indicated three contributing factors to this accident that have direct implications for human factors professionals: (a) vigilance: at least three ATCS examined their scopes and did not see other aircraft--only the one they were controlling; (b) work load and work place design: the ATCSs in the tower, though their work stations were only about 3 meters (m) apart, failed to communicate to each other the presence of their respective aircraft. The NTSB indicated some disagreement about whether the work load at the time was moderate or heavy. However, a third ATCS position, whose responsibility was to coordinate between ATCSs, was vacant. The ATCS had left, as he was allowed to do during light work loads; (c) radar advisory service: in a TRSA this can be deceptive. While ideally it should provide sequencing, advisories, and, in some cases, separation, present human-radar systems are not capable of doing so reliably. It is mandatory that pilots, on clearances of any type, *see and avoid* other aircraft when operating under visual meteorological conditions. The very concept is questionable, as there are so many limitations on the visual ability of the pilots, as well as heavy cockpit work loads that make extra-cockpit scanning a low priority item. Furthermore, operations in a radar environment contain an inherent danger: the assumption by the flight crews that they are being offered flawless, radar-based separation and, hence, can relax their scanning vigil (i.e., the inducement of false expectancies).

The preceding scenario is just a single example of what happen when various factors combine in ATC during a typical monitoring session. The amount of compiled data on such incidents and accidents, however, is startling.

CHAPTER I

INTRODUCTION

One measure of the effectiveness or safety of the National Airspace System (NAS) is the frequency of occurrence of OEs. The Aviation Safety Reporting System (ASRS) was established in 1964 to collect relevant information concerning OE occurrence and to provide a means of identifying problem areas in the system so that corrective actions could be initiated. OEs are not recorded automatically, but are based on reports from the ATCSs involved, other ATCSs, the supervisor, or pilot of one of the involved aircraft. Thus, a certain number of OEs go undetected. Concern about OEs has steadily increased based on the trend that indicates that for each successive year in the decade from 1970 through 1980, with the exception of 1973, there was an increased number of OEs. The overall relationship between the number of OEs and volume of traffic is evident in Figure 1, where the number of OEs per million operations is plotted for each year from 1970 through 1980. On the basis of these data, en route facilities as compared with TRACONs have historically handled fewer operations per error. In TRACONs, there has been a steady increase in the number of OEs per million operations for this same time period. In contrast, en route centers handled approximately the same amount of traffic per error in 1977 through 1980 as they did in 1970 and 1971. However, there was considerable fluctuation in the en route error rates during the intermediate years, 1972 through 1976.

Even though information concerning operational OEs has been gathered since 1964, there have been few studies designed to analyze the existing information. Schroeder (1980) conducted a study designed to determine the contributing factors and conditions that predispose an individual or ATC facility to have repetitions of OEs. Analyzing the database of OE reports,

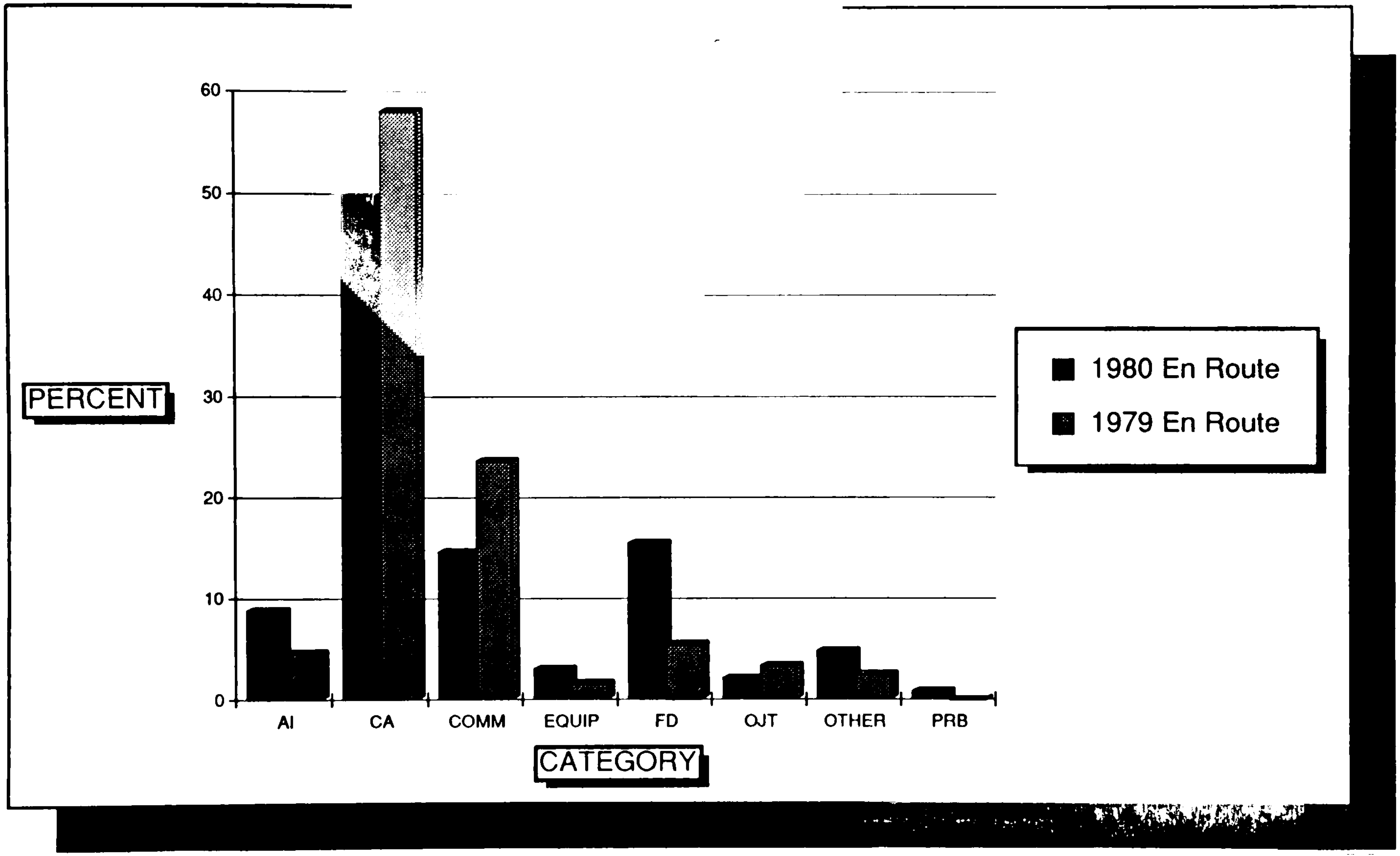

Figure 1

Direct Causal Factors of OEs

Schroeder (1980) noted that there are two subjective measures of work load typically provided--traffic volume and complexity. Of these two work load indices, Schroeder deemed traffic volume to be the more pertinent and therefore based his descriptive analyses of the data on this measure.

Information from the analysis of 1977 and 1978 OE records indicates that the largest percentage of OEs occurred under moderate work load conditions (45.01% and 47.56%, respectively). Of the three work load conditions (light, moderate, and heavy), OEs were least likely to occur under conditions involving heavy work loads (17.31% and 13.56%). A comparison of these values with information concerning OEs in 1965 and 1966, indicates that there has been a significant decline in the percentages of OEs occurring under heavy work load conditions. While the data does not include information for each year from 1965 through 1980 there is sufficient information to provide a longitudinal view of the relationship between OE rates and subjective estimates of work load. The mean number of OEs that occurred under light, moderate, and heavy work loads for each of the available years from 1965 through 1980 is presented in Figure 2. These values reflect the combined average for en route and TRACON facilities.

During the years 1965 and 1966, Kershner (1969) (cited in Schroeder, 1980) found that the majority of OEs occurred under moderate work load conditions, with the second greatest frequency occurring under heavy work load conditions. Since that time, the percentage of OEs involving moderate work loads has consistently been higher. Over the same time period there has been a marked decline in the percentage of OEs that involve heavy work loads. Correspondingly, there has been an increase in the percentage of OEs involving light work loads. Thus, while traffic volume and the complexity of the airspace system have increased significantly, a higher percentage of OEs involve light to moderate work loads. To determine, however, if a disproportionate percentage of OEs occurred under one of the three work load conditions it would be necessary to know what percentage of the ATCSs' time was spent under each work load level. Such data is unavailable. In spite of the fact that the reasons for this shift in the percentage of OEs occurring under each work load cannot be determined as yet, the trend is toward a greater frequency of OEs occurring under light and moderate work loads.

A preliminary investigation of the 1977 and 1978 OE records was conducted to determine the relationship between the subjective measure of work load (using traffic volume as a criterion) and number of aircraft handled. Unfortunately, data concerning number of aircraft was not available in a majority of the records. Under the new OE recording form, ATC facilities are required to indicate the number of aircraft being handled at the time of the OE. An average for number of aircraft was obtained for each level of work load on both TRACON and en route centers. These values are presented in Figure 3. Within each setting, the average number of aircraft for the three work load levels were significantly different ($p<.01$); there was, however, considerable overlap. For example, in TRACONs, an ATCS with five aircraft on the radarscope could, given the right set of circumstances, be involved with either a light, moderate, or heavy work load. There are obviously other aspects of the situation that become involved in the determination of this work load measure other than traffic volume alone. Comparison of the averages for TRACONs and en route centers indicated that the values for en route centers were significantly higher ($p<.01$) for each work load level. In fact, the en route average for a moderate work load was equivalent to the average for a heavy work load in a TRACON. Further investigation is needed to determine the exact relationship between traffic volume as a subjective measure of work load and the number of aircraft handled in TRACONs and en route centers.

Previous research by Kinney, Spahn, and Amato (1978) (cited in Schroeder, 1980) revealed that of the OEs that occurred in 1974 through 1976, the direct cause in 91% to 96% of the cases involved the influence of one of three factors: attention, judgment, and communications. Since 1978, the data accumulated has involved different categories for the description of causal factors. While this does not allow a direct comparison with earlier data, a review of the categorizations included indicates that these three factors are still prominent. For 1979 and 1980, in both TRACON and en route facilities, the largest number of OEs involved failure of the ATCS to initiate corrective action. The subcategories include: inattention, poor judgment, poor planning, and waiting too long to initiate action. Additionally, the categories of airspace encroachments due to lack of coordination, flight data, and communications were involved in significant numbers of OEs. Airspace encroachments describe situations where the ATCS failed to coordinate adequately with another ATC facility or position and allowed an aircraft to fly into airspace under the control of another ATCS. Included under flight data are situations in which the ATCS did not use the available flight data to analyze the traffic situation or where he/she failed to maintain a current display. A variety of problems are assumed in the category of communications, ranging from transposition of aircraft identification and altitudes to misunderstandings and failure to detect incorrect readbacks. In TRACONs, a fairly high percentage of the OEs involved the airport movement areas. Included in this category would be situations where the ATCS allowed an aircraft to take off or land too close behind another aircraft or allowed a vehicle or another aircraft to cross an active runway without approval. These data are based on review board decisions concerning the direct cause.

As already emphasized, there does not appear to be a linear relationship between ATCS work load and frequency of OEs. The majority of OEs has been found to occur when ATCSs are working under light to moderate work loads. To further aggravate the problem, a number of the near misses which were studied by the Aviation Safety Institute (1974) and others (Allnutt, 1976), however, indicated that many OEs came not during a period of intense concentration, but immediately afterward--when the work load was relatively light. While under high work load, most ATCSs perform at high levels. However, once the stressful period is over, they tend to commit OEs resulting in a possible incident or accident. This implies a relationship between intense work load and subsequent less intense periods of time. These OEs are committed by good ATCSs who have apparently no extraneous factors affecting their behavior at the moment, and yet they might fail to monitor an aircraft plainly visible on their scope, write down the wrong altitude, or make some other error (Biggs, 1979). As one FAA administrator said (Biggs, 1979),

> All of a sudden, you've got people and airplanes falling out of the sky; and you've got a happy, rested, well trained ATCS who simply made a cerebral error. You can tell me he wasn't paying attention. Well, what's attention? How do you pay attention? How do you know when you are not concentrating?

The major issue of concern in this dissertation is in discerning the nature of this post-intensive error-conducive period of work load. Correct diagnosis of the antecedents of the phenomenon, referred to as the hysteresis effect in laboratory investigations, will likely determine the type of human factors interventions prescribed in the future.

Although work load has been the subject of increased interest and research activity in recent years, one variable that may be of some significance, namely, work load history, has been consistently overlooked. Work load history refers to the level of work under which an operator performs immediately prior to the period of time when performance is being assessed.

Traditionally, temporal effects that are most often reported in the vigilance literature are described in terms of work load levels that are stable over time, and interest has focused largely on the performance changes associated with time on task. Further, when work load has been varied it has tended to be treated as a between-subjects variable, thereby producing cleaner experimental designs that permit load to be assessed independently of any confounding effects that may be due to previous work load demands.

As more emphasis is being placed on understanding the interaction between human and system performance in the context of functions that are shared, it should be recognized that operator work load may vary dramatically over time as systems attain a more dominant role in controlling the information rate to which the operator must respond. Indeed, Cumming and Croft (1973) have pointed to the fact that in many real-life tasks, task demand varies continuously over time. In such real-life situations, Cumming and Croft have argued, both variation in task demand and an individual's expectations regarding demand may play an important role in affecting performance. Consequently, the key question becomes: by how much will we be in error if we attempt to estimate work load effects on performance without taking into account work load history?

Previous investigations of the relationship between the rate of human information transmission and task demand have shown that as demand increases performance also increases, but only to a point at which the level of demand is optimal. Beyond this optimal level of task demand, the task overloads the individual and results in a performance decrement (Alluisi et al., 1957). Typically, such investigations have involved comparing performance on forced-paced serial reaction tasks under several discrete levels of task demand.

In a series of experiments reported by Cumming and Croft (1973), subjects performed a ten-choice key-pressing task, attempting to match random digits presented binaurally. The rate of presentation of the digits first increased from a minimum of one digit approximately every four seconds to a maximum rate, and then decreased to form a symmetric cycle. Different cycle times, ranging from 24 to 120 seconds were employed, and the maximum presentation rate reported was three digits/second. The results of their experiments suggest that the relationship between performance and task demand is at least partially dependent upon the time history of task demand, or work load history. Subjects' performance did not recover at the rate expected as demand was reduced.

The Hysteresis Effect

Two extreme cases of work load variations can occur: (1) a sudden increase in the number of signals that must be detected and processed following a period in which signals are infrequent, and (2) a rapid reduction in signal detection and processing demands following a period of sustained high signal work load. Informal evidence points to the performance consequences of such patterns. For instance, the first case is a broad description of the events that have become known simply as Three-Mile Island, about which little more needs to be said in the human factors literature at this time. An example of the second type--the consequences of work load reduction--comes from Allnutt (1982), who quotes anecdotal evidence of the driver of a railway train driving through a red signal in a period of low work load following a sustained period of high work load that had required intense concentration. Further evidence of sudden work load reduction comes from ATC. A number of the near misses which were studied by the Aviation Safety Institute (1974) and others (Allnutt, 1976) indicated that many OEs came not during a period of intense concentration, but immediately after, when the work load was relatively light. As noted earlier, once the period of intense work load has passed, ATCSs tend to commit OEs.

In one of the few laboratory studies to examine the effects of variable work load conditions on monitoring performance, Cumming and Croft (1973) examined choice reaction time as a function of varying the presentation rate of auditory signals.

Signal rate was varied in a cyclical manner and it was found that during the decreasing demand portion of the cycle, performance was significantly worse than in the corresponding period during the ascending phase. This performance decrement persisted beyond the high signal rates of its inception, continued through a period of reduced signal rate, and was subsequently ameliorated only after a disproportionate reduction in task demand.

Cumming and Croft (1973) have offered the explanation that this hysteresis effect occurs as a result of subjects failing to adequately match their expectancies with changing task demands. They explain that in a choice reaction time task, at levels of task demand below work overload, the operator attempts to transmit as much of the stimulus information as possible and that the operator allows an increasing proportion of errors as task demand increases. If the operator does not recognize the rate of task demand, the estimate of task demand will lag behind the actual task demand, and strategy with respect to errors will also lag. This is especially likely when the rate of signal presentation begins to slow down after having been steadily increasing, because the change will be unexpected. It is well known that reaction time is slower for events of low than of high probability, so that in the case of unexpected changes in presentation rate, the time lag could well be substantial.

Other evidence to support a perseverance of expectancies hypothesis of the hysteresis effect are extrapolated from the vigilance literature on the pre-training of subjects to expect a given signal probability during a vigilance task. Where signals rates have been switched within sessions, without the subjects being informed, the greatest decrement has been associated with the signal density being switched from high to low; the decrement was even greater than when the rate was continuously low, although the final levels of performance were not much different.

Wiener (1963) pre-trained subjects with one of three signal rates: 16, 32, or 48 signals in 48 minutes. On the second day, all the subjects received 32 signals. The group trained with 48 signals detected most, and the group trained with 16 detected fewest signals in the test session. The 16 signal group also gave fewest false responses.

A similar finding was reported by Colquhoun and Baddeley (1964, 1967). The investigators used either a low (*p*=0.02) or a high (*p*=0.18) level of signal probability during pre-training in a vigilance study and then tested their subjects either under the same probability level or under the alternative probability. The results indicated that subjects who were pre-trained at the low rate detected fewer signals and gave fewer false alarms than those pre-trained on the high signal rate, especially at the beginning of the test period. Significant decrements in target detections during the test session were found only with those subjects who had been pre-trained on the high signal rate. There was no difference in the detectability of the signals for the different groups, as measured by d', but there were significant differences in ß, a measure of the criterion level adopted by the subjects. ß was higher with those subjects who had been trained with the low signal rate, indicating that these subjects were less ready to make a positive response than those trained with the high signal rate. There was a significant increase in ß during the test session for those subjects who had been trained with the high rate and tested with the low rate. This finding lends support to the suggestion that the vigilance decrement is due to a shift in ß as the subject finds that there are fewer signals than was expected. These findings have since been replicated by Colquhoun and Baddeley (1967) and by McFarland and Halcomb (1970). Krulewitz, Warm, and Wohl (1975) have reported similar effects with event rate shifting; specifically, when switched from a high to a low event rate (signal probabilities remaining constant), the probability of detection increases, and when switched from a low to a high event rate, the subjects' probabilities of detections decreased.

Further support for the perseverance of expectancies hypothesis comes from Craig (1980). Although Craig did not shift signal probabilities intra-session, he nevertheless found that subjects informed of signal probabilities in a simple vigilance task showed less vigilance decrement than uninformed subjects. Subjects each performed for one hour on a visual vigilance task with two kinds of signals which differed in conspicuity and also in probability of occurrence. Prior to testing, one-half of the subjects were explicitly informed of the signal probabilities; the other half were not. The results indicate that prior knowledge tends to ameliorate the extent of the shift in ß and, hence, to stabilize performance over the session.

Other evidence for the contributions of perseverance of expectancies is anecdotal and has been named the *false hypothesis* in operational settings. For example, every pilot knows of end-of-tripitis--that time at the end of a flight when the level of concentration drops. The most difficult part of the flying procedure has been successfully completed, but it is all too often at this juncture that an accident occurs. Among the anecdotes offered by Davis (1958) is that of a railway driver who drove through a fog from Crewe to London--a distance of 160 miles. A few miles away from the termination of the trip the fog cleared and the sun came out. The driver relaxed, drove through a red signal and crashed into another train. Clearly, the greater part of the trip--in poor visibility--had made great demands on his concentration. The remainder of the journey must have seemed so easy, in contrast, that the driver felt he was as good as home.

Goldberg and Stewart (1980) conducted an investigation in which subjects performed an eight-choice key-pressing task, attempting to match characters presented singly under both increasing and decreasing demand. Their results supported previous research (Cumming and Croft, 1973) indicating that the relationship between the rate of information an individual is able to transmit and task demands depends, at least in part, upon the temporal history of demand. When a relatively high level of task demand was imposed, performance failed to recover at the expected rate as demand was reduced. That is, when a period of increasing task demand was followed by a period of decreasing task demand, performance failed to recover at the rate expected as demand was reduced. This finding is virtually identical to that reported by Cumming and Croft (1973). However, the perseverance of expectancies hypothesis was rejected by Goldberg and Stewart (1980). Rather, they concluded that their results suggested that at relatively high levels of demand, information enters STM more rapidly than it can be transmitted. As STM becomes overloaded, performance deteriorates even though the presentation rate continues to increase. This STM overload continues to exist as the presentation rate (i.e., task demand) begins to decrease; performance recovery at the rate that would otherwise be expected is thus prevented until demand is sufficiently reduced (i.e., until the individual is no longer overloaded).

Matthews (1986) conducted two visual search experiments that demonstrated that visual search for a signal from a number of potential signal sources in a sustained monitoring task is dependent upon previous visual load history. The primary aim of the two experiments was to examine whether the hysteresis effect could be demonstrated in a task with varying visual load that does not require the preservation of serial order information. Their objectives were to provide a critical test of the STM overload theory by employing conditions under which memory capacity was not considered to be a significant factor. Their results did not support the STM overload explanation for the performance decrement that is associated with a shift from high to low information rates. Subjects in a variable load treatment failed to show the type of decrements in accuracy that such an explanation requires, and performed consistently faster than did subjects with a constant rate history. At the same time, the failure to find a difference between cyclical and random varying load conditions also argued against Cumming and Croft's (1973) explanation that the performance decrement following high load levels is due to a failure to anticipate or recognize work load changes. This explanation would have predicted higher performance levels for the cyclical treatment (in which subjects could readily perceive the pattern of rate changes) compared with the random rate treatment. The data did support, however, the notion of strategic persistence, since it could be argued that when subjects in the two variable load groups move from a set of trials at high rate to a block at a lower rate, their higher rate of responding tends to persist. Thus, they perform the task with greater urgency even when working at much lower load levels. This explanation is consistent with Poulton's (1982) (cited in Matthews, 1986) claim that subjects exhibit strategic persistence from one experimental treatment to another in within-subject designs. Poulton has argued that in the normal case, this has led to research results that are an artifact of the within-subject design.

It is the purpose of this dissertation to assess the tenability of the role of the two competing explanations of the hysteresis effect: perseverance of expectancies and STM overload as it applies to what is most obviously a human factors issue--the hysteresis effect in ATC. Different antecedents for the effect will likely necessitate different human factors countermeasures. Thus, it is imperative to discern these antecedents.

Prior research has attributed the hysteresis effect's cause to both the untimely perseverance of signal expectancies and STM architecture limitations. This investigator has not subscribed to either theoretical stance wanting, rather, to let the current experiment indicate support for one over the other.

Hypothesis I

Based on the data presented by Cumming and Croft (1973), it is hypothesized that the hysteresis effect is due to the development of expectancies concerning signal probabilities at a given period of time. These expectancies are relevant for the time period in which they are established but may not be applicable beyond that time period. In applying this notion to a complex monitoring task, such as ATC, where the focus of the task is not merely to detect aircraft but to process these aircraft so that they revert back to a nominal treatment, the operator establishes certain expectancies early in a monitoring shift concerning the types of actions to be initiated in subsequent time periods. These expectancies may not be appropriate for subsequent time periods if, for example, weather conditions change, pilots commit errors that foul up sequencing, work loads change during the monitoring session, etc. These are very common occurrences in ATC. The question, thus, becomes what is the contribution of perseverance of signal expectancies to the hysteresis effect in ATC? Further, would the use of an effective cue to inform of forthcoming changes in signal probabilities attenuate the hysteresis effect? Formally stated, hypothesis I predicts the following:

1. If the hysteresis effect is due to an operator's inability to timely perceive a change in the work load and hence adjust expectations accordingly in a timely fashion, the effect will be observed for **any** downward shift in work load. In this experiment, this refers to both high to low and moderate to low work load shifts.
2. The effect will be attenuated through the administration of a cue that will inform the operator of a forthcoming change in task demand.

Hypothesis II

Based on the data presented by Goldberg and Stewart (1980), it is hypothesized that the hysteresis effect is due to the strain that is placed on the restricted STM architecture and thus is specific to tasks requiring the short term memorization of some characteristic of the signals. It has been noted that an operator, such as an ATCS, develops a *picture* of the session being monitored. This picture is dynamic and is stored in STM. There is empirical evidence to suggest that when some event occurs to disrupt the picture such as a system warning or a sudden downward shift in task demand, the operator *loses the picture*. This is the period of time when OEs are most likely to occur. In applying this notion to a complex monitoring task, such as ATC, where the focus of the task is not merely to detect aircraft but to process these aircraft so that they revert back to a nominal treatment, the operator must remember many pieces of information about all aircraft on the radarscope. This memorized information becomes the picture with which one functions in the present and plans for the near future. When there is an interruption in the flow of work, however, the ATCS loses the picture. That is, much of the information being maintained in STM is forgotten. This is tantamount to receiving the telephone number from an operator while in a phone booth and while you are rehearsing the number to remember it, a sudden loud fire engine passes by and you forget the information. Losing the picture is a common source of OEs for ATCSs. The question, thus, becomes what is the contribution of STM architecture limitations to the hysteresis effect in ATC? Further, considering the constraints of the STM architecture, would the hysteresis effect occur where work load histories have not approached the upper limits of the STM store? Formally stated, Hypothesis II predicts the following:

1. If the hysteresis effect is due to an overload of the STM architecture, then this pattern will occur for high to low task demands only.
2. This pattern will not be affected by the administration of a cue to indicate a forthcoming downward shift in task demand.

CHAPTER II

METHODOLOGY

Experimental Design

The design for this experiment was a 3 x (2 x 5) split-plot factorial. Three types of WORK LOAD SHIFT were used as a grouping factor for participants (high to low, moderate to low, and high to high (no shift). All participants served under both treatments of CUE (cue and no cue). All participants were assessed over a one-hour monitoring session with TIME divided into five-minute intervals. The time intervals of interest were 25-30 (tl), 30-35 (t2), 35-40 (t3), 40-45 (t4), 45-50 (t5) minutes of the one-hour session.

Participants

Forty-five undergraduate psychology students (27 males and 18 females) from the General Psychology and Physiological Psychology courses volunteered, qualified and participated in all portions of the investigation. Participants ranged in age from 18 years to 24 years.

Students registering for the investigation were screened for their suitability as participants. There were several criteria that were prerequisites for students' suitability for participation: (a) a rudimentary knowledge of the personal computer and its keyboard; (b) normal visual acuity or corrected to normal; (c) no prior experience with the task used in this investigation or any previous ATC training; (d) a mean critical composite score of 75.1% on the FAA ATCS Written Exam. This score is identical to that used by the FAA for applicants to the ATC academy. Students not passing the exam were dismissed from the remainder of the investigation and given course credit equivalent to the time spent on the exam (about one hour). Of the 145 participants that volunteered and sat for the exam, 85 failed to meet the critical composite score. Students who met the score were deemed suitable for participation in the entire experiment were given course credit equivalent to the time spent on the experiment (about five hours). The mean critical composite score of those participants that qualified for training and completed the entire experiment was 82.4%.

Forty-five suitable participants were randomly assigned to one of three possible WORK LOAD SHIFT groups (15 participants in each group) and each participant was randomly assigned as to the order of the repeated-measure variable, CUE. Fifteen students were lost to attrition and were subsequently replaced with other suitable participants.

It should be noted that the treatment of all participants was in accordance with the ethical standards of the American Psychological Association, 1990.

Apparatus

The ATCS Exam. The FAA ATCS Written Exam consists of three separate portions. The first portion is designed to assess aptitude as an ATCS; the second test is intended to assess the ability to perceive spatial relationships; and the third test is intended to assess knowledge of aviation and ATC work. Since no prior knowledge of ATC or aviation was assumed, only the first two portions of the exam was administered. A critical composite score of 75.1% of a total 100% over all portions was required to pass the exam. This is the same pass criteria that is used for actual ATCS applicants in the FAA.

The *ATCS Aptitude* portion of the exam consists of drawings that simulate a radarscope depicting characteristic patterns of air traffic. Each problem contains a drawing of a particular flight path and aircraft flying on it. A table containing information about the altitude, speed and route of flight of each aircraft accompanies each drawing. The task is to answer questions which make use of this flight information. The questions ask for identification of potential midair collisions, differences in the route of flight of the aircraft, distances between aircraft, compass headings of different aircraft and changes in the route of flight. Some preliminary instructions necessary to correctly read the information provided in the problems are given at the test site before the start of the exam.

The *Spatial Relationship* portion of the exam deals with relationships among sets of figures and among sets of letters. Two categories of questions are included. For *Letter Series* questions, participants are given a series of letters that are arranged in some definite order. To the right of each question are five suggested answer choices; each answer choice consists of a set of two letters. Participants look at each letter series and determine what its order is and then from the suggested alternatives select the set that gives the next two letters in the series in their correct order. For *Symbol Classification* questions, each item consists of two sets of symbols. Participants are asked to find the one rule that (a) explains the similarity of the symbols within each set, and (b) also explains the difference between the sets.

Psychometric assessment of the FAA ATCS Written Exam reports test-retest reliability ranging from .60 (ATCS Aptitude test) to .78 (Spatial Relationship test-Symbol Classification). Alpha coefficients range from .62 to .84. The validity of the ATCS Aptitude test is .40 for performance during training, .15 for performance on-the-job, and -.06 for progression and attrition. The validity of the ATCS Aptitude test is .65 for performance during training, .32 for performance on-the-job, and .25 for progression and attrition.

<u>Advanced Technology Learning Center Zenith Lab</u>. An instructional laboratory in the Texas Tech Library Advanced Technology Learning Center was the site for the training of 45 ATCS Developmentals (as they were referred to in the simulation)

to operate the PC-based ATC Simulator, TRACON. The laboratory consisted of 19 Zenith Z-158 PC/XTs identical to the type used in actual data collection except that the computer monitors were of a green phosphor rather than amber and were of lower resolution. Three training sessions were conducted to accommodate all participants.

Experiment site. The investigation was conducted in one 240 cm x 120 cm x 120 cm cubicle in the Texas Tech University Human Factors Laboratory located in the Department of Psychology. Every effort was made to simulate an actual TRACON environment within budget contraints. Indirect lighting was used in the cubicle (two small red and blue light sources were used to simulate the dominant spectral wavelengths in an actual TRACON environment). The level of illumination at the radar display was about 5.6 lux. This was comparable to the light level of an actual TRACON environment. The light level was calibrated with a Radio Shack brand photometer. In order to add a greater element of realism to the task, a tape recording of ATCSs at Los Angeles TRACON and accompanying TRACON operational environment noise was played continuously during each one-hour ATC session. Sound level of this recording at the participant's head location was about 62 dBA $\pm$ 3 dBA. It was not expected that this would have any effect on performance, since an earlier investigation by Thackray (1982) failed to find any significant performance effects of this specific noise type at a considerably higher (80 dBA) level. The level of sound was calibrated with a Radio Shack brand sound meter.

A pad of *Post-It* brand memo paper was provided in the cubicle. This enabled participants to make reminders to themselves about pertinent information about aircraft, airports, etc. Not surprisingly, this method is used by actual ATCSs in the field.

NYET. NYET is a public domain version of the popular Nintendo brand computer game, TETRIS. Supposedly, NYET (and

TETRIS) were developed in the Soviet Union. The object of the two-dimensional game is to properly fit falling blocks of different shapes together at the bottom of the screen. Points are earned merely by allowing blocks to fall to the bottom of the screen. By fitting the blocks together tightly, however, so that no gaps remain within a row, that row is eliminated, bonus points are earned, and the player has that much more room to maneuver falling blocks. As the game progresses, the rate at which blocks fall accelerates, thus providing the player with less time to strategically place the blocks. The game continues until the player has failed to eliminate enough rows to permit the space and the time to maneuver. Thus, the blocks build to the top of the screen and the game ends.

TRACON. TRACON is a simulation of a TRACON air traffic facility. TRACON's area of coverage for this experiment was the airspace (referred to as a sector) surrounding Los Angeles and its satellite airports. A sweeping radar line continuously scans the skies, pin-pointing each aircraft target and reporting its altitude with each scan. Airports, airways, VOR radio beacons, intersection fixes, and ILS instrument landing systems are all represented from actual FAA airspace charts. Significant ground markings such as coastlines are also shown.

Using the radarscope, the automated flightstrip display and the communications channel, participants must handle all the aircraft in their sector, keeping them on course and safely separated from each other, vectoring (guiding) them safely to their destinations, and handing them off to en route ATCSs or to tower ATCSs.

This job is complicated by pilot OEs--sometimes pilots do not hear an ATCS clearly the first time and ask for commands to be repeated, sometimes they do hear the ATCS but still do not comply. Missed approaches are common in bad weather, so even though a ATCS may have finished with an aircraft, it may pop up again announcing a missed approach, which in turn throws all sequencing off, and adds to the complexity of the scenario.

