AL-TR-90-84

2

AD-A235 944

AUTOMATIC INFORMATION PROCESSING AND HIGH-PERFORMANCE SKILLS: PRINCIPLES OF CONSISTENCY, PART-TASK TRAINING, CONTEXT, RETENTION, AND COMPLEX TASK PERFORMANCE

DTIC
ELECTE
MAY 16 1991
S B D

Arthur D. Fisk
Wendy A. Rogers
Mark D. Lee
Kevin A. Hodge
Christopher J. Whaley

Georgia Institute of Technology
School of Psychology
Atlanta, Georgia 30332

LOGISTICS AND HUMAN FACTORS DIVISION
Wright-Patterson Air Force Base, Ohio 45433-6503

April 1991

Interim Technical Report for Period November 1989 – December 1990

Approved for public release; distribution is unlimited.

DTIC FILE COPY

HUMAN RESOURCES

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

91 [illegible] 036

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

BERTRAM W. CREAM, Technical Director
Logistics and Human Factors Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
April 1991

3. REPORT TYPE AND DATES COVERED
Interim Report – November 1989 – December 1990

4. TITLE AND SUBTITLE
Automatic Information Processing and High-Performance Skills: Principles of Consistency, Part-Task Training, Context, Retention, and Complex Task Performance

5. FUNDING NUMBERS
C – F33615-88-C-0015
PE – 62205F
PR – ILIR
TA – 40
WU – 01

6. AUTHOR(S)
Arthur D. Fisk
Wendy A. Rogers
Mark D. Lee
Kevin A. Hodge
Christopher J. Whaley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Institute of Technology
School of Psychology
Atlanta, Georgia 30332

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Logistics and Human Factors Division
Air Force Human Resources Laboratory
Wright-Patterson Air Force Base, Ohio 45433-6503

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFHRL-TR-90-84

11. SUPPLEMENTARY NOTES

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

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Six series of experiments (11 individual experiments) were conducted to further extend automatic/controlled processing research to command and control mission-specific training. The issues examined in these experiments were related to retention of task-component skills; amount of practice; component training for memory-search-dependent tasks; and effects of degree of consistency, context, and task performance dependent on interactions of memory scanning, visual search, rule-based processing and acquisition of procedural knowledge. A final section of the document outlines how the present data provide processing principles which augment previous human performance guidelines that have been shown to be important for high-performance-skills training.

14. SUBJECT TERMS
automaticity
part-task training
skill acquisition
skill retention
skill transfer
training

15. NUMBER OF PAGES
288

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UL

SUMMARY

This document summarizes Phase 2 of a basic research effort investigating automatic processing theory and high-performance skills training. Research issues such as skill acquisition, skill retention, part-task training, transfer of training, context effects, and degree of within- and between-category consistency are explored. The results of this work suggest that the application of automatic processing theory to training complex skills can have an impact on skill acquisition in complex, high-performance tasks.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

PREFACE

The work documented in this report was conducted under Air Force Human Resources Laboratory (AFHRL) Contract No. F33615-88-C-0015 with the University of Dayton Research Institute and was performed by the subcontractor Georgia Institute of Technology Research Institute. This work supports an integrated research program which is developing advanced part-task training techniques based on information processing theory. Beverley A. Gable served as the AFHRL/LRG, Wright-Patterson AFB, contract monitor.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

AUTOMATIC INFORMATION PROCESSING AND HIGH-PERFORMANCE SKILLS: 2. PRINCIPLES OF CONSISTENCY, PART-TASK TRAINING, CONTEXT, RETENTION, AND COMPLEX TASK PERFORMANCE

I. OVERVIEW OF THE EXPERIMENTAL INVESTIGATION

This document details seven series of experiments (a total of 11 individual experiments) conducted to further extend automatic/controlled processing research to command and control mission-specific training. The present experiments build upon and extend an earlier investigation reported by Fisk, Hodge, Lee, and Rogers (1990). The research addresses training-program-relevant research that can be broadly categorized as (a) acquisition, (b) transfer, and (c) retention of high-performance-skilled behavior.

This document describes experiments that examine issues related to (a) retention of trained task-component skills, (b) amount of practice, (b) component training for memory-search-dependent tasks, (c) degree of consistency, (d) context, and (e) task performance dependent on interaction of memory scanning, visual search, rule-based processing, and procedural knowledge. Because of the breadth of the issues examined, each of the seven series of experiments is presented in an independent section of the document.

The second major section of the document reports the completion of an experiment partially reported by Fisk, et al. (1990). This experiment is part of a series of experiments to investigate the effects of type and amount of consistent mapping practice on automatic process development. The experiment completes the investigation of the effects of differential amounts of practice on the "strength" (degree of automatic process development) of consistently mapped stimulus items. These experiments help to assess when it is possible to reduce the amount of practice needed for a given level of skill development. To briefly summarize the findings from the previous series, the data confirm that, in general, the more consistent mapping practice persons receive, the better their performance will be at the end of the training. More important, the data suggest that it may be possible to specify how to combine training such that some training elements will benefit from the training of other elements; hence, training time can be reduced. If a "superset" can be formed during training (and that set can be formed quickly), then detection of one stimulus item seems to strengthen the entire to-be-trained set. The present experiment confirms this prediction and shows that amount of task-specific consistent mapping (CM) practice (as opposed to generalized search practice) predicts performance when a memory superset cannot be formed.

The third major section describes four experiments to examine the effect of memory-set component training on both the learning and the retention of performance in a hybrid memory/visual search task. Performance on the task was examined as a function of the amount of material to be learned (and the manner in which it is presented). Four experiments were conducted: two training and two retention experiments. In each experiment, three training conditions were used, with each condition representing different memory loads. The conditions were (a) PT2, three different memory sets of two categories each (subjects trained on one memory set before moving to the next; hence, part-task training); (b) PT3, two different memory sets of three categories each (part-task training); and (c) WT6, one memory set of six items (full-task practice). The paradigm used was the adaptive multiple frame procedure developed to test performance at each subject's perceptual processing limits. Subjects practiced for 6 days. After the initial practice, they were tested in the full tasks at various frame times. After testing, the subjects received another 6 days of practice, followed by full-task testing. In the retention experiments, subjects' performance in the full task was tested 30 days after receiving part-task or whole-task practice. The data from the experiments in this series suggest that, for tasks requiring memory-set unitization (development of a super-set), unitization can emerge through part-task training. Significantly, the retention data demonstrate that the unitization is resistant to decay with disuse regardless of whether the training was whole-task or part-task training. Most important, the retention data suggest that target strengthening benefits most from part-task training.

The fourth major section examines the influence of the degree of consistency on performance in a task that examined complex category search at each subject's individual perceptual processing limit (by use of the adaptive multiple frame paradigm). This experiment was conducted to examine the effects of degree of semantic-category consistency on performance in the highly demanding adaptive multiple frame procedure. Subjects received training on semantic-category stimuli that were either 100 percent consistent, 66 percent consistent, 50 percent consistent, 33 percent consistent, or variably mapped (VM). Subjects were first trained for seven sessions in the adaptive procedure so that they were performing at the limits of their perceptual processing ability. Following this training, subjects received 5 days of practice at a fixed frame speed which was determined for each individual as the fastest frame speed achieved during session seven. On the last day of practice all stimuli were completely consistent to provide a pure CM test of performance. This experiment assessed important characteristics of consistency effects using more complex stimuli and a much more complex processing environment than previously used. The present data coupled with those existing in the literature afford the opportunity to predict performance as a function of the degree of consistency, the complexity of the task, and the amount of practice.

The fifth major section reports data from an experiment conducted to examine the effects of within-category consistency (i.e., some elements within a category are consistent and some are not) on the processing of the entire category as well as the individual elements. It is important that, the design allowed an examination of these consistency effects on <u>both</u> performance and learning. Subjects received training on four different CM categories and on VM categories. The CM categories were either completely consistent (all words are always targets, never distractors), 66 percent consistent (i.e., six words are always targets and two words serve as both targets and distractors), 50 percent consistent (four consistent and four inconsistent words), or 33 percent consistent (two consistent and six inconsistent words). Subjects received 12 days of single-frame practice where performance, measured by reaction time and accuracy, was assessed. For 2 days following practice, subjects were tested in semantic transfer conditions where the amount of category learning (strengthening) was assessed as a function of the degree of category consistency. The data indicated that when the category was inconsistent but some words within the category were consistent, detection performance was a function of consistency at the word level. The results suggest that consistency, at any level, may be capitalized on during training to facilitate task-specific performance. The effect of "global" inconsistency, however, inhibited <u>learning</u> at the higher order category level. The learning at the category level followed the same pattern as that demonstrated for effects of degree of consistency at the elemental level (Schneider & Fisk, 1982) and for between-category degree of consistency demonstrated and reported in Section IV of this document.

The sixth major section reports an experiment that greatly extends the information obtained from a previous experiment conducted by Fisk and Rogers (1988). In the present experiment, we were interested in how quickly context could be activated to positively affect performance relative to VM performance. The experiment required 13 hours per subject to complete. All subjects received training on a completely consistent semantic category and on VM category search. In addition, all subjects received training in three context conditions where context is defined by the co-occurrence of target/distractor pairs. Although the context conditions are technically inconsistent, whenever a given target item occurred it was always paired with a given distractor category for a given context condition. (This context manipulation has been shown to positively, but temporarily, influence performance in the Fisk and Rogers experiment.) In the present experiment, we changed the context either every 1, 5, 10, or 50 trials to assess the short- and long-term performance effects on the context conditions as well as the pure CM condition. The data showed that, for this class of tasks at least, temporary salience biasing (context effects) can be seen within five exposures to the context situation. It is important that, when context was shifted every trial and the pure CM condition was embedded within this one trial cycle, we found that the context effects were minimized and performance in the pure CM condition was also compromised.

Section VII provides the results of two experiments (training and retention) using our complex dispatching task. The task is a conceptual analog of the tactical resource allocation required in real-world, battle management tasks. This experiment begins our use of complex tasks to evaluate the effects of instructional techniques on performance improvement and the transferability of our major findings to even more complex, multi-component tasks. The task has several procedural components, requires learning a substantial amount of declarative knowledge, and is very heavily rule-based. Although the task is conceptually simple, the subject must choose the optimum "driver" for a given "delivery"; the subject must learn rules associated with how to determine load level, load type, and delivery location characteristics. In addition, the subject must learn to associate 27 drivers with various "license classes" (license classification determines who can carry out the mission).

The present task requires memory scanning (subjects must hold a self-derived list of potential drivers in memory), and across trials the number of potential drivers (and hence, memory load) is manipulated, allowing data which provide information converging on issues previously addressed with more simple laboratory memory search studies. Subjects must learn rules associated with performing the task; hence, rule-based learning (necessary for most complex skill-based tasks) can be assessed. Subjects must decide when and how to optimally access help screens (a decision component), and they must also scan a display to locate the optimum driver (corresponding to standard visual search tasks).

The first experiment examined high-performance-skill development. Early in practice there were large individual differences in performance of the task. However, in line with other studies of skill acquisition (e.g., Ackerman, 1988; Fisk, McGee, & Giambra, 1988), these differences diminished with practice. Within the 10 hours of practice, all subjects increased accuracy (to ceiling), increased speed of decisions, reduced their use of help to very infrequent usage, and used only the minimum number of keystrokes required. All aspects of performance improvement followed a "power law" of practice (Newell & Rosenbloom, 1981).

The second experiment in this series examined subjects' ability to perform the complex task 60 days subsequent to their last practice session. This retention test was a surprise; subjects did not know that we would call and ask them to return. One subject had graduated, but all other subjects returned for the retention test, which consisted of another 10 days of participation; thus, we were able to examine savings and relearning scores. The data indicated that although performance declined relative to the final training session performance, the savings scores were impressive, ranging up to 82 percent. By block seven, subjects' performance had met or exceeded their final-training-level performance. Rule-based performance seemed to remain intact; however, relative to declarative information such as memory for specific names, performance declined, as indicated by the pattern of help usage.

The pattern of training and retention data clearly indicates the validity of our task for addressing complex, ecologically valid issues relevant to Air Force missions. Performance and retention characteristics followed patterns expected from high-performance-skills development.

The final experimental series, reported in Appendix A, examined one-year, long-term retention of automatic component processes. Clearly, this issue is important because situations exist where personnel are trained and then use the skill only when an emergency arises. Given this kind of scenario, we need to be able to predict the mission readiness of trainees. We also need information to predict the timeframe and the potential need for refresher training. This series of experiments gives us this information, at least for the class of tasks used herein. For completeness, the entire series of experiments, along with the retention data for retention intervals reported previously by Fisk et al. (1990), is presented in Appendix A.

In the following detailed account of the experimental investigations, each section is generally self-contained so that the reader interested in only some of the issues can turn immediately to the relevant section(s).

The final section presents one important outcome of the research program; that is, what we refer to as processing principles. Such processing principles illustrate human performance guidelines that have been shown to be important for the development of "knowledge engineering" for understanding and developing training programs for complex, operational tasks. These processing principles were developed based primarily upon the research presented in this technical report as well as AFHRL funded research reported in Fisk et al. (1990).

II. EXPERIMENTAL SERIES 1: EFFECTS OF AMOUNT OF CONSISTENT PRACTICE WHEN TOTAL TASK UNITIZATION IS NOT POSSIBLE

Introduction

This section introduces much of the background terminology associated with automatic and controlled processing theory. In addition, it outlines a strength theory approach to understanding how performance improves in consistent mapping paradigms, and discusses the rationale for the first experiment.

Automatic and Controlled Processes

A well-documented finding in the realm of attention research is that two qualitatively different types of information processing interact in the performance of most complex tasks (LaBerge & Samuels, 1974; Logan, 1978, 1979, 1985, 1988a, 1988b; Posner & Snyder, 1975; Schneider, Dumais, & Shiffrin, 1984; Schneider & Shiffrin, 1977; Shiffrin, 1988; Shiffrin & Dumais, 1981; Shiffrin & Schneider, 1977). Following the lead of Schneider and Shiffrin (1977), we will refer to these two processes as "automatic" and "controlled" processes.

Automatic processes are characterized as fast, parallel, fairly effortless, and not limited by short-term memory capacity; these processes are difficult to acquire and, once well learned, difficult to modify. Furthermore, automatic processes are not sensitive to vigilance decrements (Fisk & Schneider, 1981), alcohol intoxication (Fisk & Schneider, 1982), fatigue (Hancock, 1984), or heat stress (Hancock & Pierce, 1984).

Controlled processes, on the other hand, are generally slow, serial, attention-demanding, and limited by short-term memory capacity. (For a more detailed analysis of the characteristics of automatic and controlled processing, see Fisk, Ackerman, & Schneider, 1987; Logan, 1985; Posner & Snyder, 1975; Schneider et al., 1984; Shiffrin, 1988; Shiffrin & Dumais, 1981.)

Controlled processing components usually dominate in the performance of novel tasks. However, if major components of the task are consistent, performance can become automatized after substantial practice. A central goal of training research is to understand how, and under what conditions, performance improves. Generally speaking, an important component of many training programs involves training the consistent elements of a task (Schneider, 1985a).

In their series of experiments investigating controlled search and automatic detection, Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) demonstrated differences in performance as a function of whether training was consistent or varied. The degree of consistency in the relationship between the stimulus (or classes of stimuli) and the response requirements has been referred to as consistent or varied "mapping." In a consistent mapping (CM) situation, the individual always deals with (i.e., attends to, responds to, or uses information from) a stimulus, or class of stimuli, in a consistent manner. CM training conditions result in dramatic performance improvements (see Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977 for details), and the eventual development of performance characteristics indicative of automatic processing. Varied mapping (VM) training situations are those in which the practice is inconsistent; that is, the response or degree of attention devoted to the stimulus changes from one stimulus exposure to another. VM training conditions result in relatively little performance improvement.

Automatic Process Development

Many theories of automatic process development are based on the modal view of a strength representation of knowledge (e.g., Anderson, 1982, 1983; Dumais, 1979; LaBerge & Samuels, 1974; MacKay, 1982; Schneider, 1985b; Schneider & Detweiler, 1987, 1988; Shiffrin & Czerwinski, 1988; but see

Logan, 1988a, 1988b, for a non-strength theory). All these theories propose that some increase and/or decrease in "strength" (defined below) is responsible for the development of automaticity.

The concept of strength varies among the models, but is generally related to the role or significance of a stimulus or set of stimuli, a rule, or a connection (e.g., between nodes). For example, MacKay's (1982) strength theory is based on repeated activation, priming, reinforcement, and the resultant changes in strength among nodes. Production system models incorporate a conceptualization of strength associated with production rules. Strength is increased when a rule is invoked and weakened when application of the rule leads to error. According to Neches, Langley, and Klahr (1987), "The strength (or weight) of a production is a parameter that is adjusted to indicate the system's current confidence in the correctness and/or usefulness of that rule" (p. 39). Finally, connection system models are strength-based in that they assume that knowledge _is_ the strength of connections among units of information (for a review, see Rumelhart & McClelland, 1987).

Recently, Schneider (Schneider, 1985b; Schneider & Detweiler, 1987, 1988) proposed an eclectic strength model which is a hybrid of production system and connectionist models. According to Schneider's connectionist/control model, the development of automaticity is a function of two types of learning mechanisms: associative and priority learning, both of which are strength-based.

The associative learning mechanism alters the connection weights between input and output information such that, after sufficient training, a given input comes to evoke the associated output. Furthermore, associative learning results in the strengthening of connections among stimuli (e.g., members of a category) such that activation of one stimulus results in the activation of others.

The priority learning mechanism modifies how strongly a given message (i.e., stimulus information) is transmitted. Strength of transmission is defined as the "priority tag" of a given message. A key element of priority learning is that the increment or decrement of a priority tag is based on whether a message is important; that is, whether prior presentation of that message produced a substantial amount of subsequent processing. Important messages have high-priority tags and unimportant messages have low-priority tags.

It is assumed that consistent practice leads to continual incrementing of the priority tag for target stimuli (when detected) and decrementing of the priority for distractor stimuli. Thus, CM practice leads to a segregation of stimuli such that stimuli with high-priority tags (consistent targets) become "foreground" and stimuli with very-low-priority tags (consistent distractors) become "background." Within Schneider's hybrid connectionist model, pure automatic processing (processing without control process assistance) is not possible without sufficient priority learning. A combination of both associative and priority learning allows stimuli to be filtered and messages transmitted without control processing assistance; hence, stimuli can automatically attract attention. A common example of the presence of some stimulus or configuration of stimuli resulting in the automatic attraction of attention is the cocktail party phenomenon. This phenomenon is exemplified by the situation in which a person is listening to one conversation amid a din of background conversation yet attention is immediately drawn to another conversation when the person hears his or her own name.

Support for the Strength Theory

Many experiments have provided evidence in support of the assumption that search performance is determined by the strength of the target relative to the strength of the distractor (e.g., Dumais, 1979; Prinz, 1979). On the first trial of training, it is assumed that all stimuli have an equivalent, intermediate strength (Dumais, 1979; Shiffrin & Czerwinski, 1988; Shiffrin & Dumais, 1981). The strength of the stimuli is intermediate and not zero because the stimuli are not completely novel but are simply untrained. For example, if words or letters are used as stimuli, they are familiar but have not been previously trained to have a high strength level, at least within the experimental context (Schneider & Fisk, 1984).

By definition, each time a CM target appears in the display it is always attended to and/or responded to (except, of course, in the case of a "miss"). In this manner, the importance of a CM stimulus is increased and thus the CM stimulus becomes associated with a high-priority tag. After many trials of CM training, the high priority associated with CM targets will result in these items being transmitted without the need for serial search. Consistent distractors, on the other hand, will have a decreased strength level after practice because their appearance results in either a negative response (e.g., correct rejection) or no response at all. Therefore, CM distractors will have a very low priority. Finally, VM stimuli maintain an intermediate strength because on some trials they are targets and are attended to, whereas on other trials they serve as distractors and must be ignored. Conceptually, the priority tag of the VM stimuli increases on some trials and decreases on other trials; therefore, even after many trials of training, these stimuli will still have an intermediate strength level.

Transfer and/or reversal of CM-trained targets and distractors yields a pattern of results which supports strength-based theories of perceptual learning. For example, Rabbitt, Cumming, and Vyas (1979) found that positive transfer (i.e., no disruption in performance) occurs when previously trained CM targets are paired with new distractor stimuli. According to a strength model, this is to be expected because targets previously trained as CM targets have a higher strength relative to the novel stimuli used as distractors in the transfer condition. (As mentioned previously, novel stimuli have an intermediate strength level prior to training.)

Kristofferson (1977) demonstrated that positive transfer is also found when new targets are paired with previously trained CM distractors. In this case the CM distractors have a low strength level relative to the novel stimuli being used as targets. Although a strength theory is not explicitly formulated by Rabbitt or Kristofferson, their data provide evidence for both target learning and distractor learning in search tasks.

Dumais (1979) conducted a series of experiments explicitly examining target and distractor strength differentiation using a within-subjects design. She trained subjects in several CM conditions and then investigated the effects of target transfer (pairing trained CM targets with VM items) and distractor transfer (pairing VM items as targets with trained CM distractors). Positive transfer was demonstrated when either the CM target set or the CM distractor set remained the same and was paired with a VM set. These results demonstrated both target and distractor learning in visual search tasks.

Further evidence for both target and distractor learning in visual search has come from negative transfer (i.e., disruption in performance) found in studies that reversed the role of targets and distractors. Included in Dumais' (1979) experimental series were "partial reversal" conditions. A partial reversal is defined as a condition in which the role of either the target or the distractor set (but not both) has been reversed within a single condition. A target reversal involves using previously trained CM targets as distractors and pairing them with novel stimuli as targets. The CM stimuli, which have a high strength level, draw attention away from the new targets and serve to disrupt performance. Similar disruptions are found with distractor reversals, in which the CM distractors become targets and are paired with novel items as distractors.

The strongest reversal effects, as would be expected from a strength perspective, were found in Shiffrin and Schneider's (1977, Experiment 1) "full reversal" condition. They trained CM targets and CM distractors and then reversed the roles of both the target and distractor sets within a single condition (i.e., previous CM targets became distractors for previous CM distractors, which then became the targets). Shiffrin and Schneider found that performance in the full reversal condition was actually worse than asymptotic VM performance. The large amount of disruption is consistent with the theory that attention is actually captured by the distractors and drawn away from the targets.

Another experiment in Dumais' (1979) series compared the differences in disruption due to full reversal and to partial reversals (i.e., target reversal and distractor reversal). Her results were consistent with Shiffrin and Schneider's in that full reversal yielded a strong disruption, resulting in performance which was actually worse than asymptotic VM performance. She also found stronger disruption effects in the full reversal condition than in either of the partial reversals.

The experiments reviewed above provide supporting evidence that, within the visual search domain at least, subjects learn to attend to target information through strengthening or prioritizing that information. Furthermore, distractor information is ignored; hence, its attention-calling strength is reduced or weakened. These findings provide important information regarding the transfer of well-learned components to situations in which the use of the components remains similar (and performance is facilitated) or is reversed (and performance is disrupted). In a related manner, patterns of transfer and/or reversal allow estimation of the degree to which the components have been learned. This theoretical and empirical base was used in the present experimental series to investigate the effects of practice on the learning and transfer of components in visual search.

Overview of Present Experiment

The present experiment was an extension of research previously described in Fisk et al. (1990) conducted to investigate the effects of differential amounts of practice on the resultant strength of the CM items. A within-subjects, between-blocks design was used in which each subject received training in each of the following conditions: CM High (3,360 trials), CM Moderate (1,680 trials), CM Low (560 trials), and VM (1,120 trials). Following training, two sessions of transfer allowed a more complete specification of the effects of transfer and reversal of previously acquired automatic processes of varying strengths. The degree of disruption or transfer was measured as a function of different re-combinations of items. For example, performance in six different target reversal conditions was measured to compare the amount of disruption in a target reversal situation in which the items used as distractors (i.e., previously trained CM targets) were manipulated. The distractors were either all highly trained CM targets, all moderately trained CM targets, all low trained CM targets, or some combination of the three. Similarly, performance was measured for all combinations of distractor transfer.

We were interested in examining whether the relatively small differences between the CM High, CM Moderate, and CM Low conditions found in our previous research (see Fisk et al., 1990) were a function of the type of randomized training which may have allowed the development of a superset. In other words, it may have been possible for subjects to create a superordinate category containing all the CM target categories. Thus, though the CM High category appeared most frequently as the target, the CM Moderate and

CM Low categories may have also been activated due to associative learning; thus, they would have benefitted from training to a greater degree than would be expected given the actual number of trials. This issue is explored in greater detail later in this report.

Method

Subjects. Sixteen subjects (8 males, 8 females) participated in the experiment. The subjects were compensated monetarily for their participation: $4.00 per hour, with a $1.00-per-hour bonus for completing the entire experiment. The vision of all subjects was tested using a Snellen chart and their corrected or uncorrected visual acuity was at least 20/30 for distance and 20/40 for near (magazine print) vision.

Stimuli. Memory-set items were the semantically unrelated categories (Collen, Wickens, & Daniele, 1975) of FURNITURE, VEGETABLES, MUSICAL INSTRUMENTS, FOUR-FOOTED ANIMALS, ALCOHOLIC BEVERAGES, BUILDING PARTS, WEAPONS, EARTH FORMATIONS, UNITS OF TIME, OCCUPATIONS, BODY PARTS, RELATIVES, VEHICLES, COUNTRIES, TREES, and CLOTHING. Target and distractor items were high associates of these categories (Battig & Montague, 1969). Each category set contained eight words. Each subject received a unique assignment of categories for each condition, counterbalanced by a partial Latin square.

Apparatus. All stimuli were presented using EPSON Equity I+ microcomputers with Epson MBM 2095-5 green monochrome monitors. The standard Epson Q-203A keyboard was altered such that the '7', '4', and '1' numeric keypad keys were labeled 'T', 'M', and 'B', respectively. The microcomputers were programmed with Psychological Software Tools' Microcomputer Experimental Language (MEL) to present and time the stimulus displays and to record response behaviors. During all experimental sessions, pink noise was played at approximately 55 decibels (db) to help eliminate possibly distracting background noise. All subjects were tested in the same room at individual, partitioned workstations monitored by a laboratory assistant.

Procedure. During the first session of the experiment, the subjects completed a practice session of the experimental task. The practice session consisted of five blocks of CM trials (50 trials per block). These orientation trials allowed the subjects to become familiar with the experimental protocol and also served to stabilize the error rates. The categories used for the practice trials were not used in the remainder of the experiment.

An individual trial consisted of the following sequence of events. Subject were presented with the memory set of one category label, which they were allowed to study for a maximum of 20 seconds. Subjects were instructed to press the space bar to initiate the trial. Three plus signs were then presented in a column for 0.5 second in the location of the display set (in the center of the screen) to allow the subjects to localize their gaze. The plus signs were followed by the display set, which consisted of three words presented in a column. The subjects' task was to indicate the location of the target (i.e., top, middle, or bottom) by pressing the corresponding key (labeled 'T', 'M', or 'B'). A target (i.e., an exemplar from the target category) was present on every trial.

Subjects received the following performance feedback. After correct trials, the subjects' RTs were displayed in hundredths of a second. After incorrect trials, an error tone sounded and the correct response was displayed. Following each block of trials, subjects received their average RT and percent accuracy for that block; if a subject's accuracy fell below 90% in any block, a message was displayed encouraging a more careful response. Subjects were instructed to maintain an accuracy rate of 95 percent or better while responding as quickly as possible. After each block of trials, subjects were encouraged to take a short break to rest their eyes.

There were two phases of the experiment: training and testing. The training phase consisted of four conditions: (a) CM High - 3,360 trials, (b) CM Moderate - 1,680 trials, (c) CM Low - 560 trials, and (d) VM - 1,120 trials.

The subjects were trained for seven 1-hour sessions, each of which consisted of 24 blocks of CM training (40 trials per block): 12 blocks of CM High, 6 blocks of CM Moderate, 2 blocks of CM Low, and 4 blocks of VM. The order of the presentation of the blocks was randomized.

The testing phase of the experiment consisted of two sessions: one session of Target Reversal conditions and one session of Distractor Transfer conditions. In the Target Reversal conditions, previously trained VM sets were used as target items and the types of distractors (i.e., previously CM High, Moderate, or Low trained <u>target</u> items) were manipulated. The reversal conditions were as follows:

1. High/High Target Reversal - both distractor items on a trial were previously CM High targets.
2. Moderate/Moderate Target Reversal - both distractor items on a trial were previously CM Moderate targets.
3. Low/Low Target Reversal - both distractor items on a trial were previously CM Low targets.
4. High/Moderate Target Reversal - one distractor item was previously a CM High target and the other was previously a CM Moderate target.
5. High/Low Target Reversal - one distractor item was previously a CM High target and the other was previously a CM Low target.
6. Moderate/Low Target Reversal - one distractor item was previously a CM Moderate target and the other was previously a CM Low target.
7. New CM condition - created by pairing two of the VM sets in a consistent mapping.

The New CM condition served as a comparison condition. The six target reversal conditions were manipulated within a block and the New CM condition was presented in a separate block. In each block of 48 trials, each reversal condition was presented eight times in random order. Subjects received four blocks of target reversals followed by a block of the New CM condition (32 trials). This sequence, four Reversal condition blocks followed by a New CM block, was repeated five times within the reversal session. Subjects completed a total of 160 trials for each of the six target reversal conditions and for the New CM condition.