TRACON was designed primarily by Robert B. Wesson, Ph.D. Dr. Wesson's thesis involved applying artificial intelligence techniques to the problems of ATC. In 1977, he created a program running on a mainframe which not only simulated an Air Route Traffic Control Center's sector, but also solved separation problems in that environment and issued appropriate ATC commands to the aircraft in the sector. TRACON evolved from the simulation component of that program. TRACON was programmed in Microsoft C by Dale Young. Many of its most advanced, realistic features were suggested and tested by George Booth, a Program Manager at the FAA's Advanced Concepts Division in Washington, D.C. and by a number of local professional ATCSs at the Austin TRACON.

TRACON's screen is divided into four basic areas--the main radarscope, which represents the Los Angeles TRACON sector, the pending flightstrips bin, the active flightstrips bin, and the communications channel where ATCS-pilot dialogue and ATCS-ATCS dialogue are echoed. The screen configuration is presented in Figure 4.

The largest area of the screen consists of the radarscope. Airports are marked as circles with instrument landing system (ILS) approaches appended to them. Fixes are denoted as plus signs. The radarscope displays the aircraft that the ATCS is working. Aircraft targets move in accordance with the performance characteristics of each type of aircraft. Each aircraft is displayed as a small aircraft icon oriented in the direction of travel. Information (called a data block) about the aircraft is displayed next to it, offset by a leader line to one of the eight major compass directions (N, NE, E, SE, S, SW, W, or NW). The data block shows the aircraft's ID number, or, call letters, current altitude (in hundreds of feet), and, if it was climbing or descending to a newly assigned altitude, a small up/down arrow or slash followed by the assigned altitude.

The flightstrips bins contain the pertinent information on all aircraft that the ATCS must work. There is one flightstrip for each aircraft. The pending flightstrips bin indicates what aircraft are going to be ready for clearance sometime within the next five minutes. Within the five minutes prior to a handoff of an aircraft from the Center to the TRACON ATCS, or prior to takeoff from an airport, a flightstrip is posted in the pending flightstrips bin. A flightstrip gives information about the aircraft's type, current position (altitude, and airspeed), and flightplan. The active flightstrips bin indicates what aircraft are currently being worked by the ATCS. The flightstrips display the same information as the pending aircraft flightstrips.

The communications channel (somewhat akin to the concept of Data Link in the FAA) provides space for four lines of text and echoes all pilot-ATCS and ATCS-ATCS communications. Phraseology between pilot-ATCS and ATCS-ATCS are relatively standardized. The phraseology has been adapted from current FAA standard operating procedures. In addition to echoed text on the communications channel, TRACON generates voice synthesis output to accompany the dialogue of ATCSs and pilots.

In TRACON, as in real-life, the ATCS deals with four basic types of flights: en routes, departures, approaches and puddle jumpers. En routes are easiest to handle. These aircraft enter the ATCS's sector at one of its fixes at a flight level cruising altitude and want to exit the sector at another fix at the same altitude. If there is no other traffic, the ATCS has to do nothing more than accept a handoff from one Center ATCS and hand him off to another Center ATCS. It is possible, however, that the ATCS might have to alter the aircraft's altitude or vector it around conflicting traffic in the sector. The key constraint in this type of maneuver is that an en route aircraft must be vectored back to its original altitude, heading, and airspeed when it is handed off to Center or the ATCS will generate an OE for not adhering to the flightplan that Center expects.

Departures are still relatively easy. Once the ATCS delivers clearance for an aircraft to depart an airport, the tower ATCS clears it for takeoff and presumably it starts rolling down the runway. The TRACON ATCS does not immediately see it on the radarscope--the aircraft must have some altitude before its Mode C transponder can be detected by TRACON radar. This takes approximately one to two minutes. Then, if the ATCS does nothing, the aircraft will turn to intercept its outbound course and proceed to climb to its cruising altitude (or to the maximum vectoring altitude for the sector, whichever is higher). Just prior to reaching its coordination fix, the ATCS must hand the aircraft off to the Center. Delays arise when an aircraft's requested departure time conflicts with others at the same airport, when it would conflict with an approach in progress, or when its normal climb-out would conflict with en route or another climb-out from a nearby airport. In this case, the ATCS must choose to either hold the aircraft on the ground or clear it for takeoff and then vector it or the conflicting aircraft to provide minimum separation distance.

Arriving aircraft are difficult to handle primarily because they take so much of the ATCS's attention. After accepting the handoff from the Center, the ATCS must vector the arriving aircraft to Final Approach Fix (FAF) heading and altitude so that it is in proper position to be handed off to the tower ATCS for final descent. This involves intense concentration and great skill since the ATCS must anticipate when to vector the aircraft while accounting for other air traffic, the performance characteristics of the arriving aircraft and the weather conditions.

TRACON provides a set of parameters that can be adjusted for the specific needs of the experimenter. The wide array of options will not be described here. For replication facilitation, however, the parameters used were: sector: LAX; scenario: custom designed by experimenter; work load: 16 aircraft over 30 minutes;

weather: minimums; pilot proficiency: average; wind: 10 knots from ENE.

The general format for computer entry of an ATCS initiated action is: <aircraft id> <command> <parameter> <enter>. The first part of each command requires the ATCS to select an aircraft from either the pending or active flightstrip bins which is accomplished using the arrow keys to highlight an aircraft's flightstrip and pressing the <enter> key. The ATCS completes the command with a <command> key followed by a typed parameter followed by the <enter> key. The following command keys are available: turn right (to a heading); turn left (to a heading); climb and maintain (to hundreds of feet); descend and maintain (to hundreds of feet); handoff to tower to Center; change speed to (knots); resume normal speed/navigation; cleared direct to (fix name); hold at (fix name); report current heading and airspeed. Clearing an aircraft for takeoff or accepting a handoff is accomplished merely by highlighting the aircraft's pending flightstrip and pressing the <enter> key twice. The keyboard configuration is presented in Figure 5.

Hardware. A Zenith Model Z-158 PC/XT was used to run the TRACON software. The computer unit is an IBM compatible system that runs all DOS-based software and contains an Intel 8088 microprocessor. The video display is Hercules graphics. The Zenith ZVM-1230 monitor sports a 14-inch tube with 13 visual inches diagonal, 90-degree deflection, 0.29 mm dot pitch and the monitor output is amber monochrome. An identical monitor was placed in the cubicle directly adjacent to the experimental cubicle and was linked to the computer so that one computer sent its video output to both monitors. This enabled the experimenter to view the ATCS sessions without being intrusive. The second monitor was used solely for the purpose of videotaping each session. The computer keyboard keys were individually labeled as to their functions in the TRACON simulator.

A Speech Thing brand speech synthesizer was used to present both pilot and controller output aurally. The synthesized voice was presented through a pair of Radio Shack brand headphones to the participant. The voice was simultanteously sent through a COVOX speaker system so that a microphone could record the synthesized voice onto the videotape.

An Itcikegami Model ITC-40 television camera was used to record each one hour monitoring session for each participant. The camera was set in a fixed position on a Slick U210 tripod in front of the slave monitor in the cubicle adjacent to the experimental cubicle. A Radio Shack brand omnidirectional microphone was placed behind the TRACON work station in order to record the speech synthesized output of the commands being issued by the controller working the scenario. In this way, the experimenter was able to record, both visually and aurally, each participant's session without being intrusive. (Although analyzing videotaped sessions is a painstaking undertaking, the videotapes provided a complete, accurate, and objective method of assessing the occurrences of OEs). Forty-five RCA VHS HQ VPT294 videotapes were used to record the ATC sessions. The complete configuration of the apparatus is presented in Figure 6 where the simulator is seen on the left and the data recording equipment is seen on the right.

Procedure

Test administration. Participants who voluntarily registered for the investigation collectively arrived at a common experiment site (a classroom in the Psychology Department) for the purpose of orientation and completion of the ATCS Written Exam. The test administration took approximately one hour. Participants were told that they would be notified if they scored high enough to be asked back for the remainder of the investigation. Participants were notified about three days later. Those that did

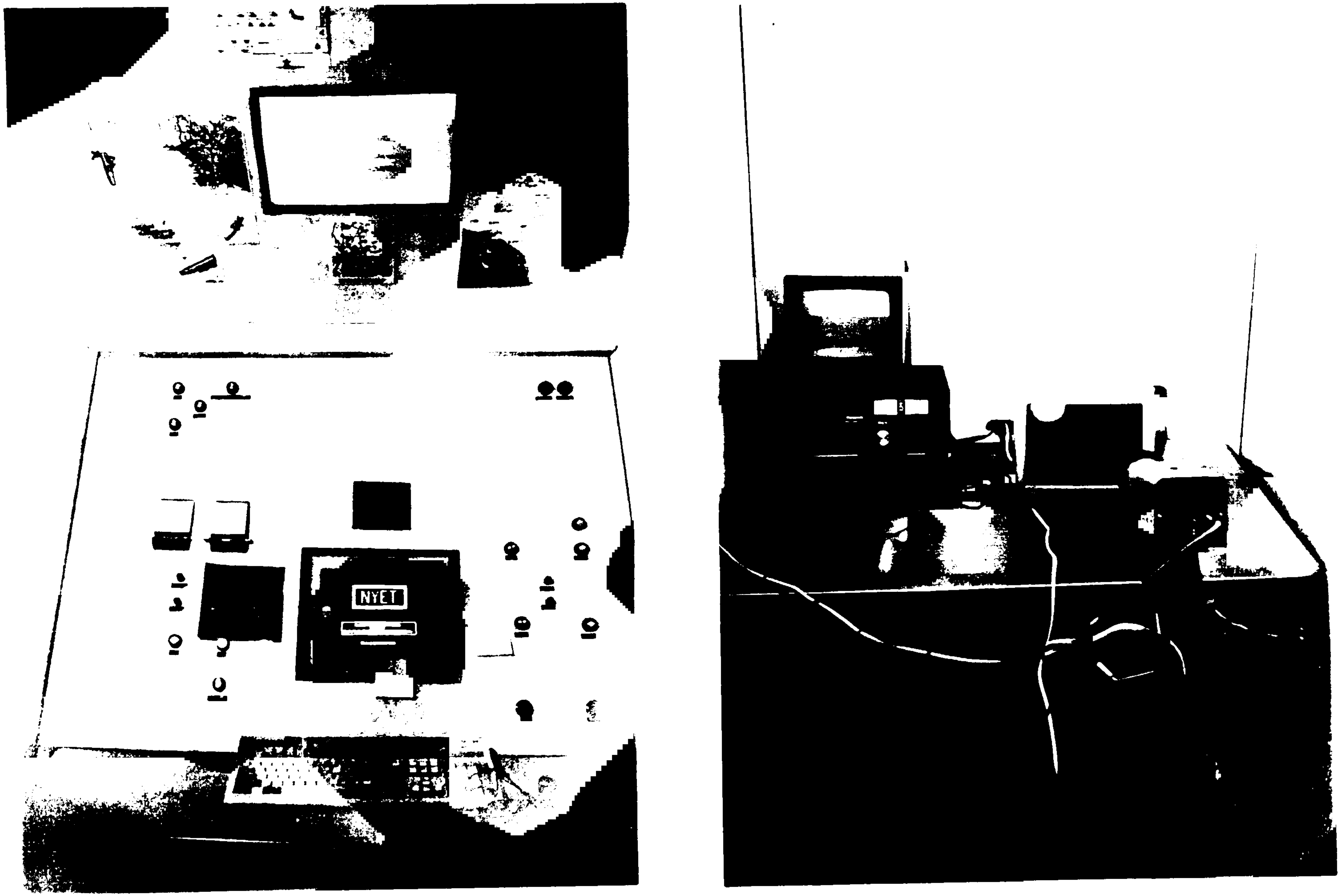

Figure 6
Equipment Configuration for Experiment Site

not qualify were given one course credit for their one hour spent. Those participants that did qualify were scheduled for training on TRACON and were subsequently referred to as Developmentals (the term used to refer to ATCS trainees).

Training. Since the Zenith laboratory contained only 19 computers, three training sessions were needed to train all participants selected for the investigation. The Zenith laboratory was reserved for three two-hour training sessions and participants who qualified were asked to arrive at the appointed time. During the training session, further orientation was given and participants were trained by first watching a complete demonstration of TRACON. Participants were then informed of what they were to do in the scenario. Each participant then worked three brief (about five minutes each) scenarios designed to illustrate the different types of flights that can enter their sector. After the three scenarios were completed, participants worked a scenario designed to integrate the components of the first three scenarios. This scenario lasted approximately 15 minutes. All participants worked the same scenario. Following this procedure, participants were scheduled for their actual shifts.

Experimental site. Each participant arrived at the experimental site at the appointed time. The experimenter briefed each participant as to the general purpose of the investigation and the procedure of the day's sessions and had each participant sign a Participant Consent Form. Participants were then asked to leave all materials outside of the experimental cubicle for the duration of their session. A participant was then seated inside the simulator and was first administered the computer game, NYET. Each participant was permitted three practice games and the score was recorded on the fourth game. Since the initial training sessions and the actual data collection sessions took place on different days, a refresher one-on-one training session was administered to each participant. This refresher training session required each participant to first watch the experimenter handle the scenario, then work the scenario with the experimenter overseeing the session and then work the scenario on their own without the experimenter's aid. The composite scenario experienced in the collective training session several days earlier was used for the refresher training session. The cumulative score for the session was recorded and assessed as a percentage of the total score possible for the session. In most cases, the participants had met the criterion for no further training needed (80%). In those few cases where the criterion was not met, further training was attempted. If additional training was not beneficial, the participant was dismissed from the remainder of the experiment and given the appropriate debriefing and course credit. Four participants (all female) were dismissed for this specific reason. They were replaced by four new participants who went through the same training procedure. Each participant was then taken out of the simulator and given a brief participant quiz to assess their knowledge of TRACON. This provided the experimenter with the opportunity to clarify specific issues of procedure with each participant.

Each participant was then taken back into the simulator and seated at the radarscope. The session began when the experimenter instructed the ATCS developmental to press the enter key. A session lasted approximately one hour. A session began with a five-minute warm-up period of 100% en route flights. About five minutes into the session, the mix of the four types of flights became more heterogeneous and continued toward maximum heterogeneity at approximately 15 to 20 minutes into the session. This level of heterogeneity was maintained for the next 20 minutes. After approximately 20 minutes of maximum heterogeneity, the traffic mix became a homogeneous of en route flights again. In the CUE treatment, this event was immediately preceded by a verbal and visual cue coming through the simulator intercom system notifying the participant of the forthcoming easing of air traffic complexity. The homogeneous traffic mix remained constant for the remainder of the session. All sessions were stopped at the end of one hour in order to standardize length of session. At the end of the session, participants were given a half hour rest period. Subsequent to the reset period, the second session was administered. Subsequent to the end of the second session, participants were debriefed and a detailed explanation of the purpose of the investigation was given. Participants were notified that upon request they could obtain the results of the investigation.

Response measures. The work load for each participant was assessed using the primary task loading method. The primary task method assesses the number of aircraft worked simultaneously as a function of task demand (defined by the degree of heterogeneity of the aircraft flightpaths) per five-minute interval. This was represented as the number of aircraft flight strips in the active flight strips bin. Five types of OEs were of interest: Handoff, Keying, Memory, Navigational, and Pilot.

CHAPTER III
RESULTS

This section presents the results that are relevant to the proposed hypotheses for the data collected. The reader is referred to appendix B for the results of the pilot investigation.

For the sake of parsimony, the results of the investigation will be discussed with regard to the OEs classification scheme defined in the methodology. Since Handoff OEs, Keying OEs, Navigational OEs, and Pilot OEs exhibited the same pattern of results, they will be discussed collectively. Memory OEs displayed a different pattern of results and will be discussed separately.

Coding of Data

The raw data generated in this investigation consisted of VHS videotapes containing the sessions of each participant. Each videotape contained the two one-hour sessions for each participant. Five major types of OEs were identified prior to videotape review. In order to quantify the raw data for which descriptive and inferential analyses could be conducted, the researcher reviewed and transcribed the 90 hours of videotape over a period of four months. The data was entered into the Microsoft Excel spreadsheet package and exported to SPSS/PC+ for statistical analysis.

As a result of this task being, to a certain extent, self-paced, the 45 participants peaked in their work load levels at different time periods. In order to strategically deal with this potential problem, a synchronization of participants' data was performed. The five-minute time period during which peak work load occurred for each participant was identified and aligned with the peak work load five-minute time period for all other participants. That five-minute time period as well as the ten minutes prior and the ten minutes subsequent to that time period were included in the analyses. Thus, the total amount of time for each participant's session that was analyzed was 25 minutes. The beginning time periods of the sessions were not analyzed since they were considered to be warm-up periods for the participants.

Analysis of OEs

The means and standard deviations for the five types of OEs are presented in Tables 1 through 5. Hypothesis I predicted that if hysteresis was due to participants' inability to timely perceive a change in the work load and hence adjust their expectations accordingly in a timely fashion, then the effect would occur for any downward shift in work load. In this investigation, this refers to both high to low and moderate to low work load shifts. The effect would be attenuated through the use of a cue that would inform the operator of the forthcoming change in task demand. Hypothesis II predicted that if hysteresis was due to participants' STM being overloaded, then the effect would occur for high to low work load only, since under moderate to low work load STM would never be overloaded. Also, the use of a cue to inform the operator of the forthcoming change in task demand should not have an impact on the hysteresis effect.
To test these hypotheses, *a priori* orthogonal contrasts were employed, followed by Analyses of Variance and *post hoc* Tukey Tests of Significance. The independent variables were type of WORK LOAD (high to low, moderate to low, high to high), CUE (cue and no cue), AND TIME (t1, t2, t3, t4, and t5) (five-minute intervals). For the sake of parsimony, Figures 7 through 12 illustrate the task loads that participants were working under across time for each of the three WORK LOAD groups.

Handoff, Keying, Navigational, and Pilot OEs. *A priori* orthogonal contrasts were employed to test the main effects that the hysteresis effect is due either to perseverance of

Table 1
Means and Standard Deviations of Handoff OEs

		CUE T1	CUE T2	CUE T3	CUE T4	CUE T5	NO CUE T1	NO CUE T2	NO CUE T3	NO CUE T4	NO CUE T5
H-L	M	5.2	4.4	5.07	5.27	1.73	5.2	5.33	4.87	0.4	0.93
	SD	1.21	1.4	1.1	2.79	1.22	2.21	1.48	2.17	0.08	0.03
M-L	M	4.67	4.4	4.27	4.33	0.4	4.4	4	4.33	1.27	0.53
	SD	1.11	2.03	2.22	1.29	0.05	0.83	2.1	2.16	0.62	0.08
H-H	M	5.13	4.6	4.8	4.87	5.07	4.87	4.53	4.73	5.27	5.33
	SD	2.77	0.51	1.01	0.74	0.88	2.92	2.85	2.89	2.02	3.09

Table 2
Means and Standard Deviations of Keying OEs

		CUE T1	CUE T2	CUE T3	CUE T4	CUE T5	NO CUE T1	NO CUE T2	NO CUE T3	NO CUE T4	NO CUE T5
H-L	M	21.1	21.7	22.5	21.5	3.13	21.4	21.1	22.1	3.2	3
	SD	1.44	1.92	2.88	2	3.27	1.06	2.75	3.31	1.01	1.51
M-L	M	11.7	10.3	13.8	13	1.73	11.5	12.9	12.5	1.53	0.2
	SD	2.58	0.98	1.37	1.46	0.67	0.92	1.55	0.52	0.13	0.04
H-H	M	20.5	21	21.8	22.5	21.9	20.5	19.8	21.4	21	20.7
	SD	2.1	3.95	1.82	1.55	3.56	2.2	3.32	3.09	3.46	1.28

Table 3
Means and Standard Deviations of Memory OEs

		CUE T1	CUE T2	CUE T3	CUE T4	CUE T5	NO CUE T1	NO CUE T2	NO CUE T3	NO CUE T4	NO CUE T5
H-L	M	0.07	0.17	0.13	6.77	0.97	0.12	0.23	0.12	6.63	1.12
	SD	0.01	0.1	0.1	1.9	0.9	0.41	0.44	0.5	1.8	1.4
M-L	M	0.33	0.5	0.33	0.13	0.07	0.33	0.45	0.45	0.12	0.12
	SD	0.5	0.5	0.5	0.4	0.01	0.5	0.73	0.72	0.41	0.4
H-H	M	0	0.07	0.4	0.37	0.97	0	0.12	0.12	0.3	0.31
	SD	0	0.01	1.9	0.9	0.4	0	0.5	0.4	1.3	0.52

Table 4
Means and Standard Deviations of Navigational OEs

		CUE T1	CUE T2	CUE T3	CUE T4	CUE T5	NO CUE T1	NO CUE T2	NO CUE T3	NO CUE T4	NO CUE T5
H-L	M	8.33	7.2	7.87	8.07	2.13	8.27	8.13	7.93	1.33	1.8
	SD	1.63	1.82	2.5	1.87	1.77	1.79	0.83	0.8	0.49	1.52
M-L	M	4.4	4.53	4.13	4.07	0.07	4	4.33	3.87	0.87	0.07
	SD	0.51	1.06	1.3	2.34	0.02	2.59	0.82	2.5	0.19	0
H-H	M	6.87	7.47	8.47	8.6	9.2	7.87	7.4	8.13	7.73	8.2
	SD	1.96	2.03	1.64	2.72	3.39	2.42	2.35	1.73	1.79	0.78

Table 5
Means and Standard Deviations of Pilot OEs

		CUE T1	CUE T2	CUE T3	CUE T4	CUE T5	NO CUE T1	NO CUE T2	NO CUE T3	NO CUE T4	NO CUE T5
H-L	M	6.67	6	6.13	5.73	0.87	6.07	5.6	6.13	0.8	0.93
	SD	0.82	1.46	2	2.31	0.36	1.98	0.74	2.56	0.07	1.28
M-L	M	2.47	2.13	2.33	2.73	0.35	2	2.4	2.2	0.4	0.07
	SD	1.77	1.41	1.18	2.28	0.09	1.13	1.5	1.27	0.06	0.02
H-H	M	4.67	4.87	5.4	5.93	5.73	4	4.4	5.07	4.8	5.33
	SD	0.9	0.92	1.24	2.22	0.27	1.47	1.88	0.8	1.21	1.18

expectancies or to STM overload. Specifically, a comparison of the means for the three time periods (t1, t2, and t3) versus the mean of t4 (the five-minute time period immediately subsequent to the shift in work load) was conducted for each work load group.

In an *a priori* test of the perseverance of expectancies hypothesis and the STM overload hypothesis, the results revealed that for Handoff, Keying, Navigational, and Pilot OEs at the no cue treatment, the high to low work load group and the moderate to low work load group did not reveal significant differences between the means of t1, t2, and t3 versus t4. This suggests the presence of the type of hysteresis effect that displays a lag between work load reduction and OE rate reduction. The mean number of OEs, however, for high to low and moderate to low work load groups at the cue treatment at t1, t2, and t3 were significantly different from the mean of OEs at t4 (F (1, 168) = 4.68, $p<0.001$). This indicates that, all else being held constant, CUE was effective in attenuating the hysteresis effect for both high to low and moderate to low work load groups. Since the lag occurred for both high to low and moderate to low work load groups and was subsequently attenuated for both of those groups, it can be concluded at this point that there is initial support for the perseverance of inaccurate expectancies hypothesis for hysteresis.

Separate 3 X (2 X 5) split plot analyses of variance were subsequently performed on the data representing Handoff, Keying, Navigational, and Pilot OEs to probe for additional evidence in support of one or the other hypothesis. Prior to ANOVA initiation, an evaluation of assumptions of normality, homogeneity of variance, skewness, and kurtosis was conducted. Results of these evaluations were deemed not to present a problem toward the straightforward analysis of the data by ANOVA. The omnibus ANOVA procedure for Handoff, Keying, Navigational, and Pilot OEs (Tables 6 through 9, respectively) revealed significant main effects for WORK LOAD, CUE and TIME. The main effects for WORK LOAD, CUE and TIME as well as the two way interactions: WORK LOAD X CUE, WORK

Table 6
Omnibus ANOVA: Handoff OEs

Source	SS	df	MS	F
Between Groups	11576.97	44		
LOAD	212.92	2	106.46	29.27*
S w. LOAD	152.76	42	3.64	
Within Groups	16709.80	405		
CUE	33.62	1	33.62	12.26*
LOAD X CUE	21.72	2	10.86	3.96*
S w. CUE X LOAD	115.16	42	2.74	
TIME	409.61	4	102.40	28.43*
LOAD X TIME	290.15	8	36.27	10.07*
S w. LOAD X TIME	605.04	168	3.60	
CUE X TIME	110.01	4	27.50	8.26*
LOAD X CUE X TIME	98.68	8	12.34	3.71*
S w. CUE X TIME X LOAD	559.31	168	3.33	
Total	28286.77	449		

* $p < .001$

Table 7
Omnibus ANOVA: Keying OEs

Source	SS	df	MS	F
Between Groups	11576.97	44		
LOAD	11276.78	2	5638.39	788.88*
S w. LOAD	300.19	42	7.15	
Within Groups	16709.80	405		
CUE	619.52	1	619.52	134.10*
LOAD X CUE	165.85	2	82.93	17.95*
S w. CUE X LOAD	194.03	42	4.62	
TIME	6802.70	4	1700.67	325.61*
LOAD X TIME	4501.62	8	562.70	107.73*
S w. LOAD X TIME	877.48	168	5.22	
CUE X TIME	1856.41	4	464.10	110.31*
LOAD X CUE X TIME	985.35	8	123.17	29.27*
S w. CUE X TIME X LOAD	706.84	168	4.21	
Total	28286.77	449		

* $p < .001$

Table 8
Omnibus ANOVA: Navigational OEs

Source	SS	df	MS	F
Between Groups	2065.84	44		
LOAD	1880.32	2	940.16	212.84*
S w. LOAD	185.52	42	4.42	
Within Groups	2954.60	405		
CUE	65.74	1	65.74	30.22*
LOAD X CUE	17.90	2	8.95	4.11*
S w. CUE X LOAD	91.36	42	2.18	
TIME	668.36	4	167.09	55.81*
LOAD X TIME	660.23	8	82.53	27.56*
S w. LOAD X TIME	503.01	168	2.99	
CUE X TIME	232.84	4	58.21	16.75*
LOAD X CUE X TIME	131.32	8	16.42	4.72*
S w. CUE X TIME X LOAD	583.84	168	3.48	
Total	5020.44	449		

* $p < .001$

Table 9
Omnibus ANOVA: Pilot OEs

Source	SS	df	MS	F
Between Groups	1063.40	44		
LOAD	963.29	2	481.65	202.08*
S w. LOAD	100.11	42	2.38	
Within Groups	1864.60	405		
CUE	67.28	1	67.28	32.76*
LOAD X CUE	9.05	2	4.53	2.20*
S w. CUE X LOAD	86.27	42	2.05	
TIME	339.53	4	84.88	37.30*
LOAD X TIME	466.77	8	58.35	25.64*
S w. LOAD X TIME	382.29	168	2.28	
CUE X TIME	118.48	4	29.62	14.49*
LOAD X CUE X TIME	51.46	8	6.43	3.15*
S w. CUE X TIME X LOAD	343.47	168	2.04	
Total	2928.00	449		

* $p < .001$

LOAD X TIME, and CUE X TIME were not subject to parsimonious interpretation, however, in the presence of the significant WORK LOAD X CUE X TIME three-way interaction (F (8, 168), $p<0.001$ (Handoff OEs: 3.71; Keying OEs: 29.27; Navigational OEs: 4.72; Pilot OEs: 3.15)). Therefore, the results of the *a prioi* tests of the main effects must be qualified.

Subsequent analyses were conducted to determine the locus of the significant three-way interaction. An analysis of simple two-way interactions (Keppel, 1982) for the significant three-way interaction revealed a significant WORK LOAD X TIME interaction at each treatment of CUE (F (8, 68), $p<0.001$ (Handoff OEs @ no cue: 7.23; Handoff OEs @ cue: 6.92; Keying OEs @ no cue: 52.73; Keying OEs @ cue: 98.25; Navigational OEs @ no cue: 12.42; Navigational OEs @ cue: 19.19; Pilot OEs @ no cue: 12.06; Pilot OEs @ cue: 18.29) (Tables 10 though 17). These significant two-way interactions indicate that performance differed at each treatment of CUE depending upon the type of work load and the particular time period. Which loads were significantly different from each other and which time periods were significantly different from each other was the focus of subsequent analyses.

Simple main effects analyses were conducted to further uncover the locus of the significant WORK LOAD X TIME interaction at each treatment of CUE (Tables 18 through 25). Of particular import to hypothesis I was whether the hysteresis effect was ameliorated by cueing. Of import to hypothesis II was whether the hysteresis effect occurred for high to high work load only or for moderate to low work load as well. It is important to recall that of the five time periods included in the analysis, peak work load occurred at t3. Immediately subsequent to t3 work load dropped sharply. Results indicate that within treatments of work load, the mean number of OEs across time periods was not all equal (F (4, 168), $p<0.001$ (Handoff OEs @ high to low no cue: 15.35; Handoff OEs @ moderate to low no cue: 22.36; Handoff OEs @ high to low cue: 19.10; Handoff OEs @ moderate to low cue: 10.88; Keying

Table 10
Simple Interactions ANOVA: Handoff OEs @ no cue TREATMENT

Source	SS	df	MS	F
Between Groups	5383.85	44		
LOAD	61.76	2	30.88	9.03*
S w. LOAD	143.60	42	3.42	
Within Groups	6496.01	180		
TIME	205.54	4	51.38	3.56*
LOAD X TIME	126.06	8	15.76	7.23*
S w. LOAD X TIME	366.40	168	2.18	
Total	11879.86	224		

* $p < .001$

Table 11
Simple Interactions ANOVA: Handoff OEs @ cue TREATMENT

Source	SS	df	MS	F
Between Groups	5383.85	44		
LOAD	172.88	2	86.44	29.20*
S w. LOAD	124.32	42	2.96	
Within Groups	6496.01	180		
TIME	314.09	4	78.52	16.53*
LOAD X TIME	262.76	8	32.85	6.92*
S w. LOAD X TIME	797.95	168	4.75	
Total	11879.86	224		

* $p < .001$

Table 12
Simple Interactions ANOVA: Keying OEs @ no cue TREATMENT

Source	SS	df	MS	F
Between Groups	5383.85	44		
LOAD	5131.05	2	2565.52	426.23*
S w. LOAD	252.80	42	6.02	
Within Groups	6496.01	180		
TIME	3377.32	4	844.33	159.69*
LOAD X TIME	2230.42	8	278.80	52.73*
S w. LOAD X TIME	888.27	168	5.29	
Total	11879.86	224		

* $p < .001$

Table 13
Simple Interactions ANOVA: Keying OEs @ cue TREATMENT

Source	SS	df	MS	F
Between Groups	5383.85	44		
LOAD	6311.58	2	3155.79	549.03*
S w. LOAD	241.41	42	5.75	
Within Groups	6496.01	180		
TIME	5281.80	4	1320.45	318.70*
LOAD X TIME	3256.55	8	407.07	98.25*
S w. LOAD X TIME	696.05	168	4.14	
Total	11879.86	224		

* $p < .001$

Table 14
Simple Interactions ANOVA: Navigational OEs @ no cue TREATMENT

Source	SS	df	MS	F
Between Groups	1038.64	44		
LOAD	865.52	2	432.76	104.99*
S w. LOAD	173.12	42	4.12	
Within Groups	1298.41	180		
TIME	303.62	4	75.90	20.40*
LOAD X TIME	369.64	8	46.20	12.42*
S w. LOAD X TIME	625.15	168	3.72	
Total	2337.05	224		

* $p < .001$

Table 15
Simple Interactions ANOVA: Navigational OEs @ cue TREATMENT

Source	SS	df	MS	F
Between Groups	1136.46	44		
LOAD	1032.70	2	516.35	209.01*
S w. LOAD	103.76	42	2.47	
Within Groups	1481.20	180		
TIME	597.57	4	149.39	54.36*
LOAD X TIME	421.92	8	52.74	19.19*
S w. LOAD X TIME	461.71	168	2.75	
Total	2617.66	224		

* $p < .001$

Table 16
Simple Interactions ANOVA: Pilot OEs @ no cue TREATMENT

Source	SS	df	MS	F
Between Groups	633.36	44		
LOAD	527.04	2	263.52	104.10*
S w. LOAD	106.32	42	2.53	
Within Groups	806.40	180		
TIME	202.87	4	50.72	22.22*
LOAD X TIME	220.12	8	27.51	12.06*
S w. LOAD X TIME	383.41	168	2.28	
Total	1439.76	224		

* $p < .001$

Table 17
Simple Interactions ANOVA: Pilot OEs @ cue TREATMENT

Source	SS	df	MS	F
Between Groups	525.36	44		
LOAD	445.31	2	222.65	116.82*
S w. LOAD	80.05	42	1.91	
Within Groups	895.61	180		
TIME	255.14	4	63.78	31.30*
LOAD X TIME	298.12	8	37.26	18.29*
S w. LOAD X TIME	342.35	168	2.04	
Total	1420.97	224		

* $p < .001$

Table 18
Simple Main Effects: Handoff OEs @ no cue TREATMENT

Source	SS	df	MS	F
LOAD	61.76	2	30.88	9.03*
TIME	205.54	4	51.38	23.56*
LOAD X TIME	126.06	8	15.76	7.23*
LOAD @ t1	2.53	2	1.27	.37
LOAD @ t2	.40	2	.20	.09
LOAD @ t3	4.98	2	2.49	1.04
LOAD @ t4	6.58	2	3.29	.99
LOAD @ t5	173.33	2	86.67	102.63*
TIME @ h-l load	133.87	4	33.47	15.35*
TIME @ m-l load	194.99	4	48.75	22.36*
TIME @ h-h load	2.75	4	0.69	0.32
S w. TIME X LOAD	366.40	168	2.18	
Total	11879.86	224		

* $p < .001$

Table 19
Simple Main Effects: Handoff OEs @ cue TREATMENT

Source	SS	df	MS	F
LOAD	172.88	2	86.44	29.20*
TIME	314.09	4	78.52	16.53*
LOAD X TIME	262.76	8	32.85	6.92*
LOAD @ t1	4.84	2	2.42	0.51
LOAD @ t2	13.51	2	6.76	1.38
LOAD @ t3	2.31	2	1.16	0.20
LOAD @ t4	202.18	2	101.09	41.03*
LOAD @ t5	212.80	2	106.40	26.60*
TIME @ h-l load	362.99	4	90.75	19.10*
TIME @ m-l load	206.75	4	51.69	10.88*
TIME @ h-h load	7.12	4	1.78	0.37
S w. TIME X LOAD	797.95	168	4.75	
Total	11879.86	224		

* $p < .001$

Table 20
Simple Main Effects: Keying OEs @ no cue TREATMENT

Source	SS	df	MS	F
LOAD	5131.05	2	2565.52	426.23*
TIME	3377.32	4	844.33	159.69*
LOAD X TIME	2230.42	8	278.80	52.73*
LOAD @ t1	836.31	2	418.16	95.45*
LOAD @ t2	1213.33	2	606.67	90.14*
LOAD @ t3	704.04	2	352.02	78.42*
LOAD @ t4	816.53	2	408.27	143.53*
LOAD @ t5	3791.24	2	1895.62	217.29*
TIME @ h-l load	4153.52	4	1038.38	196.29*
TIME @ m-l load	1419.15	4	354.79	67.07*
TIME @ h-h load	35.07	4	8.77	1.66
S w. TIME X LOAD	888.27	168	5.29	
Total	11879.86	224		

* $p < .001$

Table 21
Simple Main Effects: Keying @ cue TREATMENT

Source	SS	df	MS	F
LOAD	6311.58	2	3155.79	549.03*
TIME	5281.80	4	1320.45	318.70*
LOAD X TIME	3256.55	8	407.07	98.25*
LOAD @ t1	908.13	2	454.07	200.60*
LOAD @ t2	590.93	2	295.47	42.23*
LOAD @ t3	856.58	2	428.29	61.80*
LOAD @ t4	3492.84	2	1746.42	366.50*
LOAD @ t5	3719.64	2	1859.82	1362.43*
TIME @ h-l load	6139.28	4	1534.82	370.73*
TIME @ m-l load	2377.79	4	594.45	143.59*
TIME @ h-h load	21.28	4	5.32	1.29
S w. TIME X LOAD	696.05	168	4.14	
Total	11879.86	224		

* $p < .001$

Table 22
Simple Main Effects: Navigational OEs @ no cue TREATMENT

Source	SS	df	MS	F
LOAD	865.52	2	432.76	104.99*
TIME	303.62	4	75.90	20.40*
LOAD X TIME	369.64	8	46.20	12.42*
LOAD @ t1	118.53	2	59.27	26.29*
LOAD @ t2	78.93	2	39.47	13.83*
LOAD @ t3	165.38	2	82.69	23.28*
LOAD @ t4	184.18	2	92.09	16.86*
LOAD @ t5	688.13	2	344.07	70.47*
TIME @ h-l load	404.99	4	101.25	27.22*
TIME @ m-l load	215.55	4	53.89	14.49*
TIME @ h-h load	52.72	4	13.18	3.54*
S w. TIME X LOAD	625.15	168	3.72	
Total	2337.05	224		

* $p < .001$

Table 23
Simple Main Effects: Navigational OEs @ cue TREATMENT

Source	SS	df	MS	F
LOAD	1032.70	2	516.35	209.01*
TIME	597.57	4	149.39	54.36*
LOAD X TIME	421.92	8	52.74	19.19*
LOAD @ t1	166.58	2	83.29	15.85*
LOAD @ t2	121.91	2	60.96	26.48*
LOAD @ t3	173.91	2	86.96	26.39*
LOAD @ t4	441.64	2	220.82	136.39*
LOAD @ t5	550.58	2	275.29	277.05*
TIME @ h-l load	773.42	4	193.36	70.31*
TIME @ m-l load	239.82	4	59.96	21.80*
TIME @ h-h load	6.27	4	1.57	.57
S w. TIME X LOAD	461.71	168	2.75	
Total	2617.66	224		

* $p < .001$

Table 24
Simple Main Effects: Pilot OEs @ no cue TREATMENT

Source	SS	df	MS	F
LOAD	527.04	2	263.52	104.10*
TIME	202.87	4	50.72	22.22*
LOAD X TIME	220.12	8	27.51	12.06*
LOAD @ t1	132.40	2	66.20	43.17*
LOAD @ t2	118.53	2	59.27	35.83*
LOAD @ t3	121.91	2	60.96	26.48*
LOAD @ t4	96.40	2	48.20	9.34*
LOAD @ t5	277.91	2	138.96	137.64*
TIME @ h-l load	339.79	4	84.95	37.26*
TIME @ m-l load	65.41	4	16.35	7.17*
TIME @ h-h load	17.79	4	4.45	1.95
S w. TIME X LOAD	383.41	168	2.28	
Total	1439.76	224		

* $p < .001$

Table 25
Simple Main Effects: Pilot OEs @ cue TREATMENT

Source	SS	df	MS	F
LOAD	445.31	2	222.65	116.82*
TIME	255.14	4	63.78	31.30*
LOAD X TIME	298.12	8	37.26	18.29*
LOAD @ t1	124.04	2	62.02	25.31*
LOAD @ t2	78.40	2	39.20	18.54*
LOAD @ t3	124.13	2	62.07	21.18*
LOAD @ t4	177.60	2	88.80	57.91*
LOAD @ t5	239.24	2	119.62	116.30*
TIME @ h-l load	464.75	4	116.19	56.95*
TIME @ m-l load	71.66	4	17.92	8.78*
TIME @ h-h load	16.85	4	4.21	2.06
S w. TIME X LOAD	342.35	168	2.04	
Total	1439.76	224		

* $p < .001$

OEs @ high to low no cue: 196.29; Keying OEs @ moderate to low no cue: 67.07; Keying OEs @ high to low cue: 370.73; Keying OEs @ moderate to low cue: 143.59; Navigational OEs @ high to low no cue: 27.22; Navigational OEs @ moderate to low no cue: 14.49; Navigational @ high to low cue: 70.31; Navigational @ moderate to low cue: 21.80; Pilot OEs @ high to low no cue: 37.26; Pilot OEs @ moderate to low no cue: 7.17; Pilot OEs @ high to low cue: 56.95; Pilot OEs @ moderate to low cue: 8.78)).

As a final step in this hierarchical analyses procedure, a Tukey procedure was undertaken to determine which pairs of means differed significantly from each other (Tables 26 through 29). Of import to the hypothesis was whether the hysteresis effect occurred when no cue was administered and did not occur when CUE was administered. Also of import was whether the hysteresis effect was present for both high to low and moderate to low work load or for high to low work load only. Results indicate that for both high to low and moderate to low work loads, Handoff, Keying, Navigational, and Pilot OEs in the no cue treatment revealed no significant differences between the means of t3 and t4 even though work load shifted downward between t3 and t4. Significant differences were found, however, between the mean of t5 (ten minutes subsequent to the shift occurrence) and the means of the other 4 time periods (HSD = 3.96). These results indicate that a lag existed for OE reduction even after work load shifted downward. Results for high to low and moderate to low work loads at the cue treatment, however, reveal significant differences between the means of t3 and t4. This indicates the absence of any lag between work load reduction and OE reduction. That is, OEs reduction was proportionate to work load reduction. In the high to high (no shift) treatment, no significant differences were revealed between the means of the different time periods for either the cue or no cue treatments indicating the absence of any hysteresis effect, or any fatigue effect, for that matter. These results are presented in Figures 13 through 20.

Table 26
Tukey Procedure: Handoff OEs

	CUE													
	no cue							cue						
			t5	t2	t3	t1	t4			t5	t4	t3	t1	t2
	1.73	t5						0.93	t5					
	4.40	t2	*					1.60	t4					
h-l	5.07	t3	*					4.87	t3	*	*			
	5.20	t1	*					5.20	t1	*	*			
	5.27	t4	*					5.33	t2	*	*			
			t5	t3	t4	t2	t1			t5	t4	t2	t3	t1
	0.40	t5						0.53	t5					
	4.27	t3	*					1.27	t4					
LOAD m-l	4.33	t4	*					4.00	t2	*	*			
	4.40	t2	*					4.33	t3	*	*			
	4.67	t1	*					4.40	t1	*	*			
			t2	t3	t4	t5	t1			t2	t3	t1	t4	t5
	4.60	t2						4.53	t2					
	4.80	t3						4.73	t3					
h-h	4.87	t4						4.87	t1					
	5.07	t5						5.27	t4					
	5.13	t1						5.33	t5					

Table 27
Tukey Procedure: Keying OEs

CUE

LOAD h-l, no cue

		t5	t1	t4	t2	t3
3.13	t5					
21.07	t1	*				
21.53	t4	*				
21.67	t2	*				
22.53	t3	*				

LOAD h-l, cue

		t5	t4	t2	t1	t3
3.00	t5					
3.20	t4					
21.13	t2	*	*			
21.40	t1	*	*			
22.13	t3	*	*			

LOAD m-l, no cue

		t5	t2	t1	t4	t3
1.73	t5					
10.33	t2	*				
11.67	t1	*				
13.00	t4	*	*			
13.80	t3	*	*	*		

LOAD m-l, cue

		t5	t4	t1	t3	t2
0.20	t5					
1.53	t4	*				
11.47	t1	*	*			
12.53	t3	*	*	*		
12.87	t2	*	*	*		

LOAD h-h, no cue

		t1	t2	t3	t5	t4
20.53	t1					
21.00	t2					
21.80	t3					
21.87	t5					
22.47	t4					

LOAD h-h, cue

		t2	t1	t5	t4	t3
19.80	t2					
20.53	t1					
20.73	t5					
21.00	t4					
21.40	t3					

Table 28
Tukey Procedure: Navigational OEs

CUE

LOAD h-l

no cue		t5	t2	t3	t4	t1	cue		t4	t5	t3	t2	t1
2.13	t5						1.33	t4					
7.20	t2	*					1.80	t5					
7.87	t3	*					7.93	t3	*	*			
8.07	t4	*					8.13	t2	*	*			
8.33	t1	*					8.27	t1	*	*			

LOAD m-l

no cue		t5	t4	t3	t1	t2	cue		t5	t4	t3	t1	t2
0.07	t5						0.07	t5					
4.07	t4	*					0.87	t4					
4.13	t3	*					3.87	t3	*	*			
4.40	t1	*					4.00	t1	*	*			
4.53	t2	*					4.33	t2	*	*			

LOAD h-h

no cue		t1	t2	t3	t4	t5	cue		t2	t4	t1	t3	t5
6.87	t1						7.40	t2					
7.47	t2						7.73	t4					
8.47	t3						7.87	t1					
8.60	t4						8.13	t3					
9.20	t5						8.20	t5					

Table 29
Tukey Procedure: Pilot OEs

CUE

LOAD h-l

no cue

		t5	t4	t2	t3	t1
0.87	t5					
5.73	t4	*				
6.00	t2	*				
6.13	t3	*				
6.67	t1	*				

cue

		t4	t5	t2	t1	t3
0.80	t4					
0.93	t5					
5.60	t2	*	*			
6.07	t1	*	*			
6.13	t3	*	*			

LOAD m-l

no cue

		t5	t2	t3	t1	t4
0.13	t5					
2.13	t2	*				
2.33	t3	*				
2.47	t1	*				
2.73	t4	*				

cue

		t5	t4	t1	t3	t2
0.07	t5					
0.40	t4					
2.00	t1	*	*			
2.20	t3	*	*			
2.40	t2	*	*			