In the Distractor Transfer conditions, previously trained VM sets were used as target items and the types of distractors (i.e., previously CM High, Moderate, or Low trained <u>distractor</u> items) were manipulated. The transfer conditions were as follows:

1. High/High Distractor Transfer - both distractor items on a trial were previously CM High distractors.
2. Moderate/Moderate Distractor Transfer - both distractor items on a trial were previously CM Moderate distractors.
3. Low/Low Distractor Transfer - both distractor items on a trial were previously CM Low distractors.
4. High/Moderate Distractor Transfer - one distractor item was previously a CM High distractor item and the other was previously a CM Moderate distractor.
5. High/Low Distractor Transfer - one distractor item was previously a CM High distractor item and the other was previously a CM Low distractor.
6. Moderate/Low Distractor Transfer - one distractor item was previously a CM Moderate distractor item and the other was previously a CM Low distractor.
7. New CM condition - created by pairing two of the VM sets in a consistent mapping.

The New CM condition was included as a comparison condition. The six Distractor Transfer conditions were manipulated within a block and the New CM condition was presented in a separate block. The testing sequence was exactly the same as that used in the reversal session. Four blocks of Distractor Transfer (48 trials) were completed, followed by one block of the New CM condition; the distractor transfer session consisted of five repetitions of this sequence. Subjects completed a total of 160 trials per Distractor Reversal condition and 160 trials for the New CM condition.

Design. The within-subject independent variables were (a) Training Conditions: CM High, CM Moderate, CM Low, and VM; (b) Target Reversal Conditions: High/High Target Reversal, Moderate/Moderate Target Reversal, Low/Low Target Reversal, High/Moderate Target Reversal, High/Low Target Reversal, Moderate/Low Target Reversal, and New CM; and (c) Distractor Transfer Conditions: High/High Distractor Transfer, Moderate/Moderate Distractor Transfer, Low/Low Distractor Transfer, High/Moderate Distractor Transfer, High/Low Distractor Transfer, Moderate/Low Distractor Transfer, and New CM. The CM, Target Reversal, and Distractor Transfer conditions were manipulated within blocks whereas VM and New CM were manipulated between blocks. The dependent variables were RT and accuracy.

Results

Training Results. A one-way analysis of variance (ANOVA) was performed on the RT scores for the first session of training to assess the effect of Training Condition (CM High, CM Moderate, CM Low, VM). There was a significant effect of Training Condition, $F(3,45) = 13.78$, $p < .0001$. A Newman-Keuls comparison of the Training Condition revealed that the CM High, CM Moderate, and CM Low condition were all significantly different from VM.

To compare the effects of practice across the training conditions a 4 x 2 (Training Condition x Practice -

First/Last Session) ANOVA was conducted on the first 80 trials of each condition (in session one) and the final 80 sessions of each condition (in session seven). These data are plotted in Figure 1. This analysis revealed significant main effects of Training Condition, $\underline{F}(3,45) = 17.89$, $\underline{p} < .0001$, and Practice, $\underline{F}(1,15) = 145.66$, $\underline{p} < .0001$. The Training Condition by Practice interaction $\underline{F}(3,45) = 3.95$, $\underline{p} < .014$ was also significant. As can be seen in Figure 1, the source of this interaction is the Low CM training condition as shown by the presence of the Training Condition by Practice interaction $\underline{F}(2,30) = 5.41$, $\underline{p} < .01$ even when the VM condition is removed from the analysis.

A Training Condition x Practice ANOVA on the accuracy data yielded significant main effects of Training Condition, $\underline{F}(1,15) = 6.19$, $\underline{p} < .03$, and Practice, $\underline{F}(3,45) = 4.67$, $\underline{p} < .007$, but the interaction was not significant ($\underline{F} < 1$). The average accuracy for the CM conditions was 96 percent, which was slightly better than the VM condition (94 percent). Furthermore, there was a slight decrease in accuracy across sessions from 96 percent to 95 percent.

<u>Target Reversal</u>. A planned comparison of the means of the Reversal conditions to the New CM control condition showed a significant effect of Reversal, $\underline{F}(1,90) = 7.36$, $\underline{p} < .008$. Thus, regardless of the pairings of the items, if former CM targets (whether High, Moderate, or Low trained) were used as distractors, they were disruptive to performance. In other words, the subjects were unable to ignore the previously attended items. The accuracy scores ranged from 94 percent to 95 percent, but there were no clearly meaningful patterns of differences among the conditions.

<u>Distractor Transfer</u>. A planned comparison of the means of the Distractor Transfer conditions to the New CM control condition did not yield a significant effect of Transfer condition, $\underline{F}(1,90) = 3.24$, $\underline{p} < .076$. The accuracy scores

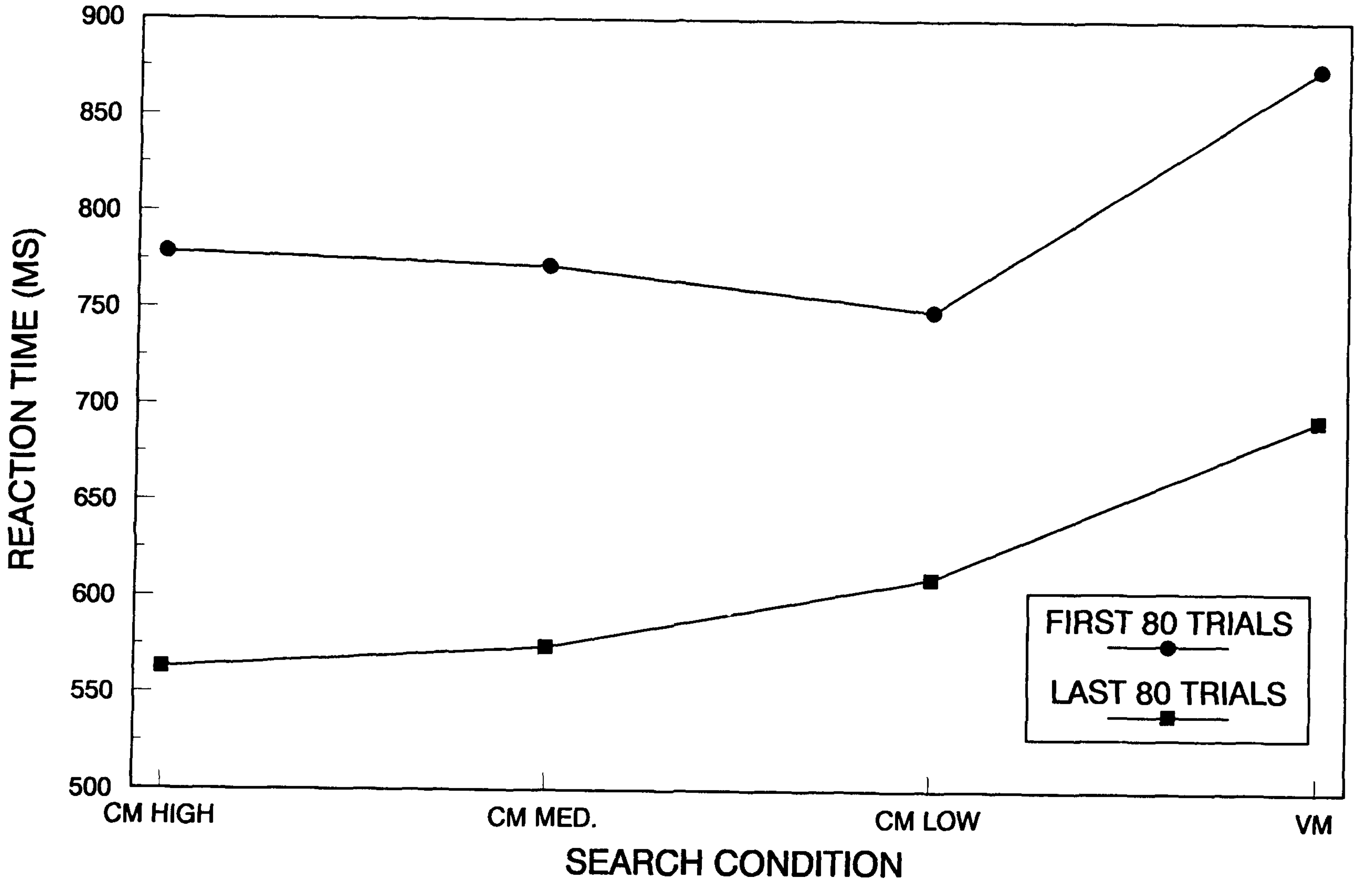


Figure 1. Reaction Time for the First and Last 80 Training Trials Plotted as a Function of Training Condition.

ranged from 94 percent to 96 percent, and there were no significant differences among the conditions.

Discussion

Fisk et al. (1990) reported that 3,150 trials of CM practice resulted in performance relatively similar to that of 1,575 trials of practice. We suggested that those results were due to the fact that subjects received practice on all conditions within a block; hence, there was the possibility that a "superset" of the memory set items was unitized during practice. In essence, we speculated that the form of practice we provided allowed the Low and Moderate training conditions to benefit from the frequently occurring High training condition due to associative learning. A major reason for conducting this present experiment was to further investigate those findings reported by Fisk et al. (1990). With the present design, because the search conditions were manipulated between blocks of trials, the effects of unitization should be at least attenuated. Unfortunately, the present findings do not allow a strong statement regarding the "unitization" hypothesis previously put forward.

It is true that the present Low CM training condition did not show the same relatively good performance (compared with the High and Medium training conditions) as that found in the Fisk et al. (1990) within-block training experiment. However, the expected "graded" effect of performance improvement across amounts of practice did not occur. The High and Medium CM training conditions did not differ even with the present experimental design.

As we will demonstrate in the following sections of this report, similar performance does not necessarily imply the same qualitative learning. However, the present data certainly suggest that fewer trials of practice than previously suggested in the literature may be needed for performance to reach a level of high proficiency. Performance may not be automatic in the sense that it may still be resource-sensitive, may still be under the control of the subject (but see our Target Reversal data), and so on. However, performance is certainly within the late phases of the associative phase of skill development (intermediate phase of skill development, see Ackerman, 1986, 1988; Anderson, 1982, 1983; Fitts, 1964; Fitts & Posner, 1967).

The present data, examined in light of the experiments reported previously which examined performance improvement as a function of practice, may have substantive implications for understanding the locus of CM performance improvements. The fact that when amount of training is manipulated between subjects, 3,000 trials of practice lead to performance superior to 2,000 practice trials and that 1,000 trials of practice lead to performance superior to that of subjects receiving 500 practice trials clearly argues that at least a partial locus of CM practice is stimulus-based. However, the previous experiment, which manipulated practice within-subjects and within blocks of trials, demonstrated that 3,000 practice trials did not result in performance superior to that obtained in 1,500 trials of practice. The present experiment replicated that latter finding using a within-subjects, between-block manipulation, thus ruling out the possibility of memory-set unitization as the major cause of that within-subjects training effect.

The present data suggest that CM practice is clearly important for stimulus-based strengthening; however, CM practice seems to facilitate performance in another important manner. Our data seem to support and extend the context activation hypothesis proposed by Schneider and Fisk (1984) as an important locus of CM training. That framework assumes that consistent exposure to the training context is a critical factor leading to performance improvement. This line of reasoning suggests that neither stimulus-based target strengthening nor consistent training context is sufficient (within the number of training trials presently provided) to lead to automatic target detection. Both are necessary for observed qualitative performance changes to be observed with CM practice. However, the present data suggest that limited target strengthening paired with strong training context will lead to performance equivalent to that achieved with moderate target strengthening. Hence, the expectations regarding improvements from part-task training may need to be lowered if part-task training provides drastically different context. Benefits from part-task training will be realized; however, those benefits will be stimulus-specific. If part-task training can be developed such that context can be activated during part-task training, then fewer exposures may lead to greater task-specific benefits.

These statements must be tempered somewhat because the present training did not examine performance after tens of thousands of practice trials. After such extensive practice, stimulus-based processing may supersede the training context. (Schneider and Shiffrin, 1977, reported subjects experiencing trouble reading subsequent to CM CRT-based letter detection training because the trained letters "popped-out" of the page. Clearly this demonstrates stimulus-based processing superseding training context; however, those subjects had received well over 10,000 trials of practice.)

More work is needed to examine this issue because it clearly has implications for cost-effective sequencing of training. The data suggest that proper sequencing may afford cost-efficient benefits by allowing the overall amount of practice to be reduced -- with similar benefits obtained by proper "packaging" of part-task training. These suggestions must be examined in more complex tasks and training environments. The issue of context seems crucial to the total understanding of CM part-task training benefits and deserves a prominent place in future research programs.

III. EXPERIMENTAL SERIES 2: THE EFFECTS OF PART-TASK TRAINING ON MEMORY-SET UNITIZATION: LEARNING AND RETENTION

Introduction

Part-task training refers to the provision of practice on specific components of a task prior to practice on the whole task. An important assumption of part-task training is that the task components can be identified, separated, and trained to improve total task performance more efficiently than training the whole task. However, as will become apparent in our review, specifying when part-task training will be effective is not always straightforward. In this introduction, the types of part-task training are reviewed. Advantages of part-task training as well as disadvantages are highlighted throughout the section.

In 1960, Adams expressed the following hopes for the future goals of part-learning research: (a) to find conditions where equal or lesser amounts of part-task practice can yield equivalent or higher levels of performance than whole-task practice; and (b) to accomplish the same goals of training or maintenance of response proficiency using part-task training for which the cost and complexity of simplified equipment will be less than for whole-task training. In the past 30 years, much of the training research has supported these hopes, at least for some types of tasks.

Types of Part-Task Training. Wightman and Lintern (1985) reviewed three part-task training methods.

1. Segmentation involves partitioning the task on temporal or spatial dimensions. Subtasks are practiced separately and then recombined into the whole task. This is comparable to teaching students to solve complex algebra problems by first training them to add, subtract, multiply, and divide.

2. Fractionation is used for whole tasks in which two or more subtasks must be executed simultaneously. For example, aircraft control during straight-and-level flight

may be partitioned into the subtasks of pitch control and roll control (Wightman & Sistrunk, 1987). Similarly, tracking tasks may be partitioned into control dimensions, perceptual and motor components, and procedural components (Wightman & Lintern, 1985).

3. Simplification involves making a difficult task easier by adjusting the characteristics of the task. For example, in a gross sense, training people to speed-read is virtually impossible unless you have first taught them to read. This type of training is related to the method of adaptive training, which will be explained later. (Note: Adaptive training usually involves simplifying the whole task, as opposed to decomposing it and training each part separately.)

Reintegrating the Trained Components. Ultimately, the entire task must be performed as an integrated whole. Wightman and Lintern (1985) defined three possible schedules for reintegration of parts, or subtasks, to the whole task. Pure part-task training involves first practicing the subtasks in isolation and then recombining them into the whole task. In the repetitive part-task training procedure, a single subtask is trained; then another subtask is added, and then another, until the whole task is being trained. Progressive part-task training is similar to repetitive part-task training except each part is first trained in isolation before being added in.

Although segmentation, fractionation, and simplification are all methods of part-task training, there are critical differences between the three techniques. In segmentation, the task is broken into its components but these tasks need not be performed simultaneously, even when the whole task is being performed. In fractionation, on the other hand, concurrent tasks are broken into components and trained separately. More careful reintegration is therefore required because there may be a crucial interrelation among components which surfaces only when the components are performed simultaneously (see Cream, Eggemeier, & Klein, 1978). Finally, simplification is most like segmentation in that components of the task are trained separately. However, segmentation methods do not involve a change in the make-up of the components, whereas simplification techniques make the task easier for training purposes by literally changing the characteristics of the task.

Each of these methods -- segmentation, fractionation and simplification -- will be explained in greater detail in the following sections, along with supporting empirical evidence for their success. Adaptive training methods and componential training approaches will also be explored in detail.

Determining what kind of part-task training to use -- if indeed, part-task training is used -- is not simple. The choice appears to be driven by the type of task to be trained. General guidelines are as follows: (a) The most successful method of segmentation has been backward chaining, in which the final segment of a task is trained prior to the sequential addition of all the preceding tasks. (b) The simplification technique is most successful for tasks which are initially very difficult to learn. By altering the task so that it is easier to perform initially, subsequent performance of the whole task is improved. Although there is evidence that simplification may not necessarily be better than whole-task training, it is often cheaper and less frustrating for trainees trying to master a seemingly impossible task at the criterion difficulty level. (c) Fractionation is the least supported method in terms of the empirical studies reported to date. The lack of support for fractionation as a viable training procedure over whole-task training is due mainly to the fact that it involves separating components which must ultimately be performed simultaneously. However, the fractionation method is beneficial if it is paired with some amount of whole- or dual-task practice.

Measurement Issues

An important consideration for the assessment of part-task training techniques is the measurement or quantification of benefits of part-task training relative to whole-task training. Wightman and Lintern (1985) proposed the use of differential transfer as a measure of the effectiveness of part-task training. Differential transfer refers to the "relative effects of equal amounts of experience with experimental [part-task training] and control [whole-task training] groups" (Wightman & Lintern, 1985, p. 271). If the differential transfer is greater than 100 percent then one may conclude that part-task training is more efficient. If it is less than 100 percent, then part-task training is less efficient than whole-task training but it does teach some skills which are useful for the performance of the criterion task: that is, it does not yield negative transfer.

. Flexman, Roscoe, Williams, and Williges (1972) expressed the importance of using the Transfer Effectiveness Ratio (TER). This measure of transfer takes into account the amount of practice on the prior tasks. The use of the TER permits a cost-benefit analysis of ground training devices. In other words, if a large amount of prior practice was necessary for positive transfer to the whole task, then the use of a part-task training procedure might not be cost-effective. Flexman et al. (1972) also warned that there are other considerations due to the complexity of measuring transfer effects. For example, simulator training transfers not only to the maneuvers in the airplane but also to other simulator maneuvers. Therefore it is important to separate the effects of transfer from simulator to simulator and those from simulator to airplane. Another consideration involves the fact that training one aspect may transfer to a totally different aspect simply because mastery of the first component allows the devotion of more time to the second component. Such confounding can be reduced by having the subjects first master a task in the simulator and then in the airplane before moving on to the next exercise.

Part-Task Training Procedures

Segmentation. Segmentation involves breaking the whole task into components which are trained separately and then recombined. One of the advantages of the segmentation method is that it allows the training procedure to focus more on the difficult components of the task, thus allowing more time to be allocated to training these components specifically. Bailey, Hughes, and Jones (1980) used a backward chaining procedure to train a dive bomb maneuver. They provided practice on the final segment of the task first and then added all the preceding tasks. These subjects reached criterion faster and had significantly fewer errors than did the control group, who had been trained on the whole task.

Wightman and Sistrunk (1987) also used a segmentation procedure similar to a backward chaining technique. They were training carrier-landing, final-approach skills using a simulator. The subjects first practiced on the terminal phase, which allowed for intensive practice on the critical elements of the task. The segmentation involved first 2,000 feet from touchdown, then 4,000 feet, then 6,000 feet (the criterion). The subjects trained under the segmented training conditions not only had more accurate performance but also showed differential transfer relative to those trained on the whole task. In fact, "...the positive effects of the chaining procedure more than compensated for the effects of smaller amounts of practice with the training task and the greater dissimilarity between training and transfer tasks" (p. 252).

Westra (1982), using a pure-part technique, trained subjects on a task involving a circling approach to landing. Subjects were first taught the straight-in approach. The results showed a superior lineup approach for these subjects. It was seen as important that there was not a significant decrement in transfer from the straight-in approach to the circling approach.

Wightman (1983) trained a straight-in carrier approach using a repetitive part-task technique. The subjects started with less distance to the approach and increased the distance, in three steps, to the whole distance. Part-trained subjects had lower errors relative to those subjects trained on the whole distance throughout the experiment. Sheppard (1984) trained the same task as Wightman, but held landing area stable. He found positive transfer but also more errors for part-trained subjects. Sheppard concluded that the mere isolation of a critical element for extended practice does not seem to be particularly useful. That is, the component chosen for prior practice must be a crucial part of the whole task.

Though all the aforementioned experiments which used some type of segmentation technique demonstrated positive transfer for part-trained subjects, the most successful procedures involved backward chaining. The importance of backward chaining may be due to knowledge of results (KR). For long tasks, earlier segments are not associated with the feedback of the end result. This is comparable to the rationale for using backward chaining in traditional learning theories; namely, well-learned task segments which occur late in the sequence may serve as feedback for earlier segments. According to Wightman and Sistrunk (1987), "...lengthy perceptual motor skills may be naturally acquired in a backward chaining progression, in which later task segments, once well learned, become the source of information feedback for earlier segments" (p. 252). Also, using this procedure, subjects are better able to associate the error feedback with the incorrect response.

Suggestions for Segmentation. The best tasks for segmentation appear to be those which have a high variability between the difficulty levels of the various components. The segmentation procedure allows the training program to focus on those tasks which have the highest levels of difficulty and therefore might require larger amounts of training. Though the segmentation procedure focuses on the most difficult components of a task, as does the simplification procedure, there are important differences between the two. In simplification, the components are, by definition, made easier to facilitate learning. In segmentation, however, increased training is provided for the difficult components but the characteristics of the task (i.e., the difficulty level) remain unaltered.

Fractionation. Fractionation may be used for whole tasks in which two or more subtasks must be executed simultaneously. The results from studies using fractionation methods are not clear-cut; that is, some of them show differential transfer while others demonstrate only equivalent performance for part- and whole-task training methods. For example, Briggs and Brogden (1954) used this technique to train a two-dimensional lever-positioning task. Using pure part-task training, they provided one part-task training group with practice on only one dimension and another part-task training group with single-task practice alternated between the two dimensions. The performance of the part-task training groups was compared to that of a group given practice on the whole task. The results showed that although there was some positive transfer for the part-trained groups, their performance was not better than that of the control group trained on the whole task. Stammers (1980) also trained a two-dimensional tracking task and his results did show positive differential transfer between part-task training and whole-task training.

Adams (1960) trained a bomb delivery task partitioned into continuous tracking parts and discrete motor responses. He did not find any difference for this training method relative to the groups trained on the whole task.

Mane (1984) used pure part-task training procedure to train a Space Fortress Game. The subjects' task was to fire missiles from a maneuverable spaceship, with the goal of destroying a space fortress while simultaneously evading the missiles being shot at their ship. The components of this task involved memory, timing, and psychomotor control. The whole-task trained subjects took longer to reach criterion and the part-task trained subjects had higher performance levels throughout. In fact, the savings (i.e., in necessary amount of practice) to criterion were more than double the time invested in pre-training.

At this point it is necessary to question the fact that there are discrepant findings from various studies using the fractionation method of part-task training. These discrepant findings are most likely due to the types of tasks involved. Wightman and Lintern (1985) delineated an important consideration for deciding when to use the fractionation method: If there is a high interaction between subtasks, part-task training will not be beneficial. Therefore, if performance on the components of the task will interact to some degree, then training them separately may not be as beneficial as training them together. However, it may still be beneficial to train the components separately for some time and pair this training with subsequent whole-task training for optimal performance. The types of tasks most frequently trained with the fractionation method are more like dual-tasks. In other words, these are actually two separate tasks which must be performed simultaneously. Schneider and Detweiler (1987) have reported that under these circumstances single-task training may be necessary, but not sufficient, for successful dual-task performance. They proposed that some level of proficiency (i.e., fast and accurate) should be reached on the single task (i.e., part-task) prior to advancing to multiple task (i.e., whole-task) training. These issues will be developed further in the sections devoted to the types of tasks which should be trained with part-task training, whole-task training, or some combination thereof.

Suggestions for Fractionation. Wightman and Lintern (1985) also offered suggestions for other manipulations within the realm of the fractionation method of part-task training: (a) more systematic partitioning; (b) follow the natural order of task; (c) concentrate on the dominant skills required for the task; (d) focus on the identifiable stages of skill acquisition (Jaeger, Agarwal, and Gottlieb, 1980, propose a possible hierarchy of stages: directional relationships -> timing -> amplitude -> coordination -> organization (spatial and/or temporal)); (e) perceptual pre-training, if this is a critical component of task and is inexpensive; and (f) time compression to allow more trials of practice (e.g., Vidulich, Yeh, & Schneider, 1983).

Simplification. Simplification is a part-task training technique that involves breaking tasks into components and training them separately. This is the type of part-task training employed in the current experiment. (Actually, the training is adaptive with the experimental groups receiving differential simplification with progressive part-task training.) The key to the simplification method is that not only are the components trained individually but they are also simplified to facilitate learning. The greatest benefit of simplification accrues mainly for tasks which are very difficult to learn. If a task is so difficult that it is seemingly impossible for a trainee to master it, making the task easier will allow novices to successfully perform it. Training can then proceed by gradually increasing the level of difficulty to match that of the criterion task. Simplification need not involve making the exact task easier but instead, training on a similar but easier task. For example, House and Zeaman (1960, cited in Wightman & Lintern, 1985) demonstrated that difficult pattern discriminations are easier to learn after practice with easier object discriminations (also see below, Gordon, 1959;

Poulton, 1974). The assumption here is that the skills learned in the performance of the easier task will transfer to a more difficult version of the task.

Briggs and Waters (1958) manipulated the component interaction of a pitch and roll tracking task. They varied the amount by which system responses on one dimension were affected by control movements on the other dimension. Subjects were trained on high, medium, and low levels of component interaction. This manipulation yielded positive differential transfer but less than 100 percent, indicating that performance was not better than whole-task training (although it was not worse either).

Poulton (1974) and Gordon (1959) trained subjects on pursuit tracking displays before training them on compensatory tracking displays. Pursuit tracking is easier than compensatory tracking but contains many of the requisite components for compensatory tracking. Their results showed improved performance relative to subjects originally trained on the compensatory displays. Although these results are generally supported by other investigators (e.g., Jensen, 1979; Roscoe, Saad, & Jensen, 1979) contradictory findings also appear in the literature (e.g., Briggs & Rockway, 1966; Simon & Roscoe, 1981)

Wightman and Sistrunk (1987) used a simplification technique to measure carrier landing final approach skills. By reducing the gross weight of the simulator, they achieved a reduced lag between a control input and the perceptible responses. Successive approximations to the true system lag were then produced in an effort to allow maximal acquisition of early proficient performance of the carrier glideslope tracking task. This manipulation of aircraft response (i.e., time lag) was not effective. In fact, transfer for low-aptitude subjects suffered as a result of training with progressive lag. Wightman and Sistrunk suggested that it is possible that lower-ability subjects may require higher levels of fidelity for control display lags between training and transfer relative to higher-ability subjects.

Overall, there is not much evidence that simplification part-task training is better than whole-task training. However, because there is also no evidence of negative transfer from this method, it might be useful if it is less expensive than whole-task training. Also, if criterion level performance is so difficult that novices would not be able to perform the task initially, then simplification is useful. For example, in teaching a novice baseball player to hit pitches, requiring this individual to practice with 90-mile-per-hour pitches would lead to minimal improvement.

Strategies for Simplification. The following suggestions for simplification methods are offered by Wightman and Lintern and are supported by the present literature review: (a) provide prior training on medium difficulty; (b) manipulate the display type (e.g., pursuit vs. compensatory); and (c) provide augmented feedback. A method based on the underlying tenets of simplification is adaptive training, which usually involves simplifying a whole task as opposed to simplifying specific components of a task. This method of training is described in depth in the following section.

Adaptive Training

McGrath and Harris (1971) offered the following definition of adaptive training: "Adaptive training is training in which the problem, the stimulus, or the task is (automatically) varied as a function of how well the trainee performs" (p. 2). Adaptive training methods are also referred to as "self-adjusting simulators," "self-organizing systems," "computer-aided instruction," and "programmed instruction."

In an adaptive system, the task starts out easy and becomes progressively harder. This approach is thought to reduce the frustration level of the subject -- an important consideration for the maintenance of the trainee's effort and motivation during practice (Schneider, 1985a). For instance, in a fixed training program if the task is very difficult, there might not be any improvement in performance for a long time. Not only is this frustrating for the trainee; it is also a waste of training time.

The adaptive system is set up to hold performance constant (e.g., at a preset accuracy level) and vary the adaptive variable. By keeping performance the same, the experimenter can use the change in the difficulty level as an index of skill. An adaptive variable is generally anything that affects the difficulty level of the task. This might include such factors as stress to the trainee (e.g., the simulated environment), characteristics of the display, display lag, information or communication load, control damping, etc. Furthermore, the adaptive variable may be varied continuously, at one of two rates (i.e., easier or harder based on accuracy) or in discrete jumps. According to McGrath and Harris (1971), the method by which the variable is changed is trivial because various methods function equally well. The choice of method will depend on the nature of the training system implementation (e.g., it is more difficult to program a method of continuous variation on a computer).

McGrath and Harris (1971) offered the following guidelines for selecting adaptive variables:

- The variable should be experimentally determined and/or selected through task analysis; the variable chosen will be unique to different training objectives and tasks.
- The variations should be easily definable or measurable.
- Consideration should be given to the ease of varying the difficulty level, as well as the nature of the difficulty dimension.

- The variable selection and the parameters of adaptive difficulty levels should be related to progress towara the training objective.
- The difficulty of the adaptive variable should be adjustable over a wide range of skill levels.
- The variables and their progressive difficulty levels should be consistent with the real-world task. This is important because, as McGrath and Harris (1971) pointed out, "...in designing an adaptive task, it makes sense to find out how the task is performed in the real-world situation, because where you begin training may not be as important [in terms of the training program design] as long as you end at the right place" (p.23). However, one must be cautious when selecting the appropriate starting difficulty level. The task must be easy enough to produce successful performance but, as we have noted elsewhere (Eggemeier, Fisk, Robbins, Lawless, & Spaeth, 1988), the final-level consistencies should be present.

Adaptive training is a form of instructor simulation in that is represents an effort to formally structure, while at the same time individualize, instruction in perceptual-motor tasks. This is important because, as McGrath and Harris (1971) pointed out, differences in motivation and background of individual instructors contribute the greatest variance in training programs.