LOAD h-h

no cue

		t1	t2	t3	t5	t4
4.67	t1					
4.87	t2					
5.40	t3					
5.73	t5					
5.93	t4					

cue

		t1	t2	t4	t3	t5
4.00	t1					
4.40	t2					
4.80	t4					
5.07	t3					
5.33	t5					

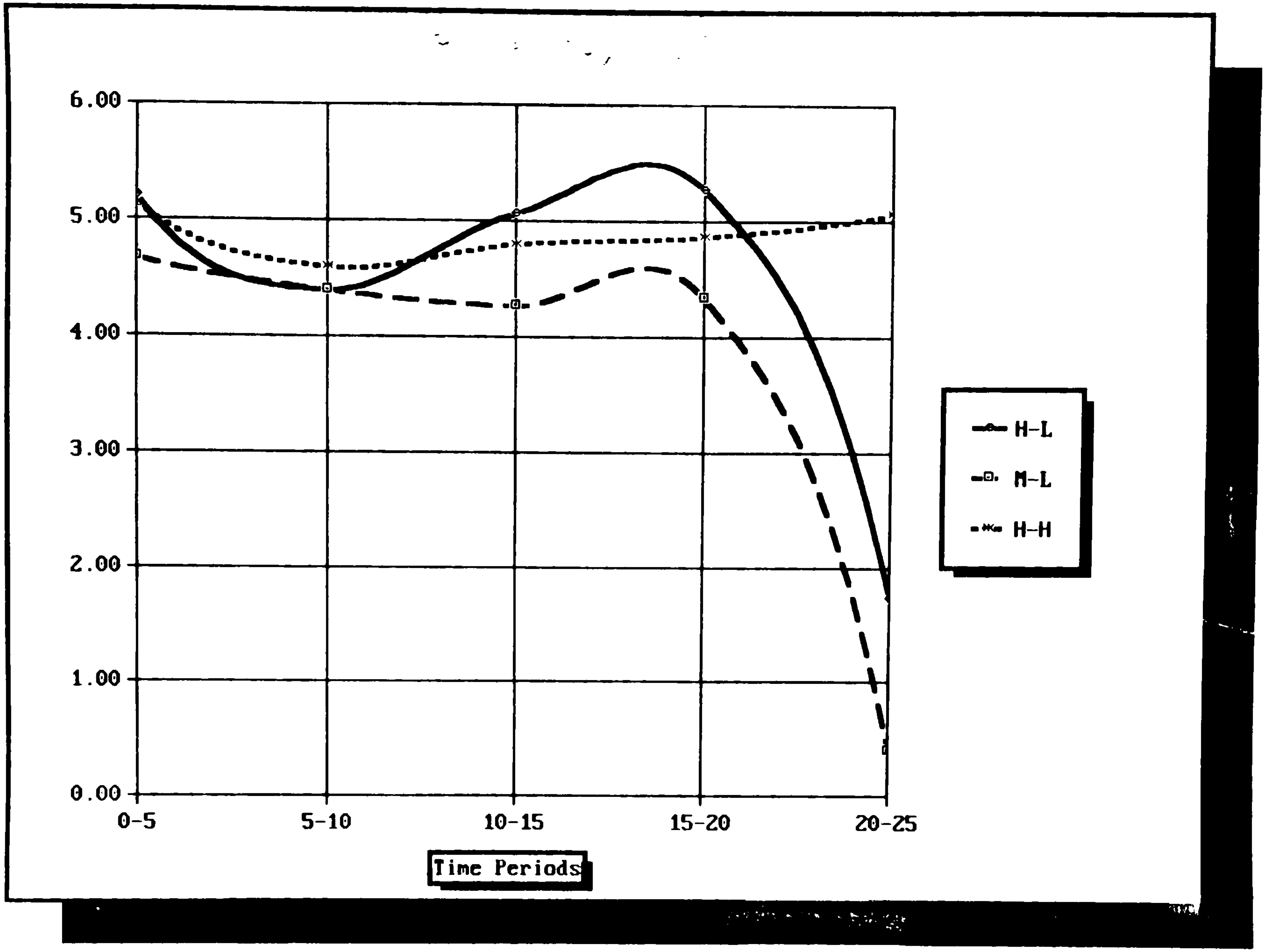

Figure 13
Handoff OEs by Work Load: no cue

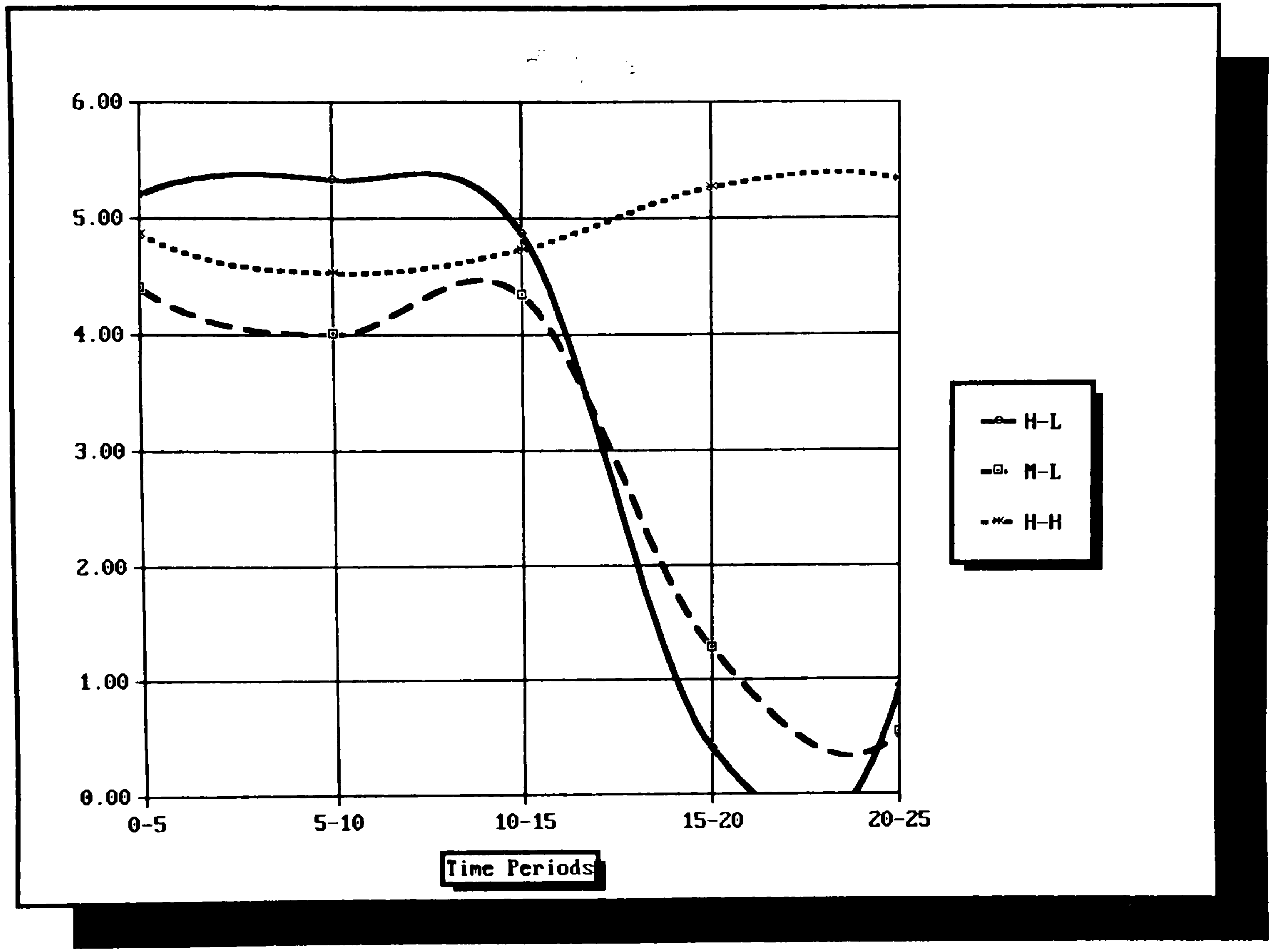

Figure 14
Handoff OEs by Work Load: cue

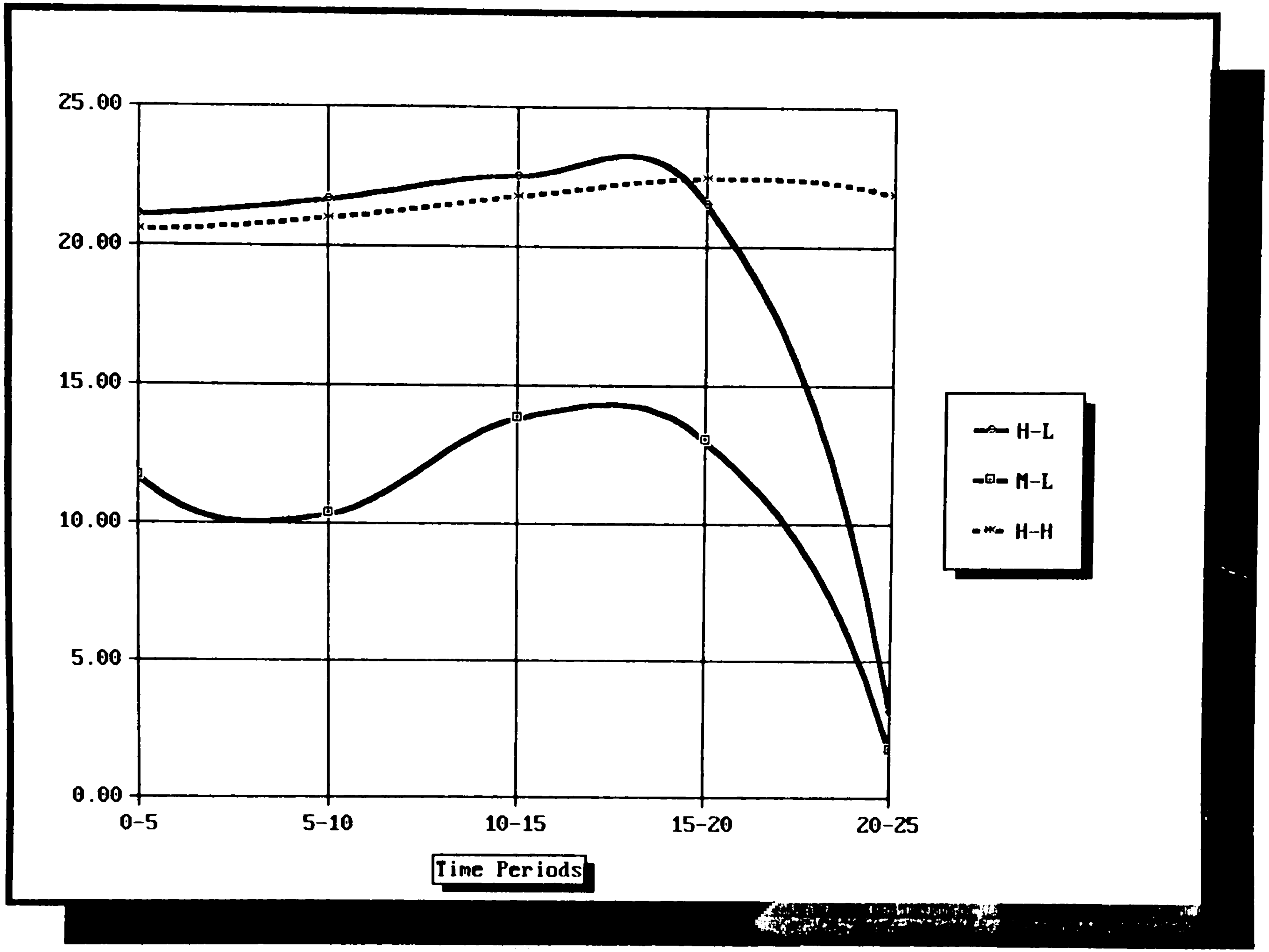

Figure 15
Keying OEs by Work Load: no cue

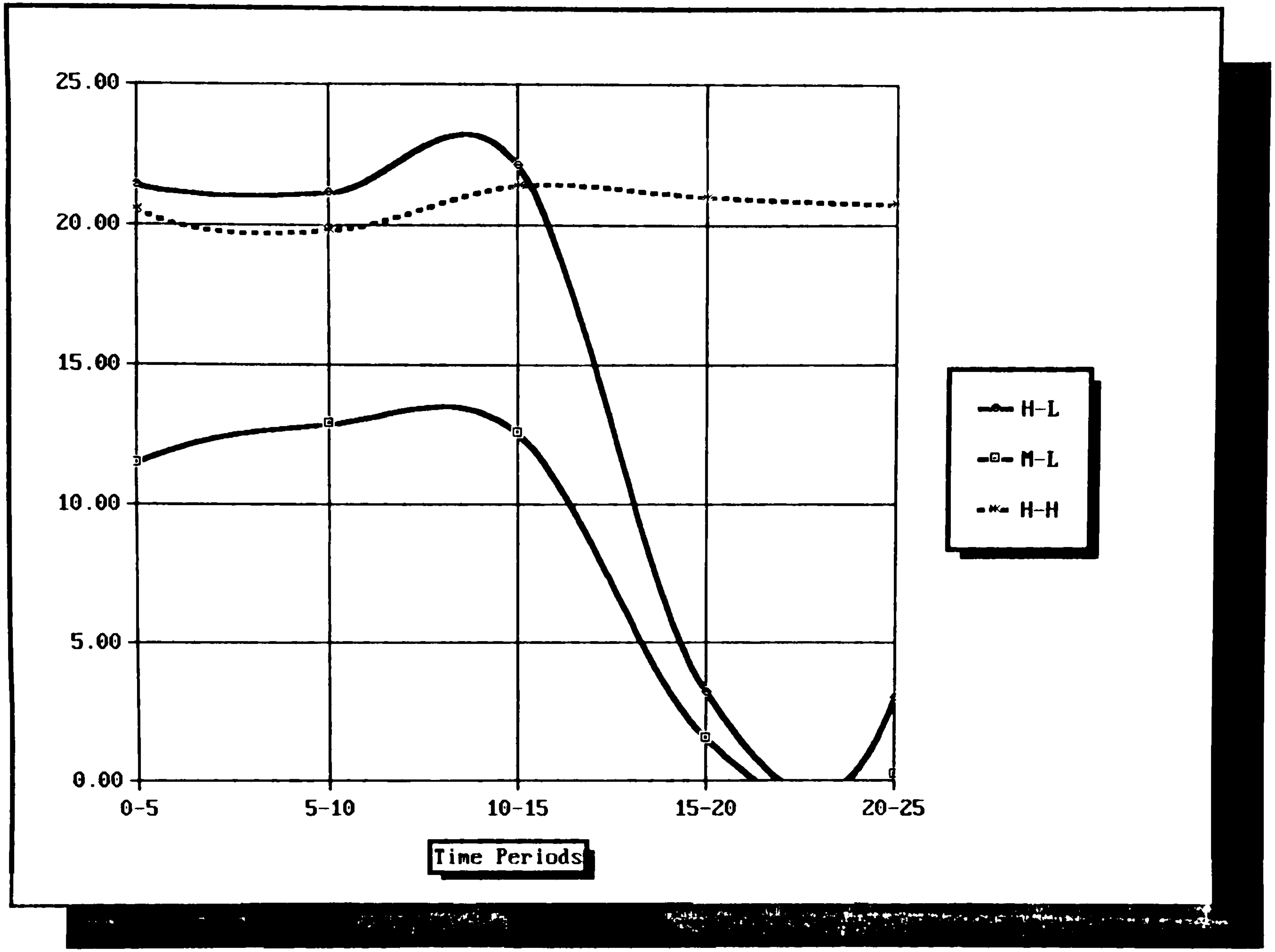

Figure 16
Keying OEs by Work Load: cue

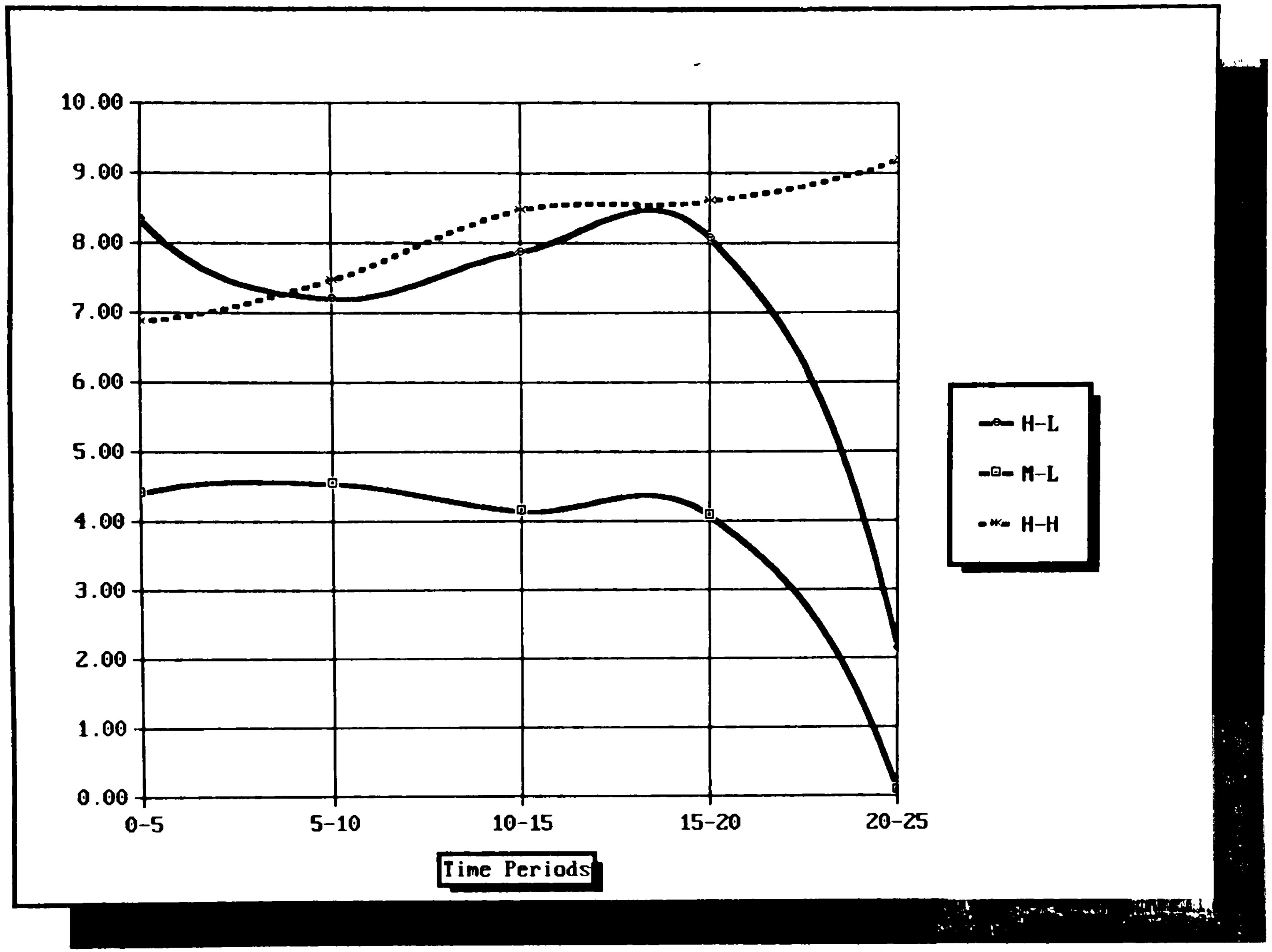

Figure 17
Navigational OEs by Work Load: no cue

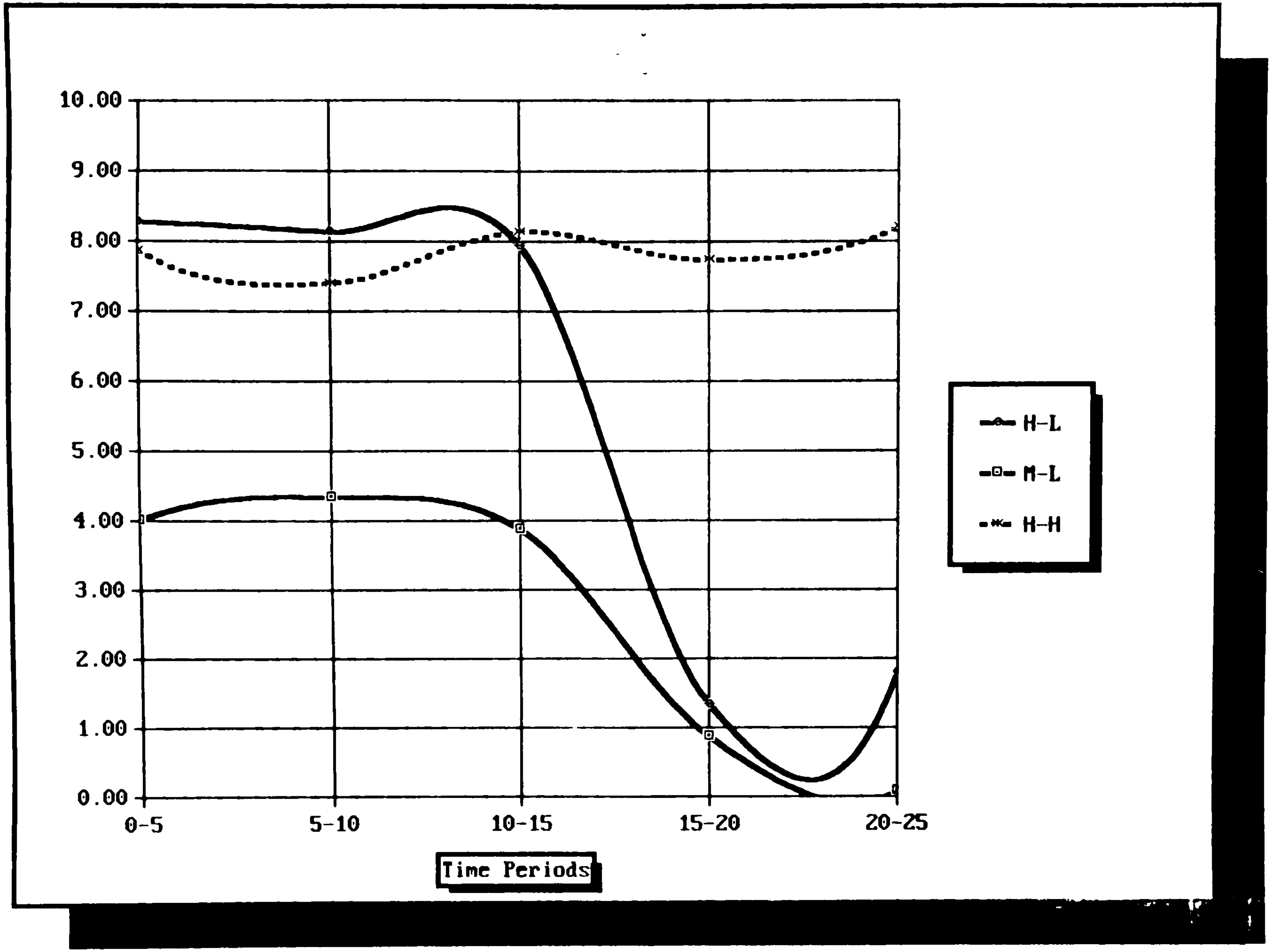

Figure 18
Navigational OEs by Work Load: cue

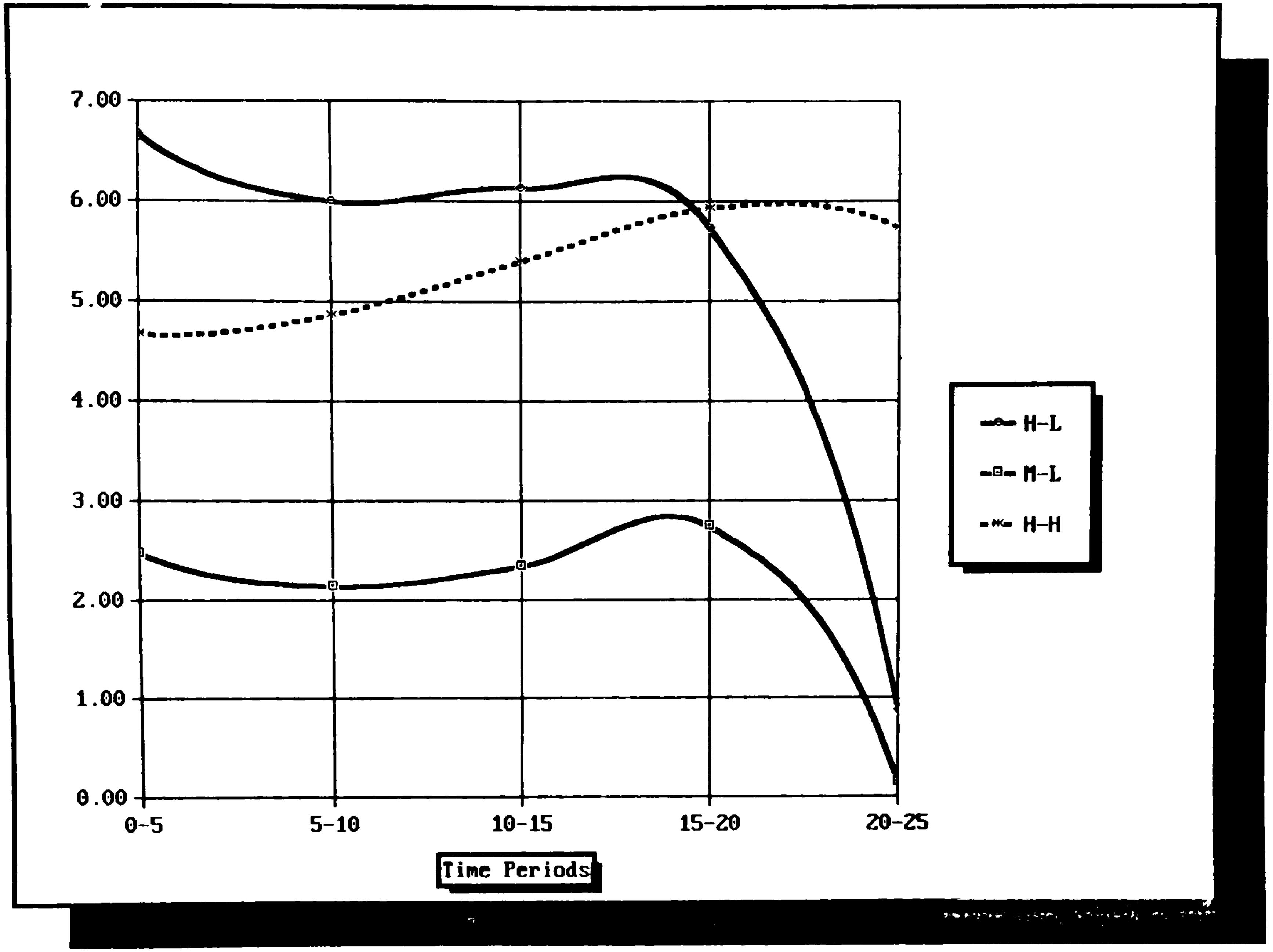

Figure 19
Pilot OEs by Work Load: no cue

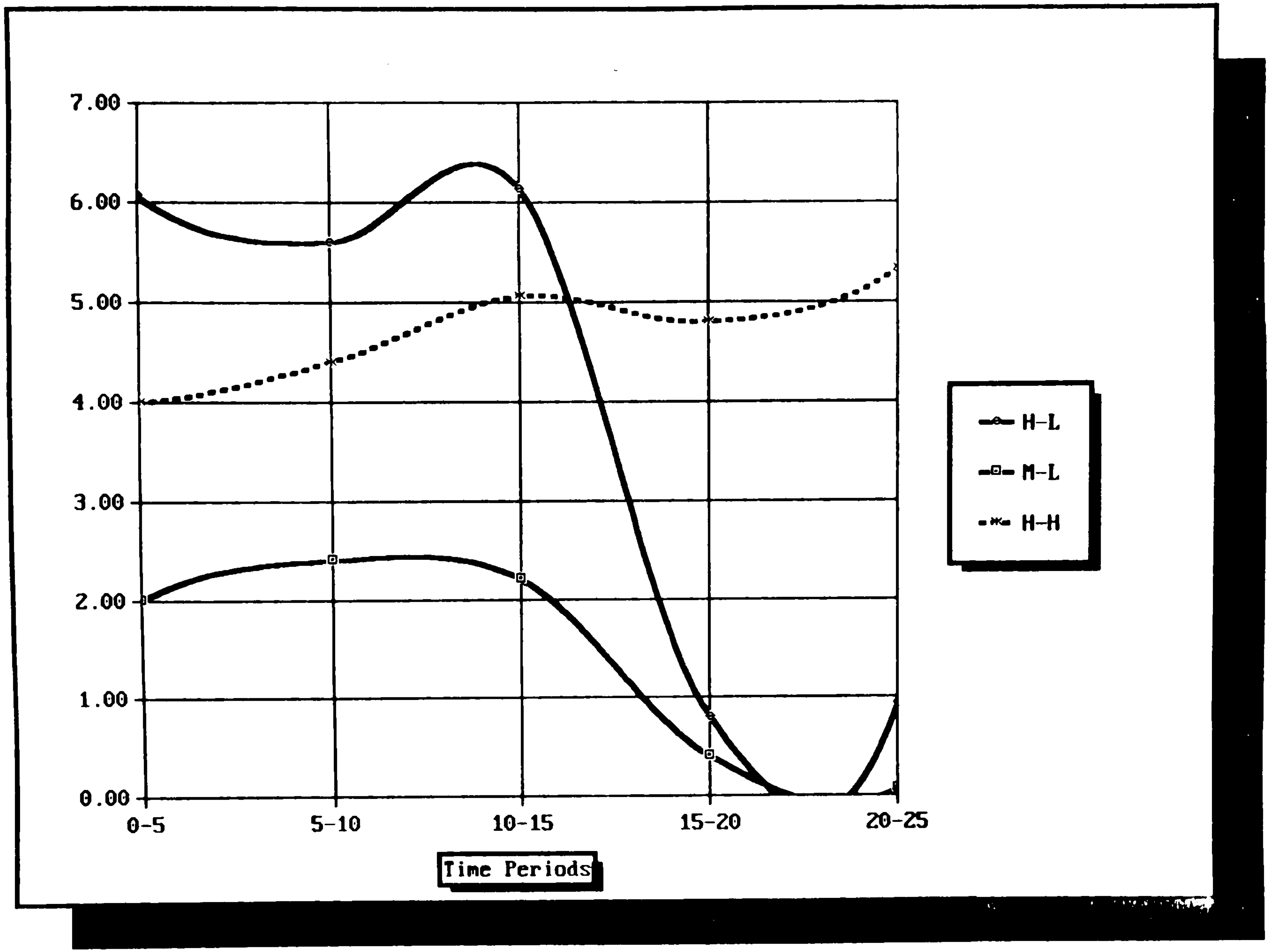

Figure 20
Pilot OEs by Work Load: cue

Finally, in order to assess the magnitude of treatment effects and possible interactions, the components of variance were examined. Results revealed that the greatest percentage of the overall population variance accounted for by each of the treatment effects was: Handoff OEs = WORK LOAD X CUE: 7%; Keying OEs = WORK LOAD: 36%; Navigational OEs = WORK LOAD: 33%; Pilot OEs = WORK LOAD: 29%.

Memory OEs. *A priori* orthogonal contrasts were employed to test the main effects for memory OEs that the hysteresis effect is due either to perseverance of expectancies or to STM overload. Specifically, a comparison of the means for the three time periods (t1, t2, and t3) versus the mean of t4 (the five minute time period immediately subsequent to the work load shift) was conducted for each WORK LOAD group. In an *a priori* test of the perseverance of expectancies hypothesis and the STM overload hypothesis, the results revealed that for memory OEs at the no cue treatment, the means of the OEs for the high to low work load group differed significantly for t1, t2, and t3 versus t4 ($F\ (1, 168) = 757769.12$, $p<0.001$). This suggests the presence of the type of hysteresis effect that displays a sudden increase in OEs immediately subsequent to a downward shift in work load. For high to low work load groups at the cue treatment, the means of OEs at t1, t2, and t3 were also significantly different from the mean number of OEs at t4 ($F\ (1, 168) = 769959.32$, $p<0.001$). This indicates that, all else being held constant, CUE was not effective in attenuating the hysteresis effect for the high to low work load group. Since the effect occurred for only the high to low group and was not subsequently attenuated for the group by cueing, it can be concluded that, based upon the hypotheses put forth in this dissertation, there is initial support for the STM overload hypothesis for the hysteresis effect for memory type OEs. Figure 21 presents the results for memory OEs.

An omnibus ANOVA procedure (Table 30) revealed significant main effects for WORK LOAD and TIME. CUE was not significant and

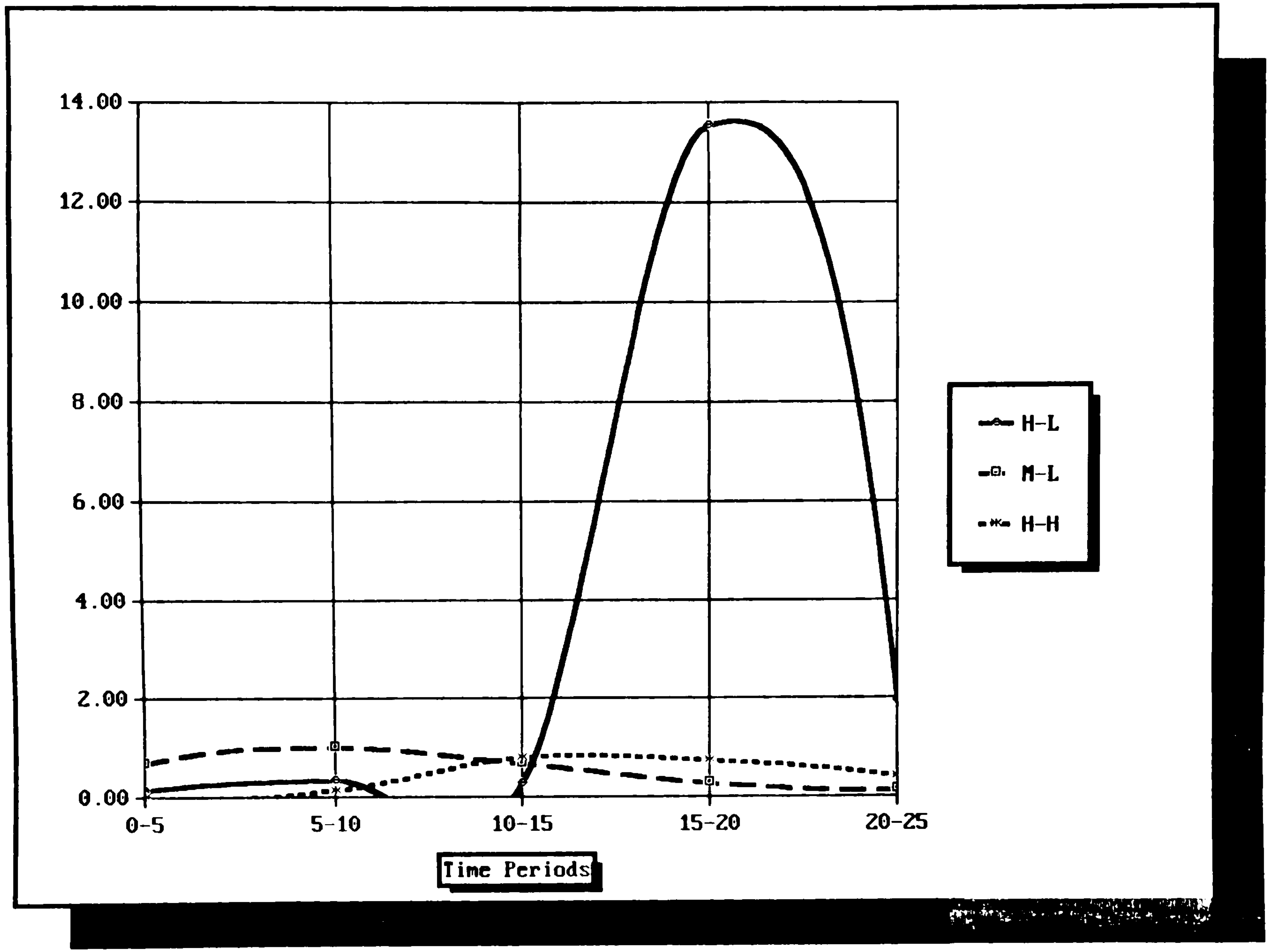

Figure 21
Memory OEs by Work Load: no cue and cue

Table 30
Omnibus ANOVA: Memory OEs

Source	SS	df	MS	F
Between Groups	218.60	44		
LOAD	190.77	2	95.39	143.97*
S w. LOAD	27.83	42	.66	
Within Groups	1241.90	405		
CUE	.06	1	.06	.10
LOAD X CUE	.12	2	.06	.11
S w. CUE X LOAD	23.72	42	.56	
TIME	337.24	4	84.31	142.34*
LOAD X TIME	679.65	8	84.96	143.43*
S w. LOAD X TIME	99.51	168	.59	
CUE X TIME	1.29	4	.32	.56
LOAD X CUE X TIME	2.96	8	.37	.64
S w. CUE X TIME X LOAD	97.35	168	.58	
Total	1460.50	449		

* $p < .001$

was not involved in any significant interactions. Therefore, the data for Memory OEs was collapsed across CUE. The main effects for WORK LOAD, and TIME were not subject to parsimonious interpretation, however, in the presence of the significant WORK LOAD X TIME two-way interaction (F (8, 168) = 143.43, $p<0.001$). Thus the results of the a prioi test of the main effects must be qualified.

Subsequent simple main effects analyses were conducted to determine the locus of the significant WORK LOAD X TIME interaction. These results are presented in Table 31. Of particular import to the hypotheses was whether the hysteresis effect occurred for all work load shifts and was ameliorated by cueing. It is important to recall that of the five time periods included in the analysis, peak work load occurred at t3. Subsequent to t3 work load dropped sharply. Results indicate that within the high to low treatment of work load, OEs across time periods were not all equal (F (4, 168) = 427.76, $p<0.001$). As a final step in this hierarchical analyses procedure, a Tukey procedure was undertaken to determine which pairs of means differed significantly from each other (Table 32). Of import at this stage of the analysis was where the hysteresis effect occurred for high to low work load. Results indicate that for high to low work load, Memory OEs revealed a significant increase between the means of t3 and t4 even though work load shifted downward between t3 and t4. Further, significant differences were found between the mean of t4 and the means of the previous three time periods. This indicates the presence of the type of hysteresis effect that displays a sudden performance decrement subsequent to a downward shift in work load. No significant differences were found across the time periods for the moderate to low or the high to high group. In the high to high treatment, no significant differences were revealed between the means of the different time periods indicating the absence of any hysteresis effect or any fatigue effect, for that matter.

Table 31
Simple Main Effects: Memory OEs no cue and cue

Source	SS	df	MS	F
LOAD	190.77	2	95.39	143.97*
TIME	337.24	4	84.31	142.34*
LOAD X TIME	679.65	8	84.96	143.43*
LOAD @ t1	1.87	2	.93	9.52*
LOAD @ t2	3.09	2	1.54	6.88*
LOAD @ t3	1.16	2	.58	.73
LOAD @ t4	850.16	2	425.08	324.97*
LOAD @ t5	14.16	2	7.08	14.79*
TIME @ h-l load	1009.51	4	252.38	427.76*
TIME @ m-l load	3.62	4	.91	1.53
TIME @ h-h load	3.76	4	.94	1.59
S w. TIME X LOAD	99.51	168	.59	
Total	11879.86	224		

* $p < .001$

Finally, in order to assess the magnitude of treatment effects and interactions, an examination of the components of variance was conducted. Results indicate that the greatest percentage of the overall population variance accounted for by effects was WORK LOAD X TIME: 52%.

A principal components analysis was conducted in order to assess the relationship between the five dependent variables. The results of this analysis are presented in Table 33. The principal components analysis revealed that four dependent variables, Handoff, Keying, Navigational, and Pilot OEs loaded highly on factor 1 while Memory OEs alone loaded highly on factor 2. This suggests that the behaviors being measured by the first four dependent variables are all in some way similar while memory OEs were in some manner measuring some different behavioral pattern. This proposition is confirmed by the similar patterns observed for Handoff, Keying, Navigational, and Pilot OEs and the different pattern observed for memory OEs.

Lastly, a Pearson *r* correlation was performed to assess the strength of relationship between the hypothesized measure of spatial ability under speed/load stress and performance overall score on TRACON. The results revealed a correlation of .13 that was not significant (p=.05). Although it is clear that the computer package does encompass some similar requirements to that of ATC work, there are other variables that preclude it from serving as a valid preditor of ATC performance.

Table 33
Principle Components Analysis of OEs

	Factor 1	Factor 2
Pilot	.94176	.18963
Navigational	.98602	.05471
Handoff	.95557	-.00431
Keying	.99080	.00969
Memory	.05716	.99700

CHAPTER IV
DISCUSSION

This dissertation asks whether the hysteresis effect is due to perseverance of expectancies or to STM overload? It was felt that discerning the nature of the effect would have a profound impact on what human factors interventions could be applied to enhance performance under such variable load situations. Clearly, the results of this investigation do not point to such a parsimonious solution. The taxonomy of OEs was developed primarily for the sake of facilitation of transcribing and coding the data. It was not expected that differences in the hysteresis effect would be manifested *as a function* of OEs type. In light of these differences, this researcher must now qualify the answer to the hypothesized question.

The hysteresis effect is a term that has been attributed to two patterns of operator performance under varying work load. The distinction has never been emphasized, however, and the two have simply been referred to collectively as hysteresis. The first pattern has been the tendency for a decrement in operator performance to begin when the operator is required to work beyond a certain work load level, and continue on even after the work load level has decreased sharply. In other words, there is a disproportionate lag between work load reduction and OE reduction. The second pattern of the phenomenon has been the tendency for the occurrence of a sudden decrement in operator performance immediately subsequent to a shift in visual load. In this case, little or no performance is observed prior to the work load shift. It is this pattern of occurrence that has been noted in operational settings such as ATC. Although not referred to as hysteresis per se in the literature, similar patterns of performance can be found in vigilance literature (i.e., Colquhoun and Baddeley, 1964) where it is referred to as a *contrast* effect. Nevertheless, these two patterns of performance meet the basic criterion to be called hysteresis effects. That is, performance at a given time period is influenced by the complexity of the task prior to that time period.

It is clear from the results that both patterns of hysteresis were observed to have occurred in this investigation. The first, the disproportionate lag in performance recovery subsequent to a downward shift in work load will be referred to from here on in as Type I hysteresis. The second, the sudden performance decrement subsequent to a downward shift in work load will be referred to from here on in as Type II hysteresis.

Overall, the results of the investigation support the position that work load history affects performance at subsequent time periods. This is particularly well illustrated by the results of the pilot investigation. As stated in the methodology section of the pilot investigation (appendix B) dissertation, experiment 1 confirmed that the pre-configured scenarios of low, moderate, and high task demands were subjectively perceived as low, moderate, and high work load scenarios. It is interesting to point out that when each of these three scenarios was worked in isolation to the others, OEs occurred in direct proportion to the demands of the task. That is, as task demands increased, so did the frequency of OEs. When the high and low work load scenarios were pasted against each other in experiment 2 of the pilot investigation, however, the frequency of OEs that were representative of the high work load scenario carried over to the first five-minute period of the low work load scenario. This constitutes, by definition, a Type I hysteresis effect. It is interesting to point out that in more conventional research, the experimenter would attempt to design an experiment that would eliminate carry-over effects. Here, however, the carry-over effect is the focus of investigation!

Hypothesis I

Hypothesis I predicted that if the hysteresis effect was due to the perseverance of expectancies established at earlier time periods, then regardless of the level of task demand prior to any shift in work load, expectancies would be established that would carry over to post-work load shift periods; in other words, a Type I hysteresis effect would occur. This carry over would occur because the operator would fail to timely perceive a change in task demand and thus would be working in accordance with the assumption that task demands have not changed from earlier time periods. It seems logical to assume that hypothesis I would inherently predict the pattern of Type I hysteresis. That is, it is a logical conclusion that the consequence of this failure to perceive would manifest itself in the form of a lag in performance recovery subsequent to a downward shift in work load. Hypothesis I further predicted that the administration of information concerning forthcoming task demand would attenuate this tendency to maintain expectancies when they are no longer appropriate. The rationale for this prediction is extrapolated from the probability matching studies of Craig (1980) and the vigilance studies of Colquhoun and Baddeley (1964, 1967) concerning pre-training signal expectancies.

The results of the current investigation do not universally provide support for hypothesis I. Non-Memory OEs (i.e., Handoff, Keying, Navigational, and Pilot OEs) demonstrate the type I hysteresis pattern (while Memory OEs demonstrate the Type II hysteresis pattern). That is, for both high to low and moderate to low work load groups, a lag in the OE rate of reduction to the work load rate of reduction is observed. Further, this lag is attenuated when cues are administered informing participants of forthcoming shifts in work load. The results for these OEs are not surprising, however, when one considers that ATCS Developmentals would naturally make a good amount of standard operating procedural errors based upon their limited training.

This would be particularly true when they are required to work under moderate to high work loads, because here it is that they are required to efficiently utilize the skills they have acquired with that limited training under load and speed stress. Therefore, the requirement to take action more quickly and accurately than is currently within their abilities stresses those abilities and, hence, performance degrades. This can be seen in the time periods prior to work load shifting downward. Performance within these pre-shift periods is clearly degraded. When the work load does shift downward suddenly, ATCS Developmentals fail to recognize this immediately. This is not to imply, that the perseverance of expectancies hypothesis is a function of inexperience. Rapid changes in work load are a very common occurrence in ATC. Even the most experienced Full Performance Level (FPL) ATCSs can and have been caught off guard. Rather, the data implies that developmentals have not honed their strategies for coping with these sudden work load changes. Matthews (1986) has raised the notion of strategic persistence. This is the notion that people develop certain strategies for coping with certain levels of task demand. He has, however, discussed the concept of strategic persistence as if it is a separate entity from perseverance of expectancies. This researcher suggests, rather, that the two are closely intertwined. That is, for as long as expectancies concerning task demand persevere, so to will the strategy for coping with that task demand. Albeit, the use of OEs rather than that of mere signal detection rates (as is traditional in simple vigilance tasks) where the OEs can be considered to be by-products of coping with perceived task demands seems to be a good indicator of strategic persistence.

Finally, this researcher puts forth the notion that the Type I hysteresis patterns observed for Handoff, Keying, Navigational, and Pilot OEs are a function of the ATCS's strategic processes for coping with sudden shifts in work load levels. It is interesting to speculate that ATCS Developmentals, not having the experience to have efficient coping strategies would be most prone to the Type I hysteresis pattern while FPLs would be less prone to the pattern. Memory OEs, on the other hand, could be viewed, not as a function of the strategy processes for coping with sudden shifts in work load levels per se, but rather, as a function of the strategic process for retaining pertinent information in STM.

Hypothesis II

Hypothesis II predicted that if the hysteresis effect was due to STM overload, then it should occur only for those tasks that load STM heavily initially and, in essence, strain the upper boundaries of the STM architecture prior to work load reduction. It would seem to be a logical assumption to make that tasks that never demand more than moderate work load initially should never strain those upper limits of the STM architecture. It would also seem to be a logical conclusion to make that when the hysteresis effect does occur for high to low task demand, that it should not be affected by information concerning forthcoming task demands (i.e., cueing) since, by definition, this effect has as its underlying antecedent, a *hard-wired* sensory-cognitive constraint and not a dynamic perceptual failure to perceive.

The data for Memory OEs demonstrate support hypothesis II as it pertains to the Type II hysteresis pattern. That is, performance did not degrade **until** work load shifted from high to low. It is important to note that only the high to low work load group demonstrated any hysteresis for memory OEs. The moderate to low work load group did not display such a pattern. Further, the provision of additional information concerning forthcoming work load demands (i.e., cueing) was not found to be effective in attenuating the hysteresis effect for Memory OEs. This would seem to indicate that this type of OE differs in some way from those discussed earlier. Additional evidence for this difference is emphasized by the different pattern of hysteresis (Type II) that was observed for Memory OEs as compared with the pattern of hysteresis that was observed for the OEs discussed earlier (Type I).

Moray (1980) (cited in Boff et al., 1986) drew attention to the apparently very limited memory span in dynamic tasks. Most laboratory research on STM has been concerned with static memory, that is, memory for words, syllables, numbers, or pictures that are presented as lists and then recalled as items in the list. The results are that under conditions where no special steps are taken to produce extraordinary effects people remember about seven items, regardless of their nature. Comparatively little work has been done on running memory, in which the observer must keep track of as much information as possible, when signals arrive in a continuous stream with no well-defined interval for recall. Such studies as do exist suggest that dynamic working memory is much less than the *magic number 7, plus or minus 2* that is the traditional memory span, let alone the striking memory for pictures (Shepard, 1967, cited in Boff et al., 1986), or the spans that appear when mnemonic systems are used (Yates, 1966, cited in Boff et al., 1986).

Work on dynamic memory suggests that observers viewing a time series do not carry more than about three items in their memory buffer, and that they probably cannot carry more. The reason for the reduction from the traditional span of eight items is unclear. Kay (1950) (cited in Boff et al., 1986) and Mackworth (1959) (cited in Boff et al., 1986) both performed an experiment in which the observer viewed a series of signals, each requiring a response in sequence, but with a delay. In Kay's experiment one of a row of lights would come on, then go out, another would come on, and the observer was then required to press the key corresponding not to the currently lit light but to the previous one. Observers found it extremely difficult to handle delays of more than one signal and impossible to handle a delay of more than two. Mackworth's results were similar.

Baker (1963) (cited in Boff et al., 1986) reported that when observers were detecting a series of signals that arrived at random intervals and trying to predict when the next would arrive they appeared to make use of only the last one or two intervals. Zeitlin and Finkelman (1975) (cited in Boff et al., 1986) also found a running span of three items. And Kvalseth found that even when provided with preview humans did not make use of more than one item to predict a series (Kvalseth, 1978, cited in Boff et al., 1986).

The restriction to a running span of three items may only apply when the time series is random and meaningless. ATCSs, for example, seem to be able to recall more information from their dynamic displays. This may be due to the fact that they have, in a sense, generated the information by their commands to the aircraft rather than merely observing the displayed information. It may also be due to some form of mnemonic coding (Yates, 1966, cited in Boff et al., 1986). But even ATCSs rapidly reduce the amount of information remembered per aircraft as the number of aircraft under their control increases (Sperandio, 1971, 1978).

With regard to Memory OEs in the present investigation, it is possible that these ATCS Developmentals have not yet developed a strategy for coping with large amounts of information that must be remembered. FPL ATCSs speak of constructing a mental picture of the traffic on their radarscope. When something happens to disrupt that picture, information is lost and OEs are committed that probably would not have occurred if the ATCS had not *lost the picture*. Whitfield and Jackson (1982) (cited in Boff et al., 1986) conducted a study in which they attempted to capture ATCS's strategies for dealing with data and organizing the picture. The method involved (a) ATC interviews and (b) verbal protocols during real-time simulations and recorded replays. The most important conclusions were that ATCSs vary in how they utilize the picture around non-routine information; they use the picture to plan, checking actual behavior against this plan; and that FPL ATCSs are more resistant to losing the picture than are developmentals.

The Type II hysteresis pattern in ATC was the impetus for this research. It seems clear from the results of the current investigation that this pattern pertains to Memory OEs. That is, subtasks that require the ATCS to store good deal amounts of information in STM are subject to Type II hysteresis. This is understandable particularly when one considers that under high work loads, ATCS have developed strategies for retaining good deal amounts of information in STM. They have, in essence, developed strategies for constructing their mental picture. Interviews with ATCSs report that sudden occurrences can cause an ATCS to lose the picture. It also seems logical that developmentals would not have these strategies well developed and as such would be more prone to losing what ever information they have stored in STM.

Limitations of the Investigation

This investigation, as do all studies, has certain limitations that need to be addressed when evaluating the results. The most pressing of these limitations involves the population sampled and the fidelity of the simulation.

The population sampled in this research was comprised of undergraduate psychology students. Although it would be ideal to say that the results obtained here generalize directly to ATCSs in the field, it would not be practical to make such a generalization. FPL ATCSs are highly skilled professionals that work air traffic day in and day out. Therefore, the number of OEs that are generated by each would, no doubt, be significantly less than the numbers generated by the sample in this investigation. Although the impetus for this research stems from reports produced by the ATC profession, it was not known whether they pertained to ATCS Developmentals, FPLs, or both. It would seem, however, that a more appropriate generalization of the population used in this investigation would be to developmentals in the field.

A second limitation deals with the use of a PC-based simulation to collect data rather than to assess data from the operational setting of ATC. Although it can be generally stated that the higher the fidelity of the simulator the greater the probability one has of including the many factors that affect performance, no simulator that this researcher knows of can match the fidelity of the operational setting. (Even the simulators at the FAA technical Center can be, at best, classified as moderate fidelity simulators.) TRACON, the simulated task to be used in the proposed investigation, is a simulator of low to moderate fidelity. Its major deficiency is that it does not account for the coordination that is inherent between ATCSs. This is a critical aspect of the ATCS's job and no ATCS can do the job in isolation. Nevertheless, the simulator served its intended purpose in demonstrating that work load history can influence subsequent performance.

Summary

The results of the present investigation, as well as the literature review on which it was based, demonstrate that work load history is a critical component to the assessment of work load performance for a given time period. Support was found for a hysteresis effect. However, the exact pattern of hysteresis observed was found to be dependent upon the type of OEs being assessed. Non-memory OEs are subject to perseverance of expectancies. Therefore, the hysteresis effect occurs for all downward work load shifts used in this investigation and can be ameliorated by cueing. Memory OEs, on the other hand, seem to provide support for hysteresis as a function of STM overload since the effect only occurs for high to low work load and the effect is not ameliorate by cueing.

It seems clear that there are two different mechanisms at work in producing the hysteresis effect. Type I hysteresis can be thought of as a *carry-over* effect that is dependent upon the awareness of the operator to perceive sudden changes in task demands. When he/she does not do so immediately, Type I hysteresis occurs. Type II hysteresis, on the other hand, can be thought of as a *contrast effect* that is dependent upon the operator's strategies for coping with sudden changes in task demands. Where coping strategies are not honed, Type II hysteresis would seem more prone to occur.

The results of the investigation have strong implications for human factors interventions. At least in the case of ATC, it can be concluded that there is support for the both explanations for the hysteresis effect dependent upon the type of OEs being considered. This information could be invaluable to the enhancement of future ATCS performance. Successful ATCS performance depends, to a great degree, on the reliable recall of relevant information in STM (e.g., aircraft, destinations, altitudes, etc). as well as prior knowledge of forthcoming changes in work loads. A clear understanding of the role of memory and establishment of expectancies in the ATCS's strategic model could lead to development of effective cueing and memory aids that could help ensure the availability of accurate essential data/information when it is needed. A memory lapse, or even a delay, during a critical point such as a downward work load shift can lead to serious consequences; such as conflicting tracks and inadequate separation between aircraft, or flight into terrain obstacles. In the same vein, a sudden, unexpected change in work load could disrupt the flow, or mental picture, that the ATCS has established to perform the job efficiently and could lead to serious consequences as well. Thus, memory and cueing aids which would efficiently facilitate recall of critical information and inform of changes in work load would be of great value.

BIBLIOGRAPHY

Allnutt, M. Human factors: basic principles. In R. Hurst and L. Hurst (Eds.), *Pilot Error* (pp.1-22). London: Granada, 1982.

Allnutt, M. Human Factors. In R. Hurst (Ed.), *Pilot error* (pp. 1-24). London: Granada Publishing, 1976.

Alluisi, E. A., Muller, P. F., and Fitts, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *Journal of Experimental Psychology*, 1957, *53*, 153-158.

Aviation Safety Institute. A quarterly report on aviation safety. Washington, DC: Author, report 74-2, July 1, 1974.

Baker, C. H. Attention to visual displays during a vigilance task: II. Maintaining the level of vigilance. *British Journal of Psychology*, 1959, *50*, 30-36.

0 Biggs, D. *Pressure Cooker*. Toronto: Geo. McLeod, Ltd., 1979.

Bisseret, A. Analysis of mental model processes involved in air traffic control. *Ergonomics*, 1971, *14*, 565-570.

Boff, K. R., Kaufman, L., and Thomas, J. P. (Eds.) *Handbook of Perception and Human Performance*. New York: John Wiley and Sons, 1986.

Broadbent, D. E. *Decision and stress*. London: Academic Press, 1971.

Buckner, D. N., and McGrath, J. J. A comparison of performances on single and dual sensory mode vigilance tasks. In D. N. Buckner and J. J. McGrath (Eds.), *Vigilance: a symposium* (pp. 220-242). New York: McGraw-Hill, 1963.

6 Colquhoun, W. P., and Baddeley, A. D. Role of pretest expectancy in vigilance decrement. *Journal of Experimental Psychology*, 1964, *68*, 156-160.

Colquhoun, W. P., and Baddeley, A. D. Influence of signal probability during pretraining on vigilance decrement. *Journal of Experimental Psychology*, 1967, *73*, 153-155.

Craig, A. Broadbent and Gregory revisited: vigilance and statistical decision. *Human Factors*, 1977, *19*, 25-36.

Craig, A. Effect of prior knowledge of signal probabilities on vigilance performance at a two-signal task. *Human Factors*, 1980, *22*, 361-371.

Craig, A. Is the vigilance decrement simply a response adjustment towards probability matching? *Human Factors*, 1978, *20*, 441-446.

Craig, A., and Colquhoun, W. P. Vigilance: a review. In C. G. Drury and J. G. Fox (Eds.), *Human reliability in quality control* (pp. 128-156). London: Taylor and Francis, 1975.

Cumming, R. W. The analysis of skills in driving. *Australian Road Research*, 1964, *1*, 4-14.

Cumming, R. W. Human factors in relation to intersection accidents. *Proceedings of the National Road Safety Symposium, Canberra, Australia*, 1972.

0 Cumming, R. W., and Croft, P. G. Human information processing under varying task demand. *Ergonomics*, 1973, *16*, 581-586.

Danaher, J. W. Human error in air traffic control system operations. *Human Factors*, 1980, *22*, 535-545.

Davies, D. R., and Parasuraman, R. *The psychology of vigilance*. London: Academic Press, 1982.

Davies, D. R., and Tune, G. S. *Human vigilance performance*. New York: American Elsevier, 1969.

Davis, D. R. Human errors and transport accidents. *Journal of Experimental Psychology*, 1958, *48*, 25-33.

Federal Aviation Administration. *Profile of operational errors in the National Airspace System, calendar year 1986*. Washington, DC: Federal Aviation Administration, Office of Aviation Safety, Safety Analysis Division, 1987.

Fisk, A. D., and Schneider, W. Control and automatic processing during tasks requiring sustained attention: a new approach to vigilance. *Human Factors*, 1981, *23*, 737-750.

Fowler, F. D. Air traffic control problems: a pilot's view. *Human Factors*, 1980, *22*, 645-653.

Goldberg, R. A., and Stewart, M. R. Memory overload or expectancy effect? 'Hysteresis' revisited. *Ergonomics*, 1980, *23*, 1173-1178.

Hurst, M. W., and Rose, R. M. Objective workload and behavioral response in airport radar control rooms. *Ergonomics*, 1978, *21*, 559-565.

Jerison, H. J. Experiments on vigilance: V. The empirical model for human vigilance. Wright Air Development Center Technical Report No. WADC-TR-58-526. Wright-Patterson Air Force Base, Ohio: Aero Medical Laboratory, Wright Air Development Center, 1959.

Jerison, H. J. Vigilance: biology, psychology, theory and practice. In R. R. Mackie (Ed.), *Vigilance: theory, operational performance and physiological correlates* (pp. 127-133). New York: Plenum, 1977.

Kahneman, D. *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Keppel, G. *Design and Analysis: A Researcher's Handbook*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1982.

Kirk, R. E. *Experimental Design: Procedures for the Behavioral Sciences*. Belmont, California: Brooks/Cole Publishing Company, 1982.

Knowles, W. B. Operator loading tasks. *Human Factors*, 1963, *5*, 155-161.

Krulewitz, J. E., Warm, J. S., and Wohl, T. H. Effects of shifts in the rate of repetitive stimulation on sustained attention. *Perception and Psychophysics*, 1975, *18*, 245-249.

Langan-Fox, G. C., and Empson, J. A. C. 'Actions not as planned' in military air traffic control. *Ergonomics*, 1985, *28*, 1509-1521.

Leplat, J., and Bisseret, A. Analysis of the processes involved in the treatment of information by the air traffic controller. *The Controller*, 1966, *5*, 13-22.

Loftus, G. R. Short-term memory factors in ground control/pilot communication. *Human Factors*, 1979, *21*, 169-181.

Mackworth, J. F. *Vigilance and habituation*. Baltimore, MD: Penguin, 1969.

Mackworth, J. F. *Vigilance and attention*. Baltimore, MD: Penguin, 1970.

Matthews, M. L. The influence of visual workload history on visual performance. *Human Factors*, 1986, *28*, 623-632.

McFarland, B. P., and Halcomb, C. G. Expectancy and stimulus generalization in vigilance. *Perceptual and Motor Skills*, 1970, *30*, 147-151.

Mohler, S. R. The human element in air traffic control: aeromedical aspects, problems, and prescriptions. *Aviation, Space and Environmental Medicine*, 1983, *54*, 511-516.

Moray, N. *Human information processing and supervisory control* (Tech. Rep. NR 196-132). Cambridge, Mass.: Massachusetts Institute of Technology, 1980.

Moray, N. (Ed.). *Mental workload*. New York: Plenum Press, 1979

Moray, N. Mental workload. *1984 International Conference on Occupational Ergonomics Proceedings*, Vol. 2, Reviews. Human Factors Society of Canada, Ontario, Canada, 1984, 41-46.

National Transportation Safety Board. *Midair collision involving a Falcon jet and a Cessna 150, Memphis, Tenn., May 18, 1978*. Washington: NTSB, Report No. NTSB-AAR-78-14, November, 1978. (b)

Norman, D. A., and Bobrow, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, *7*, 44-64.

Ogden, G. D., Levine, J. M., and Eisner, E. J. Measurement of workload by secondary tasks. *Human Factors*, 1979, *21*, 529-548.

Parasuraman, R. Memory load and event rate control sensitivity decrements in sustained attention. *Science*, 1979, *205*, 924-927.

Parasuraman, R. Vigilance, arousal and the brain. In A. Gale and J. Edwards (Eds.), *Physiological correlates of human behavior* (Vol. 2). *Attention and performance* (pp. 727-752). London: Wiley, 1983.

Parasuraman, R., and Davies, D. R. Decision theory analysis of response latencies in vigilance. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 569-582.

Pasmooj, C. K., Opmeer, C. H. J. M., and Hyndman, B. W. Workload in air traffic control. In T. B. Sheridan and G. Johnannsen (Eds.), *Monitoring behavior and supervisory control* (pp.107-117). New York: Plenum, 1976.

Posner, M. I. *Chronometric explorations of mind*. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Poulton, E. C. Arousing stresses increase vigilance. In R. R. Mackie (Ed.), *Vigilance: theory, operational performance and physiological correlates* (pp.423-459). New York: Plenum, 1977.

Poulton, E. C. The effect of fatigue upon inspection work. *Applied Ergonomics*, 1973, *4*, 73-83.

Schneider, W., and Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search and attention. *Psychological Review*, 1977, *84*, 1-66.

Schroeder, D. J. *Profile of operational errors in the National Airspace System, calendar year 1979*. Washington, DC: Federal Aviation Administration, Office of Aviation Safety, Safety Analysis Division, 1980.

Society of Automotive Engineers. Crew procedures and communication between pilots and controllers. *Fifth Aerospace Behavioral Engineering Technology Conference Proceedings*. Warrendale, PA: Society of Automotive Engineers, 1986.

Sperandio, J.C. The regulation of working methods as a function of work-load among air traffic controllers. *Ergonomics*, 1978, *21*, 195-202.

Sperandio, J. C. Variation of operators' strategies and regulating effects on workload. *Ergonomics*, 1971, *14*, 571-577.

Swets, J. A. Signal detection theory applied to vigilance. In R. R. Mackie (Ed.), *Vigilance: theory, operational performance and physiological correlates* (pp. 97-102). New York: Plenum, 1977.

Teichner, W. H. The detection of a simple visual signal as a function of time of watch. *Human Factors*, 1974, *16*, 339-353.

Thackray, R. D., Bailey, J. P., and Touchstone, R. M. The effect of increasing monitoring load on vigilance performance using a simulated radar display. *Ergonomics*, 1979, *22*, 529-539.

Thackray, R. D., Bailey, J. P., and Touchstone, R. M. The effect of increased monitoring load on vigilance performance using a simulating radar display. Oklahoma City, OK: FAA Civil Aeromedical Institute, FAA Office of Aviation Medicine, Report 77-18, July, 1977.

Thackray, R. D., and Touchstone, R. M. Effects of high visual taskload on the behaviors involved in complex monitoring. *Ergonomics*, 1989, *32*, 27-38.

Thayer, R. E. Activation states as assessed by verbal report and four psychophysiological variables. *Psychophysiology*, 1970, *7*, 86-94.

Warm, J. S. (Ed.) *Sustained attention in human performance*. New York: Wiley, 1984.

Whitfield, D., and Jackson, A. The air traffic controller's picture as an example of a mental model. Paper presented at the IFAC Conference on Analysis, Design, and Evaluation of Man-Machine Systems. Baden-Baden-FRG, 1982.

Wickens, C. D. *Engineering psychology and human performance*. Columbus, OH: Charles E. Merrill Publishing Co., 1984.

Wiener, E. L. Adaptive measurement of vigilance decrement. *Ergonomics*, 1973, *16*, 353-363.

Wiener, E. L. Controlled flight into terrain accidents: System-induced errors. *Human Factors*, 1977, *19*, 171-181.

Wiener, E. L. Knowledge of results and signal rate in monitoring: a transfer of training approach. *Journal of Applied Psychology*, 1963, *47*, 214-222.

Wiener, E. L., and Curry, R. E. Flightdeck automation: Promises and problems. Moffett Field, CA: NASA-Ames, NASA Tech. Memo. 81206, 1980.

APPENDIX A
EXPANDED LITERATURE REVIEW

The task of the ATCS requires sustained attention due to work under load at times, work over load at times, and continuous scanning, memorization, and coordinatation all of the time. Therefore, any competent literature review must take a trans-paradigmatic approach toward laying the groundwork for an investigation in this area. This researcher has chosen to discuss both the vigilance performance literature and the work load literature since the task of the ATCS is one that impinges on these concepts at different times. Included in the discussion of vigilance performance is a discussion of basic vigilance performance issues, visual search as it applies to vigilance performance, and performance issues in complex monitoring. The second portion of the literature review discusses the concept of mental work load. Lastly, the literature review ties together the two paradigms within the context of ATC.

Overview Of Vigilance Performance

Perceptual and cognitive activities often demand sustained attention if they are to be performed successfully and efficiently. When the activity is performed for a continuous period of time, the ability to sustain attention may be severely impaired. It is relatively easy to be attentive briefly for a conspicuous and predictable event, such as a traffic light changing from red to green. The task is much more difficult, however, if attention must be maintained on some source of information for the occurrence of infrequent, unpredictable events over long periods of time. The ability to maintain vigilance for such events typically declines over time, a phenomenon known as the *vigilance decrement*.

An understanding of the processes underlying the vigilance decrement is the major theoretical issues in research on vigilance. Identification of the factors controlling the overall level of vigilance, rather than decrement over time, forms a secondary though nevertheless important issue. Finally, the regulation and control of vigilance in operational settings (such as ATC and industrial inspection) constitutes an important practical issue in this area of research.

Head (1923) (cited in Boff et al., 1986) first used the term vigilance to refer to a state of maximum physiological efficiency of the central nervous system. Mackworth (1957) (cited in Boff et al., 1986) redefined vigilance in psychological terms as a state of readiness to detect and respond to certain small changes occurring at random time intervals in the environment.

There are several varieties of attention; all share in common the idea of conscious processing of information, where the source of information may be external (data driven) or internal (memory driven) (Kahneman, 1973; Norman and Bobrow, 1975; Posner, 1978). Generally speaking, the deployment of attention facilitates this processing, but there may be associated costs (Posner, 1978). Attention possesses both a selective and intensive dimension (Kahneman, 1973); attention can be directed to one or more of a number of signals or to different positions in space, and it may be deployed with more or less intensity.

The intensive dimension of attention is important in vigilance, particularly in tasks with only a single source. Such vigilance tasks involve the detection of signals presented at infrequent intervals for prolonged periods of time without rest. Accordingly, vigilance for single-source events involves sustained attention rather than selective attention (Broadbent, 1971; Jerison, 1977), whereas in multiple-source or time-shared tasks selective attention and the orienting of attention to visual space are additional important determinants of performance. Some confusion has often existed concerning the distinction between vigilance tasks and the tasks that the present proposal deals with, namely, monitoring tasks. Vigilance tasks have relatively simple, specified, unchanging signals, usually presented over a single source. Monitoring tasks have somewhat more complex signals, usually (but not always) presented over two or more sources; they include tasks where the signal may not be specified precisely but, rather, must be inferred and may change over time. Monitoring is an important component of the operator's task in many automated and semi-automated systems.

Temporal uncertainty is a principle feature of vigilance. The signals for which the observer must remain vigilant appear infrequently and at relatively unpredictable times. Search tasks, on the other hand, involve signals with little or no temporal uncertainty but with high spatial uncertainty. Signals in simple search tasks are readily discriminable, and the resolution of spatial uncertainty is the key to detecting the signal, although in complex search tasks other sources of uncertainty may affect performance. Many monitoring tasks, including those found in modern human-machine systems, are neither pure vigilance nor search tasks, often requiring the detection of signals having both temporal and spatial uncertainty.

Historical Background

Prior to World War II, there was some concern about the problems that industrial workers were having in sustaining their attention on prolonged tasks. This concern was traced to quality control situations in which workers had to observe the monotonous flow of products on an assembly line and be alert for defects (Davies and Tune, 1969). For example, Wyatt and Langdon (1932) (cited in Warm, 1984) reported variations in the performance of experienced inspectors who examined cartridge cases for flaws prior to packaging. Accuracy of inspection varied in a U-shaped manner over a four hour working period, reaching its lowest point after 90 minutes on the job. With the arrival of war, interest in vigilance was stepped up by the need to know as much as possible about the capabilities of radar and sonar operators who had to sustain attention for infrequent events under monotonous conditions.

Efforts to meet this wartime challenge were initiated at about the same time in Great Britain, the United States and Canada. In England, Ditchburn (1943) (cited in Warm, 1984) investigated the performance of lookouts and discovered that their ability to detect a target began to decline soon after their work commenced. Similarly, in the United States, Anderson et al. (1944) (cited in Warm, 1984) studied radar operators and, in Canada, Solandt and Partridge (1946) (cited in Warm, 1984) studied sonar operations and found that performance efficiency deteriorated over time in these tasks. Anderson recommended that daily operating periods not last longer than 40 minutes.

These initial investigations among others (Lindsley, 1944, cited in Warm, 1984; Baker, 1962, cited in Warm, 1984) had begun to uncover the fragility of sustained attention in human performance. However, these studies were not methodological and little attempt was made to coordinate the efforts of different workers. Also, in some cases, the focus was on fatigue and not attention (Jerison, 1970, cited in Warm, 1984). Thus, these investigations were not generally considered to be the beginning of vigilance research. They served mainly to create interest and to set the stage for more productive experimentation.

Controlled laboratory research on sustained attention is generally considered to have been initiated by a series of experiments by Mackworth (1948, 1950, cited in Warm, 1984). His studies provided several fundamental findings and set the tone for future vigilance research. Mackworth was the first to emphasize the theoretical and the practical implications of watchkeeping behavior (Davies and Tune, 1969).

Mackworth devised a simulated radar display called the clock test in which participants were asked to view movements of a black pointer along the circumference of a blank-faced clock which contained no scale markings or reference points. Once every second, the pointer would jump to a new position. From time to time, it executed a double jump, and this was the critical signal for detection. Participants were required to press a key whenever they spotted a double jump during a monitoring session which lasted for two hours.

Using the clock test in this way, Mackworth was able to assess subject performance over time. He confirmed earlier suspicions that the quality of sustained attention in monitoring tasks declines rapidly with the largest decline occurring in the first 1/2 hour. Subsequent declines in vigilance performance over the remainder of the session were much more gradual and less pronounced. This progressive decline in performance has been termed the decrement function (Dember and Warm, 1979, cited in Boff et al., 1986) or the vigilance decrement (Davies and Parasuraman, 1982).

Several general characteristics of the vigilance decrement can be discerned, regardless of the type of vigilance task used or research strategy employed (Davies and Parasuraman, 1982). The decrement occurs under all sensory modalities as well as under a combination of sensory modalities. The decrement is most pronounced in the first 1/2 hour of the task (Davies and Tune, 1969). Half of the total decrement in the task was experienced within the first 15 minutes of the vigil (Davies and Parasuraman, 1982; Warm, 1984). The data would suggest that the decrement was the result of merely sustaining attention for the purpose of detecting infrequently occurring signals over a prolonged period of time.

Research on vigilance, monitoring, and search arose in response to pressing practical concerns, but subsequently each area of research was influenced more by theoretical concerns. Yet, although the simple radar, sonar, and search tasks of World War II no longer exist, many of the operational problems they were associated with still exist in modern human-machine systems. As machines and equipment have become more complex and increasingly automated, and with the advent of microprocessor control, the role of the human operator has changed from that of an active ATCS to a decision maker and manager, a shift from active to supervisory control first characterized by Sheridan (1970) (cited in Boff et al., 1986). In today's highly automated systems, targets may be detected by instruments and controls may be executed by machine. Yet the same problems of vigilance and monitoring occur when the automated system malfunctions or some unusual infrequent treatment occurs.

Whereas military applications of research on vigilance, monitoring, and search were prominent during and just after wartime, industrial applications have come increasingly to the fore in recent years. Current research in these areas thus reflects two trends: (1) a return to a consideration of practical problems of human performance in human-machine systems following several years of consolidation of basic research; and (2) a growing interest in human performance problems in industrial and medical systems, particularly in such areas as civilian ATC, industrial inspection, process control, vehicle operation, and medical imaging.

In addition to the critical variables known to mediate operator performance in simple vigilance tasks, a number of variables may affect performance in modern automated systems. The signals that must be detected in such systems must be inferred when a situation outside the normal range of operations that can be handled by the automatic system occurs; a pilot noticing a malfunction and overriding the autopilot system represents one example. In other instances, the signal may be the complete failure of the automatic system. In most such complex task situations, the functions of vigilance, monitoring, and search were supplemented by a number of other demands on the human operator, including signal interpretation, high-level decision making, action, task scheduling and resource allocation, and other tasks. The possible effects of such multiple processing requirements on vigilance, monitoring, and search, and their interactions can only be hinted at rather than analyzed. Recent papers by Moray (1980, 1984), Wiener (1977), and Wiener and Curry (1980), consider these and related aspects of human monitoring performance in complex systems.

Measures Of Vigilance Performance

Studies of vigilance performance have employed many different discrimination, monitoring, and search tasks. As a result, a variety of measures of performance on such tasks have been used. Some measures were quite specific to a particular task or group of tasks. Others have more general applicability. Measures of the accuracy and speed of target detection are the most generally useful indices of vigilance, monitoring, and search performance.

Detection Rate

The rate of correct detection of signals (hits) is a commonly used index of continuous performance. For purposes of analysis, the period of performance is divided into a number of successive time periods or blocks, and the detection rate in each time period is computed (Parasuraman, 1986, cited in Boff et al., 1986). A typical finding is a decline in the detection rate over the course of the task, the vigilance decrement.

The reliability of the vigilance decrement can be assessed using conventional statistical procedures. If, however, the decrement in hit rate proves unreliable, alternative factors should be considered before concluding that no vigilance decrement occurred. First, an inadequate sample size or high inter-subject variability may contribute to an unreliable result (Poulton, 1973). Second, the decrement may be masked if the analysis of time-related changes in detection rate is too coarse-grained;

measures based on block intervals of 10-30 minutes, for example, may tend to hide that part of the decrement function where the most rapid changes occur (Teichner, 1974). Block-by-block analysis can be supplemented by a signal-by-signal analysis of the proportion of participants detecting each signal as a function of signal number (Jerison, 1959). Finally, if a decrement in detection rate is not obtained despite adequate sample size, low performance variability, and sufficiently fine-grained analysis, it may still be premature to infer that a decrement in performance did not take place. Although the detection rate may not have declined, perceptual sensitivity may have declined if, for example, the false detection rate increased or remained stable with time on task.

False Alarm Rate

The vigilance performance over time cannot be based on the detection rate of signals alone, but must include the false alarm rate. Errors of commission of false alarms occur when the observer reports a signal when none has occurred. Normally, the signal detection and false alarm rate show a similar decline over time (Parasuraman and Davies, 1976).

The false alarm rate may also indicate a vigilance decrement even if the hit rate shows no such decline. If the false alarm rate remains the same or increases over time, any decrement in the detection rate could be masked by the overall increase in responding. This measure is also used to determine changes in d' and ß in connection with the TSD paradigm.

Reaction Time

Reaction time is another primary measure of performance in vigilance, monitoring, and search tasks. In some tasks a signal is presented and not removed until the subject responds to its presence. In such *unlimited-hold* tasks (Broadbent, 1958, cited in Warm, 1984), reaction time is measured from the onset of the supra-threshold signal. Since the signal is virtually always detected, reaction time is the major dependent variable in such studies. In other tasks the signal is presented briefly, and reaction time is recorded as a supplementary measure to detection rate. Finally, in visual search tasks reaction time is usually measured from the time of presentation of the display to the location of the signal.

The mean reaction time to signals typically increases over time in vigilance tasks, and the vigilance decrement has been associated with either a decrement in detection rate or in detection speed (Davies and Tune, 1969). In general, reaction time can increase by several hundred milliseconds over the course of a vigilance task lasting about one hour.

Increases in mean reaction time have been found in tasks with transient, unpredictable, difficult-to-discriminate signals similar to increases in tasks with readily discriminable signals presented at predictable intervals (Boulter and Adams, 1963, cited in Warm, 1984); McCormack and Prysiaznuik, 1961, cited in Warm, 1984). In some studies both the mean reaction time and the variance in reaction time have been measured (Faulkner, 1962, cited in Warm, 1984). Even when reaction times remain stable throughout the vigilance task, an increase in the variability of the reaction time has been observed, indicating a change in the level of vigilance performance (Jerison, 1959).

Self-Report

Self-reports may provide a useful addition to objective performance measures as a source of information about observer state. Thackray, Bailey, and Touchstone (1977) found that reports of perceived boredom and monotony elicited by a standard questionnaire were correlated with the decrement during a vigilance task, reports of marked boredom being associated with greater increases in detection latency over the course of the task.

Self-reports of arousal may also provide a useful index. Thayer (1970) developed a self-report questionnaire found to correlate significantly with physiological measures of arousal. However, there were distinctions between arousal and vigilance (Parasuraman, 1983) that should be noted. The evidence for a causal relationship between lowered arousal and the vigilance decrement is not strong, so that an assessment of observer arousal level cannot necessarily be used to predict whether vigilance will decline. Consistent with this view, Thackray et al. (1977) suggested that their boredom questionnaire, which provided a fairly good predictor of the vigilance decrement, tapped attentional processes (e.g., distractibility) rather than arousal processes. On the other hand, the evidence does show an association between arousal level and the overall level of performance (Davies and Parasuraman, 1982; Parasuraman, 1983; Poulton, 1977). Since this is an important component of vigilance, especially in operational settings where marked changes in arousal level were likely (e.g., due to circadian variability, as in shift work), measurement of arousal level can be important.

Adaptive Measurement Of Vigilance Decrement

The vigilance decrement may also be assessed by measuring the change in signal discriminability or some other task parameter required to keep performance stable over time. Such adaptive vigilance tasks, in which changed in vigilance were measured by time-related changes in ask parameters (e.g., signal strength or duration) rather than task performance, have been studied by Wiener (1973).

Adaptive measurement techniques have been used widely in the analysis of performance on manual control tasks and in training research (Kelley, 1969, cited in Warm, 1984). Adaptive measurement is more easily implemented for continuous tasks like manual control than for vigilance tasks, where discrete trials were commonly used. Although not appropriate for all applications, adaptive measurement may be useful in evaluating the effectiveness of training procedures in reducing vigilance problems in the field.

Mental Work Load Measures

Evaluation of vigilance in operational settings poses particularly difficult problems of performance assessment. In many instances performance measures cannot be obtained from the primary task performed by the operator. In this case having the operator perform a secondary task may provide some information regarding the effectiveness of primary task performance. The limited-capacity theory of attention (Moray, 1967, cited in Warm, 1984) provides the rationale for the secondary task method, which has a long history of use in the evaluation of mental work load. The assumption is that performance of multiple tasks depends on a single, common pool of mental capacity, so that an assessment of spare mental capacity by measuring secondary task performance can be used to assess primary task performance. More recent findings on the interaction between primary and secondary task performance suggest a multiple-pool, limited-capacity model of attention (Wickens, 1984). Applications of the method to the assessment of spare capacity during prolonged performance have been reported by Brown and Poulton, (1961) (cited in Warm, 1984) for automobile driving and Haider, (1963) (cited in Warm, 1984) for assembly line work.

Considerable research has been devoted to the further development of the secondary task technique or measuring mental work load (Moray, 1979). These efforts have been directed mainly toward tasks in which the operator is relatively overloaded but they are also applicable in principle also to tasks where the operator is underloaded, where problems of vigilance may arise.

Theories Of Vigilance Performance

Since Mackworth's (1948, 1950) (cited in Boff et al., 1986) initial vigilance behavior investigations, a number of theories have been posited to explain the phenomenon. Generally, these theories can be grouped into three categories: (a) neurological models; (b) learning models; and (c) information processing models.

Neurological Models

Arousal Theory

Drawing from Hebb (1955, cited in Warm, 1984), Frankman and Adams (1962, cited in Warm, 1984) theorized that the low arousal state induced by monitoring tasks resulted in an attenuation of stimulation to the reticular activating system which subsequently resulted in the poor performance of other cortical functions necessary for efficient vigilance performance.

There are two types of arousal: autonomic arousal and electrocortical arousal. Autonomic arousal measurements such as heart rate and skin conductance generally have shown either no or little correlation to sustained attention. However, adrenalin has been shown to be moderately positively correlated with vigilance performance. Electrocortical arousal, on the other hand, is reflected on EEG measurements as a shift to lower wave frequencies during a monitoring task. A significant inverse relationship between the amount alpha and theta wave activity and vigilance performance has been demonstrated (O'Hanlon and Beatty, 1977, cited in Warm, 1984).

Detection efficiency is reduced by stimuli which reduce arousal such as heat (Poulton, 1977) and alcohol (Erwin et al., 1978, cited in Warm, 1984) while the detection efficiency is increased by stimuli which increase arousal such as amphetamines (Mackworth, 1950, cited in Warm, 1984) and vibration (Poulton, 1978, cited in Warm, 1984).

Arousal differences have also been noted between extroverts and introverts with the latter performing better in a vigilance task due to a higher arousal level (Harkins and Green, 1975, cited in Warm, 1984). Circadian and ultradian rhythms have also been demonstrated to play a role in vigilance performance (Blake, 1967, cited in Warm, 1984). General vigilance performance seems to improve from the morning to the afternoon (Davies and Davies, 1975, cited in Warm, 1984). The reverse is true for vigilance tasks which require the subject to use memory (Davies et al., 1980, cited in Warm, 1984).

The major problem with an arousal theory of the vigilance decrement seems to be the fact that a reduction in arousal has a number of different behavioral outcomes. While response latencies do increase with decreased arousal, the latencies of correct rejections and omission OEs do not change (Parasuraman and Davies, 1976). Furthermore, Parasuraman (1981) (cited in Warm, 1984) pointed out that pupil diameter, which is well correlated with arousal, did not correspond to performance on individual stimulus events. It can be said that overall levels of vigilance performance show a monotonic relationship to arousal level.

Habituation Theory

This theory, first proposed by Mackworth (1968) (cited in Warm, 1984) suggests that repeated stimulation, especially at high event rates leads to a decrease or elimination of the neural response amplitude (desynchronization) and a corresponding drop in vigilance performance. There is, in fact, little evidence directly supporting this position. Some experiments (i.e., Krulewitz et al., 1975) have shown that a change in stimulus event rates (dishabituation) does not necessarily increase the vigilance performance and in some cases will serve in a detrimental manner to performance.

Learning Models

Pavlovian and Hullian Inhibition Theory

Mackworth (1950) (cited in Warm, 1984) explained the vigilance decrement in terms of classical conditioning. Since Hullian conditioning theory was dominant at the time, it was natural to attribute the observed decrement in part to response inhibition, an hypothesized intervening variable producing a tendency not to repeat a response. The decrement was also attributed in part to a conditioned form of inhibition. A subject was reinforced for responding to signals during practice trials while during the actual vigilance task reinforcement was no longer present and therefore the tendency to respond to the signal was extinguished. It is, however, not clear from Mackworth's discussion whether he was referring to inhibition (1) of the overt responses to the detected signals, or (2) of the temporally preceding observing responses necessary for the detection, or both. Adams (1956) (cited in Warm, 1984) also emphasized the Hullian concept of reactive inhibition, but again it is unclear whether the inhibition referred to the detection responses or the observing responses. Such theories would appear to be generally more acceptable if the hypothesized inhibition referred to the observing responses, as detections rarely if ever drop to a zero level in an awake subject. This would be expected if inhibition applies to the observing response, for such responses were assumed to be reinforced whenever a signal was presented and recognized and therefore extinction should not typically be complete.

Skinnerian Theory (Observing Responses)

Holland (1958) (cited in Warm, 1984), has viewed vigilance performance in terms of Skinner's principles of operant behavior rather than in Hullian terms. He suggested that the responses necessary for observation of the signal and non-signal stimuli-eye movements, head movements, etc.,-continue to occur if they were reinforced by the occurrence of detectable signals, and they were extinguished in the absence of such signals.