The following situations are defined by McGrath and Harris (1971) as the most useful times or situations in which to use adaptive training systems:

- When the task is difficult enough to require extensive training.
- When the training may be computerized.
- For tasks requiring overlearning and high retention over time.

- To mechanize the instructor's adaptive function; that is, to formalize the decision logic concerning when to promote students to more difficult levels.
- To ensure standardization of the training situation.
- When the task is so difficult that it cannot be learned unless it is broken into its component parts.
- In some cases in which divided attention and time-sharing are required (Making one of the tasks easier enables the trainee to allocate more attention to the other task.)
- For perceptual-motor tasks which are initially too difficult.
- When new elements of performance are added.
- When new items of information of tasks must be mastered in addition to already demanding tasks.
- For progression from part-tasks to whole complex tasks.

Mane (1984) reported that, for adaptive training to be whorthwhile, "the transfer from one version of the task to the other should be larger than the equivalent amount of training on the target task" (p. 522). Mane provided subjects with whole-task adaptive training on the perceptual-motor components of the Space Fortress Game (see the fractionation section above for a more detailed description of the task) by gradually increasing the difficulty (according to the speed of the task). Mane proposed that reducing the pace of external events (i.e., the speed of the task) would make subjects better able to pick up the relations among the task elements. Mane used two adaptive conditions starting at differing levels of difficulty. The results showed that those subjects who were trained starting at the very slow rate showed no advantage over subjects who started out at the criterion rate (there was actually some negative transfer). However, the group that started out at the medium speed showed improved performance over that of the control group.

The results of a study by Ammons, Ammons, and Morgan (1956) showed similar effects of transferability among difficulty levels. They manipulated rotation speed by varying the difficulty level: high, medium, and low difficulty. They found benefit (i.e., positive transfer) from medium to high but not from low to medium or low to high. These results suggest that changing a fast-paced task to a very slow-paced task may violate the assumption that the relations among elements do not change. If the important relations or consistencies are different in a part-task relative to a whole-task, then it is more probable that there will be negative transfer. This may be the cause of the results found by Mane (1984) and Ammons et al. (1956) when transferring subjects from the slowest condition to the criterion task.

An important factor in an adaptive training program is the type of feedback provided. Intrinsic feedback is a natural consequence of movement or action such as kinesthetic cues. Although this type of feedback is ever-present, it is less effective in motivating performance than is augmented feedback. Augmented feedback is based on external sources of information about performance on a task.

Fitts and Posner (1967) reported the results of a study by Smode (1958). In Smode's experiment, subjects were given augmented feedback in the form of a counter which kept a running tab of their scores. The performance of these subjects was compared to that of a group of subjects who received normal feedback in the form of verbal reports of performance. The "augmented" group showed much higher performance and it was assumed that they worked at a higher level of motivation. According to Lintern and Wickens (1987), "...the evidence suggests that guidance [e.g., augmented feedback] is likely to enhance the acquisition of skills with complex stimulus-response relationships, but not those with simple or compatible stimulus-response relationships" (p. 30). They added that "where a consistent mapping is to be learned, learning is enhanced by manipulations that reduce errors in training or that reduce resource loads, while those manipulations that increase errors or resource loads retard learning. Where the mapping is inconsistent (i.e., random) or is already well-learned (i.e., compatible), these manipulations have no effect in learning." (p. 30)

Eberts and Schneider (1985) also demonstrated the value of augmented training and their studies indicate when augmented training will be most effective. Eberts and Schneider examined subjects' ability to control a continuously moving track in a second-order system. Their subjects were given different types of augmented feedback during training. Eberts and Schneider found that only augmentation that made salient the consistent relationships between control input and system output produced superior performance in solving system related control problems. Eberts and Schneider suggested that subjects only benefited by receiving consistent cues because those subjects could develop an internal (mental) model of the system. This internal model aided in control of the system when the subjects were transferred to situations different from those specifically encountered during training.

Finally, the importance of augmented feedback has been empirically demonstrated by Lintern, Thomley, Nelson, and Roscoe (1984). Using adaptive training on an air-to-ground bombing task, they found better performance in augmented-feedback training. These and other results (see Lintern & Roscoe, 1980, for a review) demonstrate that training with augmented feedback can speed skill acquisition.

Overview of the Experiments

The experiments reported in this section examined the effect of memory-set component training on both learning and retention of performance in a hybrid memory/visual search task. Performance on the task was examined as a function of the amount of material to be learned (and the manner in which it is presented). All subjects received adaptive frame-speed training so that we could examine performance at each individual subject's limits of perceptual processing (but with stimuli always presented above threshold). The part-task training groups received simplification, progressive part-task training on a hybrid memory/visual search task. The full task required detecting exemplars from six categories within a stream of 24 display items. Little, if any, emphasis has been placed on the empirical examination of part-task training in this class of tasks. It is important to understand whether part-task training will result in equivalent, worse, or better performance compared with full-task practice in tasks requiring associative learning (memory-set unitization) and automatic exemplar detection (target strengthening). We systematically examined the effectiveness of simplification using a progressive part-task training approach when full-task participation allowed performance to be guided by both target and distractor learning (Experiment 1) or just target learning (Experiment 3). This is important because many operational tasks performed by Air Force personnel require the learning of large numbers of categorized exemplars for fast, efficient detection. If building "superset" categories is not impeded by part-task training, then many of the benefits of part-task training outlined in the introduction could be realized in training this present class of tasks.

We also investigated the often overlooked issue in part-task training of the retention of the learned skill as a function of the type of part-task training. Even if part-task training is effective in producing effective performance in this class of tasks, it is crucial to know the degree to which that performance level will be retained. We may find that part-task training is effective in training associative learning and target-strengthening but also find that the learning is relatively fragile. However, the learning from part-task training may be as stable as whole-task training. In either case, an empirical evaluation of the retention of learning as a function of part-versus-whole learning is required and will provide valuable information to those engaged in training development.

Four experiments were conducted, two training (Experiments 1 and 3) and two retention (Experiment 2 and 4). In each training experiment, three training conditions were used, with each condition representing different memory loads. The conditions were (a) PT2, three different memory sets of two categories each, in which subjects trained on one memory set before moving on to the next (part-task training); (b) PT3, two different memory sets of three categories each (part-task training); and (c) WT6, one memory set of six items (full task practice). The paradigm used was the adaptive multiple frame procedure developed to test performance at each subject's perceptual processing limits. Subjects practiced for 6 days. After the initial practice, they were tested in the full task at various frame times. After testing, the subjects received another 6 days of practice, followed by full-task testing. In the retention experiments, subjects' performance in the full task was tested 30 days after receiving part-task or whole-task practice.

Experiment 1 - Combined Target and Distractor Learning

In the first experiment we examined the effectiveness of simplification, progressive part-task training relative to whole-task training when the full-task transfer afforded the subjects the opportunity to benefit from both target and distractor learning.

Experiment 1 - Method

Subjects. Eighteen undergraduate students, eleven males and seven females, were paid for their participation in the experiment, received credit for a psychology class, or a combination of the two. All subjects were tested for near vision (at least 20/40) and far vision (at least

20/30), were asked about their use of medication, and were administered three subscales (vocabulary, digit-symbol substitution, and digit span) of the Wechsler Adult Intelligence Scale-Revised (WAIS-R). The averaged WAIS-R scaled scores were representative of the average population: (a) vocabulary -- 13.00 (range 10 to 17), (b) digit span -- 12.17 (range 7 to 18), (c) digit symbol substitution -- 11.72 (range 7 to 16).

Apparatus. Epson Equity I+ personal computers were programmed with Psychological Software Tools' Microcomputer Experimental Language (MEL) to present the appropriate stimuli, collect responses and control timing of the display presentations. Standard Epson monochrome monitors (Model MBM 2095-E) connected to Epson multimode graphics adapters were used to display the stimuli. Subjects were tested at individual subject stations, with pink noise at approximately 55 decibels to mask outside noise.

Three areas of the screen were measured to calculate the appropriate visual angle data. The visual angle was determined using the average viewing distance of 46 cm from the screen. The memory-set presentations contained either two, three, or six semantic-category labels presented in a vertical column on the left side of the screen; the visual angles were approximately 1.2, 1.9, and 4.2 degrees, respectively. The target and distractor exemplars averaged six letters in length and were presented in a column of three words on the right side of the screen; the width (length) of the words subtended an average of 2.0 degrees; and the height of the three words combined also subtended 2.0 degrees.

Stimuli. The target and distractor stimuli were chosen from the taxonomic category norms compiled by Battig and Montague (1969). Six categories were used for the target sets and eight different categories were used for the distractor sets (the stimulus items were either targets or distractors; i.e., consistently mapped). The target set items consisted of words from the semantically unrelated categories (Collen et al., 1975) of COUNTRIES, EARTH FORMATIONS, FRUITS, HUMAN BODY PARTS, OCCUPATIONS, AND READING MATERIALS. The distractor set items consisted of words from the semantically unrelated categories of CLOTHING, DWELLINGS, FURNITURE, MUSICAL INSTRUMENTS, RELATIVES, UNITS OF TIME, VEHICLES, and WEAPONS. Each category in both the target set and distractor set contained six exemplars, and all words appeared in capital letters (see Appendix B for a complete list of exemplars).

Procedure. To train subjects to their visual search limit, an adaptive, multiple-frame procedure (Hodge & Fisk, 1989) was used. The procedure developed by Hodge and Fisk (1989) was based upon tasks previously used in the visual search literature (e.g., Schneider & Shiffrin, 1977; Sperling, Budiansky, Spivak, & Johnson, 1971). The main difference between the adaptive, multiple-frame procedure and the multiple-frame procedures previously reported in the literature was that the frame time (amount of time the category exemplars were presented) changed as a function of the subject's performance after each block of trials (see description presented below).

The multiple-frame procedure provided a method of presenting the subject with successive frames of stimuli (exemplars from the target and distractor categories) much like a rapid presentation from a slide projector (see Figure 2). Each frame consisted of three exemplars from either the target or distractor categories. Eight frames were presented, for a total of 24 exemplars per trial. On a positive trial, one of the 24 exemplars was a target exemplar, with the remaining 23 exemplars being drawn from the distractor categories. On a negative trial, all the words were distractor exemplars.

Frame times were "adaptively" manipulated on an individual subject basis. Adaptive frame times were set in response to the accuracy performance of each individual

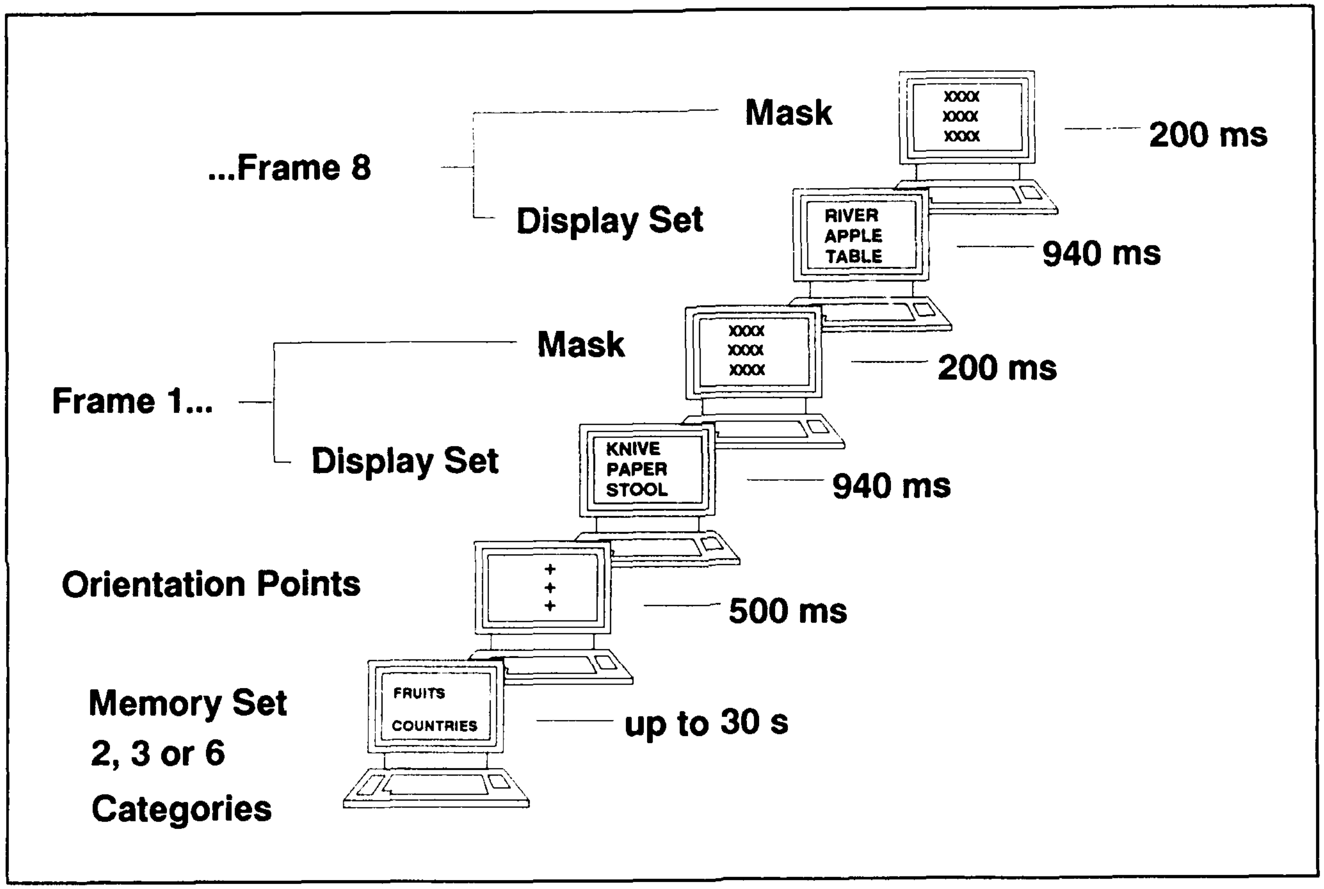


Figure 2. A Representation of the Successive Displays for the Multiple-Frame Procedure.

subject. The initial frame time was set to 940 milliseconds (ms), due to the high memory load in the six-category condition, and was the same across the three groups. Throughout training, each subject's performance (accuracy level) determined the frame speed for the following block. If the subject reached an accuracy level of 86.7 percent (26 out of 30 trials correct) or more on a block, then each frame for the next block was presented 20 ms faster. Likewise, if the subject did not reach an accuracy level of at least 73.3 percent (22/30) on a block, each frame for the next block was presented 20 ms slower. Otherwise, the speed remained the same as in the previous block of trials. The speed for the next training session for each subject was based on the frame speed and accuracy on the last block of the previous training session. The adaptive element of the multiple-frame procedure allowed accuracy to stabilize at approximately 80 percent. Frame times for the transfer sessions were based on pilot work and were set at three different frame speeds. The three speeds bracketed the mean performance of pilot subjects who had completed all the training sessions. Frame time was held constant during transfer to measure accuracy without the complications of the speed/accuracy interaction.

During the first session, subjects received written instructions on the task, were administered an eye test, and completed a practice session. The practice trials had categories other than those used in the actual experiment. The practice allowed the subjects to become familiar with the requirements of the task and the experimental environment.

Each trial consisted of the following sequence: First, the memory set (2, 3, or 6 categories depending on the between-subjects training condition) appeared on the left side of the screen in a column (e.g., FRUITS and OCCUPATIONS). After studying the category names for up to 30 seconds, the subject pressed the space bar to initiate the presentation of the frames of exemplars. Subsequent to the frame sequence initiation (pressing the space bar) and prior to the display of the frame sequence, three plus signs were presented in the center of the screen for 500 ms to allow the subject to fixate on the area of the screen where the category exemplars appeared. A frame consisted of three exemplars presented in a column on the screen, followed by a column of X's to mask the word presentation. The mask was used to prevent the potential timing inaccuracies associated with phosphor decay. A sequence of eight frames appeared on the screen in succession, much like a slide projector on high speed. On a positive trial only one of the 24 exemplars was from the target categories (e.g., Apple): the other exemplars presented were distractors. The position of the target within a frame (top, middle, or bottom of the column) was selected randomly. Likewise, the target occurred on a randomly determined frame, with the restriction that it had to occur between Frames 2 and 7, inclusive.

The subjects were required to note either the position of the target within the column of words (i.e., top, middle, or bottom) by pressing the corresponding key (labeled 'T', 'M', or 'B'), or that no target was present by pressing the 'NO' key (labeled 'N'). The top, middle, bottom, and "NO" keys corresponded to the 7, 4, 1, and 5 keys on the number pad, respectively. Subjects could respond at any point during the 8 frames and for up to 4 seconds after the final frame was presented.

The subject was then presented with a Likert-type scale to assess the degree of certainty concerning the choice of target presence or target absence. The scale ranged from 1 to 5, where 1 represented "Absolutely certain no target present"; 5 represented "Absolutely certain a target present"; and 3 represented "Guess." The 2 and 4 each represented an intermediate value between a guess and absolute certainty. This provided signal detection information regarding the criterion used for making visual detections.

A subject was provided with feedback after each trial and block. For successful trials, the words "correct response" appeared. For unsuccessful trials, a tone sounded and the words "incorrect, target in TOP (MIDDLE, BOTTOM) position" appeared if the incorrect position was given, or "no target present" appeared if a position was selected and no target had been presented. After each block, the percentage of correct trials and frame time were shown to the subject for all completed blocks during the session.

During training, subjects searched for an exemplar (target) from either two, three, or six categories (between subjects) against a background of distractor exemplars (words from eight categories semantically unrelated to the target categories). During transfer, all subjects searched for a target exemplar from the same six trained categories against a background of the same distractor exemplars.

Design. The 15 sessions of the experiment were broken into seven different stages which occurred chronologically: (a) Subject Orientation/Practice (one session); (b) Training I (six sessions); (c) Transfer I (one session); (d) Training II (six sessions); (e) Transfer II (one session); and, (f) Refresher (one session). Of primary interest was the performance of each group during Transfer I and Transfer II and the improvement of each group from the first transfer session to the second. Memory-set size was manipulated between subjects (i.e., two, three, or six categories). The primary dependent variables recorded during training were frame time (speed) and accuracy level; during transfer, frame times were held constant (see below) and accuracy level was of primary interest.

The practice session allowed subjects to become familiar with the task and to perform 3 blocks of trials with 30 trials per block, for a total of 90 trials. Training sessions consisted of 10 blocks of trials with 30 trials per block, for a total of 3,600 trials (1,800 trials for each of the stages: Training I and Training II). An average of 20 percent of the trials were negative (target absent) during each session. The actual number of negative trials varied between 5 and 7 out of 30 trials on any given block, with the mean being 6 negative trials per block for the entire session.

Three training conditions were manipulated between subjects (see Table 1 for an outline of the category training sequence for each training condition): PT2 - two categories in the memory set; PT3 - three categories in the memory set; and, WT6 - all six categories in the memory set. In condition PT2, subjects trained with two categories during each training session. After two training sessions, the categories changed to the next set of two categories. After six sessions of training, the PT2 subjects had received equal training on each of the six categories (i.e., on the average each category served as the target category an equal number of times). In condition PT3, subjects trained with three categories during each training session. After three training sessions the categories changed to the next set of three categories. Likewise, after six sessions of training, the PT3 subjects had received equal training on each of the six categories. Condition WT6 differed from the other two conditions in that all six categories were trained throughout the six sessions of training. As with conditions PT2 and PT3, the training on WT6 was equivalent in the average number of times each category was the target category. For conditions PT2 and PT3, the assignment of the categories to search days was counterbalanced across subjects by a partial Latin square.

In both Transfer I and Transfer II, all subjects completed 270 trials with the same six trained categories in the memory set (a total of 540 transfer trials for the experiment). Three blocks (30 trials per block) were run at each of the following frame speeds: 180 ms, 220 ms, and 260

Table 1. Category Training Sequence for Experiment 1 and Experiment 3.

	TRAINING						TRANSFER
Day	1/8	2/9	3/10	4/11	5/12	6/13	7/14
PT2	Category 1 Category 2	Category 1 Category 2	Category 3 Category 4	Category 3 Category 4	Category 5 Category 6	Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6
PT3	Category 1 Category 2 Category 3	Category 1 Category 2 Category 3	Category 1 Category 2 Category 3	Category 4 Category 5 Category 6	Category 4 Category 5 Category 6	Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6
WT6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6	Category 1 Category 2 Category 3 Category 4 Category 5 Category 6

ms, for a total of nine blocks. There were six negative trials per block (20 percent). The same target and distractor categories from training were used for both transfer sessions.

After the second transfer session an additional session was conducted as part of the retention phase of the experiments. The procedure for the refresher session was identical to the procedure of the transfer sessions.

Experiment 1 - Results and Discussion

Training. Mean frame times and accuracies for each training session were aggregated across subjects. Accuracy stabilized close to 80 percent after four sessions of training as a result of the adaptive procedure used. Mean Frame Times for all three conditions decreased over training sessions according to a normal power function (see Figure 3). A fit of the power function to each of the Training Conditions yielded $r^2 = .96$ for PT2, $r^2 = .98$ for PT3, and $r^2 = .96$ for WT6. Subjects' average Frame Time (aggregated across conditions) decreased from 879 ms after the first session to 216 ms in the last session of training.

Transfer. A repeated measures ANOVA was calculated to compare accuracy across Training Conditions (PT2, PT3, and WT6), Frame Speeds (180, 220, and 260 ms), and Transfer Sessions (I and II). In addition, for the part-task training groups (PT2 and PT3) tests were performed to assess accuracy differences between categories learned early in training (temporal order) versus those learned later in training (i.e., the PT2 group learned two categories during the first two sessions, then was not exposed to those categories again until the transfer session; likewise, the PT3 group learned three categories during the first three sessions, and then did not see them until the transfer session). Frequency data for the randomly chosen Target Positions, Target Categories, and Frame Numbers (on which the Target Exemplar appeared) were analyzed and appear in Appendix C.

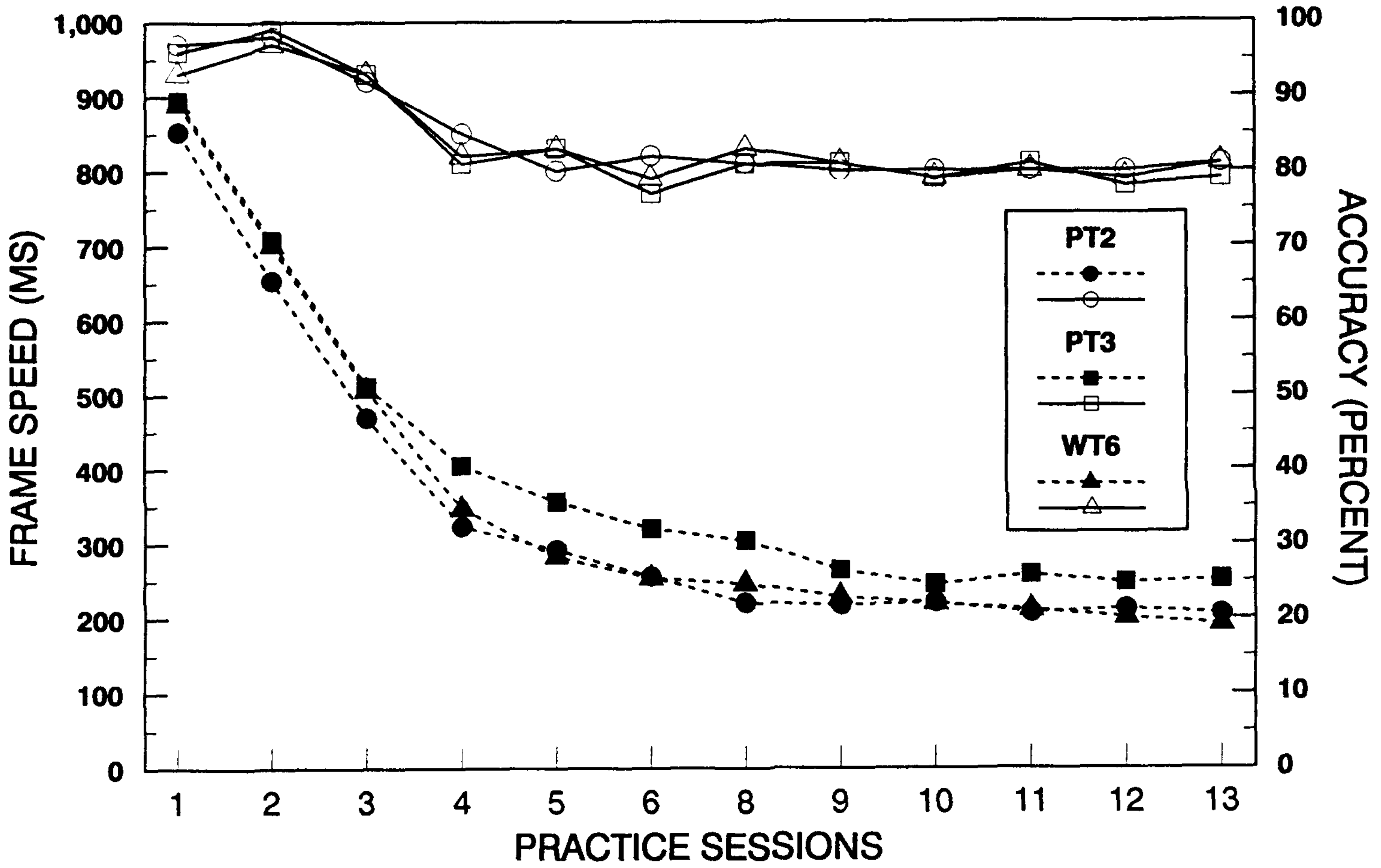


Figure 3. Frame Speed and Accuracy for Each Training Condition as a Function of Practice Session for Experiment 1.

Mean accuracy was determined for each Training Condition (PT2, PT3, WT6) across three different Frame Speeds (180 ms, 220 ms, 260 ms), and two Transfer Sessions (I and II). A 3 x 2 x 3 (Training Condition x Transfer Session x Frame Speed) repeated measures ANOVA was performed on the accuracy data and is summarized in Table 2. The main effect of Transfer Session was significant, $\underline{F}(1,15) = 13.67$, $\underline{p} < .0022$, $MS_e = .0327$, reflecting the improvement in accuracy after 6 additional days of consistent training. Also, the effect of Frame Speed was significant, $\underline{F}(2,30) = 33.05$, $\underline{p} < .0001$, $MS_e = .0020$. However, neither the main effect of Training Condition nor the higher-order interactions reached significance (specific values for the ANOVA are in Table 2). A Newman-Keuls test (alpha = .05) showed significant differences among all three Frame Speeds. A power test (Á=0.05, n=6, f=0.77, u=2; Cohen, 1977) on the Training Condition data revealed power equal to 0.75.

An analysis of Temporal Order was performed to test for the possibility of an effect due to category training sequence. That is, Temporal Order x Frame Speed ANOVAs were conducted to determine if the order in which the categories were learned in the PT2 and PT3 conditions had an effect on transfer performance. For PT2 the main effect of Temporal Order was not significant, $\underline{F}(2,10) < 1$, nor was the interaction of Temporal Order by Frame Speed, $\underline{F}(4,20) < 1$. Similarly, for PT3 the main effect of Temporal Order was not significant, $\underline{F}(1,5) < 1$, nor was the interaction of Temporal Order by Frame Speed, $\underline{F}(2,10) = 1.51$, $p = .266$, $MS_e=.0036$. This result indicates that whether a category was learned early or late in training did not make a significant difference during transfer.

Certainty scale data were collected for each trial after the subject made a target selection, but before the trial feedback. Because little difference in subjects' use of the certainty scale was found among Training Conditions,

Table 2. Summary of ANOVA for Experiment 1: Transfer Data

SOURCE	df_{num}	df_{den}	MS	F
Group	2	15	.0378	1.21
Speed	2	30	.0022	38.90***
Session	1	15	.0327	13.67**
Group x Speed	4	15	.0022	< 1
Group x Session	2	15	.0033	< 1
Speed x Session	2	30	.0024	2.33
Group x Speed x Session	4	30	.0024	< 1

* $p < .05$
** $p < .01$
*** $p < .0001$

the results will not be discussed in detail; they are presented in Appendix D.

Discussion. In this experiment, we did not find a difference among Training Conditions, indicating that the part-task training groups learned the categories as well as the whole-task group. Apparently, though there is no deficit for learning only a portion of the six categories during a training session, there is also no advantage. This finding is not necessarily surprising based on previous experiments which reported no benefit for part-task training (Adams, 1987; Adams & Hufford, 1961; Briggs & Brogden, 1954; Briggs, Naylor, & Fuchs, 1962; Briggs & Waters, 1958; McGuigan & MacCaslin, 1955). In fact, the most important finding from the present experiment with respect to part-task training decisions for this class of tasks may be that the part-task groups performed as well as the whole-task groups. Adams (1960) compiled statistics on basic research in the part-task training domain and found a training advantage of whole-task over part-task training of two to one.

Differences in mean accuracy between the two transfer sessions reflect the improvement after 6 more days of training. Performance at all three frame speeds differed, providing a range of measurements on which to compare the training conditions. Subjects found the fastest frame speed (180 ms) to be quite difficult during Transfer I, a 6 days of training were not sufficient for subjects to become as accurate at the fastest frame speed (see Figure 4). Pilot work predicted difficulty at very fast presentation rates, but challenging frame speeds were chosen intentionally to avoid a ceiling effect during Transfer II.