Attempts to measure observing responses directly have met with mixed success; they were generally difficult to measure, especially in ways that do not interfere with the observing or vigilance conditions. As Mackworth (1964) (cited in Warm, 1984) reported, however, signals were missed as often when the subject emitted the observing response as compared to an absence of the observing response.

Information Processing and Criterion Shift Theories

Filter Theory

Broadbent (1958) (cited in Warm, 1984) suggested that the monitor's information--handling capacity is limited and that information is selected by a filter biased to receive information from some sources and reject it from others. It is also biased to reject the same or very similar chunks of information repetitively presented and to accept novel information. The net result is a decrement in hits and an increase in reaction times over time when observers were monitoring channels of information where there is considerable repetition or little variation in output.

The theory has been modified (Broadbent, 1971) to heed data presented by Moray (1969) (cited in Warm, 1984) and others, which indicate that the filter, like most filters in the real-life, is relative rather than absolute (i.e., the filter accepts some kinds of information more readily than other kinds, rather than simply accepting and rejecting different kinds of information). While the filter concept can be used to explain a wide variety of data, it is difficult to test its validity as opposed to that of other concepts.

Expectancy Theory

Signal probability has the most potent effects both on the absolute value of the decision criterion (ß) and on the changes in decision criterion over time. The effect of signal probability suggest a role for expectancy in the interpretation of the criterion increment, if expectancy is related to apparent shifts in the observer's subjective probability of signal occurrence over time.

The expectancy model was first suggested by Baker (1959), who suggested that participants typically expect temporal patterns very different from those than they actually encounter, and that the discrepancy between their expectations and the schedule actually encountered is the principal cause of the results in a vigilance experiment. The theory predicts that expectation, and therefore, performance, should be optimal at the mean inter-signal interval. Indeed, Mowrer (1940) (cited in Warm, 1984) had demonstrated much earlier in a simple reaction time situation that reaction time is least at the mean inter-signal interval where the foreperiod is varied over a number of trials. However, Baker (1959) has demonstrated that participants may have expectations regarding the mean inter-signal interval and its variability, but the expectations may be unrelated to the level of vigilance performance. Baker also suggested that knowledge of results may act to inform participants about the signal schedule. Signal-detection performance has been found to be better when signals occur with greater regularity (that is with less variability around the mean inter-signal interval), but generally only when the inter-signal interval is quite short--of the order of a few seconds (Smith et al., 1966, cited in Warm, 1984). This is consistent with findings that humans are not very precise estimators of time except at very short intervals. In any event, beneficial effects of knowledge of results and false knowledge of results did not differ, even when signals were regular, an effect hard to reconcile with Baker's interpretations of knowledge of results effects in terms of expectancy.

The optimal criterion for a vigilance task, assuming the costs and values of detection outcomes are equal and symmetric, is $(1 - p)/p$, where p is the *a priori* signal probability. Since the observer is monitoring a low-probability event, p was much less than .5, the criterion was much greater than 1, and thus the observer's (positive) response rate were less than the signal rate. The observer who monitors responses to signals in an attempt to estimate the future occurrence of signals will always underestimate the true signal probability. Thus the criterion will subsequently be made more stringent to conform to the (lower) estimate of signal probability. Making the criterion more stringent will result in a lower hit and false alarm rate, leading to further revision of the criterion, and so on in a *vicious circle* (Baker, 1959; Broadbent, 1971; Davies and Parasuraman, 1982). The result is a steady increase in the response criterion.

Craig and Colquhoun (1975) (cited in Warm, 1984) suggests that a major part of the decrement observed in many vigilance studies may be due to inappropriate expectancies developed in the pre-task period. Nevertheless, within-session fluctuations in expectancy, as outlined by the viscious-circle hypothesis, will lead to the small criterion shifts over time that were typically found in vigilance studies.

One implication of the viscious-circle expectancy theory is that the criterion should decrease over time if the *a priori* signal probability is greater than .5 (assuming observers over-respond initially). Williges (1969) (cited in Boff et al., 1986) had observers monitor a display for targets presented either with a low (.16) or high (.84) *a priori* probability. In addition, in a pre-training session observers were given either an accurate or inaccurate *set*, signals being presented at either an appropriate (p=.16 or p=.84) or inappropriate (p=.5) rate. The results conformed to the effects on the criterion predicted by the expectancy model. Given an accurate set, ß increased in the low signal probability treatment and decreased (although not significantly) in the high signal probability treatment. When observers were misled as to the appropriate signal probability and told it is .5, ß did not vary over time. These results were consistent with the assumption that participants use self-feedback to adjust their criteria during a session of continuous performance. Williges (1973) (cited in Boff et al., 1986) confirmed these results in a further study in which, in two separate conditions, targets were either presented infrequently (p = .1) or at the same rate as non-targets (p = .5). In a third treatment p changed from .1 to .5 halfway though the session. ß increased over time for the p = .1 treatment, remained stable for p=.5, and rose initially and then stabilized for the shifted probability treatment.

Although it is clear that expectancy effects do occur, the mechanisms though which they occur were not at all clear. For example, the effects could be identified with a change in observers' criteria for responding, with probability matching (Craig, 1978), or with a change in the level of adaptation to a specific stimulus or set of stimuli (Bevan and Turner, 1966) (cited in Boff et al., 1986).

Probability Matching

Craig (1978) proposed an alternative explanation of the criterion increment, a variant of expectancy theory, suggesting that criterion shifts reflect an attempt by the observer to match response rate to the signal rate, or a tendency toward probability matching. In reanalyzing data reported in 30 studies of vigilance in the literature, Craig found that in 14 of 20 studies reporting within-session data, the response rate is greater than the signal rate at the beginning of a session but approached the signal rate at the end of the session. Craig computed the ratio of the total number of responses (hits plus false alarms) to number of signals presented (R:S). In those studies in which the R:S ratio is less than 1 at the beginning of the session, the ration declined (but not significantly) thereafter. Craig suggested that these results show that observers approach probability-matching behavior at the end of a vigilance session.

Whether these group findings support the notion that individual criterion shifts were attributable to probability matching is problematical. There are four areas of difficulty. First, Dusoir (1974) (cited in Warm, 1984) and Thomas (1975) (cited in Warm, 1984) found that although group data were often indicative of the possible use of probability matching, individual participants show large deviations from probability matching. Thomas proposed a modification of the basic probability matching model to account for a systematic deviation from probability matching (e.g., undermatching,) but Dusoir (1980) (cited in Warm, 1984) found that in a threshold detection task participants did not consistently match, undermatch, or overmatch signal probability, so that neither the original nor modified versions of the model fitted more than a proportion of the participants tested. Departures from probability matching were also noted in another study by Craig (1980) in which participants monitored a display for two signals differing in discriminability and in *a priori* signal probability. A second difficulty is that probability matching implies that response probability should be invariant over different levels of discriminability, but the data contradict this prediction (Dusoir, 1974, 1980, cited in Warm, 1984). Third, although Craig (1978) found that the R:S ratio declined from a mean of 1.7 to a mean of 1.0 in 14 of 20 studies, it did so in only 56% of the 96 experimental conditions examined. In the other 42 cases the ratio was less than 1.0 initially and declined from 0.8 to 0.7. It is not clear whether all instances of significant vigilance decrements were restricted to the 54 cases for which R:S was greater than 1.0. Finally, if probability matching is employed, then presumably participants should adjust their response rates downward (when R:S>1.0) or upward (when R:S<1.0). However, Craig found that in the 42 cases where R:S was initially less than 1.0, the ratio did not approach 1.0 at the end of the session, but declined further.

Signal Detection Theory (TSD)

Egan et al. (1961) (cited in Warm, 1984) are usually credited with being the first to suggest that the vigilance decrement is due to a change to a more conservative criterion for responding, which results in a decline in both hits and false alarms. An earlier suggestion that the vigilance decrement is in fact a criterion shift was made by Howland (1958) (cited in Warm, 1984), but his formulation is less quantified and formalized than that of Egan et al. (1961). In any event, a change in criterion should be analyzable in terms of the theory of signal detection (TSD) (Green and Swets, 1966, cited in Warm, 1984).

One reason why TSD had not been applied earlier to the analysis of vigilance was that TSD experiments typically delimit the intervals during which participants were allowed to respond, while vigilance experiments do not. Egan et al. (1961) (cited in Warm, 1984) suggested that by observing the operant rate of responding at intervals following signal presentation, the times during which responses should be considered hits could be effectively determined, as could the times when they should be considered false alarms. Still another approach could be to use data of this kind to determine the inter-stimulus intervals at which to present stimuli.

Numerous experiments have attempted to apply TSD methodology to the vigilance situation. In experiments by Broadbent and Gregory (1963a, b) (cited in Warm, 1984) and by Loeb and Binford (1964) (cited in Warm, 1984), observers rated their confidence that signals were being detected. The findings regarding both the certainty ratings and the computed TSD indices (d′ and ß) indicated that there was a progressive increase in conservatism in responding as a function of time-on task. This result could be taken as evidence for a TSD theory of vigilance and against other theories (e.g., filter theory, observing response theory, and neural habituation theory), though it could be argued, and often has been, that these other mechanisms mediated changes in performance as measured by TSD indices. It is not entirely clear just why a criterion shift should occur as a function of time-on-task. Possibly observers come to realize that they were often responding to non-signals and raise their criteria accordingly.

Changes in the form of the receiver operating characteristic cumulated over increasingly stringent response categories have been noted for individual data from several earlier experiments in a recent article by Craig (1977). Craig concluded that the data supported TSD in a majority of the cases, but not in all, and he suggested that the remaining data might better be explained either by other signal detection models, such as those of Luce and Green (1972) (cited in Warm, 1984) or McGill (1967) (cited in Warm, 1984), or by the lack of reliability of vigilance data. Loeb (1978) (cited in Warm, 1984) has suggested that the latter alternative is much more probable; i.e., relatively high reliability should not be expected given (a) the small number of signals presented, (b) the small amount of training as compared with typical TSD experiments, and (c) the very low degree of certainty of participants in typical vigilance experiments.

Nevertheless, TSD interpretations have been suggested in other studies. For example, Williges (1976) (cited in Boff et al., 1986) manipulated parameters such as signal probability, payoff, and knowledge of results; he found performance consistent with TSD and the ideal observer hypothesis. According to his idea Williges (1969) (cited in Boff et al., 1986), an observer initially adopts some intermediate ß and later shifts it in the direction of an optimal criterion. If signal probability is less than 0.5, which is almost always true in the vigilance situation, this should results in an increase in ß, as it did in his studies.

Other experiments on the effects of differential payoff and feedback on performance have yielded ambiguous and complex data. Payoff has been found to influence ß, but not d', and increasing costs for false alarms impaired performance, while changing payoff for hits had no effect (Levine, 1966, cited in Boff et al., 1986). It may be that the symmetrical payoff matrix employed is better suited to a conventional detection situation (where signal and non-signal stimuli were equal in probability) than to the vigilance situation (where the signal events were much lower in probability than non-signal events). Also, since human detectors can be expected to have their own values (not entirely monetary), their performance may not be totally determined by the differential (but typically unscaled) payoff set by the experimenter.

The Ideal Observer

The studies of Williges (1969, 1971, 1973) (cited in Boff et al., 1986) support the expectancy view of the criterion increment when interpreted within the normative framework of TSD. Williges (1973, 1976) (cited in Boff et al., 1986) took this interpretation one step further. He suggested that over time the observer's behavior approaches that of the ideal observer. Williges proposed that since criterion values approach the optimal ß value toward the end of a vigilance session, the observer's performance actually represents an attempt to reach optimal behavior; this he termed the vigilance increment (Williges, 1976, cited in Boff et al., 1986).

The ideal observer hypothesis suggests that if observers are provided with sufficient exposure to the task and given training to develop appropriate expectancies about signal occurrence, they will eventually adopt response criteria that are optimal, and then the chances of any further vigilance decrement are minimized. The idea that response criteria can be stabilized, given appropriate training and sufficient practice, provides a useful general principle with application not only to vigilance performance but also to performance on a range of other tasks where criterion shifts are obtained over both short and long intervals of time, as in tasks involving monitoring and supervisory control (Moray, 1976, cited in Boff et al., 1986).

Information Theory and Channel Capacity

The vigilance situation may be viewed as a choice reaction-time task, with relatively infrequent signals to which the subject responds actively by pushing a button or a key, and relative frequent non-signals for which the subject's expected response is to do nothing. If this is the case, then other models of signal processing, such as those derived from information theory (Dember and Warm, 1979, cited in Warm, 1984), may be employed to explain some facets of vigilance behavior. For instance, Kulp and Alluisi's (1967) (cited in Warm, 1984) finding that vigilance performance is a decreasing function of signal-response uncertainty fits into this framework. Similarly, both signal rate and regularity effects have been explained in terms of a metric of subjective uncertainty (Smith et al., 1966, cited in Warm, 1984). This sort of interpretation might also partially explain the previously cited findings of Hawkes et al. (1964) (cited in Warm, 1984) that overall detection is poor and a decrement in hits occurred when a vigilance task is part of a complex of tasks, since channel capacity may have been approached or exceeded on their complex task, and any facilitating effect of greater arousal might thereby have been counteracted.

Parasuraman (1979), after surveying the literature and attempting a systematic classification, has suggested that situations in which a true decrement (a decline in d') occurs generally involve a memory load for the subject. Presumably, such a load imposes a continuous strain in that it requires a sustained effort; the result is a cumulative decrease.

Fisk and Schneider (1981) hypothesize that there are two kinds of information processing--an automatic mode, involving parallel search, not limited by STM, and a controlled-processing mode, involving high attentional effort and limited comparison rate serial search, which is quite susceptible to the effects of task load (Schneider and Shiffrin, 1977). Fisk and Schneider (1981) suggest that a vigilance decrement should occur only with the latter, controlled-processing, situation.

Adaptation Level

Still another general information-processing model--Helson's adaptation-level theory--has been applied by Bevan (1965) (cited in Warm, 1984) to the vigilance situation. According to this view, vigilance depends upon both a subject's state of arousal and his expectancies in rather complex interactions: (a) to some extent, his expectancies determine his arousal, which is maximum when expectation is ambiguous, and (b) both arousal and expectancy influence the effects of stimulus variables, situational background variables, and personal residual variables, as in other applications of adaptation level theory (Helson, 1964, cited in Warm, 1984). This model incorporates both arousal and expectancy, but it combines them in a way not entirely predictable from their usual formulation.

Vigilance Task Classifications and Stimulus Parameters

Since Mackworth's initial investigations, a plethora of experimental tasks has been used in the study of vigilance behavior. A number of task dimensions can be specified which might be considered to be special features of vigilance situations. As described by Jerison (1970) (cited in Warm, 1984) and by Warm (1977) (cited in Warm, 1984) they include the following: (1) the task is prolonged and continuous, often lasting for 1/2 hour or more; (2) the signals to be detected are usually clearly perceivable when the observer is alerted to them, but would seem weak to most observers because they are not compelling changes in the observers' operating environment; (3) the signals to be detected occur infrequently, aperiodically and without forewarning; and (4) the observer's response typically has no effect upon the probability of appearance of critical signals. It should be emphasized that all vigilance tasks do not conform to these characteristics in a hard and fast manner. There are exceptions, as for example tasks in which critical signals appear quite frequently (e.g., Jenkins, 1958, cited in Warm, 1984) and tasks of very brief duration (e.g., Davies, 1968, cited in Warm, 1984). These four dimensions, however, seem to capture the special characteristics of most vigilance tasks.

Discrete-Dynamic

In some cases, vigilance tasks have involved relatively simple displays in which observers are required to detect the onset or conclusion of a discrete stimulus event. In such cases, all stimulus or non-stimulus occurrences are critical signals for detection. In most cases, however, more complex, dynamic displays have been used in which participants are required to observe a stream of repetitively presented neutral stimulus events for specified changes that constituted critical stimuli. Mackworth's clock test is an example, as is a case in which participants must listen to repeated pulses of acoustic stimulation of fixed duration for occasional longer pulses. In these situations only some stimulus events are critical signals.

The discrete-dynamic distinction is important for two reasons. First, the decrement function is more pronounced for dynamic tasks compared to discrete tasks (Davies and Tune, 1969). Secondly, the presentation rate of the neutral events or background rate has been shown to play a critical role in vigilance performance (Dember and Warm, 1979, cited in Boff et al., 1986); Parasuraman, 1979).

Sensory-Cognitive

Two types of signals can be used in vigilance tasks which are distinguished along a sensory-cognitive dimension (Davies and Tune, 1969). In the former, signals and non-signals vary along a physical dimension. The jump of the pointer in Mackworth's experiment (1950) (cited in Boff et al., 1986), 0.3 inches versus 0.6 inches, would be an example of a sensory discrimination. In the latter, the discrimination is more of a symbolic nature. Bakan (1959) (cited in Boff et al., 1986), for example, asked participants to detect a sequence of three consecutive odd digits.

Performance on cognitive tasks differs in important ways from vigilance experiments employing sensory tasks (Sprague, 1981, cited in Boff et al., 1986; Lysaght, 1982, cited in Boff et al., 1986). In a cognitive vigilance task, signal conspicuity takes on a different meaning compared to sensory tasks. In order to make a differentiation between signals and non-signals, certain mental operations are required since there is no inherent physical difference between the two types of signals. Signal conspicuity then refers to the number and complexity of the mental operations involved in the discrimination.

Criterion shifts have been found ubiquitously in a variety of detection, discrimination, and monitoring tasks. Performance changes arising from sensitivity shifts, however, are found only in some vigilance tasks. Inter-task correlations in vigilance performance are low for some task pairs and high for others (Davies and Tune, 1969). Some researchers have interpreted these findings as implying that vigilance performance is completely task specific (Buckner and McGrath, 1963). However, patterns of consistency in vigilance performance occur if a taxonomy is used to classify different types of vigilance task (Parasuraman, 1976; Parasuraman and Davies, 1977, cited in Boff et al., 1986).

A variety of tasks have been used in research on human performance, generating large amounts of data. Using the data for predictive or generalization purposes is limited because classification techniques for describing the important features of the tasks used have not been fully developed. Although isolated attempts have been made to identify the common features of vigilance tasks (Bergum, 1966, cited in Boff et al., 1986; McGrath, 1963, cited in Boff et al., 1986), very little attention has been paid to the development of a task classification system for vigilance and monitoring tasks. Parasuraman and Davies (1977) (cited in Boff et al., 1986) identified some of the more important task categories: sense modality, source complexity, response type, coupling, signal duration, event rate, attention requirement, stimulation value, and task abilities.

Parasuraman and Davies (1977) isolated four categories from the categories afore stated as being of particular importance for the specification of sensitivity decrement and the analysis of inter-task correlations in performance: sense modality, source complexity, the event rate, and signal discrimination type.

Of the four categories, signal discrimination type is considered to be of particular importance. Parasuraman (1979) distinguished vigilance tasks on the basis of whether signals required simultaneous or successive discrimination. In successive discrimination, the signal is defined as a change in some feature of a repetitive standard (non-signal), the standard value being absent when the nonstandard (signal) is presented; thus, a successive comparison of a change in a standard value held in memory must be made. In simultaneous discrimination, signal and nonsignal features are present within the same stimulus event or at the same time, as in the detection of a target item in a display containing other distractor items. The important distinguishing feature between these two signal discrimination types is that the successive discrimination signal imposes an additional load on STM since signal and nonsignal features are not presented simultaneously. The successive/simultaneous distinction has been found to be an important task dimension in vigilance (Parasuraman and Davies, 1977, cited in Boff et. al., 1986). In more recent work, Davies and Parasuraman (1982) have suggested a task taxonomy that distinguishes between types of vigilance tasks (simultaneous versus successive) and between event rates (high versus low). They hypothesize that a vigilance decrement during a successive discrimination task with a high event rate is due to a decrease in the sensitivity of the observer [d'] (i.e., habituation). For any of the other combinations of event rates and type of task, they hypothesize that the vigilance decrement is due to an elevation of the observer's response criterion (ß).

Complex Monitoring

As a direct result of the introduction of automation technology in many work environments, humans and computers often have to serve as joint monitors of critical events, processes, or entire systems. Such human-computer monitoring is now an important aspect of many modern systems, including commercial aircraft and ATC.

User interaction with such systems may be influenced by various factors related to human attention, among other perceptual and cognitive processes. For example, ATC represents a human-computer monitoring system in which vast numbers of aircraft in the immediate airspace have to be tracked. ATCSs in the system have to monitor complex symbolic displays that present coded information representing aircraft call name, airspeed, altitude, and so on in a dynamic manner. Effective separation of aircraft is possible only insofar as the ATC system does not overburden the ATCS's limited selective attention and memory capabilities, and only if the ATCS is able to remain vigilant for potentially dangerous conditions.

The monitoring function in ATC and related systems is usually shared among the human operator, an automated subsystem, and other instruments. A representation of the monitoring functions is shown in Figure 22. The distal variables represent the objects (e.g., aircraft in an ATC system) that must be monitored. A variety of information sources are provided to the human operator: instrument panels providing transduced versions of information about the distal stimuli (e.g., radar), the outputs of a computer monitoring system, information sources associated with other operator tasks, and occasionally the distal stimuli themselves. Information is presented on these sources at rapid rates and is dynamically updated. Attention must be allocated selectively to one or more of these information sources; sustained attention to a single source or multiple sources is also required. The human operator may elect to switch in to one, two, or all three (or none) of these information sources.

Human-computer monitoring represents a prototypical task of human-computer interaction. The importance of this task is likely to increase in the future as the extent of automation increases. Modern computer systems are capable of presenting vast amounts of information to the human user. As information load has increased and systems have become more complex, automation has been touted as a means of reducing the probability of human error in human-computer interaction. For example, the new Advanced Automation

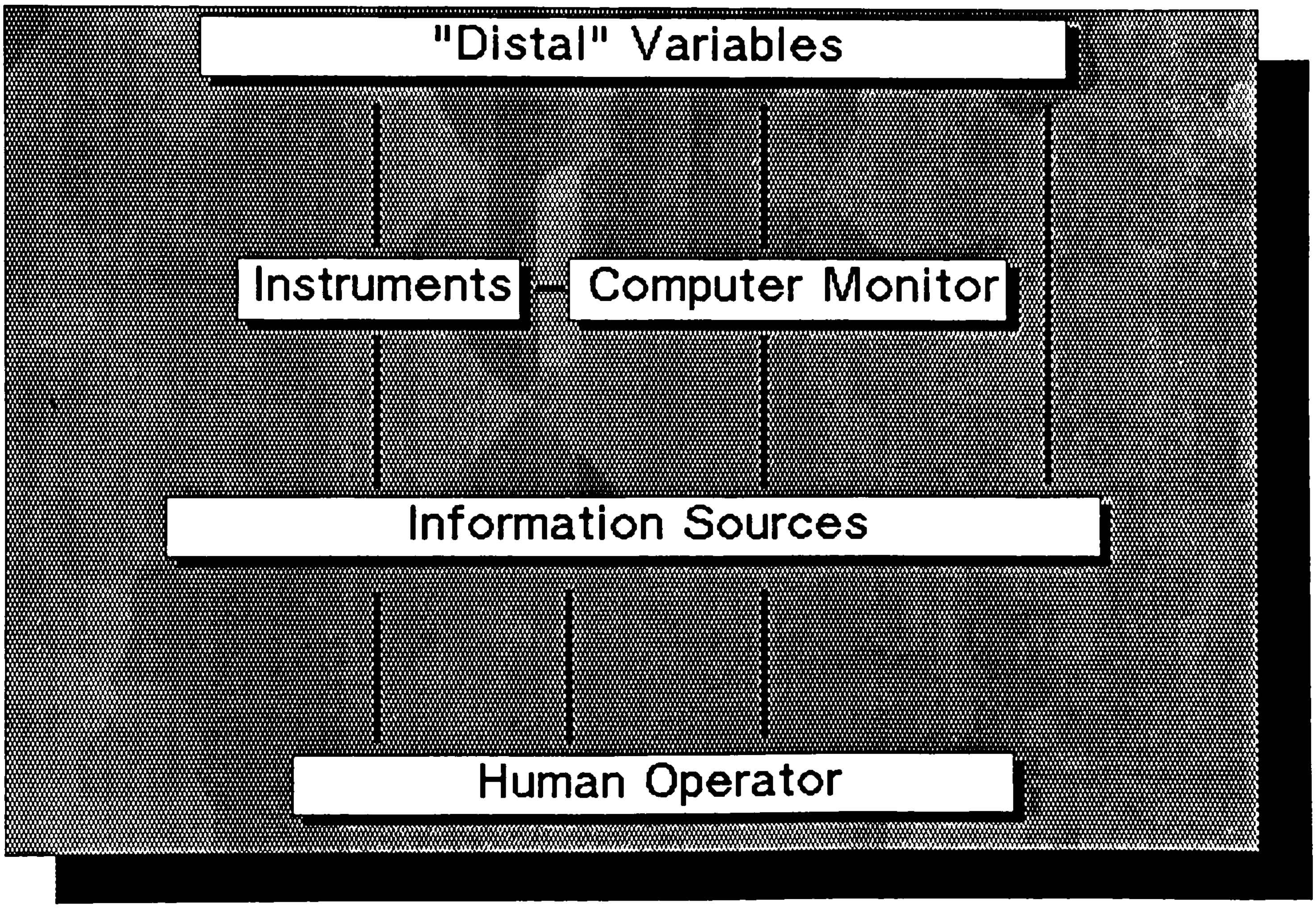

Figure 22
Monitoring Functions in ATC

System being developed for the FAA will automate many ATC functions now done manually and will provide new information to the ATCS on potential conflicts between neighboring aircraft. Automation of en route ATC is also in the planning stages; such a system will provide information on potential flight path conflicts between one aircraft and another and will present proposed resolutions.

It has become clear in recent years that the impact of automation on total system performance demands further examination and understanding. As Wiener and Curry (1980) have noted, the major issue in cockpit automation is no longer whether certain flight deck functions can be automated but whether they should be. One of the principal issues to be considered is the impact of automation on human attention capabilities; for example, the ability of human operators to monitor system failures or emergency events in highly automated environments.

Automation and related technological developments are driving much current research on human-computer interaction. Research on attention and vigilance also began in response to technological developments but to an older technology than computers--radar. The successful deployment of improved radars during World War II is thought to be critical to the defense of Britain. However, problems were encountered. Instances were reported of airborne radar operators missing the presence of potential submarine contacts on the radar screen or misidentifying Spanish fishing vessels and large whales as enemy submarines. Despite a high level of motivation and skill, the radar operators exhibited the vigilance decrement.

Despite this history, technological developments soon overtook the scope of laboratory research on attention and vigilance. Most laboratory studies have investigated vigilance for simple sensory events such as tones and light flashes presented over a single source. With some exceptions (e.g., Adams, 1963), studies of complex displays (Parasuraman, 1986) and cognitive vigilance (Warm, Howe, Fishbein, Dember, and Sprague, 1984, cited in Boff et. al., 1986) have been carried out only recently. This contrasts markedly with the multi-source, information-rich environment of current computer systems. In modern systems signals may be detected by instruments and responses carried out by robots, both under the control of an intelligent executive computer. Operators may no longer directly sense the critical signals they must watch for. Sensors may feed incoming information to preprocessors and then to automated subsystems that detect and classify such information. Rule-based expert systems may even interpret or diagnose the information. The users of today's human-machine systems are thus gradually being relieved of much of the active control once required in extended observation.

Attentional factors may still affect system performance despite this change in the nature of operator control. A potential attention problem may still remain because the human may be required to detect an infrequent but critical treatment or the failure of the automated system itself. The changing role of the human operator from active cpntroller to passive monitor has not eliminated the vigilance problem but merely changed it (Wiener and Curry, 1980).

As noted previously, many studies of vigilance have used fairly simple single-source monitoring tasks that differ from the multi-source tasks found in typical automated work environments. In early studies in which more complex displays were used little or no decrement in detection rate over time was found (Adams et al., 1961, cited in Boff et. al, 1986; Broadbent, 1951, cited in Boff et. al, 1986; Howland, 1958, cited in Boff et. al, 1986; Jerison and Wing, 1957, cited in Boff et. al, 1986). This led to the view that the vigilance decrement is found only with very simple displays, and provided ammunition for those critics who contended that the decrement is not found in real settings, where generally complex rather than simple displays are to be found

(Kibler, 1965, cited in Boff et. al, 1986). The common view developed that (a) the vigilance decrement does not occur in complex monitoring situations because such situations are sufficiently stimulating to prevent changes in vigilance over time (Adams, 1963, cited in Boff et. al, 1986), and (b) in the few cases in which such a decrement is observed it is too small to be of any practical consequences.

The work of Adams et al. (1961) (cited in Boff et. al, 1986) has often been quoted in support of this position. Adams et. al developed a vigilance task that represented a significant departure from the traditional simple vigilance task used since Mackworth's (1948) (cited in Boff et. al, 1986) wartime research. The task used symbolic stimuli, and target detection is not based on the sensory qualities of the stimuli. The task is meant to simulate an air defense situation and consisted of a circular display containing alphanumeric symbols each representing a moving aircraft. The symbols moved slowly across the screen (although they appeared almost static because a 1000-mile areas is meant to be simulated), and signals (position and configuration changes of a symbol) persisted for 20 seconds. Participants monitored the display for occasional symbol changes over a 3-hour session.

Adams et al. (1961) (cited in Boff et. al, 1986) varied the nonsignal target density so that at any time the circular display contained a total of either 6 or 36 nonsignal symbols. Although detection latency was significantly longer when signals were presented in the more dense display, the changes in response time with time at work were small and not significant. These and other related results obtained by Adams and colleagues were interpreted as supporting the view that the vigilance decrement does not occur with complex tasks because they are sufficiently stimulating to prevent changes in arousal over time (Adams, 1963, cited in Boff et. al, 1986; Frankmann and Adams, 1962, cited in Boff et. al, 1986) and that, in the few cases where such a decrement is observed, it is too small to be of any practical consequence.

The results of other studies of complex monitoring do not entirely support the view that participants can monitor complex displays without significant performance decrement. The tasks and procedures used in these studies vary so widely that it is not possible to classify individual tasks along a dimension of task complexity. Nevertheless, the weight of the evidence suggests that task complexity does not affect the vigilance decrement. Recent studies of vigilance with complex multi-source displays provide evidence for the existence of the vigilance decrement in complex displays. The studies have examined simulated ATC (Parasuraman, 1986; Thackray, Bailey, and Touchstone, 1979), ship navigation (Schmidtke, 1966, cited in Boff et. al, 1986), industrial inspection (Saito and Tanaka, 1977, cited in Boff et. al, 1986), surveillance (Tickner and Poulton, 1973, cited in Boff et. al, 1986), and train driving (Haga, 1984, cited in Boff et. al, 1986). As reviews of these and other studies indicate (Craig, 1984, cited in Boff et. al, 1986; Parasuraman, 1986), the results do not entirely support the view that human operators can monitor complex displays without significant performance decrement.

Three possibilities may account for the failure of early studies (e.g., Adams et al., 1961, cited in Boff et. al, 1986) to confirm the existence of a vigilance decrement in complex monitoring tasks. First, individual differences in complex task performance may be so large that performance decrements over time may not reach statistical significance. Second, performance on complex tasks may be already poor at the start of the vigilance session, so that performance cannot get much worse with time (Davies and Tune, 1969). Third, it is possible that performance decrement does occur in the form of a decrement in sensitivity. This could occur even if the decrement in detection rate is negligible if, for example, the false detection rate increased with time.

Howell, Johnston, and Goldstein (1966) (cited in Boff et. al, 1986) carried out a series of studies on vigilance effects on complex monitoring. Howell et al. had participants detect specified signals among alphanumeric stimuli on a computer-generated CRT display. The display consisted of an 8 X 8 matrix of cells. Each cell in the matrix contained either an asterisk or an alphanumeric symbol. The signal is an addition to or a deletion of one of the symbols in successive displays of the 8 X 8, 16, or 32 symbols per display), signal frequency (30 or 75 signals per hour), and monitoring time (two hours). There is a significant increase in mean detection latency with monitoring time for the highest-density (32 symbols), low signal frequency (30 signals) treatment, but not for the other conditions. Similar findings were also reported by Thackray, Bailey, and Touchstone (1979). Thackray et al. simulated an air-traffic control task in which alphanumeric symbols (representing aircraft identification, airspeed, and altitude) had to be monitored for the occasional occurrence of a signal. The signal was specified as an aircraft at a particular, critical altitude. These critical signals appeared, in different conditions, in displays containing either 4, 8, or 16 noncritical stimuli, each representing an aircraft. There was a significant increase in mean detection latency with monitoring time for the highest-density, low signal frequency treatment, but not for the other conditions.

In the study by Howell et al. (1966) (cited in Boff et. al, 1986), the mean detection latency increased from 8 to 22 seconds from the beginning to the end of a two-hour monitoring period. This increase in mean response time is considerably greater than that reported by Adams et al. (1961) (cited in Boff et. al, 1986), and in a further analysis of their data Howell at al. showed that the increase in mean reaction time primarily reflected an increase in the slower rather than the faster reaction times. A similar result was obtained by Thackray et al. (1979). The pattern of an increasingly skewed reaction time distribution has also been reported in previous studies with unlimited-hold signals. Observers performing unlimited-hold tasks in which signals are present for very long periods or until detected do not show changes in correct or false detection rates over time. However, detection latencies increase and the vigilance decrement is characterized by very long reactions, the frequency of which increase with time, in a manner suggesting periodic lapses of attention or *blocks*, as originally postulated by Bills (1931) (cited in Boff et. al, 1986). The observation that similar results were obtained with complex monitoring tasks indicates that task complexity per se cannot be the factor responsible for the finding that decrements in such tasks were occasionally not found. Yet, at the same time, the added visual search and scanning requirements of these task could interact with factors responsible for the decrement in an as yet unspecified manner. Howell et al. (1966) (cited in Boff et. al, 1986) suggested that the occurrence of the decrement in the high-density treatment is due to the need for display scanning, which is not present in simple vigilance tasks.

The results of studies of complex monitoring show that performance decrements do occur in such tasks. This suggests that the same factors that affect the vigilance decrement and the overall level of vigilance in simple detection and discrimination tasks also affect performance on complex monitoring tasks. At the same time, other factors may also be important in understanding complex monitoring performance.

For real-life or operational setting tasks, assuming that a minimum level of efficiency can be identified for a particular operational task, vigilance may be deficient if a vigilance decrement occurs, or the overall level of vigilance is too low, or both. There are four possible vigilance performance profiles that might be encountered in an operational setting: (a) efficiency might be satisfactory to begin with, but declines to a substandard level subsequently; (b) a vigilance decrement might occur but is operationally insignificant; (c) no vigilance decrement might occur, but the level of vigilance is below the minimum level; and (d) performance might be satisfactory throughout the work period.

According to this analysis, therefore, identifying deficiencies in vigilance in operational settings is not just a matter of finding performance decrement over time. A vigilance decrement may occur, but it may be irrelevant operationally. On the other hand, the failure to obtain a decrement does not indicate the lack of a problem of vigilance. Thus evaluation of vigilance requires assessment of both vigilance decrement and the level of vigilance.

In a compilation of a number of studies of vigilance that have employed either operational tasks or simulations of operational tasks, Davies and Parasuraman (1982) wished to emphasize whether the study reported a significant vigilance decrement over time as a function of subject experience with the task. The participants tested in these studies were either unpracticed volunteers (usually students), practiced volunteers (students or military personnel), or actual operators drawn from some occupational or military group (usually sonar/radar operators or industrial inspectors). The sample of studies illustrates two important points: (1) trained and practiced volunteer participants, as well as experienced operators, were just as likely to show a performance decrement (or not) as were naive, unpracticed, or untrained operators; and (2) in any situation requiring sustained performance on a task, there may or may not be a vigilance decrement, regardless of whether the task is artificial, a simulation of a real task, or an actual operational task.

The first point suggests that the motivational differences that undoubtedly exist between participants in laboratory studies and actual operations cannot be the sole reason for the lack of vigilance decrement in operational settings, as suggested by Nachreiner (1977) (cited in Boff et. al, 1986). Furthermore, practice can improve vigilance performance, but practice alone is insufficient to abolish the decrement (Davies and Tune, 1969). Poulton (1973) (cited in Boff et. al, 1986) reviewed a number of studies of vigilance relevant to industrial inspection performance. He noted several instances of studies employing repeated sessions in which reliable within-session decrements in detection rate were found following extensive practice, although in many cases the decrement is less marked in the later session. Practice can improve the level of efficiency in both simulated sonar tasks (Colquhoun, 1975 , cited in Boff et. al, 1986) and in industrial inspection jobs (Thomas, 1962, cited in Boff et. al, 1986; Wiener, 1975, cited in Boff et. al, 1986), but although it may attenuate the vigilance decrement, it does not necessarily eliminate the decrement.

With regard to the second point, it seems clear that vigilance decrements can occur in operational tasks. Admittedly, there is little evidence of decrements in actual operational settings, but the existence of significant decrements in operational task performance in experimental conditions suggests that it would be unwise to conclude that such decrements cannot occur.

The complex monitoring tasks considered thus far have limited face validity for many of the tasks in modern automated and semiautomated systems because of their pre-dominant use of discrete, independent trials, with specified signals. An additional dimension of tasks in modern human machine systems such as ATC may involve continuous rather than discrete stimuli and signals that are not uniquely specified (e.g., a 100-Hz alarm) but must be inferred from the system state and other information (e.g., an increase in boiler temperature and pressure over baseline conditions).

Some of the properties of this type of inferential monitoring were recently developed into a preliminary taxonomy of the different information-processing activities involved in supervisory control (Moray, 1980). Some of the activities that the supervisory ATCS must perform that Moray has discussed include intake of information, interpretation of information, decision making, incorporation of information into permanent storage, extrapolation and prediction, and generation of information and action. Because systems are increasingly being fully or partially automated, the practical importance of this type of monitoring is also increasing, although the amount of research directly relevant to the fundamental issues in this area is still small.

Even less is known of the role of vigilance in inferential monitoring. Does error detection become less reliable with time? Does the effectiveness of a warning device signaling the failure of a component in a semiautomated system vary with the operator's level of vigilance? These and related issues are only just beginning to be explored. The recent increase in research interest in mental work load (Moray, 1979) has been stimulated, in part, by the requirement to measure and reduce work load levels in complex semiautomated systems. But extremely low levels of work load can be as pernicious as high work load. Furthermore, if, as is commonly held, extremes of work load constitute a source of stress (Caplan, Cobb, Franch, Harrison, and Pinneau, 1975, cited in Boff et. al, 1986), then, as Thackray (1981) has pointed out, efforts to reduce work load through increased automation may have the ironic effect of replacing one source of stress with another.

A system where inferential monitoring is involved can be either machine controlled (automated) or operator controlled (semi-automated). In the automated treatment the observer has only to monitor the system; in the semiautomated treatment the observer both controls and monitors the system. An important issue in inferential monitoring is whether monitoring performance is better in the fully automated or in the semiautomated treatment. Another important issue is whether information on the system being monitored should be presented continuously or only at discrete periods of time. The effects of these two factors--whether information is presented continuously or discretely, and whether the observer actively controls the system or only monitors it--have been examined in some recent studies of monitoring performance. Ince and Williges (1974) (cited in Boff et. al, 1986) examined the effects of continuous display presentation and active control on monitoring performance. Observers were required to perform a visual tracking task and also to detect slow changes in the control dynamics. The change in control dynamics could occur at a rate of 3, 6, or 9% per second. It might be expected that participants' response time would be fastest at the highest signal rate, the 9% per second treatment. In fact, the opposite is true, participants being relatively sluggish to respond when the control dynamics changed quickly and relatively swift when the change occurred slowly. No changes in response time over time at work were reported in this study, but such data were reported in a study by Wickens and Kessel (1979) (cited in Warm, 1984). In this study operators had to detect not gradual but abrupt changes in the order of control dynamics; in addition, participants either actively controlled the system or only monitored it. Wickens and Kessel found that detection speed is slower in the monitoring than in the active control treatment. However, no decrement in performance over time occurred in either treatment.

These results show that inferential detection can be performed without a vigilance decrement, whether the event being detected is an alarm or a change in some parameter of the control process being monitored. As in other complex monitoring tasks, however, a deficiency in the level of vigilance, rather than vigilance decrement, could lead to substandard performance on such tasks.

Spatial Uncertainty And Search

The temporal uncertainty associated with the appearance of a signal is a major attribute of pure vigilance tasks. This is not true of pure search tasks in, which spatial uncertainty is the major factor of interest. Two issues regarding search and vigilance are important: (1) the effects of search on vigilance performance, and (2) the effects of vigilance on search performance. More is known about the effects of search on vigilance.

Search can have both beneficial and detrimental effects on vigilance performance. Performance may benefit because the vigilance decrement may be reduced in combined search and vigilance tasks, although it must be noted that no systematic analysis of performance on such tasks has been carried out to confirm if it is the search requirement per se or some other aspect of the task that is responsible for lowered decrement.

Search can benefit performance if the display is constructed so as to direct the observer's search along regular paths. This is the reasoning behind the use of rotating sweep lines in radar displays such as the PPI radar studied by Mackworth (1950) (cited in Boff et. al, 1986) and Baker (1958) (cited in Boff et. al, 1986). The PPI radar uses a circular display with a radial sweep line rotating clockwise around the center of the display. The radar operator's task is to search for targets in noise, usually brief-duration pips that appear in the vicinity of the rotating sweep line. In modern radar displays, such as those used in ATC, the targets may be represented as alphanumeric symbols on a graphics display. Thackray and Touchstone (1980) examined the effects of a rotating sweep line on detection performance with such a display. The sweep line rotated clockwise around the display once every six seconds. Observers monitored the display either with or without the sweep line. The stimuli were rectangular blips (representing the location of aircraft), adjacent to which were two alphanumeric symbols, the aircraft identification symbol and a six-digit number giving the aircraft ground speed and altitude. The critical signal is a change in displayed altitude to a value greater than 550 or less than 150. Thackray and Touchstone found that reaction time to critical signals increased over a two-hour period of monitoring. However, although there is a slight reduction in reaction time with the introduction of the sweep line, the effect is not statistically significant. Nevertheless, the use of the sweep line can be considered beneficial since the performance decrement over time is reduced.

Search can also harm vigilance performance. Performance may suffer for two reasons. First, although vigilance performance may not decline, the level of performance with a display requiring extensive search may be unacceptably low. Second, although the overall level of search and vigilance performance may be acceptable, the detection of particular items in the display, particularly those on the periphery of vision, or those carrying low-value information, may deteriorate if the search requirement is increased. In free search, where the observer is free to search the display in any manner (as in real-life operations), the particular search path chosen may tend to neglect certain areas of the display. Most often, observers' scan paths tend to avoid the edges of the display, a phenomenon known as the edge effect (Baker, Morris, and Steedman, 1960, cited in Boff et. al, 1986; Enoch, 1959, cited in Boff et. al, 1986). Enoch found that observers made more fixations of the central areas of a display (aerial photographs) in detecting targets than of the edges of the display, regardless of the size of the display. Baker et al. extended this finding further and showed that detection of centrally located targets is faster than detection of targets located near the edge of a circular display. The edge effect has also been observed in studies of experienced industrial inspectors searching complex products for small flaws. Schoonard, Gould, and Miller (1973) (cited in Warm, 1984) found that inspectors tended not to fixate on the edges of slides of integrated circuits and thus tended to miss faults that occurred at these locations.

Only a few studies have examined the effects of prolonged performance on the edge effect. In the Schoonard et al. study no decrement in inspection efficiency over time is reported, regardless of the location of the flaws. Baker (1958) (cited in Boff et. al, 1986) found that for radar pips presented equiprobably at 16 different locations on a simulated PPI display, periphery targets were detected less frequently than those presented centrally. During a prolonged session lasting about an hour, observers missed 10-15% more targets if they were presented at outer locations than if they were presented at inner locations on the circular display. However, there is no vigilance effect found, and no change in the edge effect over time is reported. Colquhoun and Edwards (1970) (cited in Boff et. al, 1986) required observers to detect a target disk in a display containing six horizontally spaced disks. A distinct edge effect is found, the detection rate of targets presented at the outer locations being lower than those presented centrally. There is no significant changes in the edge effect with time on task.

These studies demonstrate that targets presented at peripheral parts of a display were less well detected than more centrally presented targets in prolonged search tasks. Prolonged performance does not, however, lead to a further deterioration in the detection of peripheral targets. In general, therefore, search may depress the overall level of vigilance performance, but it does not affect the vigilance decrement.

Overview Of Mental Work Load

To deal with the topic of work load adequately is difficult. Not only is there no single, commonly accepted definition of work load, there are many conflicting concepts of what work load is (Hart, 1985, cited in Boff et. al, 1986; Kantowitz, 1985a, cited in Boff et. al, 1986; Moray, 1979) and often the term is used without any definition at all. There is also the difficulty of differentiating work load from stress which has many features in common with work load. Often the term work load and stress are used interchangeably. There are two problems of definition. As an input, work load is represented by stimuli that load the operator in the sense of bearing a burden. Work load is also the operator's internal experience of difficulty and discomfort, recognition of experiencing a load, and the strategy to overcome it. As an output, work load affects not only performance (and can be measured) but, when the operator is part of a larger system, impacts (usually negatively) on the system itself. Work load can be viewed, then, as some feature of the system (even when it is one's own incapacity) that forces the operator to work harder, as the operator's feeling of being stressed and having to work harder, and as the effect of all the preceding that cause him to make errors.

Other indices of multidimensionality are that the stimuli initiating the work load process can be either outside the operator or within oneself. They can be physical or mental, can be expressed overtly or internally, and can be modified by the operator's intelligence, attitudes, experiences and knowledge of consequences. They can, additionally, result in performance modifications, such as a change in problem solution strategy. Any stimulus can be a load-producing stimulus, almost any response can be the operator's response to the recognition of load.

Often a failure to perform is easy to explain: the task is evidently beyond the capacities of the operator. The demands imposed on the operator by the task have become larger than they are when performance is at its best. It is at this theoretical juncture that the concept of work load is normally introduced. That is, work load is invoked to account for those aspects of the interaction between a person and a task that cause task demands to exceed the person's capacity to deliver. The concept is needed only for those cases in which the required performance of a specified quality is clearly within the performer's current repertoire.

The current concept of mental work load implies that limitations exist in the information processing structures, making it difficult for a person to fully use the information processing apparatus in the service of the target task. The key assumption is that to perform a task the operator uses effectors through which the overt responses are made as well as sensors through which information is gathered. The path between sensors and effectors is an elaborate information processing apparatus with structural properties and a limited capacity. These limitations on the capacity of the information processing system must be measured and modeled if we are to account for performance failures attributed to mental work load.

In other words, mental work load may be viewed as the differences between the capacities of the information processing system that are required for task performance to satisfy performance expectations and the capacity available at any given time. Task difficulty is thus manifested by a differences between the expected and the actual performance. The level of expected task performance is established by the level of performance of the same task under the least demanding circumstances.

By defining work load in terms of the limitations on the capacity of an information processing system, the close affinity between the literature concerned with work load and the literature that focuses on attention is highlighted. In both bodies of literature, the prime concern for investigators is to assess, and possibly explain, performance limitations that are manifested despite an apparent ability of the individual to perform a task. Thus, the psychologist attending the cocktail party (Cherry, 1957, cited in Boff et. al, 1986) and listening to at least two concurrent conversations is capable of following either. Yet, one conversation may come to predominate at the expense of the other. In fact, the content of one of the conversations is often ignored. This failure to converse is noteworthy, and it is related to the discussion of work load because the person is clearly capable of switching attention from one conversation to another depending on the level of interest the ignored conversation promises.

The ability to switch from task to task implies that the contending tasks are all within the person's repertoire when performed singly. The limitation, as in the case of work load, appears to be in the system's capacity to deal with multiple demands. As with work load, investigators have suggested that the limit on attention reflects constraints inherent in the structure and organization of a central limited processor.

The notion of a central limited processor can be traced to the concept of attention discussed by such pioneers of psychology as James (1890), (cited in Boff et. al, 1986) and Titchner (1908) (cited in Boff et. al, 1986). Another important influence has come from communication engineering after World War II (Broadbent, 1958, cited in Boff et. al, 1986; Miller, 1956, cited in Boff et. al, 1986). James and his contemporaries equated selective processing and attention with the mechanism that regulates consciousness. James' discussion of the limits of consciousness implied, in modern terms, that the mechanisms operate as if they are a central limited processor. It is so structured that it can attend to one single event at any one time.

The effects of communication engineering on psychological theory can be seen by examining the effect of information theory on experimental psychology during the 1950s and 1960s. Psychologists attempted to explain human information processing in terms of the flow of information within the operator. In formal information theory, information is transmitted to the extent that the appearance of a message reduces the prior probability of a response in the response set. The greater the change between the prior and the posterior probability of the response, as a result of the presentation of a message, the greater the amount of information transmitted by the message.

The human processing system is likened to a communication channel that processes messages and transmits information from a stimulus set to a response set (Attneave, 1959, cited in Boff et. al, 1986). The communication channel is defined by its capacity to transmit information between sender and receiver, and it is characterized by a number of quantifiable parameters, the most crucial of which is that of channel capacity. One of Shannon and Weaver's (1949), (cited in Boff et. al, 1986) insights is that channels can vary in their capacities and that these differences can be quantified within the framework of information theory. A channel displays its full capacity if it poses no reduction in the transmission of information between sender and receiver. Degradations of channel capacity are measured as decrements in information transmission. If the uncertainty of the receiver is reduced at a lower rate than would result from the information available from the sender, channel capacity is assumed to have been degraded.

A variety of experimental work designed to test the applicability of information theory in the analysis of human performance has been undertaken. The common purpose of these studies is to model a variety of tasks in terms of their information transmission characteristics. The studies are designed explicitly to assess the capacity limitations of the human information processing system. Task difficulty is equated in these models with the channel's communication load. The larger the required capacity, relative to available capacity, the greater the communication load.

A paradigm in which information theory appears to be particularly useful is the absolute judgment task. The subject is presented, on each trial, with one of a set of stimuli and is asked to identify the stimulus. Stimuli may vary along a single dimension or along several dimensions. The investigator is generally interested in determining the number of different stimuli that the subject can identify. Models of human channel capacity assess that the human communication capacity of transmission reaches its limits at about 2.5 bits. All investigators seem to find the limit on channel capacity to be at about 2-2.5 bits of information per second.

An important consequence of the view of the human as a processor with an upper limit on information transfer is the recognition that different tasks impose different demands on this processor and therefore load it to different extents. Thus, the construct *work load* developed within the framework of the information theory *zeitgeist* of the 1950s. Of even greater consequence is the concept of a leftover or spare capacity. If capacity is measurable, and if different tasks consume different amounts of capacity, then some tasks consume less than full capacity and must leave a residual of spare capacity, a hypothetical quantity that might be measurable. For example, if capacity is limited to about 2.5-3 bits of information per second, and the demands imposed by a certain task are 2 bits/sec, the processor can be said to have 1 bit of spare capacity.

General Theories of Work Load

Attempts to develop models of the limited processor in terms of strict information theory were abandoned by the late fifties in favor of models postulating internal mechanisms that operate on information and determine channel capacity. The emphasis was on postulating and demonstrating bottlenecks in information flow. The work load construct played a minimal role in these efforts as they were primarily viewed by practitioners as studies of attention. Yet, because this work provided a framework within which the work load construct developed, this researcher will review briefly a couple of prominent bottleneck models and trace their influence on the development of work load as a hypothetical construct.

Single-bottleneck views promote an information flow model that comprises several processing mechanisms, one of which is more constrained than the others. The processing capability of the mechanism then sets the limits for the entire system in compiling task demands.

The most influential single-bottleneck model was proposed by Broadbent (1958) (cited in Boff et. al, 1986). Broadbent's model proposed several general principles that made it possible to describe the information flow in the human information processing system and that serve as anchors for much of the current research in the field of work load assessment. He suggested that the path between stimulus and response might be viewed as three successive stages and described the functional properties of each stage in terms that could be applied to the analysis of the task components. He assigned a cost function to the performance of mental operations, identifying operations that are more or less costly to the system. Finally, he emphasized the study of selective attention as an inherent part of the study of work load by suggesting how the efficiency of selection might affect the whole system's work load level.

Another single-bottleneck model that focuses on the structure on the central processor as a source of the limitations of the system is Welford's model (1967) (cited in Boff et. al, 1986). The key observation that Welford's theory is designed to explain is that the response to a second stimulus is delayed if the subject has not yet responded to a stimulus just presented. The shorter the interval between the first and second stimulus, the longer the response to the second stimulus is delayed. Welford, like Broadbent, postulated a three-stage model of the information flow within the organism and located the bottleneck of processing in the limited capacity of the central processor. Welford attributed the delay to the limitations on the operation of a central decision mechanism that can process only one task at a time. The time required by the mechanism to process one task is labeled the psychological refractory period, as a countermatch to the term refractory phase used by Teleford (1931) (cited in Boff et. al, 1986) to describe delays in response of a synapse or a nerve to successive stimulation.

Welford's main thesis is that performance is limited by the operation of a single-channel decision mechanism. This mechanism can deal with data of only one signal, or a group of signals, at a time, so that data from a signal arriving during the reaction time to a previous signal have to wait until the decision mechanism becomes free. The decision mechanism is frequently occupied by feedback from execution of the movements or termination of the response; therefore additional delays may occur even when a signal arrives shortly after the response to a previous signal.

The most comprehensive attempt to employ an energy metaphor in the analysis of work load, and of attention, can be credited to Kahneman (1973) in which he attributes a system's failure to perform to a shortage in the supply of what he calls processing resources. Resources is a label applied to a single undifferentiated pool of energizing forces necessary for task performance.

Kahneman (1970) viewed the amount of resources available at any time as limited, but the limit varied with the level of arousal, according to the classical inverted-U function relating effectiveness of performance to arousal. Changes in the level of arousal and consequent changes in capacity are assumed to be controlled by feedback from the execution of ongoing activities; a rise in these activities causes an increase in the level of arousal, effort, and attention.

An important construct in Kahneman's model is the mechanism responsible for the allocation policy. This mechanism directs and supervises the allocation of resources and is influenced by enduring dispositions, momentary intentions, and the feedback from ongoing activities.

In a structural model, failures to perform occur when a mechanism is required to carry out incompatible operations. In an energy-oriented model, such as Kahneman's capacity model, decrements in performance are due to demands of two concurrent activities exceeding the available capacity. A structural model therefore implies that the interference between two tasks is specific. In a capacity approach it is nonspecific and depends only on the total demands of the two tasks.

Data from a variety of experimental conditions have shown that performance on some tasks interferes with one type of task, but not with another, while a third group is equally affected when paired with members of the first two groups (Ogden, Levine, and Eisner, 1979). Other studies have shown that some manipulations of task variables within the same pair of concurrently performed tasks affect both tasks, while others degrade performance on one task only and cannot be compensated by shifting resources from the performance of the shared task (e.g., Gopher and Navon, 1982, cited in Boff et. al, 1986; Gopher, Brickner, and Navon, 1982, cited in Boff et. al, 1986; Wickens and Kessel, 1981, cited in Boff et. al, 1986). These results are inconsistent with the notion of a single undifferentiated pool of processing energy. Rather, they suggest the existence of several more specific sources of interference and competition.

The evident weakness of the single-resource model led to the development of multiple-resource models, according to which the human system is best modeled as possessing a number of processing mechanisms, each requiring its own supply of resources. The capacity of each of the structures, that depended on the level of arousal and its own specific dependence on this level, could be deployed at any moment among a number of tasks. Thus, there is continuing competition for resources between tasks that overlap in resource needs (Gopher and Sanders, 1984, cited in Boff et. al, 1986; Norman and Bobrow, 1975; Sanders, 1983, cited in Boff et. al, 1986).

Techniques For The Measurement Of Work Load

The most fundamental assertion regarding the measurement of work load is that work load is an attribute of the loop between an operator and a task. Work load is a hypothetical construct intended to capture limitations on the operator's information processing apparatus as these are viewed from the perspective of some assigned task. The critical implication of this assertion is that it is not particularly meaningful to measure work load in an open loop. One cannot specify work load associated with a task without reference to the operator, and one cannot relate work load to an operator without this operator's being in the matrix of the task-operator loop.

Relevant operator and task characteristics having an important effect on work load characterize a closed loop. The operator's capacities and skills can be assessed and the structure of the task designated *a priori* or ascertained by means of a task analysis. It is critical to distinguish between measurements of subject capacity or analyses of a task's structure and measurements of work load. The critical distinction lies in the degree to which work load measures are made in the context of, and in reference to, the interacting combination of subject and task. Strictly speaking, work load must be defined anew for each subject, and for each set of prevailing circumstances. One can assume, however, that an individual remains (subject to the effects of training and practice) essentially stationary. Thus, generalizations can be made across occasions.

In most tasks, the operator must deploy one or more processing facilities, and performance is related monotonically to the level of deployment of these facilities. Task difficulty can be expressed in terms of the demands on these facilities by the structural properties of the task. The interaction between task and operator is manifested by the degree to which these facilities can be made available given their inherent limitations. The assessment of these interactions is the assessment of work load.

Physiological Measures

A class of direct measures of mental work load is the ensemble of psychophysiological measures. These measures are obtained by recording, in general noninvasively, signals generated by the activity of some bodily system.

In general, the assumption underlying this work is that as the demands for mental effort increases various bodily systems are activated, or aroused in the process of marshalling resources in the service of this increased effort. This arousal may be manifested through increased cardiovascular activation. It may also be manifested through activation of the parasympathetic system so that pupil dilation is evident. The signals are readily recordable and can be obtained with minimal disruption to the performance of the task. The recording of cardiovascular activity requires merely the attachment of a few electrodes to the body. Pupillary activity can be recorded without any attachments to the body, though the equipment required to monitor the pupil may be somewhat cumbersome (Hamilton, Mulder, Strasser, and Ursin, 1979).

The validation of these physiological measures of work load has been based on the recording of changes in the measure under conditions in which control variation of work load is induced. A key problem in the interpretation of these data and in the utility of these measures of work load is related to the specificity of the response. In a way, the problem is quite similar to the problem one encounters in the attempt to use psychophysiological measures as indexes of deception. It is quite clear that the stress associated with deception does manifest itself in an ensemble of observable physiological changes reflecting changes in arousal. It is, however, equally clear that exactly the same changes may be observed in connection with stress caused by many factors other than deception. Thus, it is not the physiological measure per se but rather the context within which it is recorded that determines the value of any psychophysiological measure. The investigator must create a setting within which the physiological changes, and the arousal they signify, can be interpreted in an unambiguous fashion.

Primary and Secondary Task Measures

Objective methods are performance oriented, that is, they do not rely on the operator's opinion or on his physiology. Since these methods emphasize output responses to these input stimuli, objective measures describe the outputs or consequences of task performance. Viewed in this sense, any task output can be used under certain experimental circumstances to reflect work load.

Measures can be taken of primary task performance (in which no comparison is made of the task performance with the performance of any other task) and of tasks in a secondary situation. Primary task performance is not experimentally controlled; secondary task performance is-more or less-controlled.

If a single task is being performed, it is possible to assume that variation in performance of that task reflects changes in work load. Any task can therefore be used to measure work load. If work load is made sufficiently high, degradation in performance over time will inevitably occur. Primary task performance is not sensitive to moderate or low levels of work load, largely because the operator adapts his performance strategy to changes in work load. Work load can only be measured on a comparative basis because there is no absolute measure of it. Hence, primary task performance as a measure of work load is inadequate when no other task treatment exists with which it can be compared. Moreover, because there are no controls over the performance of the primary task, it is impossible to estimate the influence of possible contaminating factors such as fatigue or motivation or any other intrusive factor.

In this measurement paradigm, the work load task associated with a given task, the primary task, is measured by assigning the operator another task to perform concurrently with the primary task. The operator is told that this new task is secondary in importance. The primary task must be performed to the best of the operator's ability, even if this means neglecting performance on the secondary task. Fluctuations in performance of the secondary task are therefore assumed to reflect fluctuations in work load associated with the primary task. An assumption underlying this technique is that performance on any task depends on the measure of the resources allocated to that task. It is further assumed that the pool of resources is fixed in magnitude. From these two assumptions, it follows that there is a reciprocal relationship between the performance in each of the two tasks. Deterioration in the performance of the secondary task must be due to an increase in the resources drawn to maintain the level of performance on the primary task. So presented, the class of measures assumes a general, undifferentiated pool of resources. It is possible to extend this logic to a system characterized by pools of specific resources by selecting secondary tasks assumed to draw from one or another specific pool (Knowles, 1963).

Knowles (1963) suggested the following criteria for selection of a secondary task: noninterference with the primary task, ease of learning, self-pacing (to allow the secondary task to be neglected, if necessary), continuous scoring, compatibility with the primary task, sensitivity, and representativeness.

Subjective Task Measures

Subjective measurements of work load are made whenever participants are asked for a direct estimate of the work load they experience during the performance of a task. The term work load is rarely used when instructing the participants; instead they are usually asked to report the difficulty of the task. However, the target is the experience of difficulty during execution of the task. Thus, the participants are judging the interactions between themselves and the system.

It has been common practice to use subjective assessment as a backup for other measures. Physiological and objective measures are compared with the subjective, and the latter are used as a confirmation of the former. There is a directness to the phenomenologic experience that one does not find in the other indices.

The subjective experience of work load helps to validate other work load indices because performance and physiological measures are subject to contamination, can be interpreted as being caused by other factors, or are uninterpretable. Of course, subjective expressions are not necessarily clean either; there is possible contamination by the effects of motivation, internal and external criteria of what is appropriate behavior, etc.

A variety of techniques (e.g., magnitude estimation, paired comparisons) have been applied in work load judgments, but the rating scale is the most frequently used. Other less structured subjective methods that have been used are questionnaires and structured and unstructured interviews.

Mediating Factors in Performance Efficiency Under Varying Work Load Levels

In the performance of a skilled task such as piloting an aircraft, driving an automobile or controlling air traffic, information processing strategies and operator expectancies play important roles. An ATCS, for example, may enhance his decision-making capacity by employing a workable set of probabilities, based on past experience, concerning the occurrence of events in one's environment, and by determining effective policies of work load shedding under conditions of high task demand (Cumming, 1964, 1972).

One might hope for high intercorrelations between different measures of work load. But mental work load is a matter of great complexity and it seems reasonable to anticipate that many factors and variables will contribute to it (Moray, 1982, cited in Meister, 1985). It is not just that different resources may be needed for particular tasks or task components, it is also a question of the strategic deployment of those resources, the operator gearing efforts to objectives that may vary at different levels of mental work load. This is well illustrated by a study by Tulga (cited in Meister, 1985), which testifies to the importance of the operator's own assessment of mental work load, and shows the complex relation between subjective and objective load. Tulga's study is concerned with a theoretical model of time stress (in effect the rate of arrival of items for processing), a factor that is related to mental work load. An operator is presented with a display depicting queues comprising items of different values, which had to be dealt with so as to maximize the value of the items serviced. Initially, as the rate of arrival rose (and objective mental work load increased), subjective load also increased, and performance is efficient. But a treatment of overload is reached beyond which performance deteriorated, yet subjective load decreased. It appears that this is because the operator no longer sought to maintain perfect performance. Thus subjective mental work load may depend not only on objective load, but also on the operator's criterion as to the accuracy, speed, and precision of the performance to be produced.

It is clear from Tulga's study that the operator's objectives--and strategies for achieving them--may influence the experienced degree of work load. It is arguable that the subjective monitoring of performance by the operator contributes in an important way to the setting of achievable task goals. Another study which illustrates this is Sperandio's (1978) discussion of field research into the performance strategies of French civil ATCSs. As is commonly assumed, one of the objective indices of mental load for an ATCS is the number of aircraft in the sector for which the ATCS is responsible. This simplifies the matter of work load since the task confronting the ATCS increases in difficulty, not just because their are more problems to deal with, but also because the complexity of the individual problems also increases as the available airspace decreases. Acceptable solutions to the problem of avoiding collisions become harder to devise. A further complication is that the nature of the work varies from one control station to another, and the problems vary also, depending, for example, on whether aircraft are passing through the ATCS's sector or whether they are landing. In the light of these problems the number of aircraft is a suitable objective work load measure for practical purposes since it is convenient and reliable.

Sperandio (1978) observed three different ways of dealing with the overall task, depending on how many aircraft are in hand. When there are between one and three to be controlled, full details of the aircraft's course, altitude, speed, and other characteristics of the flight are taken into account and optimized flightpaths are calculated. for between four to six aircraft, a change in this strategy is noted as the problem as a whole became more complex. The solution might require all aircraft to adopt uniform speed and stereotyped flightpaths, but although the spacing between them might be optimized, their individual flightpaths could not be. It seems that the collection of aircraft are treated as individuals within a configuration. For seven or more aircraft simultaneously, there is a problem arising from the saturation of the airspace, and aircraft have to wait in stacks, each one having flight characteristics (e.g., speed and descent path) very much like adjacent aircraft, and having to queue until the ATCS can deal with it individually.

In practice, the separation criterion (ensuring that aircraft are at a safe distance from each other) is the most important consideration. Nevertheless, there are other goals for the efficient management of ATC, including fuel consumption (i.e., the very recent Avianca Airlines accident in Cove Neck, New York) and punctuality. It can be appreciated that this exacerbates the difficulty of specifying mental work load in objective terms. Moreover, the existence of these multiple goals undoubtedly serves to prompt shifts in the control strategy, as when the number of aircraft controlled falls to a sufficiently low number of flightpaths to be optimized in terms of fuel economy and time of arrival.