An analysis of Temporal Order was performed to test for the possibility of an effect due to category training sequence. If subjects performed more accurately during transfer on categories learned just before transfer, then the sequence of the training may have been suspect, and

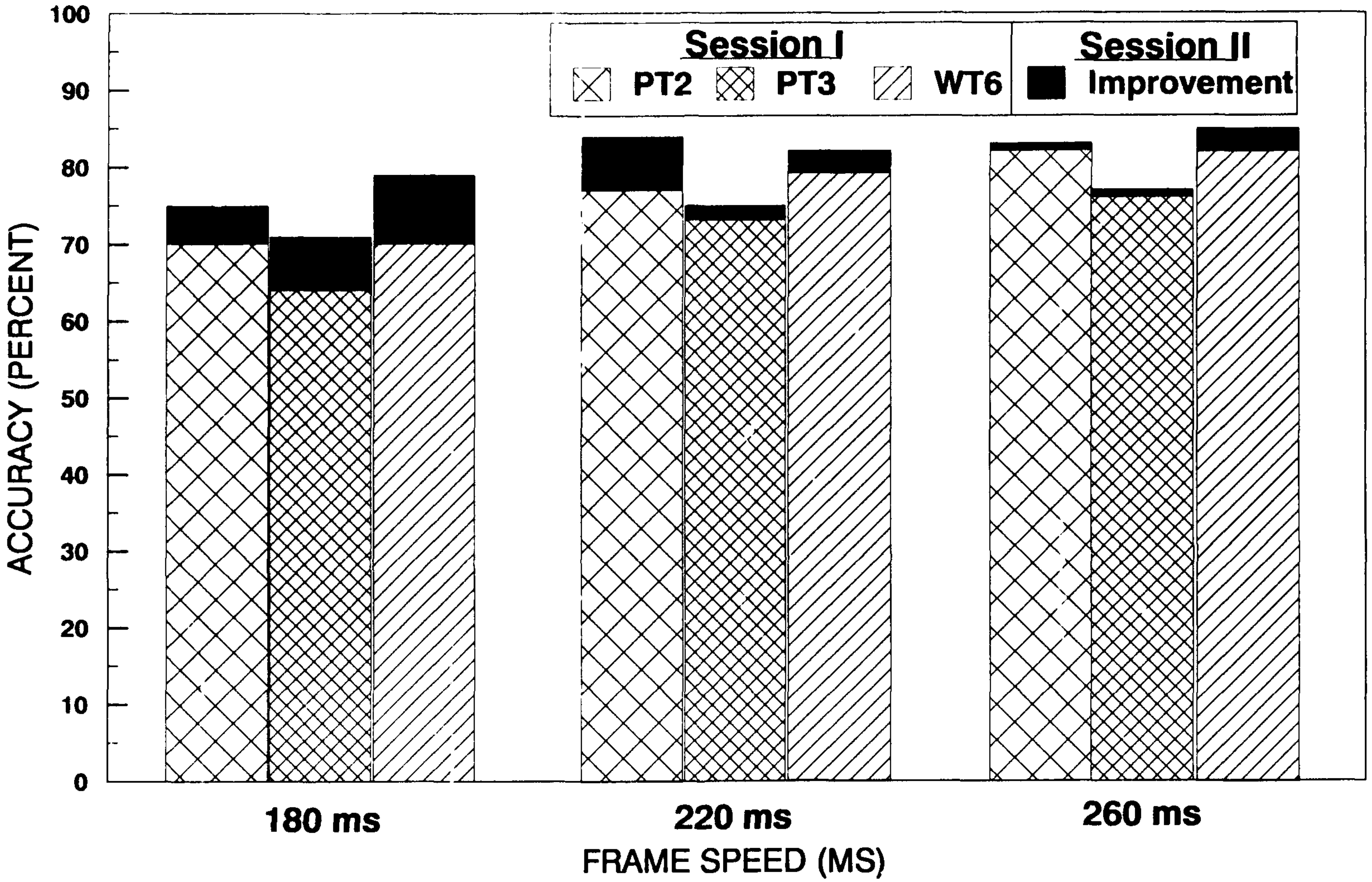


Figure 4. Accuracy for Each Training Condition as a Function of Frame Speed for Transfer Session 1 and 2, Experiment 1.

interpreting the results would have been a formidable task. The finding of no difference in learning order lessens the likelihood of a massed/distributed learning effect.

Experiment 2 - Retention Performance, Combined Target/Distractor Learning

Although the part-task training was as effective as whole-task practice in leading to efficient performance, the issue remains whether the unitization (and/or target-distractor learning) is as stable in the part-trained groups as the whole-task trained group. This issue of retention of the trained performance level is examined in the second experiment.

Experiment 2 - Method

The second experiment was a continuation of Experiment 1. The same subjects participated in one retention session 30 days following the end of Experiment 1. The session was identical to the previous transfer sessions.

Experiment 2 - Results and Discussion

Thirty-Day Retention. Following Experiment 1, subjects ran through an additional transfer session (Transfer III, day 17) and were asked to return 30 days later for another session (Transfer IV). Transfer III and IV consisted of exactly the same categories and frame speeds as the two transfer sessions of Experiment 1. One subject (PT2 condition) did not return for the 30-day session and those data are eliminated from the analyses. The mean accuracy data are reported in Table 3.

A 3 x 2 x 3 (Training Condition x Transfer Session x Frame Speed) repeated measures ANOVA was performed on the accuracy data. These analyses are summarized in Table 4. A main effect of session was found, $F(1,14) = 11.87$, $p < .0039$, $MS_e = .0024$, reflecting the small performance decline (4 percent) over the 30-day interval. The effect of Frame Speed was significant, as previously found in Experiment 1, $F(2,28) = 32.24$, $p < .0001$, $MS_e = .0020$, but the interaction of Frame Speed x Training Session did not reach

Table 3. Mean Accuracy for Transfer Sessions from Experiments 1 and 2

	PT2			PT3			WT6		
Frame Speed	180	220	260	180	220	260	180	220	260
Transfer Session III	77	82	85	71	77	81	78	84	85
Transfer Session IV (Retention)	75	79	84	68	70	77	73	80	81

Table 4. Summary of ANOVA for Experiment 2: Retention Data

SOURCE	df_{num}	df_{den}	MS	F
Group	2	14	.0420	1.12
Speed	2	28	.0020	32.24***
Session	1	14	.0024	11.87**
Group x Speed	4	14	.0020	< 1
Group x Session	2	14	.0024	< 1
Speed x Session	2	28	.0016	< 1
Group x Speed x Session	4	28	.0016	< 1

* $p < .05$
** $p < .01$
*** $p < .0001$

significance, F < 1. As indicated by the ANOVA, the decline in performance was relatively stable across frame speeds. The percentage decline of accuracy was greatest for the 220 ms Frame Speed (4.7 percent), followed by 180 ms (3.3 percent), and finally 260 ms (3 percent). Again, no differences were found among Training Conditions, F(2,14) = 1.12, p < .3537, MSe = .0420, replicating the finding of Experiment 1. None of the higher order interactions reached significance (all Fs < 1).

Discussion. These results provide two important pieces of information. First, the accuracy level across the Training Conditions remained statistically equivalent after the retention interval; second, the performance level after 30 days remained higher than that for Transfer I of Experiment 1. These data provide additional support that part-task training is no different than whole-task training for this class of tasks. If differences were found it could be argued that the methods of training led to differential levels of performance at retention. However, because the structure of performance at retention was identical to that at the end of training, it is unlikely that this argument is tenable. It has been shown that for hybrid memory/visual search tasks the greatest decline in performance occurs during the first 30 days following training. After this initial decline, performance tends to stabilize (see Appendix A). This allows an empirically based prediction of skill decay for retention intervals up to a year (for hybrid memory/visual search tasks).

Experiment 3 - Assessment of Pure Target Learning

Experiment 1 demonstrated effective learning under part-task training conditions. Unfortunately, with only those data we cannot separate the effects of target learning from distractor learning. In the next experiment, we evaluate the effectiveness of our part-task training regimen when full-task performance is dependent on only target-set learning. Hence, we attempt to replicate our findings from

Experiment 1 and isolate the training effects on target learning.

Experiment 3- Method

Subjects. Eighteen undergraduate students, ten males and eight females, were paid for their participation in the experiment, received credit for a psychology class, or were given a combination of the two. All subjects were tested for near vision (at least 20/40) and far vision (at least 20/30), were asked about their use of medication, and were administered three subscales (vocabulary, digit-symbol substitution, and digit span) of the WAIS-R. The averaged WAIS-R scaled scores were slightly higher than those for the average population being: (a) vocabulary -- 15.50 (range 9 to 19), (b) digit span -- 12.67 (range 9 to 17), (c) digit-symbol substitution -- 13.61 (range 9 to 18).

Design and Procedure. This experiment was identical to Experiment 1 except (a) the distractor categories were switched at transfer; (b) different target categories were used; and (c) no refresher session was used for the retention phase.

Two sets of distractor categories were compiled. The assignment of distractor category sets to subjects was counterbalanced so that half of the subjects in each condition trained with one set and transferred to the other set. A set consisted of 8 categories with 6 exemplars in each category for a total of 48 distractor exemplars. (Experiment 1 used only one distractor set throughout the training and transfer sessions.)

Eighteen categories were used for the present experiment. The same guidelines were followed for category and word selection as used in Experiment 1: (a) semantically unrelated categories (Collen et al., 1975), (b) exemplar length between four and seven letters, and (c) target exemplars of high to moderately high production frequency (high item dominance) ranking (Battig & Montague, 1969). The target categories were FRUITS, OCCUPATIONS, BODY

PARTS, COUNTRIES, CLOTHING, and MUSICAL INSTRUMENTS. The categories used as distractors during training were (a) Set 1 -- TOOLS, BUILDING PARTS, VEHICLES, WEAPONS, METALS, and COLORS; and (b) Set 2 -- READING MATERIALS, DWELLINGS, SPORTS, RELATIVES, UNITS OF TIME, and EARTH FORMS.

Experiment 3 - Results and Discussion

Training. Mean frame times and accuracies for each training session were aggregated across subjects. Accuracy followed the same pattern as the results of Experiment 1 and stabilized at approximately 80 percent after four sessions of training. As before, mean Frame Times decreased for all three conditions according to a normal power function (see Figure 5). A fit of the power function to each of the Training Conditions yielded $r^2 = .97$ for PT2, $r^2 = .98$ for PT3, and $r^2 = .96$ for WT6. Subjects' average Frame Time (aggregated across conditions) decreased from 872 ms after the first session to 219 ms in the last session of training.

Transfer. Equivalent ANOVAs were performed on Experiment 3, as were previously performed on Experiment 1, and appear in Table 5. Mean accuracy was determined for each Training Condition across Frame Speeds and Transfer Sessions. A 3 x 2 x 3 (Training Condition x Transfer Session x Frame Speed) repeated measures ANOVA was performed on the accuracy data. The main effect of Session was significant, $F(1,15) = 30.95$, $p < .0001$, $MS_e = .0022$, reflecting the improvement in accuracy after 6 additional days of consistently mapped practice. The main effect of Frame Speed also reached significance, $F(2,30) = 58.37$, $p < .0001$, $MS_e = .0016$. A Newman-Keuls test (alpha = .05) showed significant differences among all three Frame Speeds. No difference was found among Training Conditions, $F(2,15) = 1.24$, $MS_e = .0399$, replicating the finding of Experiment 1. None of the higher order interactions reached significance (all F's < 1). A power test (Á=0.05, n=6, f=0.76, u=2; Cohen, 1977) on the Training Condition data revealed power equal to 0.73.

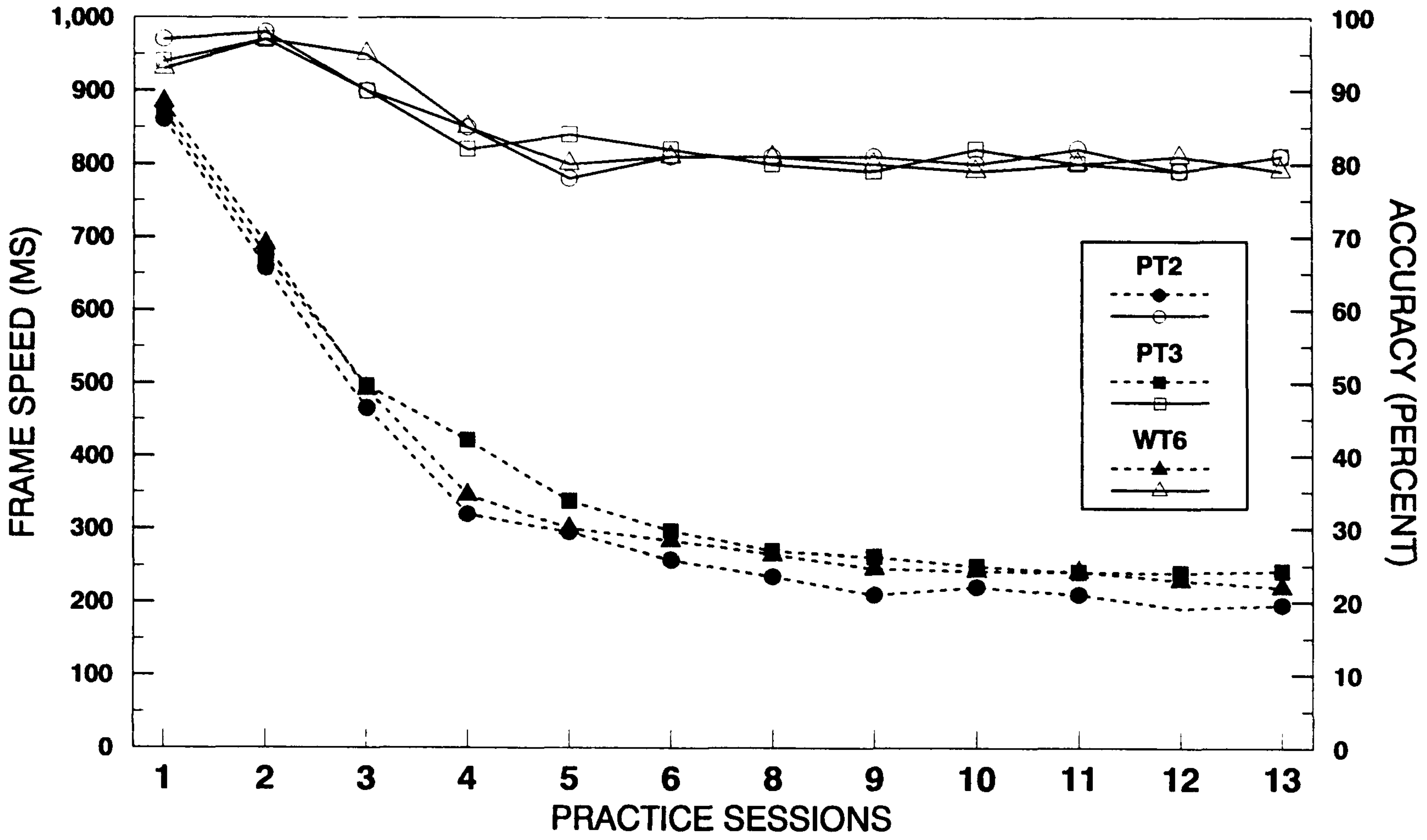


Figure 5. Frame Speed and Accuracy for Each Training Condition as a Function of Practice Session for Experiment 3.

Table 5. Summary of ANOVA for Experiment 3: Transfer Data

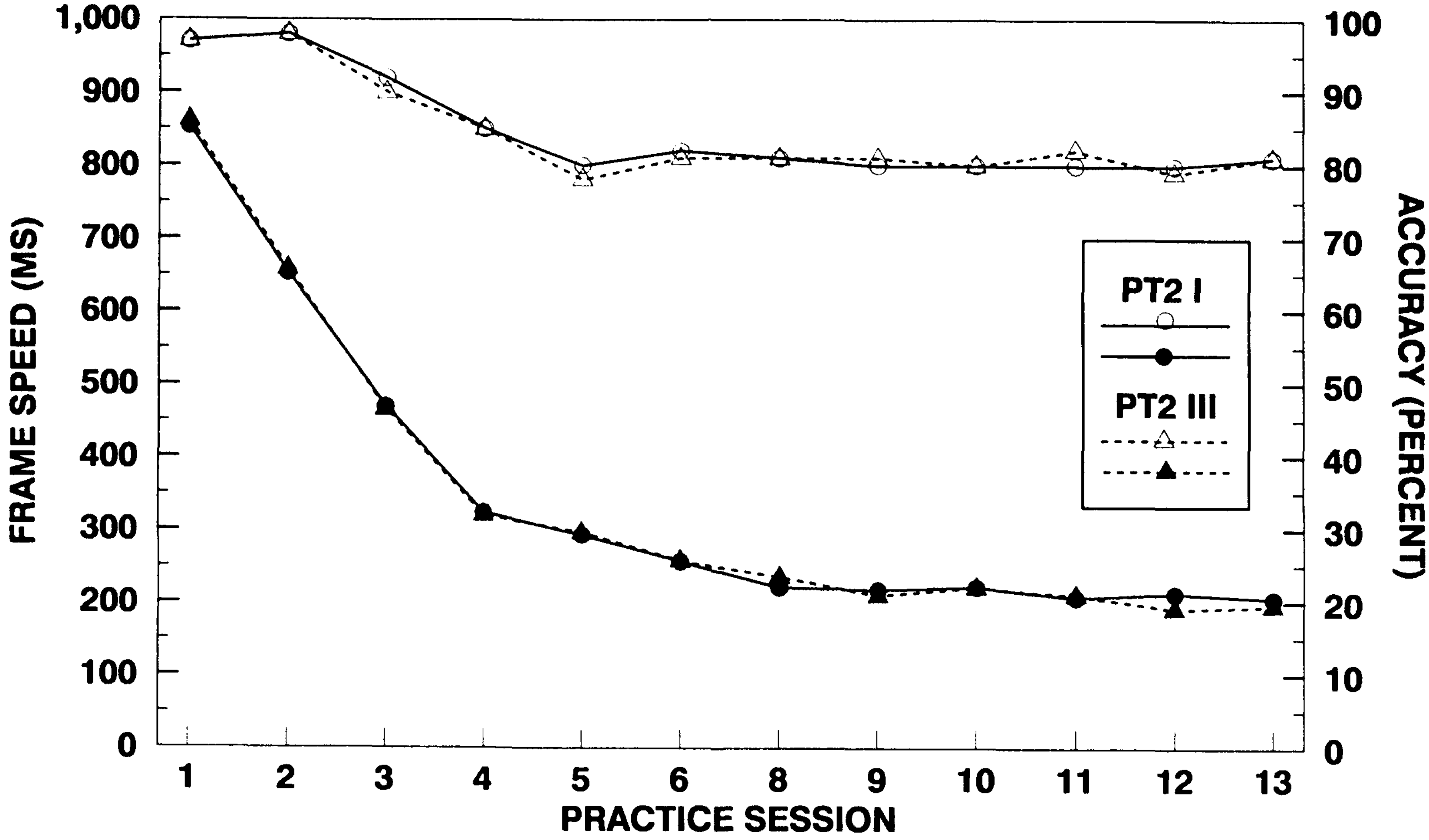
SOURCE	df_{num}	df_{den}	MS	F
Group	2	15	.0399	1.24
Speed	2	30	.0016	58.37***
Session	1	15	.0022	30.95***
Group x Speed	4	15	.0016	< 1
Group x Session	2	15	.0022	< 1
Speed x Session	2	30	.0022	< 1
Group x Speed x Session	4	30	.0029	< 1

* $p < .05$
** $p < .01$
*** $p < .0001$

An analysis of Temporal Order tested for the possibility of an effect due to category training sequence. That is, Temporal Order x Frame Speed ANOVAs were conducted to determine if the order in which the categories were learned in the PT2 and PT3 conditions had an effect on transfer performance. For PT2 the main effect of Temporal Order was not significant, $\underline{F}(2,10) < 1$, nor was the interaction of Temporal Order by Frame Speed, $\underline{F}(4,20) < 1$. Similarly, for PT3 the main effect of Temporal Order was not significant, $\underline{F}(1,5) < 1$, nor was the interaction of Temporal Order by Frame Speed, $\underline{F}(2,10) < 1$. These results indicate that whether a category was learned early or late in training did not make a significant difference during transfer.

Discussion. Experiment 3 attempted to replicate the results of Experiment 1, and in addition examined the issue of distractor learning. The results of Experiment 3 mirrored the results obtained in Experiment 1. The notable exception was the difference in mean transfer session accuracy between the two experiments. Training performance (frame speed and accuracy during 12 sessions of adaptive training) was almost identical between the experiments as shown for each of the Training Conditions in Figures 6, 7, and 8. Comparing the transfer performance for the two experiments, the mean transfer accuracy (aggregated across Frame Speeds) was 74.8 percent versus 64.2 percent during Transfer I, and 79.0 percent versus 69.2 percent during Transfer II, for Experiment 1 and Experiment 3, respectively. This 10.6-percent difference for Transfer I and 9.8-percent difference for Transfer II were likely the result of switching distractor categories at transfer.

This large difference in transfer accuracy supports previous findings of distractor learning (Dumais, 1979; Rogers, 1989). During CM practice, subjects strengthen consistent target categories, as well as weaken consistent distractor categories (i.e., subjects learn both target and



<u>Figure 6.</u> Frame Speed and Accuracy for the Two-Category Training Condition as a Function of Practice Session Comparing Experiment 1 and Experiment 3.

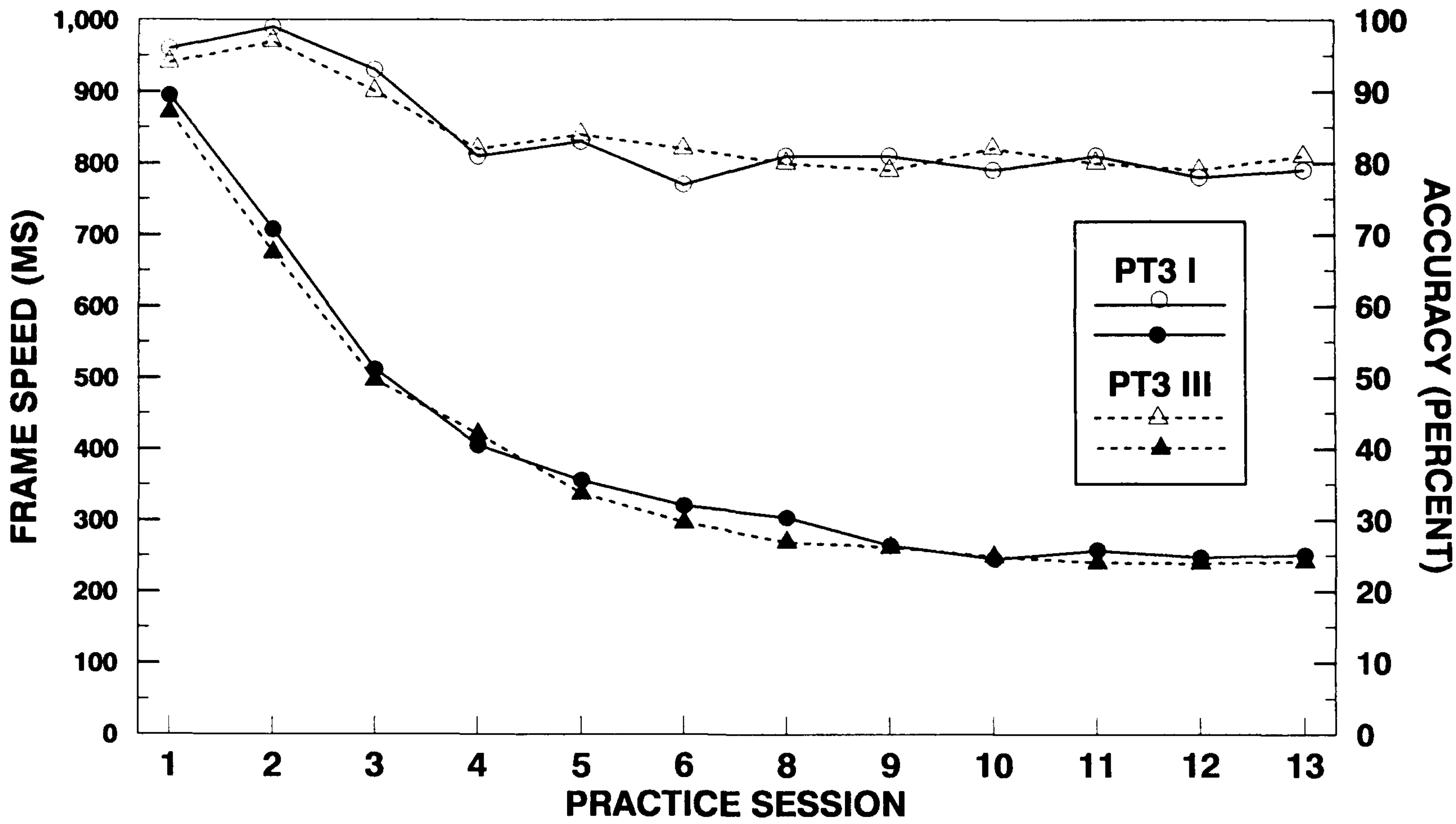


Figure 7. Frame Speed and Accuracy for the Three-Category Training Condition as a Function of Practice Session Comparing Experiment 1 and Experiment 3.

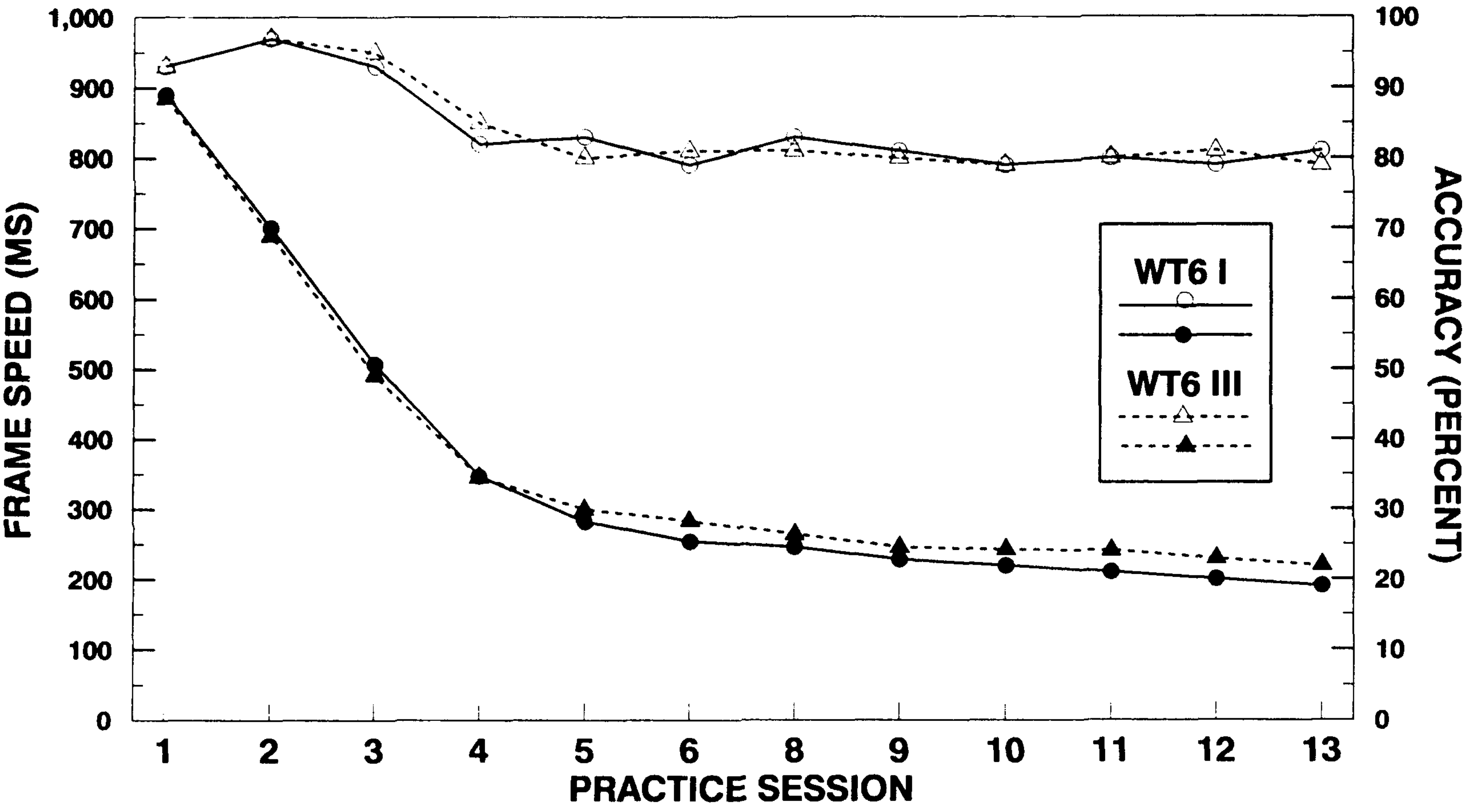


Figure 8. Frame Speed and Accuracy for the Six-Category Training Condition as a Function of Practice Session Comparing Experiment 1 and Experiment 3.

distractor sets). The distractor categories were not changed at transfer in Experiment 1, but were changed at transfer in Experiment 3. The results of Experiment 3 are important because they demonstrate that part-task training for this class of tasks will not lead to decrements, relative to whole-task practice, for tasks that allow for both target and distractor learning (Experiment 1) or merely target learning (Experiment 3). The present results indicate that the findings of Experiment 1 were not due simply to distractor transfer effects (Dumais, 1979; Rogers, 1989).

As in Experiment 1, no differences were found among Training Conditions. Replication of Experiment 1 with one of similar experimental design such as Experiment 3 lends greater credence to the previous finding of no difference among Training Conditions.

Experiment 4 - Retention of Target Learning

We found very good retention for all training groups in Experiment 2. However, that retention could have been due to target learning, distractor learning, or the combined effect of both target and distractor learning. In the next experiment, we examine retention of the trained detection skill developed in Experiment 3 in a way that isolates the target learning characteristics of the learning.

Experiment 4 - Method

The fourth experiment was a continuation of Experiment 3. The same subjects participated in one retention session 30 days following the end of Experiment 3. The session was identical to the previous transfer sessions.

Experiment 4 - Results and Discussion

Thirty-Day Retention. The mean accuracy data are reported in Table 6. A 3 x 2 x 3 (Training Condition x Transfer Session x Frame Speed) repeated measures ANOVA was performed on the accuracy data. The effect of Frame Speed was significant, as previously found in Experiment 3, $F(2,30) = 41.20$, $p < .0001$, $MS_e = .0021$, but the interaction

Table 6. Mean Accuracy for Transfer Sessions from Experiments and 4

	PT2			PT3			WT6		
Frame Speed	180	220	260	180	220	260	180	220	260
Transfer Session II	62	68	73	59	66	74	70	73	78
Transfer Session (Retention)	66	68	73	65	72	74	68	69	75

of Frame Speed x Training Session did not reach significance ($\underline{F} < 1$). Again, no differences were found among Training Conditions ($\underline{F} < 1$). However, the important finding was the significant interaction between Training Condition and Session, $\underline{F}(2,15) = 3.77$, $\underline{p} < .05$ $MS_e = .0030$. The source of that interaction can be seen by examining Table 6. The performance of the two part-task training groups did not decline (in fact performance in those conditions slightly improved) whereas the whole-task training group's performance did decline. These data are important as they suggest that the part-task training allowed for development and/or better retention of target activation.

Experimental Series 2: General Discussion

The present series of experiments was designed to address how part-task training influences skilled detection performance in a task requiring both associative learning and distractor and/or target learning. The important results were as follows: (a) Part-task training and whole-task training did not lead to differences in transfer performance; (b) performance for the three training conditions did not differ even after a 30-day retention interval; and (c) distractor learning had a large effect on transfer performance for both part- and whole-task learning; but, the effects seen for situations where performance can benefit from both target and distractor learning were replicated when target learning was isolated.

Although a direct statistical comparison of the data of Experiment 1 and Experiment 3 was not made, the methodological differences between the two experiments were minimal. Therefore, a superficial comparison of the two experiments was made. Experiment 1 used 36 exemplars from six target categories and 48 exemplars from eight distractor categories. Experiment 3 used 36 exemplars from six target categories and 48 exemplars from six distractor categories. The target categories were not the same across experiments (although there was overlap), but were chosen using the same criterion (i.e., less than 20 percent related) from the list compiled by Collen et al. (1975).

The principal difference between the experiments was the use of two sets of six distractor categories for Experiment 3 (one for training and the other for transfer), compared to one set of six distractor categories for Experiment 1. Because this was the only major difference in methodology, the data from the two experiments were compared for major trends. The 10.2-percent decline in overall accuracy between transfer in Experiment 1 and transfer in Experiment 3 is probably due to the switching of distractor categories in the latter experiment. As already mentioned, this follows the findings of previous researchers (Dumais, 1979; Rogers, 1989). A comparison of Experiment 1 and Experiment 3 shows a difference in overall accuracy due to distractor learning; however, the results of the two experiments led to the same pattern of results. Therefore, it would seem that distractor learning did not interact with Training Conditions for this set of experiments.

The implications of these findings for skill acquisition in the present class of tasks are that part-task training will yield similar results compared to whole-task training. This may be important under circumstances when it is more cost-effective to break a task into sub-components (Adams, 1987). Of course this is true only if the sub-components are developed following the guidelines discussed below.

The retention data from Experiment 4 allow us to argue that the part-task training was in fact superior to whole-task training in terms of target strengthening. Distractors used in training were changed for whole-task testing and retention. During training, and in Experiments 1 and 2, the distractor always remained the same; hence, subjects could benefit from distractors learning as well as target learning. When subjects could not benefit from both distractor and target strengthening, and had to rely solely on target strengthening then retention performance showed some decay only in the whole-task training group. These results argue that the load produced by the whole-task condition facilitated (necessitated) memory-set unitization but did not facilitate, to the same level as the part-task training group, target strengthening.

We now turn to a review of part-task training issues that are directly relevant to implementation of the present findings into part-task training situations.

Task Decomposition

According to Schneider (1985a), it is not necessarily the best strategy to structure the task components to be trained in a form equivalent to final execution. (Although not explicit on this point, Schneider was referring to the early stages of training.) For this to be the most effective method, the following assumptions would have to hold: (a) The real-world makes consistent elements of the task easily recognizable and presents those elements in an optimally sequenced manner; (b) training should be done at attentional capacity limits; (c) the influence of errors is unimportant and, related to that, frustration does not reduce effort or learning; (d) there is little transfer from component training to whole tasks. More often than not, these assumptions will not hold when training the task in a whole, final-criterion-level manner. Therefore, Schneider proposes that we should decompose the task, train on each component, and then reintegrate those components into the whole task.

Frederiksen and White (1989) have recently outlined an approach to training based upon principled task decomposition. Their approach focuses on the "decomposition of the task domain from the perspective of the inherent structure of the task, its human information processing demands, and the characteristics of expert performance. The decomposition identifies the top-level goals of experts and the strategies, skills, and knowledge developed by them in pursuit of those goals" (p. 1). Frederiksen and White reported the critical value and need for carefully designed and sequenced activities, set in the context of interactive simulations, for the facilitation of learning within a variety of domains: (a) Reasoning about electrical circuits (White & Frederiksen, 1985, 1986a, 1986b); (b) understanding Newtonian dynamics (White, 1981, 1984); and (c) teaching reading and arithmetic (Feurzeig & White, 1983; Frederiksen, Warren & Rosebery, 1985a, 1985b).

Frederiksen and White (1989) trained subjects to perform the Space Fortress Game, which involves concurrent and coordinated use of perceptual and motor skills, conceptual and procedural knowledge, and high-level strategies. This is the same task trained by Mane (1984), which was reported in the introduction to this section. In this game, the subjects must deploy missiles from their spaceship to destroy a space fortress while simultaneously avoiding the missiles directed at their ship. This particular study focused more on the cognitive aspects of expertise as opposed to motor skills.

An important facet of this approach is that they also tried to identify skills which may not be actually present in expert performance but are necessary precursors to the acquisition of that skill. Many of the subgoals developed were not proper part-tasks in that they involved activities and types of feedback not actually present in the criterion game. However, they allowed for the development of particular knowledge, heuristics, or strategies important for skilled performance. The specific sub-games which they trained were motor skill games, ship control knowledge games, strategic games, integration games, and information games. The order of training for developing the sub-skills was motor skills, ship control heuristics, and strategy development.

The principles they used for designing these special training tasks included the following: (a) Constrain the task so that it requires the component skill. Achieving this constraint is non-trivial in that typically game tasks can be performed via several alternative means. (b) Eliminate irrelevant complexities from the task. In other words, constrain the task so that it requires only the component skill and no other superfluous skills. (c) Clearly represent the phenomena to be learned. For example, in the ship control sub-games (designed to teach subjects how to control the ship's trajectory while maintaining low speeds) a speedometer was provided which was an additional indicator of the spaceship's speed of motion. (d) Provide immediate and high-quality feedback. For instance, scores can show the subjects that they either succeeded or failed. Also seeing the speed of the ship can provide the subject with more information about how he or she failed. (e) Provide a clear explanation of the strategy, game principle, or skill to be developed in the training task.

In their first experiment, Frederiksen and White (1989) focused on training the optimal strategy for one goal of the game: how to hit the fortress without being hit by the fortress. Two groups of subjects were tested. The control group received training only on the criterion task (the Space Fortress Game). The experimental group was given training on the set of sub-games listed above (i.e., knowledge, strategy, skill and motor games). On the last day of the experiment, the experimental subjects performed the criterion game. The results were as follows: (a) Experimental subjects successfully developed the skills, knowledge, and strategies they were taught; (b) training resulted in substantial differences between training groups in their methods of controlling the ship; and (c) skill and strategy differences resulted in higher performance on the criterion game and in a higher rate of improvement with continued practice.

A portion of their second experiment explored ability differences. (They also investigated other factors of performance which are not germane to the present discussion.) The results showed that (a) subjects differing in ability (low, middle, and high thirds on the screening test) also differed significantly in mean game scores on the final day; (b) experimental training entirely eliminated differences in performance between the middle and high ability groups; and (c) subjects in the low ability group showed the greatest improvement.

Frederiksen and White (1989) also gave subjects a set of transfer tasks (in both Experiments 1 and 2) to determine if the experimental subjects had attained skills and knowledge which were more generalizable and transferable than that of the control subjects. The results indicate that the experimental group developed a more generic knowledge and skill base. They concluded that "When the componential training tasks have been designed to reflect the strategic character of expert performance and the high integration of skill components in such performance, the effects of such transfer are superior to those of training based upon practicing the whole task for a comparable amount of time" (p.34).

In an earlier paper, Frederiksen et al. (1985b) focused on a componential approach to training reading skills. The specific constraints of their design may be extrapolated to the training designs of other tasks. They proposed that: "The critical test for a component-centered approach to developing complex skills lies in demonstrating that individual components are trainable in such a way as to affect global, integrated performance of the skill...meeting this criterion requires the development of a comprehensive sequence of training environments that increasingly come to involve the full complement of skills characteristic of domain expertise" (p.331).

Frederiksen and his colleagues emphasized that it is crucial to specify a model of component interactions and identify those components having critical functional linkages to other components. A skill hierarchy may then be developed in which the components are ordered according to their importance for improving other skills, which are then placed higher in the hierarchy. For example, in their training program directed toward reading, training in perceptual encoding preceded training in decoding, which in turn preceded training in context utilization. "...by definition, training tasks higher in the skill hierarchy build upon a larger and larger repertoire of automatic processing components (p. 332). "An extension of the skill hierarchy to encompass components of comprehension on an analysis of their functional linkages represents important steps in demonstrating the feasibility of a component centered approach to training a complex cognitive skill such as reading" (p. 336).

Suggestions for When to Use Part-Task Training

One recurring theme becomes evident from a review of the literature; namely, procedural items or psychomotor tasks will benefit greatly from part-task training. For example, Battiste (1987) investigated the effects of part-task training on the psychomotor portion of a supervisory control simulation known as "popcorn." He gave the part-task training group prior practice on the psychomotor portion, which consisted of control and movement of the cursor with the magnetic pen and pad. His results yielded three important effects: (a) The part-task group learned the task faster; (b) the part-task group's scores and task times continued to improve while the whole-task group's did not; and (c) the speed of response increased significantly for the part-task training group whereas almost no improvement for the whole-task-trained group. Battiste concluded that "Part-task training was particularly effective because the subjects were taught a learnable, consistent task component which was an integral, busy part of the overall task" (p. 1368).

In a similar vein, Vidulich et al. (1983) used a massed practice procedure to train the visual/spatial skills which are part of the controller task for in-flight refueling. Using a compressed time procedure, they were able to provide the subjects with a higher number of trials on this portion of the task than subjects normally receive when the task is trained in real time. The results showed that subjects who received the compressed-time training (and therefore more trials) were subsequently more accurate. These results provide empirical evidence that increased practice on procedural and/or psychomotor tasks will result in improved performance. That is, when this task was trained in real time, subjects received fewer practice trials and their performance suffered.

Flexman et al. (1972) also demonstrated benefits of massed practice for procedural items. They provided isolated practice on four procedural exercise: (a) cockpit familiarization; (b) cockpit check; (c) starting procedure; and (d) run-up check. The benefit of massed practice on these tasks is exemplified by the benefits accrued for the starting procedure: normally, trainees are allowed only one trial per scheduled flight, but massed practice on the simulator provided benefit without any cost to the equipment of the aircraft.

Folds, Gerth, and Engelman (1987), in training complex tracking tasks, also found initial advantages for subjects who were part-task-trained on the target acquisition task. This prior training allowed subjects to become well acquainted with the typical dynamics of the task.

Flexman et al. (1972) reported that the magnitude of savings (i.e., the percentage of errors as well as the amount of time and number of trials necessary to reach criterion performance was less for part-task-trained subjects relative to whole-task-trained subjects) was related to the difficulty of the maneuver. For example, rated climbing, descending turns, steep turns, and stalls were the most difficult maneuvers in the experiment and these showed the highest percent of transfer from part-task training. Similarly, Briggs and Naylor (1962) trained a three-dimensional compensatory tracking task by separating the task into three one-dimensional tasks. They manipulated the difficulty levels of the tracking tasks, and the results of this study showed that the higher difficulty yielded greater differential transfer.

Adams (1960) offered the following tentative principles, which still hold today, for the design and use of part trainers:

1) Part trainers should be used whenever part-task training, plus the added integrative whole-task practice required to learn the interactions among the parts, costs less than whole-task practice to achieve a criterion of proficiency.
2) Part trainers can be used unequivocally for response sequences which do not have to be performed in a concurrent, time-shared relationship with other responses in the whole task.
3) Part trainers may be effective for the maintenance of proficiency in procedural response sequences which are performed concurrently with continuous responses.
4) Part trainers, being so much simpler than the whole task, are less difficult and yield measures of response proficiency which are spuriously high. They should not be used for proficiency measurement purposes.

Evaluating Part-task Training: A Caution.

Wightman and Lintern (1985) proposed a type of validation technique to test the success of part-task training. They claimed that if a backward transfer method is used in which the whole task is trained and then followed by a test of the isolated critical components, the feasibility of using part-task training will be evident.

Salthouse and Prill (1983) reported results of this type of measurement. They trained subjects to perform a task which required the judgment of the temporal intersection of two trajectories. After training, they measured performance separately for two of the components: temporal and spatial information. None of the measures of component effectiveness exhibited significant practice effects, despite large differences in overall level of performance. Salthouse and Prill therefore concluded that the components of this particular task were both necessary and sufficient for successful performance. Though this conclusion may be true, it is not a relevant criticism for the use of part-task training in other situations. It is likely that practice under dual-task conditions of sufficient difficulty will preclude learning to perform one of the tasks alone (Nissen & Bullemer, 1984). Thus, one would not expect better performance on task components if subjects had been trained to perform them in conjunction with the rest of the components. It is very likely that the components are interdependent and these results demonstrate that for certain tasks, part-task training is not possible, or at the very least, must be paired with whole-task training.

Suggestions for When to Use Whole-Task training

Klapp, Martin, McMillan, and Brook (1987) have stated that the relative effectiveness of part- versus whole-task training depends on the type of task. They trained subjects to press two telegraph keys, one with each hand, each with a different fixed period of repetition. They found that training this task was much more effective if whole-task training was used rather than part-task training. They concluded that "...it appears that whole-task training may be best for tasks that require temporal coordination of the component responses" (p.129). They further proposed that whole-task training will be more effective than part-task training, but only if an integrated and unified conception of the task is encouraged. For example, they suggested that for flying a standard helicopter (which requires coordinated movement of both hands and both feet), training the individual hand and foot movements may not be as effective as whole-task training which encourages the subject to view the task as a unified whole.

Folds, Gerth, and Engelman (1987) trained subjects to perform a complex tracking task. This particular task encouraged anticipation and was found to benefit from whole-task practice. Their results showed that the dual-task organization of the whole-task group was far better organized than in the part-task group. They concluded that "Tasks which do facilitate response organization, and which must be performed in dual-task conditions, may benefit from training in the dual-task conditions. The response organization which is promoted by single-task practice may be inappropriate for the combined demands of the dual task" (p. 350). This conclusion was echoed by Lintern and Wickens (1987): "Component training generally inhibits the development of task integration skills, and this is particularly true for the case of difficult tasks and high subtask integration" (p. 33) Naylor and Briggs (1963) similarly hypothesizedthat as complexity is increased for relatively highly organized tasks, training the whole task should work better than training parts of the task.

Combined Part/Whole-Task Training: The Most Usual Situation

Many of the tasks shown to require whole-task training will, in most cases, benefit from some amount of part-task training. Schneider and Detweiler (1987) proposed that both types of training may be necessary, although neither may be sufficient, for optimal performance. In fact, single-task training to a criterion level of performance may be crucial. However, after a certain level of skill is reached, continued single-task training may be inefficient. Schneider and Detweiler also advocated the consideration of the amount of single-task practice provided. This is related to the point made by Lintern and Wickens (1987) with regard to task integration skills, which they proposed may be inhibited by single-task training.

The importance of providing dual-task performance may be related to the idea of a time-sharing "ability" advocated by several researchers (e.g., Gopher & North, 1974; Jennings & Chiles, 1977). Jennings and Chiles (1977) proposed that there is a "reliable source of variance that contributes to performance of complex tasks, but is independent of simple task performance on the constituent tasks." The concept of time-sharing abilities has been recently explored further by Rieck, Ogden, and Anderson (1980). They proposed that because there is evidence for single-task proficiency (e.g., Freedle, Zavala, & Fleishman, 1968) and time-sharing skills (e.g., Gopher & North, 1974), it should be possible to investigate the relative effectiveness of each type of practice. Rieck and her colleagues varied (between subjects) the amount of single- and dual-task practice and measured subsequent performance on a dual task. The single task consisted of a single-dimensional discrete compensatory tracking task and the additional task was a digit classification task. They also measured transfer to a dual-task which consisted of the discrete tracking task paired with a delayed digit recall task. Their results indicated that those subjects who had received more dual-task training had better overall performance. They concluded that dual-task practice was more efficient in the development of time-sharing skills. Furthermore, in the transfer phase, subjects who had received prior dual-task training performed better. Rieck et al. (1980) suggested that general time-sharing skills improve with practice.

Beginning a training program with single-task (or part-task) training and then proceeding to dual-task (or whole-task training) may be the most efficient training method. It is possible to take what is known about effective part-task training methods and used it in the first phase of a training program. For example, as reviewed above, procedural or psychomotor tasks often benefit highly from part-task training. Similarly, simply allowing subjects to become familiar with the specific dynamics of a task (e.g., Folds et al., 1987) results in improved performance. After subjects have been allowed to become proficient on the specifics of single tasks it would then be possible to provide training under whole- or dual-task conditions. Subjects would then be able to learn the necessary strategies for pairing the components of a task or for performing two tasks simultaneously. However, if the integration of the task is reliant on a highly organized structure between the tasks, then less part-task training should be provided. If the amount of necessary organization is low, more part-task training could be provided with a smaller subsequent amount of whole-task training.

Future Research

Two important questions remain: (a) What implications do these experimental results have for future research in the area of hybrid memory/visual search tasks? and (b) What additional experimental designs would address these issues?

Although difficulties were predicted with the high comparison load for the whole-task subjects, they apparently encountered little difficulty with a comparison load of 18. (Comparison load in this case refers to the number of categories in the memory set multiplied by the number of exemplars in a given frame.) The results of these experiments imply that subjects may be able to simultaneously learn a much larger number of categories in a multiple-frame paradigm than previously thought. In addition, the results imply that part-task training may be beneficial in refresher courses for tasks involving visual search (air traffic controlling, computer operators). Refresher courses could include a greater amount of practice on individual groups of subtasks, without showing a deficit when the tasks are reintegrated. Concentration on the more important subtasks would allow more cost-effective refresher training to be developed (Wightman & Lintern, 1985).

A number of alternative designs are possible to test the hypotheses set forth by the above experiments. To test the limits of comparison load, a replication of the above experiments could be performed substituting four, six, and twelve categories for the three training conditions. This would provide an upper comparison load of 36 rather than 18.

A second alternative would be to change the training from a specific number of sessions and blocks to a design where subjects train until they reach a preset criterion. A comparison could then be made on the number of blocks required to reach criterion. Transfer sessions would occur after the subject had reached criterion on each of the subtasks (or in the case of the whole-task group, when they reached the one preset criterion). A large number of subjects would be necessary for this design because the variance would probably be higher than that in the experiments presented above.

A third alternative emphasizes the adaptive nature of the training used in Experiment 1 and 3. Rather than training which begins at a relatively slow frame speed (940 ms), a much faster frame speed (100 to 200 ms) might be used. The advantage of this design is that subjects are pushed to their mental limits from the very beginning. (A similar concept was suggested in Wightman & Lintern, 1985.) Obviously, there is a disadvantage if the subject is not able to learn the categories due to the difficult frame speed.

Finally, a design which trains each of the part-task categories between blocks (two categories on block one, another two categories on block two, etc.), rather than between sessions, may yield different results. In addition, a "transfer" session could be included at the end of each session to test reintegration of the categories.

IV. EXPERIMENTAL SERIES 3: PERFORMANCE IMPROVEMENT AS A FUNCTION OF DEGREE OF BETWEEN SEMANTIC-CATEGORY CONSISTENCY

Introduction

Practice alone does not improve performance, but *consistent* practice does improve performance (Schneider & Fisk, 1982). The validity of this statement has been well documented in the training literature (e.g., Fisk, Oransky, & Skedsvold, 1988; Schneider, 1985a; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). However, an important issue not thoroughly investigated involves the *degree* of consistency necessary for improved performance (but see Schneider & Fisk, 1982). This issue was examined in the present study using a high-speed perceptual learning task in which consistency involves whether target items are responded to (attended or ignored) in the same manner across situations. We assessed performance improvements at varied levels of consistency: 100 percent consistent, 67 percent consistent, 50 percent consistent, 33 percent consistent, and 13 percent consistent.

An important point is that levels of consistency were manipulated while the subjects were performing the task *at their perceptual limits*. We used a high-speed multiple-frame word search task in which the stimuli were presented above threshold but very briefly. Furthermore, the duration of the stimuli was decreased according to each subject's accuracy level (as accuracy increased, stimulus duration decreased, thus increasing the difficulty of the task). This paradigm allowed us to assess the ability of subjects to take advantage of consistency levels in a high-speed, perceptually demanding task.

Background

Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) conducted a series of experiments which clearly demonstrate the importance of consistent practice for performance improvement and automatic process development. They demonstrated differences in performance which varied according to whether training was consistently or variably "mapped." More precisely, in consistent mapping (CM) training the individual always deals with (i.e., attends to, responds to, or uses information from) a stimulus, or class of stimuli, in the same manner. CM training conditions result in dramatic performance improvements (see Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977 for details) and the eventual development of performance characteristics indicative of automatic processing. Varied mapping (VM) training situations are those in which the practice is inconsistent; that is, the response or degree of attention to the stimulus changes from one stimulus exposure to another. VM training conditions result in little performance improvement.

Performance principles based on an understanding of consistent practice have been applied to designing training programs for a variety of domains including map reading skills (Fisk & Eboch, 1989), instructional systems design (Fisk & Gallini, 1989), in-flight refueling (Eggemeier, Fisk, Robbins, Lawless, & Spaeth, 1988), and air traffic control (Kanfer & Ackerman, 1989).

The majority of applications-oriented research has been based on an assumption of perfect consistency; mely, the assumption that the stimuli are <u>always</u> attended to, responded to, or classified in exactly the same manner in all situations. Unfortunately, in real-world settings perfect consistency may be unattainable. For example, a stimulus may result in a given outcome only in a proportion of cases. However, it may be important for a trainee to quickly execute responses in those critical cases. For example, certain types of cloud formations may generally (but not always) be used to forcast severe weather and navigator must be prepared to respond to the probability of severe weather even if it occurs only 70 percent of the time. Thus, the cloud formations are not perfectly consistent as predictors of severe weather but only 70 percent consistent. It is important, from a training perspective, to understand the level of consistency which will yield improved performance (i.e., faster and/or more accurate responding) in this type of situation.

The issue of degree of consistency was investigated in the laboratory by Schneider and Fisk (1982) using a relatively simple letter search task. Subjects were required to search for a single letter in a series of displays, each of which contained four letters. The degree of consistency was manipulated to be 100 percent (traditional CM task), 67 percent, 50 percent, 33 percent, or 13 percent (traditional VM task). With extensive training (6,720 trials) there was a functional relationship between degree of consistency and percent correct. The 100 percent and 67 percent consistent conditions showed the greatest improvements in performance with practice while the 50 percent condition showed a moderate level of improvement. The 33 percent and 13 percent conditions showed the least improvement and did not differ statistically from each other. In the second experiment in this series, Schneider and Fisk demonstrated that there was also a functional relationship between degree of consistency of training and dual-task performance. The more consistent conditions yielded better dual-task performance (i.e., when performed concurrently with a VM task).

Schneider and Fisk's data suggest that degree of consistency is an important factor in training and that a task need not be 100 percent consistent for improvement in performance to occur. This finding has implications for real-world situations which may not be perfectly consistent; that is, practice will still be beneficial even at less-than-perfect levels of consistency.

The present experiment was designed to replicate and extend the Schneider and Fisk results. A multiple-frame word search task was used, thereby increasing the amount of semantic processing required of the stimuli (Schneider and

Fisk used letter search). Furthermore, the timing of the stimulus presentation was adapted to each individual's perceptual ability level. A fairly low criterion was used (75 percent) for increasing the presentation rate. As a result, the subjects were challenged to perform at their perceptual limits. This design has obvious implications for training situations which involve high-speed tasks and require processing at a level higher than the 'featural' level of briefly presented stimuli. The issue of interest was whether the subjects would be able to take advantage of the consistency levels present in the task even under <u>time-stress</u> situations requiring semantic processing.

The experiment consisted of two phases of training followed by a test phase. The first phase was the adaptive training phase, during which the presentation of stimuli was a function of each individual subject's accuracy level. The goal was to train subjects to perform near their perceptual limits (but above threshold). The stimulus speed was adjusted after every block of 95 trials according to the following criteria: If accuracy rate was above 75 percent for a block, the stimuli in the next block were presented 25 ms faster; if it was below 60 percent, the stimuli in the next block were presented 25 ms slower; if accuracy was between 60 percent and 75 percent, the presentation speed did not change in the next block. There were a total of 3,325 trials of training in this phase of the experiment. Performance improvement during this phase was measured by increasing stimulus speeds.

Following the adaptive training phase of the experiment, the subjects received 2,125 trials of training at a fixed rate of stimulus presentation. The adaptive training in the first phase served to adjust the speed of stimulus presentation according to the abilities of each subject and the purpose of the fixed rate training was to provide subjects with the opportunity to practice at that level. The stimulus presentation speed for this phase was different for each subject and was the fastest presentation speed attained by the subject during the last session of adaptive training. With the fixed rate of stimulus presentation, accuracy rate was the primary dependent variable.

There were five training conditions which varied along the dimension of consistency. Consistency is operationally defined as the number of trials in which a word appears as a target relative to the number of trials in which the same word appears as a distractor. This ratio was manipulated by holding constant the number of times a word appeared as a target in each condition and manipulating the appearance of words as distractors.

To determine if the appearance of items as distractors in the same block was affecting performance levels, a CM test was conducted at the end of the fixed rate training. That is, each of the degree conditions was tested in the situation where the items were presented only as targets and never as distractors.

The performance predictions for this experiment are straightforward. First, during the adaptive training phase, the stimulus speed should increase for all subjects. Due to the experimental design (all manipulations were within-block), the stimulus speed will necessarily increase for all conditions at the same rate. However, during the fixed rate training, the primary measure of performance is accuracy and if subjects are able to "tune-in" to the consistency of the conditions, there should be a functional relationship between consistency of training and accuracy rate. That is, performance should be better for the 100 percent consistent condition and decreasing for the other degree conditions. Based on previous findings (Schneider & Fisk, 1982), it was expected that the 33 percent consistent condition would not differ from the VM (13 percent consistent) condition. Finally, the CM test should yield a similar pattern of decreasing performance across the conditions of previously decreasing consistency.

Method

Subjects. Fifteen subjects, nine males and six females, participated in the experiment. Subjects received course credit and/or monetary compensation for their participation ($4.00 per hour, with a bonus of $1.00 per hour upon completion of the study). Vision was tested for all subjects, and their corrected or uncorrected visual acuity was at least 20/30 for distance and 20/40 for near (magazine print).

Stimuli. The targets and distractors consisted of the following nine words which were pre-tested to be equally confusable: ORGAN, PANSY, SATIN, SHACK, ROBIN, RIFLE, SPEAR, OCEAN, PEACH. The assignment of words to conditions was counterbalanced across subjects.

Equipment. EPSON Equity I+ microcomputers were programmed with Psychological Software Tools' Microcomputer Experimenter Language (MEL) to present the appropriate stimuli, collect responses, and control timing of the display presentations. Epson MBM 2095-5 green monochrome monitors were used to present the stimuli. The standard Epson Q-203A keyboard was altered such that the '7', '4', '1', and '5' numeric keypad keys were labeled 'T', 'M', 'B', and 'N', respectively. During all experimental sessions, pink noise was played at a sound pressure level of approximately 55 db to help eliminate possibly distracting background noise. All subjects were tested in the same room, at individual, sound-attenuated workstations, and were monitored by a laboratory assistant.

Procedure

Multiple Frame. An individual trial in the multiple frame procedure utilized in the present experiment consisted of the following sequence of events. The subject was presented with the memory set of one word, which he/she was allowed to study for a maximum of 20 seconds. Subjects were instructed to press the space bar to initiate the presentation of the frames. Three plus signs were then presented in a column for .5 second in the location of the display set (in the center of the screen) to allow the subject to localize his/her gaze. The plus signs were followed by eight frames; each frame consisted of a display set (three words presented in a column) and a visual mask (three rows of X's positioned in the same location as the words to prevent continued perceptual processing of the stimuli). The duration of the display set was a function of each individual's performance (referred to as stimulus speed, see below), but the duration of the visual mask was 200 ms for all subjects throughout the experiment. Therefore, in each trial, 24 words (eight frames x three words per frame) were presented and the subject was required to search for the memory-set word.