One approach to the measurement of performance, and to the effects of work load, is to record those occasions when mistakes or unintended actions are made. Slips of action are common enough in everyday life (Reason, 1979, cited in Meister, 1985), and everyone can tell of such incidents as when they got into their car intending to drive to the supermarket and ended up at their office! The incidence of performance errors is noted by Langan-Fox and Empson (1985) in a study of military ATCSs at a Radar Approach Control (RAPCON). Most of the time such mistakes were noted by the individual concerned, or by a colleague, and the great majority of errors recorded in this study were inconsequential. They do, however, constitute a measure of primary task performance.

Data and information related to actual ATC OEs can be of use in two basic ways. It can point to error types and causal factors, and it helps reality-test information derived from the research literature.

Monitoring Performance and Work Load in ATC

ATCS's in the NAS must perform many complex tasks under increasingly heavy work loads in order to accomplish the critical functions required for safe and efficient flow of air traffic. Many of the most important elements of the ATCS's task are accomplished cognitively, based on a strategic planning model that is sustained by several sources of information including memory. Successful ATCS performance depends, to a great degree, on reliable recall of relevant information (e.g., aircraft, destinations, altitudes, etc). and long-term memory (e.g., ATC procedures, aircraft characteristics, route structure, etc). Simultaneously, the ATCS must visualize aircraft positions and maintain minimum separations. A memory lapse, or even a delay, during a critical point on control operations can lead to serious consequences; such as conflicting tracks and inadequate separation between aircraft, or flight into terrain obstacles.

A number of factors are at work in the efficient control of aircraft: the air crew, ATCS's, airline personnel, prevailing conditions, and the electro-mechanical status of the aircraft. A system error usually involves some combination/interaction of the above factors. Although such OEs cannot always be traced to specific ATCS's mistakes, substantial evidence of ATCS's OEs exists. In a study on OEs in the NAS (FAA, 1987), 1352 OEs were reported, each error representing an infraction of the required separation between two different aircraft or between an aircraft and terrain, obstacles, or obstructions; about 96 percent of OEs are attributable to human error (as opposed to equipment malfunction, etc). In addition, OEs undoubtedly are underestimates of ATCS mistakes. Most ATC OEs are recoverable, not resulting in separation infractions.

Other studies support the importance of human error in ATC and give some indication of the types of OEs. For example, Danaher (1980) states that more than 90 percent of system OEs stem from mistakes in attention, judgment, and communications by ATCS's and their supervisors. Fowler (1980) also says that specific examples of crashes and near-midair collisions point to system-induced OEs caused by radar and human information-processing limitations, as well as communications problems. In addition, a report by the Society of Automotive Engineers (1986) suggests that STM failure in ATC communications and control, operations contributes directly to ATC OEs. Finally, a review of ASRS reports strongly suggests that STM failures often are factors in near-miss incidents; in some cases the evidence for memory lapse is conclusive.

One of the few experimental studies on STM, per se, in ATC settings was conducted by Bisseret (1971), who based his work on an ATC strategy model developed earlier by Leplat and Bisseret (1966). The model hypothesized an organizational sequence of ATC operations for aircraft under his control, the ATCS utilizing such attributes as flight path, separation (longitudinal and lateral), speed, and direction in making decisions at successive choice points. Any specific aircraft hypothetically is classified as conflict or no conflict by the ATCS. In his later experiment, Bisseret used the model to classify flight strips into two classes, critical and non-critical pairs, based upon the likelihood of conflict in the near future. Subject ATCS's were given simulated air-traffic control problems, the experimenter periodically removing specified strip boards; the subject was asked to recall the information on the strip. The results demonstrated better ATCS memory for strips pre-classified by the experimenter as critical. In addition, increases in traffic and decreases in ATC experience degrades performance on the recall task. Finally, Bisseret states that it is likely that performance will improve with increasing time-on-shift, although there appears to be nom actual test for this in the experiment. His logic is that this kind of organizational memory takes time to reach maximum capability.

A study of pilot memory in ATC-pilot communications by Loftus (1979) may indirectly shed some light on the question of ATC STM. Utilizing simulated communications tasks, he found retention declines rapidly after 15 seconds. He also found that place information was remembered well, frequency information poorly, and transponder information in between. He concludes that number of ATC-pilot messages should be minimized and that chunked encoding enhances performance in high memory-load situations.

These are important points, which correspond with the observations of many researchers and members in he ATC community.

Sperandio (1971) recorded ATCS's cognitive strategies as a function of traffic. He presented ATCSs with traffic on a radar display, the amount of approach traffic varying in number from four to eight aircraft. Subjects' strategies were recorded. The results, further summarized in Sperandio (1978), indicate that ATCS's traffic-handling strategies vary by traffic load. With lighter traffic, the ATCSs used direct routing, nursing the aircraft, having to account more for speed, altitude, type of aircraft, etc. As the number increased toward six, the ATCSs adopted more uniform speeds and stereotyped flight paths; however, the maximum number of total data were requested in this intermediate case. Finally, with six-to-eight aircraft the number of data requested per aircraft decreased. In general, it appears that ATCSs tend to intentionally refine their control with light traffic, giving themselves more work in doing so. In medium traffic they use more standard routings but also manipulate courses and speeds. With high traffic, they use highly standardized routing and reduce data load.

Whitfield and Jackson (1982) conducted a study in which they attempted to capture ATCS's strategies for dealing with data and organizing the picture. The method involved (a) ATC interviews and (b) verbal protocols during real-time simulations and recorded replays. The most important conclusions were that ATCSs vary in how they utilize the picture around non-routine informations; and they use the picture to plan, checking actual behavior against this plan.

The studies by Bisseret (1971) and Sperandio (1971 and 1978), mentioned earlier, have indicated that there is some indication that increasing traffic, respectively, can degrade memory for strips and influence ATCS strategies. Many studies of work load, per se, use ratings to measure the effects of work load. Hurst and Rose (1978) recorded the effects of 28 variables on TRACON ATCS activity levels as rated by ATCS observers. Communication time and peak traffic were most highly correlated with total activity, these variables being frequently cited in other work load studies. However, neither factor accounted for more than 15% of the activity variance, although Hurst and Rose (1977) (cited in Hurst and Rose, 1978) had found peak traffic to account for 50 percent of the variance in a similar study in ARTCCs. Another study, by Pasmooj, et al. (1976) found that ATC ratings of task difficulty correlated with traffic density, as did total communications time and heart rate.

A few studies have attempted to focus on the effects of work load upon cognitive activity. For example, Thackray, Bailey, and Touchstone (1979) gave subjects signal detection tasks over a period of two hours, utilizing low to high levels of target density. In the low-density conditions there were no appreciable increases in response time over the two-hour period. However, in the high-density treatment there were increasing latencies, which the researchers interpreted as a degradation in attention. Thackray and Touchstone (1989) had subjects perform simulated ATC tasks under high work load, presenting them periodically with simple (e.g., easily detected alphanumeric change) and complex (e.g., possible conflict) events. Detection times increased over time for the more complex, but not the simpler events; this is consistent with the findings of the first experiment, although the former varied density, not task complexity. But there were no apparent differences over time or between complexity levels on performance of a memory task, the authors concluding that attention is the critical behavior affected. However, there are two points worth noting here: (a) attentional responses (which did degrade) are critical to the learning of material to be held in STM, and (b) STM was measured only for low-complexity events, limiting the generality of the findings.

Finally, Langan-Fox and Empson (1985) performed a study in which ATCS observers recorded OEs of working ATCSs under variable work loads during actual operations. The subjects were performing Radar Director and Radar Approach tasks. The Director ATCSs are responsible for directing aircraft to the airfield for visual or instrument approaches; the approach ATCSs are responsible for all those aircraft within a 30 mile radius requesting service. The data indicated that Director ATCSs made substantially more memory OEs than did approach ATCSs. This was especially true during moderate-to-high work load. During light work load the Director ATCSs actually were slightly less error-prone. The authors attributed the overall higher error rate of the Director ATCSs to the forced pace and time stress under which they work. This might also explain their relatively lesser error rate during low work load. The forced pace, even if slow, enforces a discipline and level of attention on them, reducing the types of OEs which might result from inactivity and boredom. Finally, the kinds of memory OEs noted were (a) failure to remember a communication and (b) reverting to earlier plans (which had been changed).

Although stress and STM have not been studied together directly in the ATC environment, stress has been linked to many aspects of cognitive performance in the basic research, and it deserved brief mention here. Smith (1980) summarizes much of the research evaluating ATC stress, these studies using subjects' responses/scores on anxiety scales as measures. In general, Smith found that ATCs are well within normal limits on every subjective indicator of anxiety and appear to experience less anxiety than found among people in other work settings. He did find, that anxiety levels increase across an eight-hour work shift and are higher on difficult shifts than on easy ones. Another study, by Mohler (1983) examines the findings on several studies in which various physiological measures of stress (e.g., hypertension, blood pressure, illness, etc). were used. Although the findings were not compared to the population at large, the conclusions should be noted. Smith states that possible impairments of ATCSs' mental functions due to stress-related factors such as illness, fatigue, and alcohol use constitute a threat to air safety. More directly related to the cognitive and memory issues at hand, he says that shift changes are often in violation of circadian sleep-wake cycles and place additional stress on ATCSs. Finally, he also states that a reasonably work load level is necessary to produce sustained arousal levels, and that low work load and boredom can be as stressful as excessive work load.

APPENDIX B

PILOT INVESTIGATION

Rationale

A pilot investigation was undertaken to confirm the following assumptions: (1) the scenarios constructed to be low, moderate, and high in workload were subjectively assessed as such; (2) the hysteresis effect would indeed occur in this simulation; (3) the cueing strategy was effective in altering OE generation.

Methodology

Participants

Ten students from the Psychology Department (six male, four female, aged 18 to 23) voluntarily participated in the pilot investigation. Participants did not have prior experience with the TRACON simulation package. Participants had normal or correctable to normal vision and were all familiar with a personal computer keyboard configuration.

Apparatus

The same experimental site apparatus was used for the pilot investigation; specifically, the TRACON workstation, the TRACON simulation package, the Zenith Z-158 computer, etc. In order to confirm subjective assessment of these workloads by participants, a Modified Cooper-Harper Workload scale was administered.

Procedure

Training. Participants were given a brief but comprehensive course on using TRACON. Emphasis was placed on allowing participants enough practice time to be able to use the package with moderate proficiency. Participants were trained until they achieved a 80% proficiency overall rating on sample scenarios.

Experiment 1. The purpose of this experiment was to confirm that the scenarios constructed to be low, moderate, and high in workload were subjectively assessed as such. Three pre-configured scenarios were given to participants-low, moderate, and high workload. Each participant worked each scenario and then was asked to give a rating of that scenario on the Modified Cooper-Harper Scale. Each of the three scenarios lasted about 15 minutes. The mean number of aircraft worked per five-minute period of time was recorded.

Experiment 2. The purpose of this experiment was to confirm that the hysteresis effect would indeed occur in this simulation. Participants were given a pre-configured high work load scenario that shifted downward in work load after 15 minutes. This scenario consisted of nothing more than the high task demand scenario from experiment 1 pasted against the low task demand scenario from experiment 1. The number of handoff errors was chosen to be a representative measure of air traffic OEs. The number of handoff errors was recorded per five-minute period. The scenario lasted about 40 minutes but only the first 30 minutes were the focus of the analysis.

Experiment 3. The purpose of this experiment was to confirm that the cueing strategy was effective in altering OE generation. Each participant was given a pre-configured scenario that was designed to remain relatively stable in complexity over the session. The scenario was about 15 minutes in length. At two distinct points in the scenario (t1 and t2) a verbal indication by the experimenter of either a forthcoming change work load level was given or a continuation of the current work load state was given, respectively. CUE for the downward shift in task demands was administered first and CUE for a continuance of the current task demands was administered second. In order to effectively demonstrate the efficacy of the CUE treatment, participants were told of the increase or decrease in probability of a particular type of maneuver being required--in this case, a handoff. Participants worked the scenario until its completion.

Results

Experiment 1

Figure 23 represents the mean number of aircraft worked per five-minute time period over an entire 15 minute session for each of the three pre-configured scenarios. The light work load scenario permitted participants to work the greatest number of aircraft per five-minute period of time while the heavy work load scenario permitted participants to work the least number of aircraft per five-minute period of time. This data for each of the participants was correlated with the Modified Cooper-Harper Scale of Work Load Assessment ratings each participant gave for each of the scenarios. The pre-configured work load scenarios correlated negatively with the ratings given them by the participants (-.80, p=.005; -.84, p=.002; -.91, p=-.91; for high, moderate, and low, respectively). That is, the lower the mean number of aircraft worked per five-minute time period across a session, the higher the number given the scenario on the Modified Cooper-Harper rating scale. This confirms the notion that work load complexity, as defined by the mixture of different types of subtasks, is an important determinant of subjective work load.

Experiment 2

Figure 24 represents the results of a 30 minute high to low work load session given to the ten participants for confirmation of the hysteresis effect as the baseline measure. It should be noted that the 30 minute session was constructed by pasting the 15 minute confirmed high and low scenarios together. Since it was felt that including all types of errors in the measure would most likely produce a a great amount of noise and would therefore impede interpretation of the data, only one type of participant action item was recorded. The action item to be measured as an indicator of performance was the number of additional information accesses (using the <alt-I> key combination) that were requested

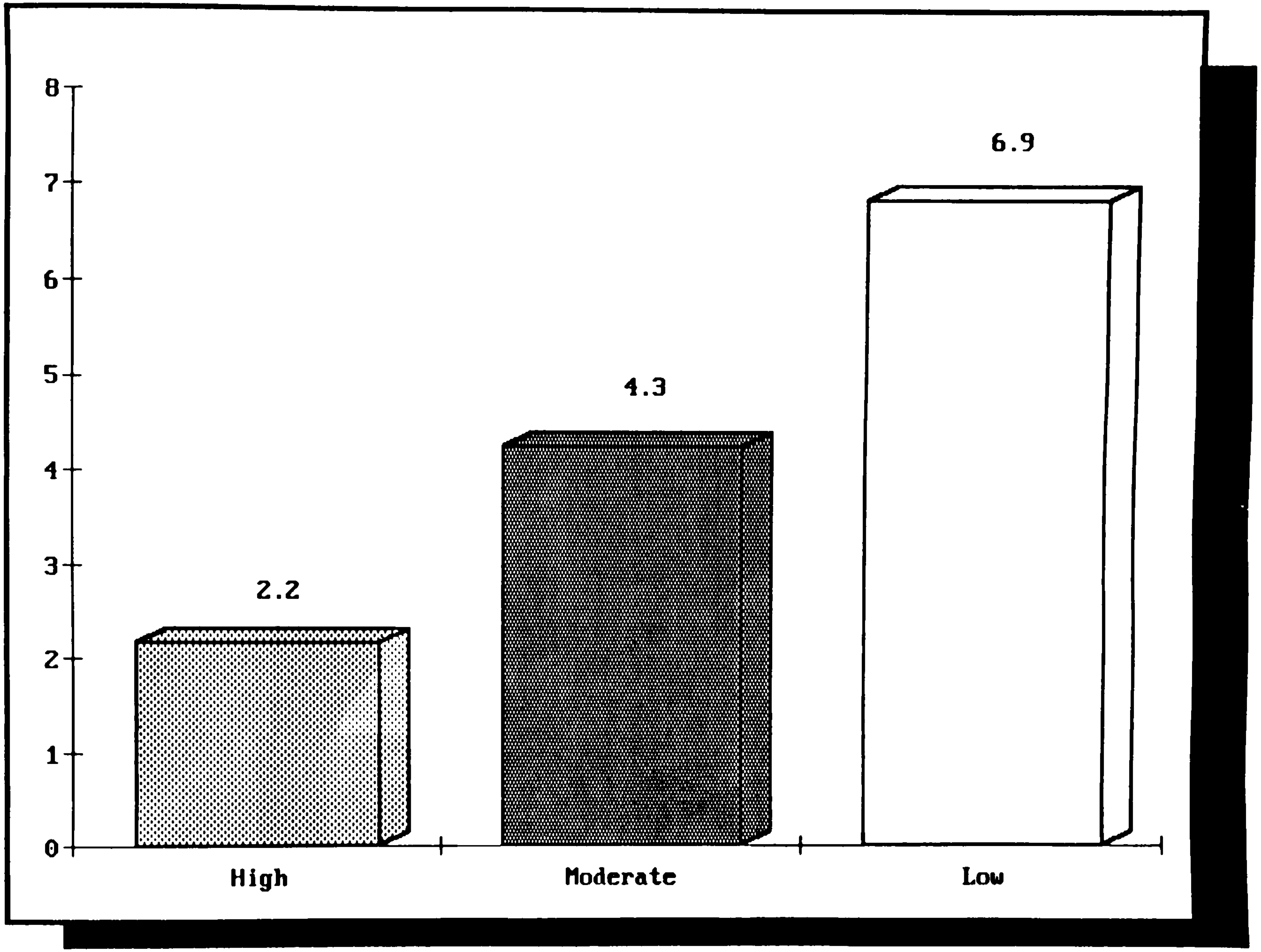

Figure 23
Mean Number of Aircraft Worked in a Session

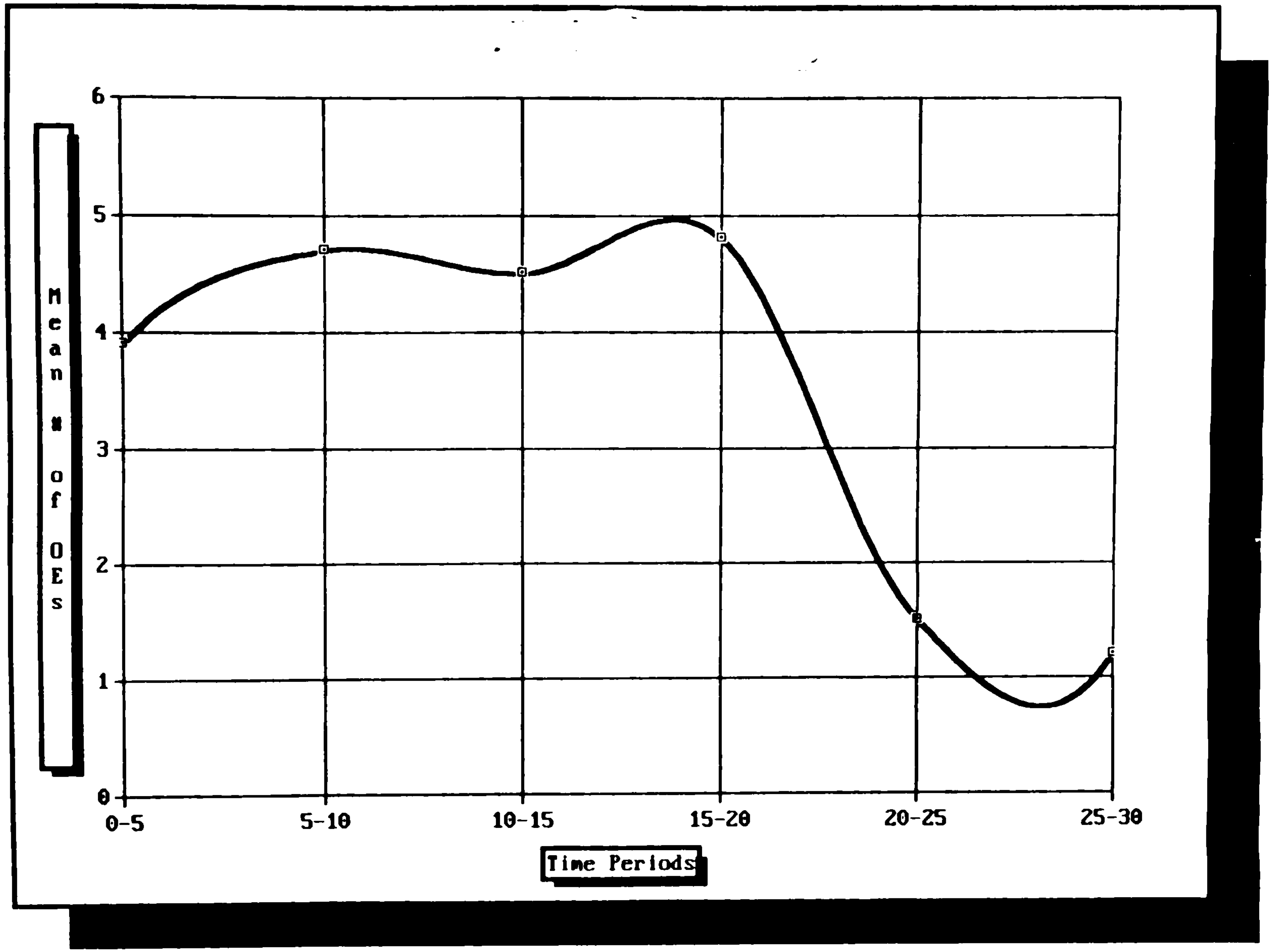

Figure 24
Mean Number of Additional Information Accesses

per five-minute time period. Although not an OE per se, the additional information access is an indicator of what information was forgotten and how often it was forgotten. Forgetting of pertinent information could result in the generation of OEs. The results are presented in Table 34. An omnibus ANOVA confirms a significant difference between additional information accesses across time periods, (F (5, 45) = 15.55, p<0.001). Paired T-tests for comparisons between time periods indicates that additional information accesses failed to drop off immediately after work load shifted downward at t3 (means for t1, t2, t3, t4, t5, and t6 are 3.90, 4.70, 4.50, 4.80, 1.50, 1.20, respectively). This confirms the presence of the hysteresis effect as a baseline measure with which one can work.

<u>Experiment 3</u>

The numerical results for this experiment are presented in Table 35. Figure 25 represents the results of the confirmed 15 minute high work load scenario as it was affected by the cueing strategy. Here, the performance measurement used was the mean number of handoff errors generated per five-minute period across the session. At t2, the cue was given that the overall number of actions to handoff aircraft would increase over the handoff rate of the previous five minutes. Though this was not the case, participants displayed a significant number of early handoff errors (mean = 4.2) over the previous five minutes (mean = 2.1). At t3, the cue was given that the overall number of actions to handoff aircraft would remain the same over the previous five minutes as it had been the previous five minutes. Participants displayed a significantly less number of *aircraft not handed off* outcomes as well as a significantly less amount of early handoff errors (mean = 1.8) as opposed to the five minutes previous (mean = 4.2). This seems to confirm the efficacy of the cueing strategy that was to be used in the main investigation.

Table 34
Omnibus ANOVA: Additional Information Accesses

Source	SS	df	MS	F
TIME	135.53	5	27.11	15.55*
S w. TIME	78.47	45	1.74	
Total	214.00	50		

* $p < .001$

Table 35
Omnibus ANOVA: CUES

Source	SS	df	MS	F
TIME	34.20	2	17.10	20.34*
S w. TIME	15.13	18	.84	
Total	49.33	20		

* $p < .001$

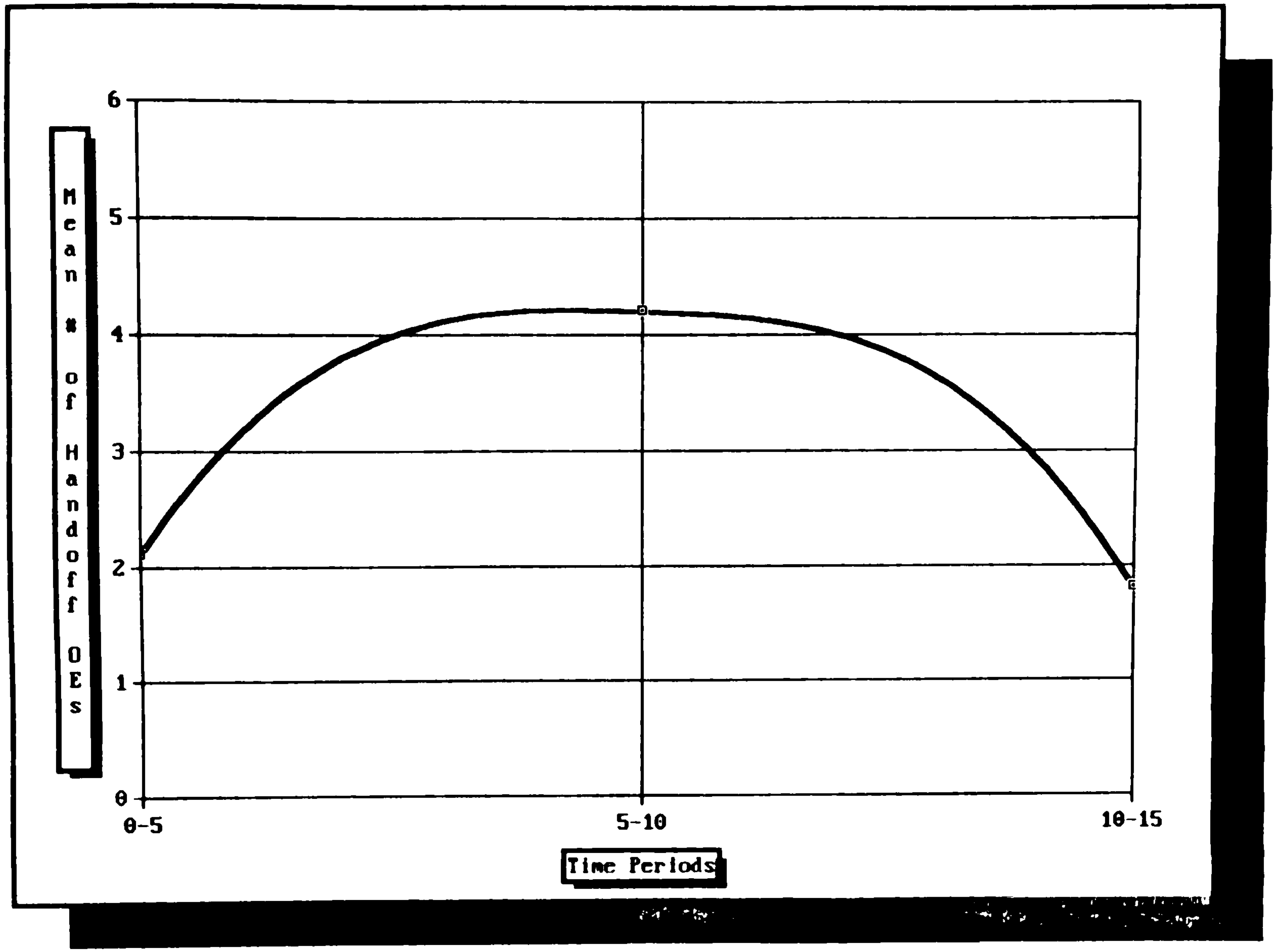

Figure 25
Mean Number of Handoff Errors Affected by Cueing

APPENDIX C

GLOSSARY OF TERMS

AIR TRAFFIC CONTROL: A service provided for the purpose of: (1) preventing collisions between aircraft and on the maneuvering area between aircraft and obstructions and (2) expediting and maintaining an orderly flow of air traffic.

COORDINATION FIX: The fix in relation to which facilities will hand off, transfer control of an aircraft, or coordinate flight progress data. For terminal facilities, it may also serve as a clearance for arriving aircraft.

DEVELOPMENTAL: An ATCS trainee.

EN ROUTE CENTER: The ATC facility responsible for aircraft separation and coordination while in cruising flight between departure and destination airports.

FINAL APPROACH FIX: The designated fix from or over which the final approach (IFR) to an airport is executed. The FAF identifies the beginning of the final approach segment of the instrument approach.

FIX: A geographical position determined by visual reference to the surface, by reference to one or more radio NAVAIDS, by celestial; plotting, or by another navigational device.

FLIGHTPLAN: Specified information relating to the intended flight of an aircraft that is filed orally or in writing with an FSS or an ATC facility.

FLIGHTSTRIP: A strip of paper containing the current flight plan parameters of the aircraft.

HANDOFF: An action taken to transfer the radar identification of an aircraft from one ATCS to another if the aircraft will enter the receiving ATCS's airspace and radio communications with the aircraft will be transferred.

HANDOFF OEs:

<u>Handoff Omission</u>: A Controller, having vectored a departing or en route aircraft to the terminating fix, fails to handoff the aircraft to the Center thereby generating a handoff error. Handoff omissions cause the flight strip information of the aircraft to not be passed along to the ARTCC. --OR-- A Controller, having vectored an arriving aircraft to the Final Approach Fix, fails to handoff the aircraft to the tower thereby causing the aircraft to overshoot the airport since the pilot was not given clearance to land by the local controller.

Handoff Error: A Controller, having had to vector a departing or en route aircraft to a new altitude and/or airspeed while on route to the terminating fix, fails to correct the aircraft's altitude and/or airspeed prior to handoff OR a controller hands off an arriving aircraft at the incorrect Final Approach Fix altitude or vector heading thereby generating a missed approach OR a controller, having vectored an arriving aircraft to the wrong airport hands him off to that airport.

Handoff Conflict: A Controller hands off an en route or departing aircraft at the terminating fix at a conflicting altitude or heading.

Early Handoff: A Controller attempts to handoff an en route or departing aircraft before the Center Controller can see the aircraft on the radarscope OR a controller attempts to handoff an arriving aircraft before the pilot has been given all of the approach vectors to the Final Approach Fix..

Late Handoff: A Controller attempts to handoff an arriving aircraft well after it has passed the outer marker.

Instrument Flight Rules: Rules governing the procedures for conducting instrument flight. Also a term used by pilots and controllers to indicate type of flight plan.

KEYING OEs

Syntax Input Error: A Controller does not issue all four portions of the communications syntax or issues all four portions out of sequence.

Parameter Error: A Controller directs an aircraft to descend/ascend to an altitude that is below/above the minimum/maximum vectoring altitude for the sector OR a controller directs an aircraft to change airspeed to a value that is below that aircraft's stall speed OR a Controller directs an aircraft to change airspeed to a new value that is above the speed limit for the sector.

MEMORY OEs:

Resume Normal Navigation: A Controller, having issued to an aircraft vector heading and/or airspeed commands around potentially conflicting traffic, fails to direct that aircraft to resume his original heading and airspeed. This results in the aircraft deviating greatly from flight plan and possibly entering conflicting traffic and/or a handoff error.

Resume Assigned Altitude: A Controller, having issued to an aircraft vector altitude commands around potentially conflicting traffic, fails to direct that aircraft to resume flight plan approved altitude. This results in the aircraft entering conflicting traffic and/or a handoff error.

Redundant Commands: A Controller issues the same command more than once to the same aircraft even though the aircraft is in the process of executing or has already executed the command.

NAVIGATIONAL OEs:

Vector Altitude Error: A Controller directs an arriving aircraft to descend/ascend and maintain a altitude that is incorrect at the Final Approach Fix OR a Controller fails to recognize the need to vector an arriving aircraft to the appropriate Final Approach Fix Altitude OR a controller fails to recognize the need to assign potentially conflicting aircraft to new altitude(s) OR a controller inadvertently assigns an aircraft an altitude change that places it in potential conflict with other aircraft.

Vector Airspeed Error: A Controller directs an arriving aircraft to change airspeed to a new value that places it in potential conflict with other aircraft.

Vector Heading Error: A Controller directs an arriving aircraft to a heading not within the approach heading parameters for that airport OR a Controller mistakenly directs an aircraft to turn to a relative number of degrees rather than an absolute number of degrees OR a Controller directs an aircraft to a heading that places it in potential conflict with other aircraft OR a Controller directs an aircraft to turn the wrong way to a given heading OR a Controller issues a vector heading to the wrong airport.

Vector Conflict Error: A Controller fails to detect a potential conflict between aircraft.

Vector Transposition Error: A Controller confuses the flightstrips of two aircraft with similar call signs and issues the command to the wrong aircraft.

Vector Off Radar: A Controller fails to detect a non-enroute aircraft straying out of sector boundaries and off radar.

PILOT OEs:

Pilot Deviation Error: A Controller fails to detect an aircraft executing a command incorrectly. For example, a controller directs an aircraft to turn right to heading 340° and the aircraft turns left to heading 340°.

Pilot Transposition Error: A Controller fails to detect an aircraft initiating an action that was directed at another aircraft. For example, a controller directs United Airlines flight 984 to descend and maintain to an altitude of 3000 feet and United Airlines flight 233 initiates the action instead.

Pilot Acknowledgement Error: A Controller fails to detect a pilot's failure to acknowledge a command. A failure to acknowledge the Controller's command will mean that the aircraft will not initiate action on the command.

TERMINAL RADAR SERVICE AREA: Airspace surrounding designated airports wherein ATC provides radar vectoring, sequencing and separation on a full-time basis for all IFR and participating VFR aircraft. Service provided in a TRSA is called Stage III service. Pilot participation is urged but not mandatory.

VECTOR: A heading issued to an aircraft to provide navigational guidance by radar.

VISUAL FLIGHT RULES: Rules that govern the procedures for conducting flight under visual conditions. The term VFR si also used in the United States to indicate weather conditions that are equal to or greater than minimum VFR requirements. In addition, it is used by pilots and controllers to indicate type of flight plan.