On positive trials (i.e., target present), the subject's task was to indicate the location of the target word (i.e., the word previously presented in the memory set). A response of top, middle, or bottom was made by pressing the corresponding key labeled 'T', 'M', or 'B'. On negative trials (i.e., target absent), the subject indicated the absence of a target by pressing the key labeled 'N'. A response could be made at any time during the presentation of the eight frames; that is, as soon as a target word was located, the subject could respond and thus terminate that trial. Otherwise, the subjects were allowed 4 seconds following the end of the presentation to make a response. Target words were never located in the first or eighth frame (although the subjects were not told this). Aside from this restriction, the frame in which the target appeared was random, as was the vertical position in the display.

The subjects received the following performance feedback. After each correct trial, the message "CORRECT!" was displayed. If the subject "missed" the target or input the wrong location of the target, then the message "ERROR, _ was present in the _____ position" was displayed simultaneously with a tone. If the subject input a location when there was not a target present (i.e., a false alarm), the message "ERROR, there was no target present" was displayed along with an error tone. Following each block of trials, a message was displayed showing for that block both the average percent accuracy and the frame speed. At this time, the subject was encouraged to take a break to stretch or look around the room.

Adaptive Procedure. An adaptive procedure was used in which the presentation rate of the stimuli (i.e., stimulus speed) was a function of each individual's accuracy rate. The goal was to train subjects to perform near their perceptual limits (but above threshold). The stimulus speed was adjusted after every block of 95 trials (the mask speed was held constant at 200 ms). If the subject's accuracy rate was above 75 percent for a block, the stimuli in the next block were presented 25 ms faster. If it was below 60 percent for a block, the stimuli in the next block were presented 25 ms slower. If, however, accuracy was between 60 percent and 75 percent, the presentation speed did not change in the next block.

Training Conditions. There were five training conditions which varied along the dimension of consistency. Consistency is operationally defined as the number of trials in which a word appears as a target relative to the number of trials in which the same word appears as a distractor. This ratio was manipulated by holding constant the number of times a word appeared as a target in each condition and manipulating the appearance of words as distractors. Degree Condition 1 - 100 percent consistent (CM); the ratio of target appearance to distractor appearance was 10:0 (i.e., the word never appeared as a distractor). Degree Condition 2 - 67 percent consistent; the ratio of target:distractor presentation was 10:5. Degree Condition 3 - 50 percent consistent; the ratio of target:distractor presentation was

10:10. Degree Condition 4 - 33 percent consistent; the ratio of target:distractor presentation was 10:20. (Note: Conditions 1 through 4 will be referred to collectively as the degree conditions). VM Condition - 13 percent consistent; the ratio of target:distractor presentation was 9:61.

Practice. Practice consisted of two blocks of trials. Each block consisted of 85 positive (target present) trials and 10 negative (target absent) trials. The subjects were offered the opportunity to take a short break before the thirtieth trial, before the sixtieth trial, and again at the end of each block. All subjects began the first practice block at a stimulus speed of 500 ms. These orientation trials allowed the subjects to become familiar with the experimental protocol. The words used for the practice trials were not used in the remainder of the experiment.

Sessions 1 - 7: Adaptive Training. Each adaptive training session consisted of five blocks of practice and lasted approximately 1 hour. Within each block, there were 85 positive trials and 10 negative trials, for a total of 475 trials per day. The subjects were offered the opportunity to take short breaks within blocks: before the thirtieth trial, before the sixtieth trial, and again at the end of each block. All subjects began with a stimulus speed of 450 ms. The choice of the beginning stimulus speed was determined by pilot data from six subjects. These subjects were tested for three sessions each (320 trials per session) at stimulus speeds of 300 ms, 400 ms, 450 ms, and 600 ms. The stimulus speed chosen for the present experiment was the speed at which the pilot subjects could perform during the first session which was above chance but below ceiling. Performance for 450 ms was 87 percent for the first session of the pilot testing.

During these sessions, the adaptive procedure explained above was used. Throughout all seven sessions, the rate of stimulus presentation was a function of the accuracy of each individual subject. There were a total of 3,325 trials of training in this phase of the experiment; 350 trials per each of the degree conditions and 1,575 trials for the VM condition. Dueto the within-block presentation of the conditions, all conditions were presented at the same frame speeds. Thus, differing improvements in accuracy across conditions were not confounded with differing stimulus speeds.

Sessions 8 - 12: Fixed Rate Training. Following the adaptive training phase of the experiment, the procedure was changed such that an adaptive procedure was no longer used. The adaptive training had served to adjust the speed of stimulus presentation according to the abilities of each subject (to a criterion of 75 percent accuracy). The purpose of the fixed rate training was to provide subjects with the opportunity to practice at that level. The stimulus presentation speed for this phase was different for each subject and was the fastest presentation speed attained by the subject during Session 7. This presentation speed became the constant rate of presentation for the next five sessions. The remaining details of the procedure were the same as described above. Each subject completed a total of 250 trials of each degree condition and 1,125 VM trials during this phase of the experiment (for a total of 2,125 trials). During these sessions each subject was working at the limits of his/her own perceptual ability as determined by the adaptive training sessions.

Session 13: Pure CM Test. This session consisted of a pure CM test of the conditions. That is, each of the degree conditions was tested in the situation where the items were presented only as targets and never as distractors. There were three blocks of the CM test, with 45 trials in each block (40 positive trials, 10 per degree condition, and 5 negative trials). These blocks were presented at the same rate of stimulus presentation used during the fixed rate training on Sessions 8 through 12.

Design

All manipulations were within subjects. The primary independent variable was Degree Condition, based on degree of consistency (100 percent consistent, 67 percent consistent, 50 percent consistent, 33 percent consistent, and VM - 13 percent consistent).

During the adaptive training phase (Sessions 1-7), stimulus speed was the primary dependent variable. However, accuracy was the primary dependent variable during the fixed rate training phase (Session 8-12), as well as for the pure CM test of performance (Session 13).

Results: Adaptive Training

Stimulus Speed.

Stimulus speed was the primary dependent variable during the adaptive training phase. These data are presented in Figure 9 (the bottom-most line indexed by the right axis). A one-way analysis of variance (ANOVA) was conducted to test the effect of Session (1 through 7). As is clear from the figure, there was a significant effect of training session, $F(6,84) = 1764.72$, ($p < .0001$). A Student-Newman-Keuls analysis revealed that Sessions 1, 2, 3, and 4 were all significantly different from each other (each one better than the last) indicating steady improvement. Increases in speed asymptoted at Session 5 and did not change significantly for the remaining sessions. This asymptote is partially due to the fact that the system could not reliably present stimuli faster than 100 ms. Consequently, we imposed 100 ms as the lower limit on the stimulus speed. Eleven of the subjects reached this limit and the remaining subjects asymptoted at 125 ms.

Accuracy.

Also plotted in Figure 9 are the accuracy rates for each of the conditions as a function of session during the adaptive training phase. A Degree Condition (100 percent, 67 percent, 50 percent, 33 percent, and 13 percent consistent) x Session (1 through 7) ANOVA was conducted. The effect of Session, $F(6,84) = 75.46$, $p < .0001$, was significant because accuracy decreased during the first

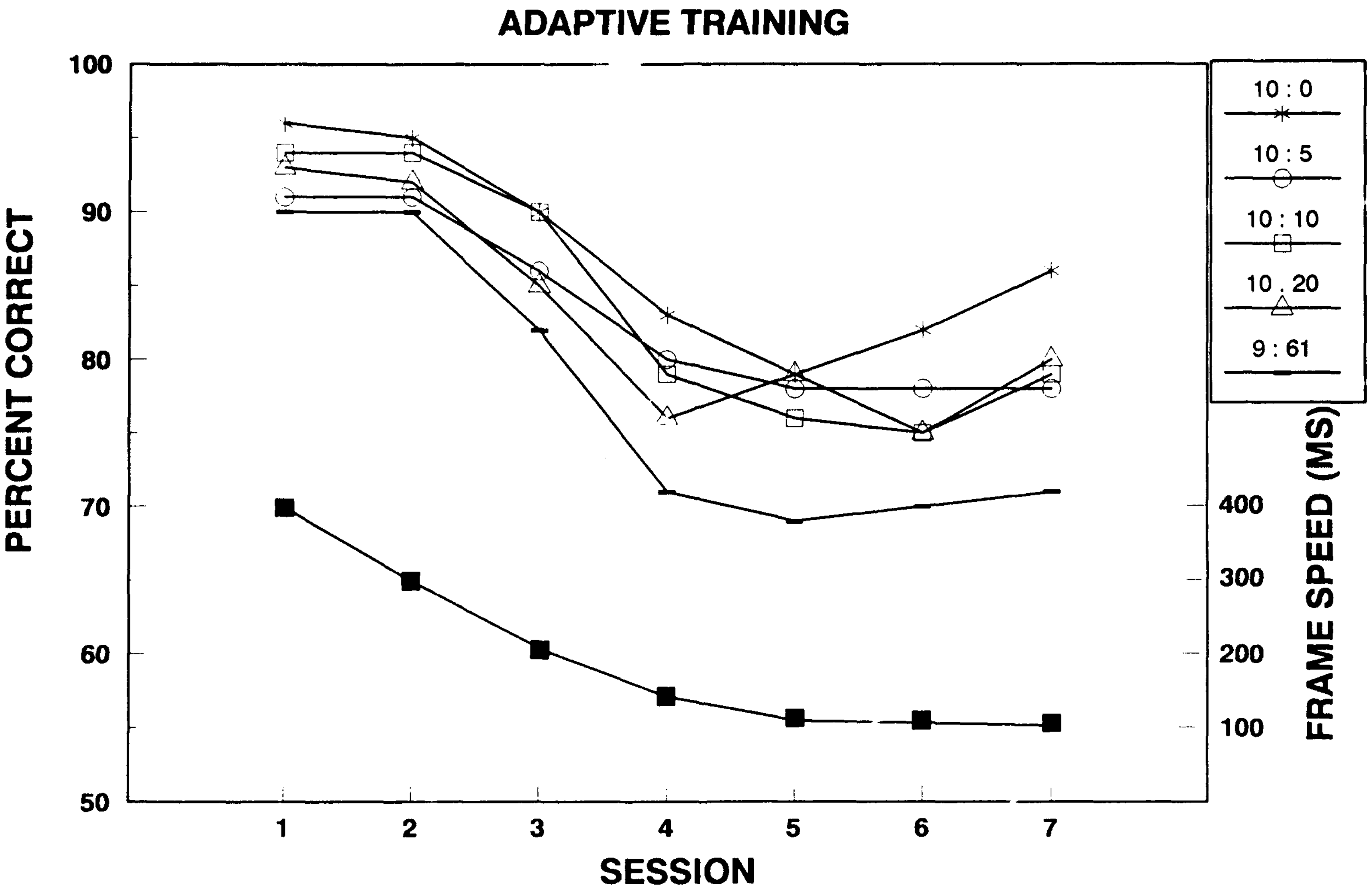


Figure 9. Mean Accuracy Rates and Frame Speeds for Each Condition Plotted as a Function of Each Session of Adaptive Training.

three sessions as stimulus speed was increasing. The overall accuracy rates were stable across sessions 4 through 7. However, the interaction of Degree Condition x Session was marginally significant, $\underline{F}(24,336) = 1.51.$, $\underline{p} < .06$. The source of this interaction is the fact that during sessions 4 through 7 (when stimulus speed had stabilized as reported above) the 100 percent consistent condition began to improve. Further analyses demonstrated that the Degree Condition x Session interaction was not significant in the first three sessions [$\underline{F}(8,112) = 1.05$] but it was marginally significant across sessions 4 through 7, $\underline{F}(12,168) = 1.77$, $\underline{p} < .06$. Simple effects analysis revealed that the effect of session (during Sessions 4 through 7) was significant only in the 100 percent condition, $\underline{F}(3,42) = 4.13$, $\underline{p} < .01$.

Summary of Adaptive Training Results. As predicted, subjects were able to increase the presentation speed at which they were able to perform the task. Stimulus speed decreased steadily for the early sessions and then asymptoted at Session 5. During the later sessions, accuracy rates were generally stable across the conditions -- with the exception of the 100 percent condition, which began to improve.

Results: Fixed Training

Stimulus Speed. Stimulus speed during the fixed training phase was no longer an adaptive function of accuracy rate but was fixed at a constant rate which was individually determined; that is, the fastest stimulus speed obtained during the final session of adaptive training became that individual's stimulus speed for this phase of training. The average speed during this phase was 106 ms (range 100 to 125).

Accuracy. The data for the fixed training phase are presented in Figure 10. A Degree Condition (100 percent, 67 percent, 50 percent, 33 percent, and 13 percent consistent) x Session (8 through 12) ANOVA yielded significant main effects of Degree Condition, $\underline{F}(4,56) = 2.93$, $\underline{p} < .03$, and

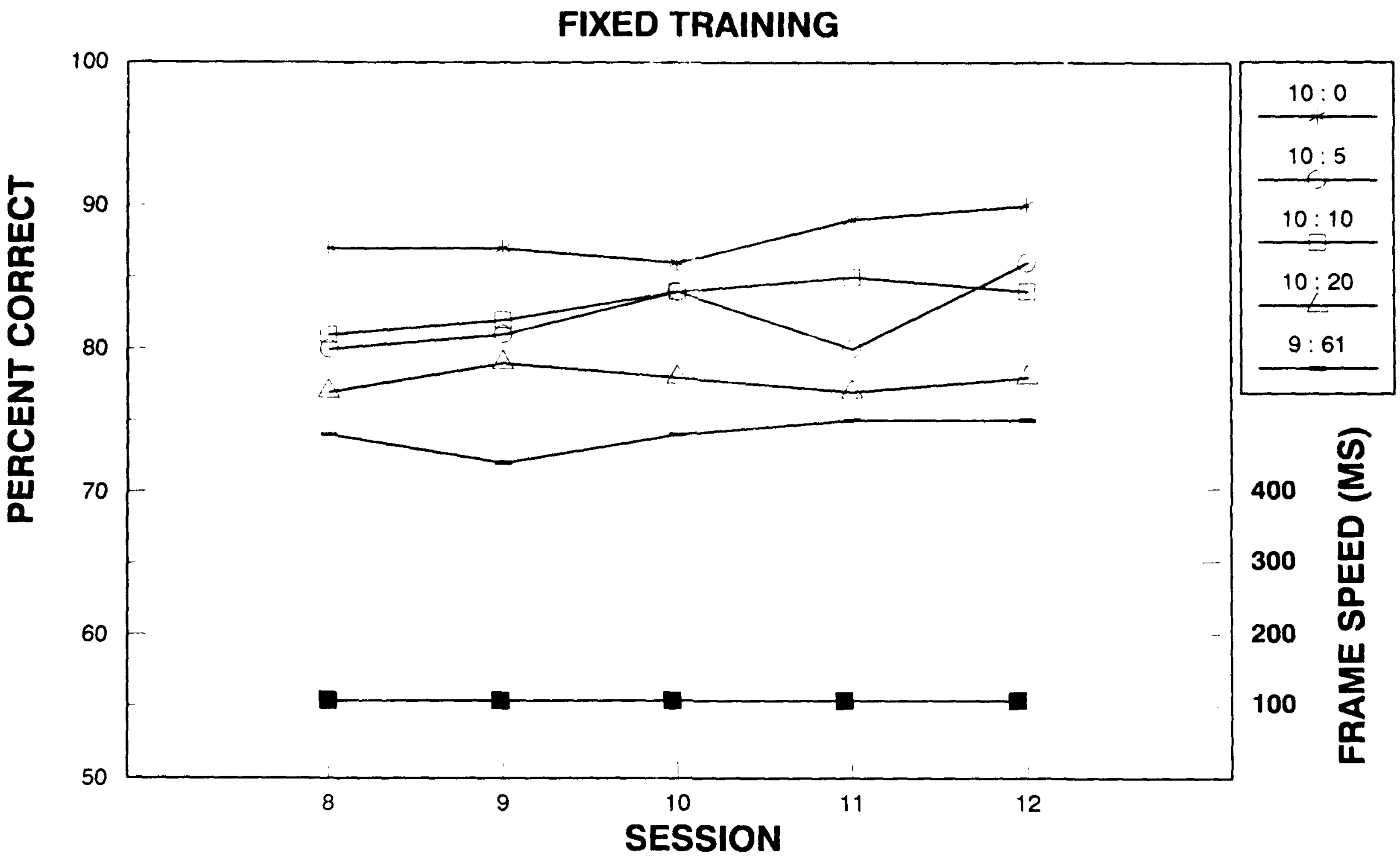


Figure 10. Mean Accuracy Rates and Frame Speeds for Each Condition Plotted as a Function of Each Session of Fixed Training (i.e., constant frame speed).

Session, F (4,56) = 2.62, p < .04. Student-Newman-Keuls analyses revealed that the 100 percent condition was superior to the other conditions; the 67 percent and 50 percent conditions were equal to each other, and slightly better than the 33 percent condition, which was in turn slightly better than the VM condition. (The figure shows that these comparisons collapse across training sessions.) Comparisons of the sessions revealed that Sessions 8 and 9 were significantly worse than Sessions 10, 11, and 12, which did not differ (thereby suggesting asymptotic performance).

To assess final-level performance, a one-way ANOVA was conducted on Session 12 data to determine the differences between the Degree Conditions (100 percent, 67 percent, 50 percent, 33 percent, and 13 percent consistent). The main effect of Degree Condition was significant, F(4,56) = 3.96, p < .007. A series of planned comparisons revealed the following pattern of effects: 100 percent consistency was superior to 33 percent consistency and 13 percent consistency (VM), but not different from 67 percent consistency or 50 percent consistency; both 67 percent consistency and 50 percent consistency were superior to the VM condition, but not different from 33 percent consistency and not different from each other; and 33 percent consistency was not better than VM.

Summary of Fixed Training Results. As is evident in Figure 10, throughout the fixed training phase there was a functional relationship between degree of consistency and accuracy performance. This is supported by the fact that across these sessions the 100 percent consistency condition always yielded superior performance; the 67 percent and 50 percent consistency conditions were slightly worse, followed by the 33 percent condition and the 13 percent condition (VM). This pattern follows our original predictions. However, assessment of final level performance revealed that the 67 percent and 50 percent consistency conditions were not different from the purely consistent condition. This is an important finding. It suggests that even in a high-speed, perceptually demanding task, the subjects were able to benefit in terms of performance improvement as a function of the degree of consistency present in the task.

Results: CM Test

Stimulus Speed. The same stimulus speed was used during the CM test as was used during the fixed training phase.

Accuracy. A one-way ANOVA was conducted on the Degree Conditions (100 percent, 67 percent, 50 percent, and 33 percent; there was not a VM condition in this session). The main effect of Degree Condition was significant, $F(3,42) = 3.17$, $p < .034$. The series of planned comparisons yielded a very similar pattern to that observed in the final session of fixed training. The 100 percent consistent condition was superior to the 33 percent consistent condition and the remaining comparisons were not significantly different. The contrast results are presented in Table 7 with the results of the contrasts for the final session of fixed training.

A Session (12 vs. 13) x Degree Condition (excluding the VM condition in Session 12) ANOVA was conducted in order to directly compare the accuracy performance in the final fixed training session relative to the CM test session. These data are presented in Figure 11. The main effect of Session was significant, $F(1,14) = 47.67$, $p < .0001$, and the main effect of Degree Condition was marginally significant, $F(3,42) = 2.64$, $p < .06$. The interaction of Session by Degree Condition was not significant [$F(3,42) = 1.27$]. As is evident in Figure 11 all of the Degree Conditions improved somewhat from the final session of fixed training (where words appeared as both targets and distractors) to the CM test sessions (where the words appeared only as targets). The marginally significant effect of Degree Condition further supports the idea of a functional relationship between accuracy performance and degree of consistency.

Table 7. Contrasts for Fixed Training and CM Test Sessions

Final Session Fixed Training

Contrast	DF	F Value	p Value
10:0 vs. 10:5	1,56	1.04	0.3124
10:0 vs. 10:10	1,56	1.68	0.1998
10:0 vs. 10:20	1,56	7.06	0.0103
10:5 vs. 10:10	1,56	0.08	0.7820
10:5 vs. 10:20	1,56	2.68	0.1072
10:10 vs. 10:20	1,56	1.85	0.1796
10:0 vs. 9:61	1,56	12.73	0.0007
10:5 vs. 9:61	1,56	6.49	0.0136
10:10 vs. 9:61	1,56	5.15	0.0271
10:20 vs. 9:61	1,56	0.83	0.3661

CM Test Session

Contrast	DF	F Value	p Value
10:0 vs. 10:5	1,42	3.41	0.0717
10:0 vs. 10:10	1,42	1.14	0.2925
10:0 vs. 10:20	1,42	8.91	0.0047
10:5 vs. 10:10	1,42	0.61	0.4387
10:5 vs. 10:20	1,42	1.29	0.2619
10:10 vs. 10:20	1,42	3.68	0.0618

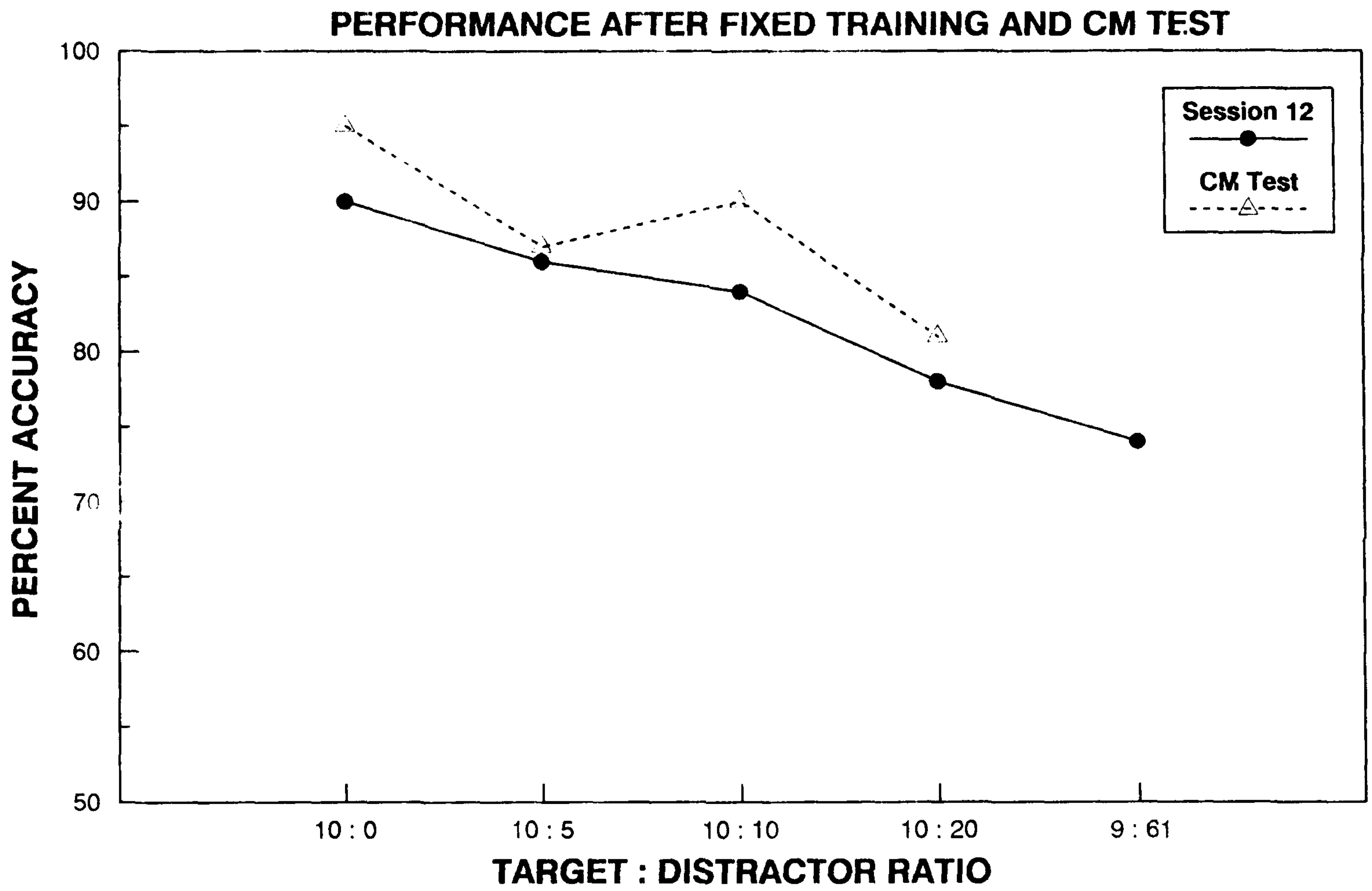


Figure 11. Mean Accuracy Rates for Each Condition for the Final Session of Fixed Training (Session 12) and the CM Test Session.

Summary of CM Test Results. The similarity between the pattern of results of the CM test and that for the final session of fixed training suggests that benefits of training with greater consistency are stable across sessions as well as across training situations (see Figure 11). During both the adaptive and the fixed training phases, the same words appeared as both targets and distractors (except in the 100 percent consistent condition). Changing the task such that the words appeared only as targets (i.e., the CM test) changed the experimental context but did not change the pattern of results (although overall performance did improve).

Discussion

The present data support the prediction based on previous research (Schneider & Fisk, 1982) that detection accuracy in search/detection tasks is a monotonically increasing function of degree of consistency and amount of practice. The present results are important because they extend what was previously known about automatic process development in situations with less than perfect consistency. The present paradigm employed a task which is a conceptual analog of real-world, high-performance perceptual processing tasks and requires automatic detection to occur at a more global level than an individual stimulus feature.

Schneider and Fisk (1982) examined effects of degree of consistency on automatic process development by using a relatively simple, single-letter detection task. They found that large amounts of practice in a VM condition produced little improvement in performance. They also found that consistent practice resulted in little benefit to performance until a substantial number of trials had occurred. Schneider and Fisk found that a ratio of 10 stimulus occurrences as a target to 20 stimulus occurrences as a distractor led to little performance improvement. Their results suggested that consistency is a necessary condition for automatic process development. Their results further demonstrated that learning is not the result of process execution but rather, a function of consistent executions of a process. Unfortunately, from the perspective of application to more complex real-world tasks, the consistent feature in the Schneider and Fisk experiments was a letter shape. It was not known whether degree of consistency effects were operational in a task where a higher-order consistency existed even though the elemental features (e.g., specific letter shapes) were not consistently mapped.

The present experiment resolved those questions that were unanswered from the original degree of consistency study. In the present experiment we found that, once subjects were performing at their limits of perceptual processing, performance improved as a multiplicative function of degree of consistency and practice. In fact, throughout the fixed training phase, there was a consistent functional relationship among practice, degree of consistency, and detection performance. The 100 percent consistency condition always yielded superior performance, the 67 percent and 50 percent consistency conditions resulted in intermediate performance and the 33 percent and 13 percent consistency conditions led to poor performance. The 33 percent and the 13 percent consistent conditions did not improve throughout the fixed frame time evaluation phase of the experiment.

The present data do support the fact that consistency is necessary for performance improvement even in tasks requiring complex, high-speed visual search, with consistency defined as a combination of lower level features (i.e., with consistency defined by high-order feature combinations).

The present experiment also places limits on what can be defined as training context, at least for search-detection-type tasks. (At least the present experiment allows a determination of what kind of contextual information will or will not bias performance.) In the present experiment, the initial context could be defined as the degree of consistency manipulation. In the pure CM testing phase, we changed the task (and thus, one could argue, the context) such that the words in the 67, 50, and 33 percent consistent training conditions became completely consistent. This manipulation demonstrated that such a change did not produce a change in the pattern of results. Performance in the previously inconsistent conditions did not immediately return to the level of the 100 percent consistent condition, nor did performance in those conditions deteriorate. Either of those findings would have forced us to argue that consistency at the stimulus level (higher-order in this case) was less important than the overall context within which the stimuli were presented.

In summary, the present data allow for greater confidence in a qualitative statement about the effect of degree of consistency and practice on performance <u>across a range of tasks</u>. Thus, individuals who must design training programs can have some confidence in statements about relative levels of performance improvements, given that the degree of component consistency level can be determined. The present data, coupled with the Schneider and Fisk (1982) data, also suggest that if a quantitative statement about performance levels based on degree of consistency is desired, then task-specific factors such as the level of consistency (e.g., elemental versus global), the type of task (e.g., high-speed perceptual detection, visual scanning of a static display), and so on must be considered.

V. EXPERIMENTAL SERIES 4: GLOBAL VERSUS LOCAL CONSISTENCY: EFFECTS OF DEGREE OF WITHIN-CATEGORY CONSISTENCY ON LEARNING AND PERFORMANCE

Introduction

It is a common observation that there is a gradual transition from novice performance to the skilled activities of the expert. Schneider (1985b) has suggested that this gradual transition represents a change from performance dominated by controlled processing to the development and increased use of automatic processes. An illustration of this transition is the changes that occur when first developing a new skill such as learning to ride a bicycle. At first most, if not all, of the novice's attentional resources are consumed in attending to the details of the task. Attention is devoted to each distinct movement. Gradually, with practice, the task becomes dominated by automatic processes and less attention must be devoted to components of the task. Controlled processing can be used to plan ahead, to talk to one's riding partner, or to think about the day's activities. Clearly, the development of automatic component processing, though not sufficient for skilled performance, is necessary for the novice to become an expert. It is important, therefore, to understand factors that facilitate, as well as inhibit, the development of automatic processing.

Several researchers have demonstrated that practice must be consistent for subjects to benefit from training (Fisk et al., 1987; Schneider & Fisk, 1982). Traditionally, consistent practice is said to occur when stimuli are dealt with in the same manner from stimulus exposure to stimulus exposure. In other words, consistent practice occurs when the stimuli and responses are consistently mapped; that is, across training trials the individual makes invariant responses to stimuli (or classes of stimuli). If individuals receive VM training (i.e., the stimuli require responses that change across time) automatic processing will not develop and performance will not dramatically improve with practice (Schneider & Shiffrin, 1977).

Support for the value of consistency (as traditionally defined) in development of automaticity can be found in research using memory and visual search paradigms (e.g., see Fisk & Schneider, 1983; Schneider & Shiffrin, 1977). However, consistency need not occur at the individual stimulus level to benefit performance. Recently, Durso, Cooke, Breen, and Schvaneveldt (1987) compared performance improvement with practice on a traditional CM letter search task to improvements on a "digit detection" task. Their digit task differed from both traditional CM and VM search tasks. Their digit task required subjects to respond to the largest digit in a display (largest in terms of ordinal property; that is, 9 is larger than 8, 8 is larger than 7, etc.). The digit task was not consistently mapped in the traditional sense because a given digit was not always responded to when it appeared on the screen. For example, the digit 7 is largest and responded to when digits 6 and below are on the screen but it is ignored when the digit 8 or 9 is in the display. Durso et al. found results in the digit task that were comparable to the CM letter search task; that is, an overall reduction in reaction time and an attenuation of comparison load effects with practice.

At first glance, the Durso et al. (1987) research calls into question the need for consistency in training. However, Fisk, Oransky, and Skedsvold (1988) explored whether relationships among stimuli might generate task-relevant consistencies by manipulating the consistency of relationships among stimuli. Their experiments demonstrated the facilitating role of "higher-order" or "global" consistency in developing skill-like performance. Fisk et al. furthered the understanding of consistency in complex tasks by demonstrating that in conditions where subjects could utilize higher-order consistencies (relationships), normal CM practice effects occurred even when the individual stimuli were not always mapped to a particular response.

The present experiment was conducted to examine the interaction between consistency at the "global" versus the "local" level. This is important because, although the effect of high-order consistency on overall task performance is now known, the influence of higher-level inconsistency on learning lower-level task elements remains unknown. Global-level consistency is defined as higher-order or situation-specific consistency such as the consistency defined by relationships among stimuli (Durso et al., 1987; Fisk, Oransky, & Skedsvold, 1988; Myers & Fisk, 1987). Local-level consistency is defined as stimulus-specific consistency.

Fisk and Schneider (1983) and many other investigators have provided information on both the CM/CM (GLOBAL CONSISTENCY/LOCAL CONSISTENCY) and the VM/VM (GLOBALLY INCONSISTENT/LOCALLY INCONSISTENT) conditions. In the Fisk and Schneider experiments, the CM condition is considered CM/CM (in terms of the global/local distinction) because categories and words from the CM categories appear only as targets. In the VM condition, it is considered VM/VM because categories (global level) and words (local level) from the VM categories appear as both targets and distractors.

The Fisk, Oransky, and Skedsvold (1988) studies provide data for the CM/VM situation. In those relational learning studies, the consistency is maintained at the global level even though the individual stimuli are inconsistent. In the present experiment, we were particularly interested in the VM/CM condition; that is, we specifically examined the effect of inconsistency at the global level when local level processing was consistent. Consistency at the global level of processing was manipulated by varying the consistency of mapping at the semantic category level. Consistency at the local level of processing was manipulated by varying whether specific words appeared as both targets and distractors (VM) or merely as targets (CM).

In the present classification, a semantic category (e.g., "articles of clothing") may be consistent (CM/CM) because all the exemplars appear only as targets. Conversely, a category (e.g., "human body parts") may be inconsistent at both the global and local levels (VM/VM because all the words in that category are used as targets and distractors. Finally, at the global level, a category may be inconsistent because some exemplars are used both as targets and distractors but some of the words from that category may be used only as targets (VM/CM); hence, consistency is maintained at the local level for some stimuli.

Three potential patterns of results could occur for performance improvement in the VM/CM conditions. We could find similar performance for the consistent and inconsistent words in the VM/CM categories. This result is unlikely in light of the findings by Schneider and Fisk (1982), in which improvement (over VM performance) was found for letters of differing degrees of consistency. However, a finding of no difference between the CM and VM words (within the VM/CM condition) would shed light on the influence of higher-order inconsistency, at least for laboratory perceptual learning tasks. Second, the improvement found for the CM words may be influenced by the degree of category consistency. This result would show an important interaction between category (top-down) and word (bottom-up) learning. Finally, within the VM/CM condition processing of the CM words may not be influenced by inconsistencies at the category level (shown by superiority over the VM words) which would imply use of consistency at the highest level possible within a given situation (in this case the local or word level).

Another important issue relevant to the present study has to do with the transfer of learning that occurs in a search task. In this case we are interested in how well people, upon being trained to a certain group of words from one category, will detect a new word belonging to that same category.

Schneider and Fisk (1984) examined the possibilities and found the following. In the first of four experiments, they studied the latency to detect words from a category of varying sizes (i.e., from 4 to 12 exemplars). The results showed an overall improvement in performance for CM conditions, but there was no significant effect for the number of exemplars in a category. The second experiment examined the transfer of trained to untrained items. They found positive transfer that was in fact significant (60 percent to 92 percent). The relationship between transfer effects and exemplars was that the more exemplars there were in a category during training, the better the transfer performance was. The third experiment demonstrated that the more subjects practiced the task, the less sensitive they were to resource costs under consistent mapping conditions; however, performance in the VM condition did not benefit from training. Finally, in the fourth experiment (under high workload), the effects of practice given CM training still produced positive transfer to untrained exemplars. Therefore, practice affects processing at the level of the category feature node.

Integrating these above-mentioned principles--degree of consistency, category search effects, and transfer effects--we used a methodology whereby automatic processing is evaluated at the level of varying degrees of within-category consistency. That is, we were interested in the degree of within-category consistency on performance and the amount of learning. In essence, we were asking if, given that categories differ in the percentage of consistent category members, learning will in fact differ at the category level. As an extension of previous studies investigating the effects of consistency on automaticity, this study additionally requires subjects not only to detect the presence of a word but also to make a semantic category judgment as to whether the word belongs to the category presented in the memory set. Thus, subjects are operating at the level of semantic processing and not at the level of simple detection and identification.

Thigpen and Fisk (1988) suggested that learning should take place at the level of the stimulus (local level) to facilitate performance when category (global) inconsistency is present. If this is true, then some difference in transfer performance should be observed across the within-category consistency conditions.

Method

Subjects. Nine subjects, six males and three females, participated in this study. Subjects were paid $4.00 per hour, with a bonus of $1.00 per hour upon completion of the experiment. All subjects were students at the Georgia Institute of Technology. All subjects were administered subscales of the Wechsler Adult Intelligence Scales (Wechsler, 1981). The subscales included the digit-span, digit-symbol substitution, and vocabulary tests. The scaled scores for the vocabulary test ranged from 9 to 19, with a mean score of 14.33. The scaled scores for the digit-span test ranged from 7 to 18, with a mean score of 11.78. The scaled scores for the digit-symbol substitution test ranged from 11 to 19, with a mean score of 13.44. All subjects had normal or corrected to normal vision--at least 20/30 for distance and 20/40 for near vision.

Apparatus. Epson Equity I+ microcomputers equipped with Epson MBM-2095-E monochrome monitors (green phosphor, 50-Hz refresh rate) with Epson multimode graphics adapters were programmed to present the task and collect data. The microcomputers were programmed with Psychological Software Tools' Microcomputer Experimental Language (MEL) to present and time stimulus displays and to record response behavior. Pink noise set at approximately 55 db was provided to minimize distractions. Subjects were positioned at different computer stations in the same room. Subjects were required to manually respond to stimuli by indicating whether a target was in the top, middle, or bottom position. Subjects responded using the '7', '4', and '1' keys on the numeric keypad, which were labeled 'T', 'M', and 'B', respectively; there was one-to-one correspondence between target position and response-key position. The index finger was used to make responses. Subjects were told to keep the index finger positioned over the 'M' key when not responding. Reaction time was measured in milliseconds.

Stimuli. Nine semantic categories with eight exemplars from each category were used during the training portion of the experiment. Four additional exemplars were added to each category during the transfer phase, for a total of twelve exemplars per category. The categories were MUSICAL INSTRUMENTS, ANIMALS, VEGETABLES, BODY PARTS, WEAPONS, EARTH FORMS, RELATIVES, UNITS OF TIME, and CLOTHING. The exemplars chosen were all high associates of their respective categories as indicated by the Battig and Montague (1969) norms. In addition, care was taken to ensure that the categories were semantically non-overlapping, as described by Collen et al. (1975). The stimuli were counterbalanced across conditions and subjects using a Latin Square.

Design. All conditions were manipulated within-subjects. The dependent variables were reaction time and accuracy. The primary independent variable was the ratio of consistent to inconsistent words within a category (defining the degree of within-category consistency). The ratios of consistent to inconsistent words (C:I) were 8:0, 6:2, 4:4, 2:6, and 0:8. Consistency conditions were manipulated between trials. Consistency Condition 1 was completely consistent (8:0). That is, all eight words within that particular category were consistently targets. For Consistency Condition 2 (6:2), one category was selected in which the first six words were consistently mapped and the last two words were variably mapped. For Consistency Condition 3 (4:4), the third category chosen was consistent 50 percent of the time. The first four words in that category were consistent and the last four were inconsistent. Consistency condition 4 (2:6) contained a category for which the first two words were consistent and the remaining six words were inconsistent. Finally, condition 5 was a traditional VM condition in which all of the words within the remaining five categories were inconsistent.

Each 1-hour session consisted of 17 blocks of 48 trials: 16 VM trials, and 8 trials of each of the four Consistency Conditions (8:0, 6:2, 4:4, and 2:6) per block; thus subjects completed 136 trials of each Consistency condition and 272 trials of VM per session. The training phase of the experiment lasted for 12 sessions for a total of 9,792 trials. During each transfer session, there were 22 blocks of trials with 36 trials in each block. Subjects completed two sessions of transfer.

Procedure. Subjects were trained for 12 days on four CM/VM conditions. Subsequent to training, subjects participated in 2 days of transfer.

The temporal sequence of an individual trial was as follows. Subjects were presented with a category label shown on the left side of the screen. They were told that they would be required to search for an item within the category presented. Subjects were allowed 20 seconds to study the category. When they were ready to initiate the trial, they were instructed to press the space bar. At this time, three plus signs (+) aligned in a vertical column appeared for .5 second. The plus signs were positioned in the location of the probe words to follow (i.e., in the center of the screen). The plus signs were to act as orientation points so that the subjects could focus their eyes on the area where the words would appear. Following the plus signs, the probe display for that trial was presented consisting of three words presented in a column. The first letter of each word occurred in a location corresponding to one of the orientation plus signs. On each trial, one of the words in the probe display was a member of the previously presented category. Subjects had to decide which word was the target word and to respond accordingly by pressing one of three keys on the numeric keypad that corresponded to the position of the word on the screen: 'T' (top), 'M' (middle), 'B' (bottom). Subjects had up to 6 seconds to respond. Subjects were instructed to respond as quickly and accurately as possible. Following each trial feedback was provided. If the response was correct, the word 'CORRECT!' appeared at the bottom of the screen. If the response was incorrect, a tone sounded and the word 'ERROR', along with the correct response (i.e., the target word), was presented at the bottom of the screen. Immediately following the feedback, the next category appeared and the subject could again initiate the trial by pressing the space bar.

At the end of each block (48 trials), subjects were given feedback on their mean accuracy and correct trial reaction time for that block, as well as a running account of their mean accuracy and correct reaction time for each of the completed blocks of trials. Subjects were also told that at this time they could take a break and resume working when they were ready.

During the training phase, subjects were trained in five conditions which varied in the degree of within-category consistency. The level of consistency is denoted by the ratio of CM words to VM words within a particular category. Consistency condition 1 was 100 percent consistent in that all eight words in the category appeared only as targets giving a ratio of 8:0 (CM:VM). Consistency condition 2 was 67 percent consistent, with a ratio of 6:2. Consistency condition 3 was 50 percent consistent, yielding a ratio of 4:4. Consistency condition 4 was 33 percent consistent with a CM:VM ratio of 2:6. Condition 5 was a pure VM condition, with zero words serving as completely consistent targets (i.e., the CM:VM ratio was 0:8).

In the transfer phase, four new exemplars were added to each of the trained categories in the Consistency Conditions. A new CM condition was added which was created by consistently pairing two of the VM categories. During transfer all words were consistent.

Results: Training Phase

Correct trial reaction times for each Consistency condition across the 12 sessions of training are presented in Figure 12. The data in this figure represent the average reaction time for all words within a category; that is, for the 6:2, 4:4, and 2:6 Consistency conditions, RT represents an average of the CM and VM words within that condition. When performance is examined in this manner, it is clear that performance improved as a function of the degree of within-category consistency. That is, the VM condition showed the least amount of improvement and RT decreased (i.e., performance improved) as a function of practice and degree of within-category consistency.

A Search Condition (Consistency condition 1 [8:0], Consistency condition 2 [6:2], Consistency condition 3 [4:4], Consistency condition 4 [2:6], VM) x Practice (Sessions 1 through 12) ANOVA was performed on these reaction time data (one subject's data for Session 5 and Session 6 were lost; hence, the analyses used a correction for unequal number of observations within those two sessions). There were significant main effects of Search Condition, $\underline{F}(4,32) = 4.16$, $\underline{p} < .01$, and Practice, $\underline{F}(11,86) = 22.31$, $\underline{p} < .0001$. The Search Condition x Practice interaction was also significant, $\underline{F}(44,344) = 1.92$, $\underline{p} < .001$. The results from this overall analysis support previous findings that there is in fact a significant difference in the mean RTs among varying degrees of consistency.

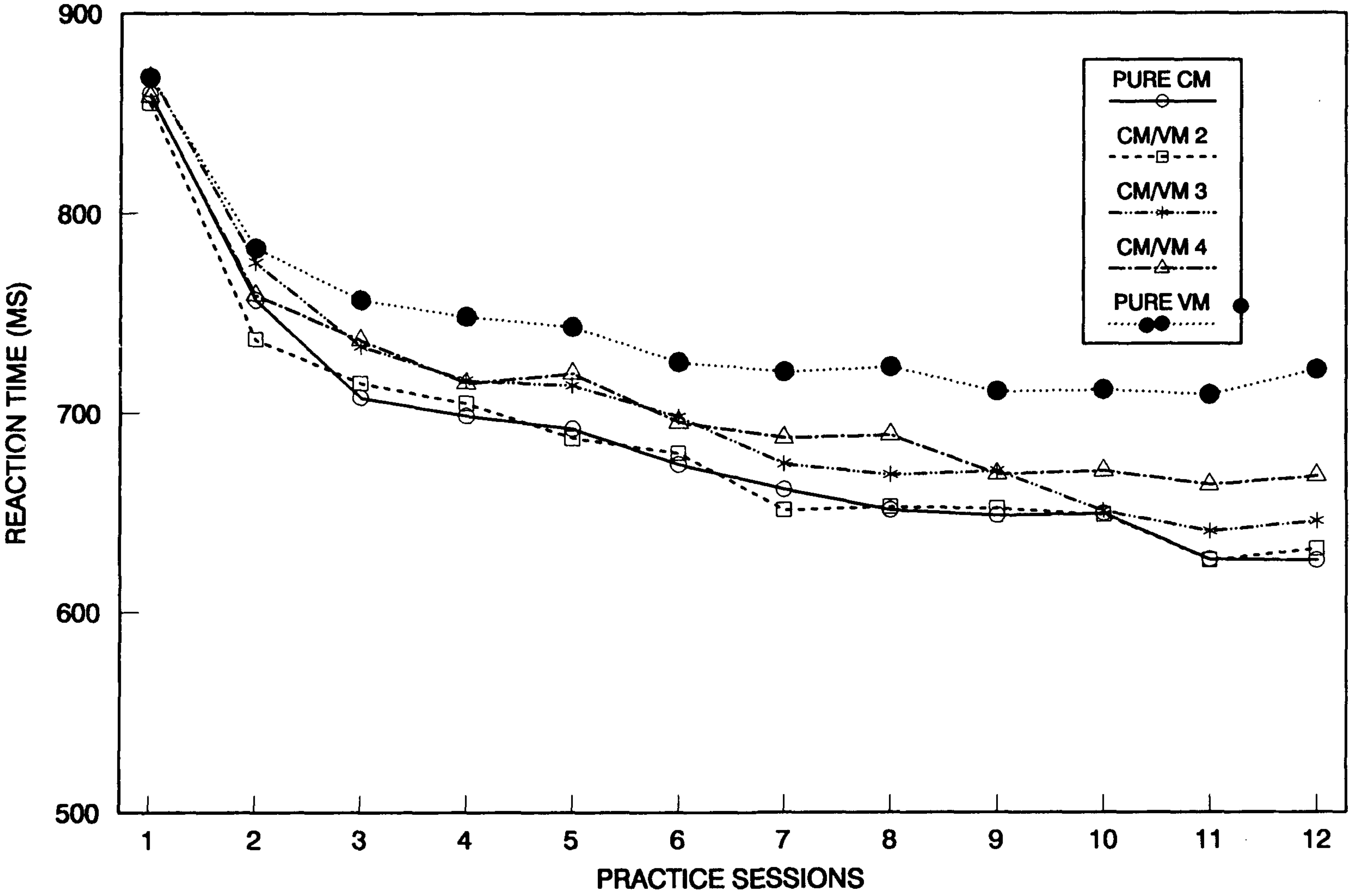


Figure 12. **Reaction Time for Each Training Condition Plotted as a Function of Practice Sessions**

Consistent Words versus Inconsistent Words. The above degree of consistency effects may have occurred because performance was affected at the global (or category) level or because performance was affected at the local (or word) level. If performance was affected at the category level, then the degree of category inconsistency should affect the detection time of the CM words within the category. If performance was affected at the local, word level, then all CM words, regardless of the degree of within-category consistency, should have equivalent reaction times. If the latter is correct, then the above findings of poorer performance as a function of degree of within-category consistency would have occurred because as consistency decreased (from 8:0 to 0:8), the averaged RT of each condition would be represented by more VM (slow) words and fewer CM (fast) words.

Statistical analysis and an examination of Figure 13 show that the latter explanation is correct. That is, Global consistency did not affect RT performance on the CM words at the local level. The main effect of Practice was significant, $F(11,86) = 28.71$, $p < .0001$. However, there was little effect on the RTs of the CM words across the different degrees of within-category consistency conditions. That is, the main effect of Search Condition was not significant, $F < 1$, nor was the Search Condition x Practice interaction, $F < 1$.

When we separately examined the VM words as a function of the degree of within-category consistency (see Figure 14), we found that once again the degree of consistency at the global level did not affect performance at the local level. The main effect of Search Condition and the interaction of Search Condition x Practice were both insignificant, with Fs < 1.

When the 6:2, 4:4, and 2:6 conditions were analyzed separately, the pattern of results was the same. CM words differed from VM words, $F(1,8) = 24.32$, $p < .004$, and there

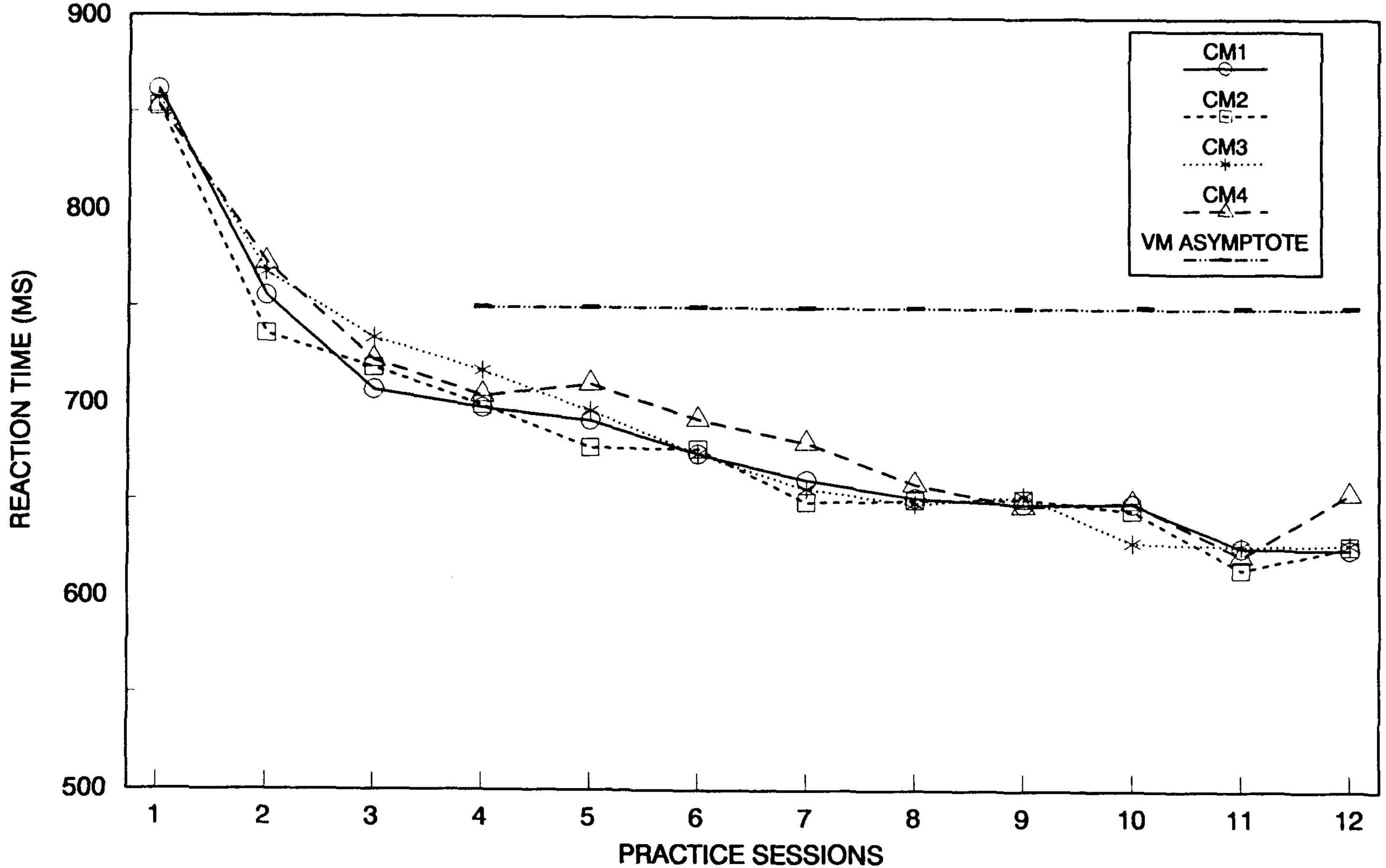


Figure 13. Reaction Times for Consistent Exemplars Only at Each Degree of Within-Category Consistency, Plotted as a Function of Practice Sessions.

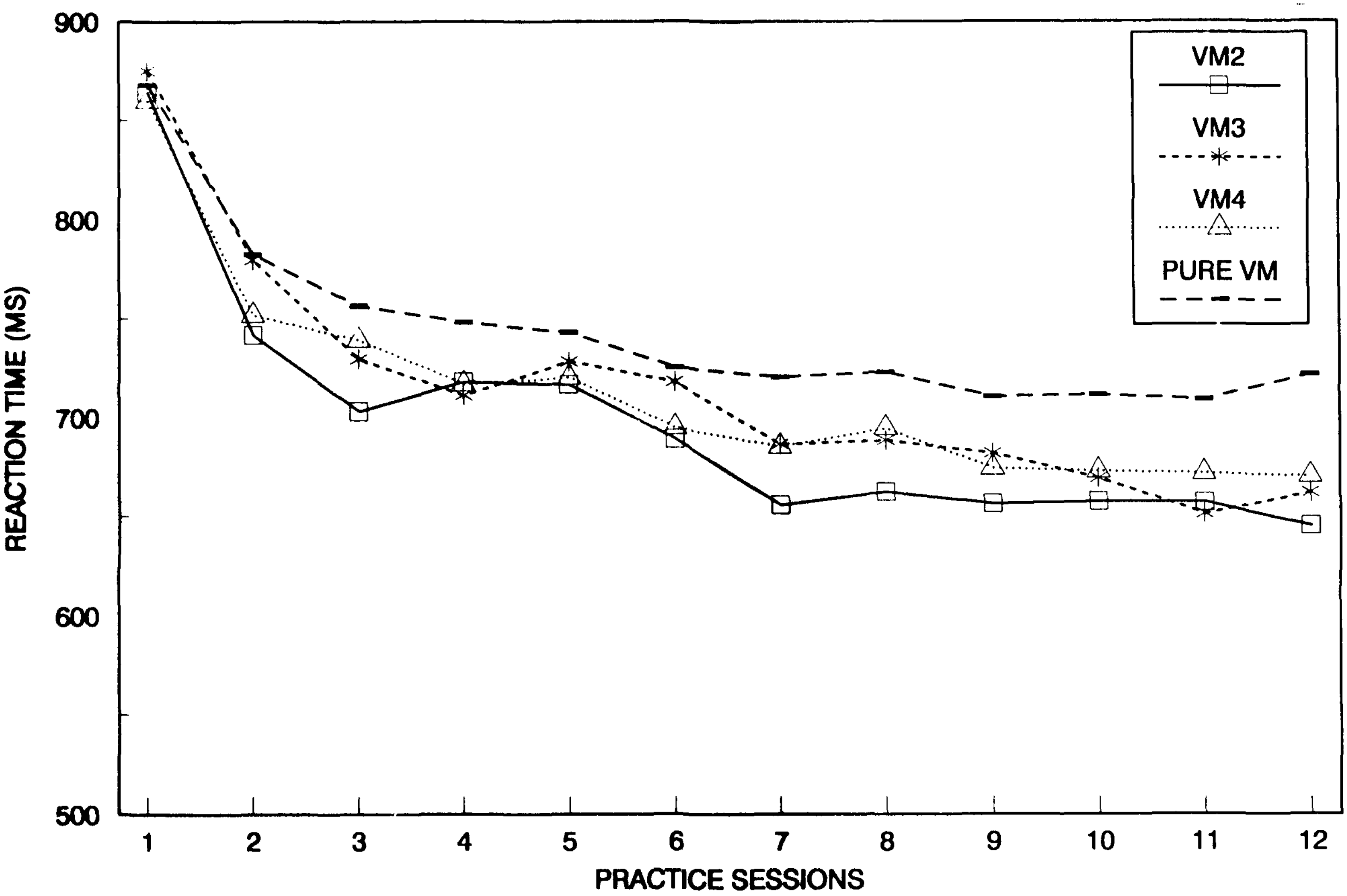


Figure 14. Reaction Time for Each VM Condition Plotted as a Function of Practice Sessions.

was an interaction between Word Type (CM versus VM) and Practice, $F(2,20) = 16.87$, $p < .0003$.

Results: Transfer Phase

The previous analyses examined the effects of degree of within-category consistency on performance. We next examine the effects of within-category consistency on learning at the category level. By examining performance on the untrained words from the trained categories (i.e., transfer performance), we can assess learning.

Figure 15 presents the correct trial reaction times as a function of search condition and trained versus untrained category exemplars. As can be seen, when compared to the New CM condition (baseline) RT for untrained exemplars increases as the degree of within-category consistency decreases. The main effect of Search Condition was significant, $F(8,64) = 2.59$, $p<.05$.

To most conservatively examine transfer we examined the proportionality of change between the detection performance of the trained exemplars compared with the untrained exemplars relative to the New CM condition (see Roscoe & Williges, 1980, for a general discussion). As a measure of transfer we used the following equation to compute percent of transfer: Transfer = (New CM RT - Untrained Exemplar RT) / (New CM RT - Trained Exemplar RT) x 100. The averaged transfer scores as a function of condition were 41.0 percent (8:0 condition), 11.0 percent (6:2 condition), 4.0 percent (4:4 condition), and a negative 81 percent and for the 2:6 condition. Transfer for the 8:0 condition was positive and significantly different from zero, $t(8) = 12.87$. Percentage transfer was not different from zero for the 6:2 or the 4:4 condition (ts < 0). Unexpected negative transfer was observed for the 2:6 condition. These transfer data suggest that even a moderate degree of within-category inconsistency will attenuate learning at the global or, in this case, semantic-category level.

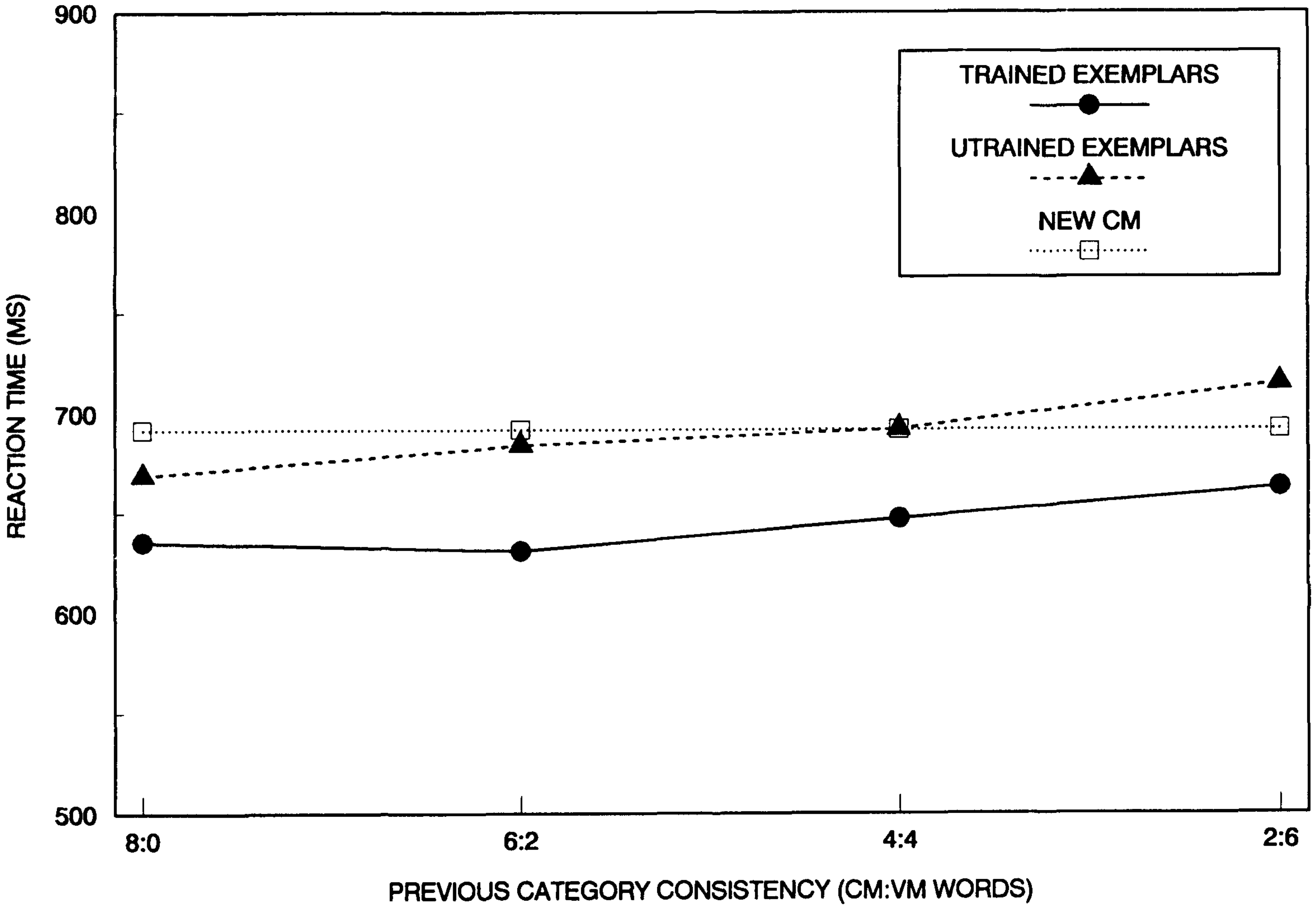


Figure 15. Reaction Time for Each Transfer Condition Plotted as a Function of Previous Category Consistency.

Discussion

Given the similarity of detection performance of the CM words across the conditions of within-category consistency (i.e., across the 8:0, 6:2, 4:4, and 2:6 conditions), we argue that consistency, at any level, may be capitalized on during training to facilitate task-specific performance. With the present experimental design, inconsistency at the category level did not inhibit detection of the consistent exemplars within those inconsistent categories. These data can be interpreted within a framework in which consistent training is assumed to enhance the strength of attention for each target stimulus (cf. Dumais, 1979; Fisk, Lee, & Rogers, in press; Rogers, 1989). "Attention strength" is related to learning invariant features, invariant relationships among stimuli, higher level verbal codes, etc.

Previous research has supported this global view of attention strengthening (Durso et al., 1987; Fisk, Oransky, & Skedsvold, 1988) by showing that complete stimulus-response consistency need not be present for attention strengthening to occur. However, the present data suggest, as might be expected, when invariant higher-level relationships do not exist, the locus of performance improvement will be at the stimulus level if the stimulus is consistently mapped to a response. Indeed, as Duncan (1986) has pointed out, "...the important question is...at what level consistency affects learning" (p. 283). The present data aid in answering that question by demonstrating that attention strengthening, and therefore task-specific performance improvement, will not be disrupted by inconsistencies at a level above the to-be-responded-to stimulus.

However, task-specific performance improvement is not the only issue at hand. Learning can be more broadly defined as the ability to transfer to situations <u>related</u> to the trained task. The present transfer phase of our experiment demonstrated that "global" inconsistency can have disastrous effects on more generalizable learning. Transfer was a direct function of global, or within-category consistency; however, only the completely consistent category resulted in statistically significant transfer. These data are in line with the Schneider and Fisk (1982) degree of consistency data in which degree of consistency at the element (single letter) level defined the highest order of learning. In their experiment, only the 100 percent consistent condition resulted in statistically significant improvement over the course of the experimental training session; also, as with our present findings, a functional relationship was discovered between degree of consistency and detection performance. Hence, we must modify our original statement: Consistency at any level will be used to facilitate task-specific performance improvement; however, transferable learning (learning that is not stimulus-level-based) can occur only at the level of highest-order consistency.

The present results should send a message of caution to those designing training for rich, complex tasks. Real-world tasks are composed of many different levels of consistency. If lower-level consistencies are known to interfere with higher-level performance (e.g., automatically detecting certain letters when trying to read), then care must be taken to ensure that the lower-level consistencies are made as non-salient as possible during training on the high-order skill. If care is not taken, then trainees may focus on irrelevant aspects of the task or, worse yet, incorrectly learn aspects of the task. If learning is desired at a higher order than the task-specific performance training, then care must also be taken to ensure that the higher-order consistencies are present and recognizable.

VI. EXPERIMENTAL SERIES 5: THE TEMPORAL NATURE OF CONTEXT AS A FACILITATORY MECHANISM FOR PERFORMANCE IMPROVEMENT IN VISUAL SEARCH

Introduction

In this section of the report, we expand on previous research (e.g., Eggemeier et al., 1988; Fisk & Eboch, 1989; Fisk & Gallini, 1989; Fisk etal., in press; Fisk & Lloyd, 1988; Fisk, Oransky, & Skedsvold, 1988; Fisk & Rogers, 1988; Myers & Fisk, 1987) that has examined the incorporation of the concept of consistency and the principles of automatic/controlled processing theory (see overview in Section II) into training real-world, complex skills. The underlying goal of this research effort has been to uncover generalizable, guiding principles for understanding both the role and the training of consistent components of tasks classified as high-performance skills (Schneider, 1985a). In the present section, we report an experiment conducted to better understand situation-specific, contextually driven consistency (Fisk & Rogers, 1988).

Fisk and Rogers (1988) investigated the issue of situation-specific context using a semantic category search task in which context was defined as the combination of target and distractor sets. That is, a given category was the target set <u>only</u> in the context of another particular category as the distractor set. For example, "ANIMAL" words might be the target set if "WEAPONS" are the distractors; however if animal words are paired with "VEGETABLES" the "ANIMAL" words are distractor items. Thus, the experimental context defines whether a particular set of items is attended to or ignored. Fisk and Rogers found that in the absence of traditional consistency, context can play an important role in facilitating performance. The results from their experiment showed that performance in the context conditions improved more than performance in the VM condition, which indicated some benefit of context in the absence of total consistency. However, the context conditions improved less than performance in the pure CM condition. Fisk and Rogers also found that the performance improvement seen in the context conditions seemed to be the result of a temporary "salience" biasing. Subsequent analyses of Fisk and Rogers' data suggested that the context effect seemed to occur within five exposures to the context situation; that is, previously acquired context effects that were incompatible with a new to-be-performed task could be eliminated and new context effects built up very quickly.

The above-mentioned analyses of the Fisk and Rogers (1988) data have important implications for training and for the assessment of why performance improves (e.g., on-line computer-aided evaluation of automatic process development). Unfortunately that examination of the data was conducted in an ad hoc fashion. The present experiment was designed to directly test the temporal buildup of context effects and the strength of those effects as a function of time (trials) within a given context before cycling to a different context.

Why should one be interested in context effects when attempting to develop high-performance skills? Before addressing this question, we should diverge and review consistency and practice improvement.

It is well known, and indeed a truism, that practice is required to improve performance in most behavior we would classify as skilled. However, the fact that practice, in and of itself, does not lead to skilled performance has been well documented (e.g., see Fisk et al., 1987; Schneider & Fisk, 1982; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Performance improves as a function of both the consistency and the amount of practice (Schneider & Fisk, 1982). However, what does it mean to say that practice is consistent? To investigate the role and value of consistency of practice, we have used laboratory tasks in which consistent practice refers to those situations that allow the individual to always deal the same way with either (a) a specific stimulus (Schneider & Fisk, 1982); (b) a category or class of stimuli (Fisk & Schneider, 1983); or (c) relationships among a set of stimuli (Fisk, Oransky & Skedsvold, 1988). Traditionally, consistent practice has been referred to in the literature as consistent mapping (CM) because the mapping between a stimulus (or set of stimuli) and a given response is consistent (Schneider & Shiffrin, 1977).

The other general class of training situations, varied mapping training conditions, are those in which practice is inconsistent; that is, the individual cannot attend to or respond to a stimulus in a consistent manner from one stimulus exposure to another. Several investigations (e.g., Ackerman, 1986; Fisk & Schneider, 1983; Logan, 1978; Myers & Fisk, 1987; Schneider & Shiffrin, 1977) have demonstrated that varied mapping training situations lead to much less improvement than consistent mapping training, especially if the tasks incorporate complex stimuli or training situations.

One important piece of information that has recently received some attention relates to the ability of subjects to use situation-specific context to mediate "consistency" and, therefore, to show performance improvement in the absence of traditional consistency or to cue the use of well-developed automatic processes. More generally, we need to better understand the characteristics of how situation-specific context can facilitate the development of what would otherwise be competing automatic processes. This, in fact, appears to be a characteristic of skilled performers. For example, observation of skilled air-intercept controllers suggests that responses made to pilots in one context (an intercept with the goal to simply identify an aircraft) are different from those made in other contexts (such as an intercept with the goal to defend friendly air space from hostile intruders). Skilled controllers can develop reasonable "situational awareness" by listening to pilot transmissions and knowing the intent of the mission. Pilot to controller communication in one context (e.g., when identifying a commercial airliner) generates a different situational awareness than the same message in a different context (e.g., defending against hostile aircraft). Context is also important in activating behavior sequences. For example, it is estimated that fighter pilots are much more successful after some number of actual combat encounters with the enemy even if their performance was superior in training and realistic combat simulation. It has been argued that the actual combat situation elicits different internal and external contextual cues when compared with simulated combat training exercises. (This is not meant to imply that the training is ineffective; but rather, that proper use of context may add to the effectiveness of training.)

These casual observations suggest to us that some benefit could be derived from practice in seemingly inconsistent situations if the context is consistently maintained. Of course, to establish the validity of our observations we must empirically evaluate our intuitions. The following experiment was designed to closely parallel our real-word observations and to test the ability of context to activate or bias automatic processing.

Current formal modeling of automatic/controlled processing also suggests the importance of the following experiment. The hybrid connectionist model developed by Schneider and his colleagues (see Schneider, 1985b; Schneider & Detweiler, 1988; Schneider & Mumme, 1987) suggests that, given sufficient context, performance in the context conditions should at least be better than in a VM condition. Also, given sufficient training, the context conditions may even closely approximate a pure CM condition in terms of detection performance. Briefly, controlled processing may bias performance for a given context but, within each context, the processing may become automatic. Unfortunately, the speed with which context effects will emerge has yet to be determined.

Overview of Experiment

In the present experiment we specifically examined the effects of context throughout training when context was modified every trial, every five trials, every ten trials, or every fifty trials. Actually, three independent context environments were cycled every one, five, ten, or fifty trials; hence, the ability to train context effects was examined.

The experiment we performed to test this issue was divided into two parts. First subjects were trained to detect categorically distinct words (taxonomically distinct by semantic category) in the background of another set of categorically distinct words. For example, subjects might search for COLOR words, with words naming ANIMALS as distractors, and then search for ANIMALS as targets, with words naming BUILDING PARTS, used as distractors. In the third context condition, subjects might search for words naming BUILDING PARTS with COLOR words used as distractors. Subjects also were trained in standard (pure) CM and VM conditions. Akso, the categories were distinctive and did not overlap semantically with one another. Subjects were assigned to one of four context cycle conditions: Context as described above was changed (i.e., cycled) every one, five, ten, or fifty trials.

After training, the specific combinations of target and distractor sets encountered by the subjects during training were modified (i.e., new target/distractor combinations were created). This transfer test was conducted to determine the subjects' ability to detect the trained stimuli in different target/distractor pairs. An important reason to be concerned about this type of transfer performance is to determine the degree of category-pair learning (i.e., the degree to which the relationships among the previously paired stimuli were learned).

We anticipate that the situational context provided by the "cycle" conditions will facilitate search performance to some degree. However, an answer to the question concerning the actual extent of performance facilitation that we might find does remain open. The nature of the functional relationship between number of context trials before cycling to a different context and performance improvement also remains an open question.

Method

Subjects. Thirty-two undergraduates, 17 males and 15 females, participated in this experiment. All subjects were tested for corrected or uncorrected visual acuity of 20/30 (far vision) and "20/40" (near vision). All subjects reported English as their native language.

Apparatus. All stimuli were presented using EPSON Equity I+ microcomputers with Epson MBM 2095-5 green monochrome monitors. The standard Epson Q-203A keyboard was altered such that the '7' '4', and '1' numeric keypad keys were labeled 'T', 'M', and 'B', respectively. The microcomputers were programmed with Psychological Software Tools' Microcomputer Experimental Language (MEL) to present and time the stimulus displays and to record response behaviors. During all experimental sessions, pink noise was played at approximately 55 db to help eliminate possibly distracting background noise. All subjects were tested in the same room at individual, partitioned workstations which were monitored by a laboratory assistant.

Stimuli. The semantically unrelated (Collen et al., 1975) categories of ANIMALS, VEGETABLES, UNITS OF TIME, COUNTRIES, BODY PARTS, WEAPONS, EARTH FORMATIONS, and CLOTHING were used as stimuli. Eight high-associate exemplars (Battig & Montague, 1969) were chosen from each category to serve as target and distractor stimuli. Each category contained eight words, four to seven letters in length.

Procedure. During the first session, subjects were administered an eye test, as well as the vocabulary and digit span subscales of the WAIS. They were then given an orientation session which consisted of three blocks of CM trials (50 trials per block). In these practice trials, the subjects searched for exemplars from categories that were not used in the actual experiment; that is, the categories COLORS and BIRDS. The purpose of the practice session was to orientate the subjects to the experimental procedures and to minimize the error rates before the subjects began the actual experiment.

An individual trial consisted of the following sequence of events. The subject was presented with the memory set of one category label, which he/she was allowed to study for a maximum of 20 seconds. The subject was instructed to press the space bar to initiate the trial. Three '+' signs positioned in a column were then presented for .5 second in the location of the display set (in the center of the screen) to allow the subject to localize his or her gaze. The display set consisted of three category words presented in a column and the subject's task was to indicate the location of the target (i.e., top, middle, or bottom) by pressing the corresponding key (labeled 'T', 'M', or 'B'). A target was present on every trial.

Training Sessions. There were five training conditions presented in the following order for each subject (the representation A(B), for example, refers to Target Set A displayed with Distractor Set B): Context 1 - A(B); Context 2 - B(C); Context 3 - C(A); CM - D(E); and VM - FGH(FGH). The specific ordering of the categories, A through H, was different for each subject and was counterbalanced by a Latin Square. For example, Context 1 [A(B)] for a particular subject might consist of FRUITS as targets with MUSICAL INSTRUMENTS as distractors. In the second condition, the distractors of Context 1 would now be the targets and there would be a new set of distractors; that is, MUSICAL INSTRUMENTS (ANIMALS). Context 3 would have ANIMALS as the target category and FRUITS as the distractors (the target category in Context 1). In the pure CM condition, the targets never appeared as distractors in any other part of the experiment and the distractors never appeared as targets; for example, FURNITURE (BODY PARTS). In the VM condition both the targets and the distractors were chosen from the same set of categories; e.g., WEAPONS, EARTH FORMATIONS, CLOTHING.

<u>Cycle Conditions</u>. Eight subjects were randomly assigned to one of four cycle conditions. The cycle was simply the number of trials that each search condition was presented in succession before cycling to the next search condition. Thus, the four cycle conditions differed in that each search condition was presented either one time, five times, ten times, or fifty times in succession. For example, subjects in the Cycle 1 condition (search condition changes every trial) would receive search condition A(B) on the first trial, then B(C) on the second trial, then C(A), then CM, and then VM. On the sixth trial the cycle would begin again (i.e., A(B), B(C)...). Subjects in the Cycle 5 condition would receive 5 trials of A(B), then 5 trials of B(C), followed by 5 trials of C(A), 5 trials of CM, and 5 trials of VM. Subjects in the Cycle 10 condition would receive 10 trials of A(B), then 10 trials of B(C), followed by 10 trials of C(A), 10 trials of CM, and 10 trials of VM. Subjects in the Cycle 50 condition would receive 50 trials of A(B), then 50 trials of B(C), 50 trials of C(A), 50 trials of CM, and then 50 trials of VM. (See Table 8 for a comparison of the cycle conditions.)

After each block of 50 trials (for all groups), subjects were encouraged to take a short break. All subjects received an equal number of trials (200) of each search condition within each session of the experiment.

<u>Table 8</u>. Progression of Cycle Training Conditions

Cycle Condition

CYCLE 1	CYCLE 5	CYCLE 10	CYCLE 50
$A(B)_1$ $B(C)_1$ $C(A)_1$ $D(E)_1$ $FGH(FGH)_1$	$A(B)_1$ . . . $A(B)_5$	$A(B)_1$ . . . $A(B)_{10}$	$A(B)_1$ . . . $A(B)_{50}$
Repeat After 5 Trials	$B(C)_1$ . . . $B(C)_5$	$B(C)_1$ . . . $B(C)_{10}$	$B(C)_1$ . . . $B(C)_{50}$
	$C(A)_1$ . . . $C(A)_5$	$C(A)_1$ . . . $C(A)_{10}$	$C(A)_1$ . . . $C(A)_{50}$
	$D(E)_1$ . . . $D(E)_5$	$D(E)_1$ . . . $D(E)_{10}$	$D(E)_1$ . . . $D(E)_{50}$
	$FGH(FGH)_1$ . . . $FGH(FGH)_5$	$FGH(FGH)_1$ . . . $FGH(FGH)_{10}$	$FGH(FGH)_1$ . . . $FGH(FGH)_{50}$
	Repeat After 25 Trials	Repeat After 1 Block	Repeat After 5 Blocks

Subjects trained for 11 sessions of 1,000 trials each session (20 blocks of 50 trials per block). All subjects completed a total of 11,000 trials (2,200 per condition).

Transfer Session. After training, the subjects were placed in the transfer phase of the experiment. At the beginning of the transfer phase, the subjects were informed that the conditions were going to change and that the categories would appear in different pairings. The testing procedure used in the transfer phase of the experiment was the same as the procedure used in the training phase. The transfer conditions were Context 1 Reversal - B(A); Context 2 Reversal - A(C); Context 3 Reversal - C(B); CM Reversal - E(D); and New CM - F(G), which was a control condition formed using stimuli from the VM sets of the training phase. Each subject completed a single session of transfer. There were a total of 1,000 trials in the transfer session (200 trials per transfer condition).

Performance Feedback. Subjects received the following performance feedback. After each correct trial, the subjects' reaction time (RT) was displayed in hundredths of a second. After each incorrect trial an error tone sounded and the correct response (the correct target word) was displayed for .8 second. Following each block of trials the subject was given his/her average RT and percent accuracy for that block. If a subject's accuracy fell below 90 percent the computer displayed a message which instructed him/her to respond more carefully. (Subjects were encouraged to maintain an accuracy rate of 95 percent or better while responding as quickly as possible.) Each day subjects were shown their performance for the previous session and encouraged to improve upon it.

Design. The within-subjects independent variables were (a) Training conditions: Context 1, Context 2, Context 3, CM, and VM; (b) Transfer conditions: Context 1 Reversal, Context 2 Reversal, Context 3 Reversal, CM Reversal, and New CM. The between-subjects independent variable was the Cycle condition--either 1, 5, 10, or 50 trials. The dependent variables were RT and accuracy.

Results

Each cycle condition was first analyzed separately to assess the benefits of the context conditions relative to CM and VM for each cycle time. Thus we will present a separate results section for each cycle condition. Following these results we will present the analyses that directly compare the cycle conditions with each other.

Results: Cycle Condition 50

Training data. Reaction time (RT) for correct trials from both the training (Sessions 1 to 11) and transfer (Session 12) phases of the experiment are shown in Figure 16 for the Cycle 50 condition. A Training Condition (Context 1, Context 2, Context 3, CM, VM) x Practice (Sessions 1 through 11) ANOVA was conducted. The main effects of Training Condition, $F(4,28) = 17.95$, $p < .0001$, and Practice, $F(10,70) = 30.29$, $p < .0001$, and the interaction between Training Condition and Practice, $F(40,280) = 1.97$, $p < .0009$, were statistically significant. Multiple comparisons were conducted among training conditions for performance at the end of training (i.e., final 200 trials per condition). The CM condition differed from VM, $F(1,28) = 74.32$, $p < .0001$, and the CM condition was significantly faster than all of the Context conditions ($F(1,28) = 13.37$, $p < .001$; $F(1,28) = 14.84$, $p < .0006$; and $F(1,28) = 21.10$, $p < .0001$, for comparisons with Context 1, Context 2, and Context 3, respectively). In addition, VM was significantly slower than all the Context conditions, $F(1,28) = 24.64$, $p < .0001$, $F(1,28) = 22.74$, $p < .0001$, $F(1,28) = 16.22$, $p < .0004$, for comparisons with Context 1, Context 2, and Context 3, respectively. None of the Context conditions differed significantly from each other.

An examination of the subjects' accuracy did not reveal trade-offs across conditions that would interfere with the interpretations of the reaction time data. Accuracy was 95

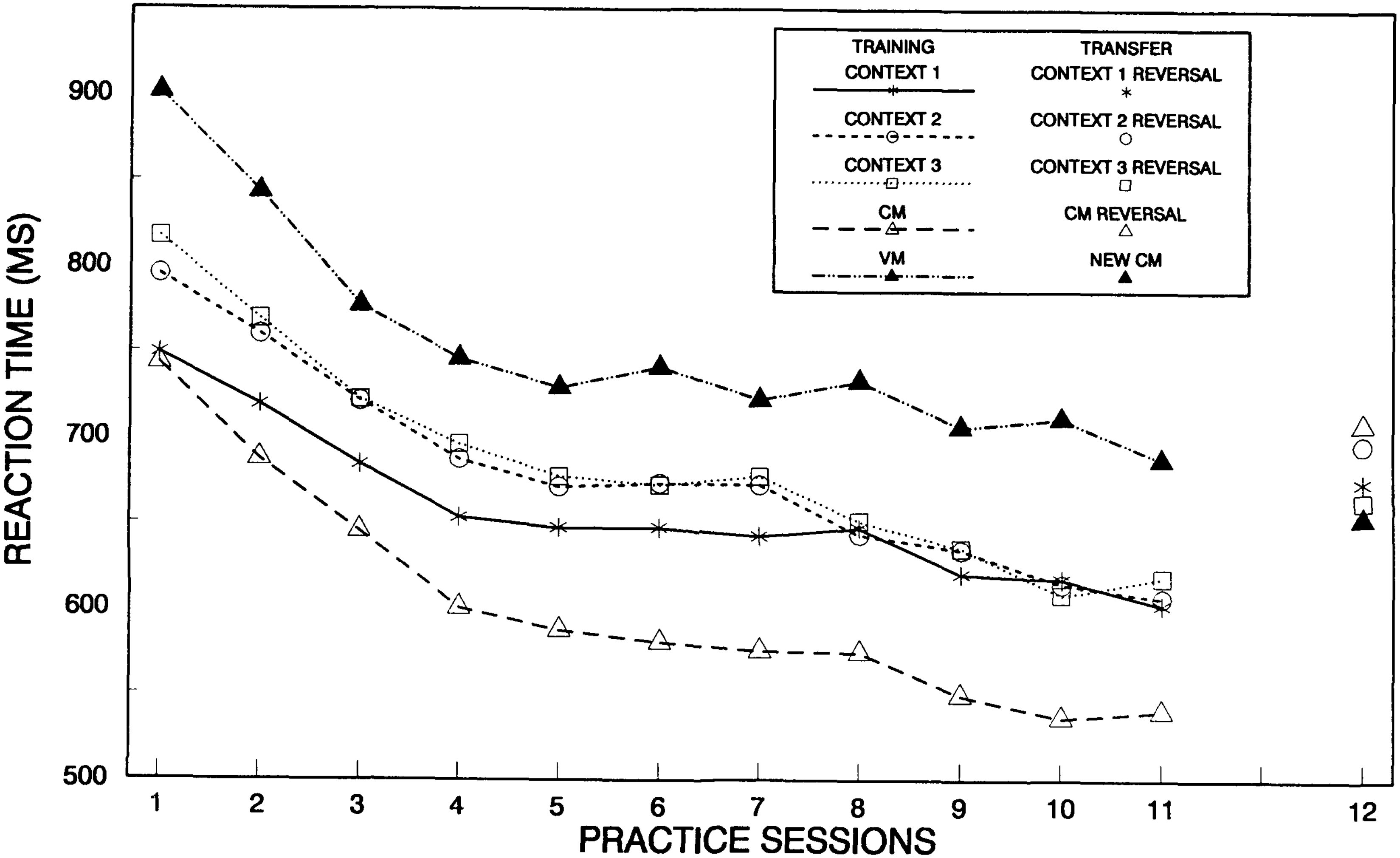


Figure 16. Reaction Time for Each Search Condition Plotted as a Function of Practice Session, for the Cycle 50 Condition.

percent in the CM condition, 92 percent in the VM condition, and 94 percent across all of the Context conditions.

Transfer data. A one-way ANOVA was conducted to test the effect of Transfer Condition (Context 1 Reversal, Context 2 Reversal, Context 3 Reversal, CM Reversal, New CM). There was not a significant effect for either RT, $F(4,28) = 1.44$, $p = .25$ or accuracy, $F(4,28) = 1.87$, $p = .14$. Thus, though there was a clear separation between the CM, Context, and VM performance at the end of training, there were no differences among conditions at transfer.

To test the effects of transferring subjects to the reversal conditions, separate comparisons were made between final level-training RT and transfer RT for each condition (the difference scores are presented in the first column of Table 9). The comparisons were significant for Context 1, $F(1,63) = 9.36$, $p < .0033$, and Context 2, $F(1,63) = 15.12$, $p < .0002$, and approached significance for Context 3, $F(1,63) = 3.71$, $p < .0586$. The difference between Training RT and Transfer RT for the CM condition was also significant, $F(1,63) = 52.02$, $p < .0001$. The new CM condition was not significantly faster than previous VM, $F(1,63) = 2.12$, $p = .15$.

Discussion: Cycle Condition 50

The training data from the Cycle 50 condition corresponded to our predictions: Performance in the Context conditions was superior to that in the VM condition but not as good as the CM condition. This result suggests that 50 trials were clearly sufficient to allow a temporary biasing of the salience of target and distractor items. It is important, however, that 50 trials were not sufficient to allow a "mimicking" of CM performance.

The transfer data suggest that there may be a greater amonut of learning than was apparent in the Fisk and Rogers (1988) experiment. Recall that they did not find significant reversal disruption effects for the context conditions.

Table 9. Effects of Transfer (Transfer RT - Training RT)[a]

	Cycle 50	Cycle 10	Cycle 5	Cycle 1
Context 1 Reversal	72	83	59	58
Context 2 Reversal	90	89	36	88
Context 3 Reversal	45	70	42	55
CM Reversal	168	199	161	193
New CM	-34	-5	0	-13

[a] A positive score denotes disruption in performance (i.e., an increase in RT) whereas a negative score indicates an improvement in performance (i.e., a decrease in RT). The scores are in ms.

Results: Cycle Condition 10

Training data. RT for correct trials from both the training (Sessions 1 to 11) and transfer (Session 12) phases of the experiment are shown in Figure 17 for the Cycle 10 condition. A Training Condition (Context 1, Context 2, Context 3, CM, VM) x Practice (Sessions 1 through 11) ANOVA revealed that the main effects of Training Condition, $F(4,28) = 9.69$, $p < .0001$, and Practice, $F(10,70) = 28.08$, $p < .0001$ were significant, as was the interaction between Training Condition and Practice, $F(40,280) = 1.69$, $p < .0008$. Multiple comparisons were conducted among training conditions for performance at the end of training (i.e., final 200 trials per condition). The CM condition differed from VM, $F(1,28) = 47.78$, $p < .0001$ and the CM condition was significantly faster than all of the Context conditions, $F(1,28) = 6.00$, $p < .0208$, $F(1,28) = 14.35$, $p < .0007$, and $F(1,28) = 14.13$, $p < .0008$, for comparisons with Context 1, Context 2, and Context 3, respectively. In addition, VM was significantly slower than all of the Context conditions, $F(1,28) = 19.91$, $p < .0001$, $F(1,28) = 9.76$, $p < .0041$, $F(1,28) = 9.95$, $p < .0038$, for comparisons with Context 1, Context 2, and Context 3, respectively. None of the Context conditions differed significantly from each other in terms of performance.

Accuracy was 96 percent in the CM condition, 94 percent in the VM condition and 94 percent across all the context conditions.

Transfer data. A one-way ANOVA conducted to test the RT effect of Transfer Condition (Context 1 Reversal, Context 2 Reversal, Context 3 Reversal, CM Reversal, New CM) yielded a significant effect of Transfer Condition, $F(4,28) = 3.09$, $p < .0316$. The New CM condition was significantly faster (73 ms) than in the CM Reversal $F(1,28) = 8.79$, $p < .0061$. The Context conditions did not differ from each other in terms of RT. A similar analysis conducted on the accuracy

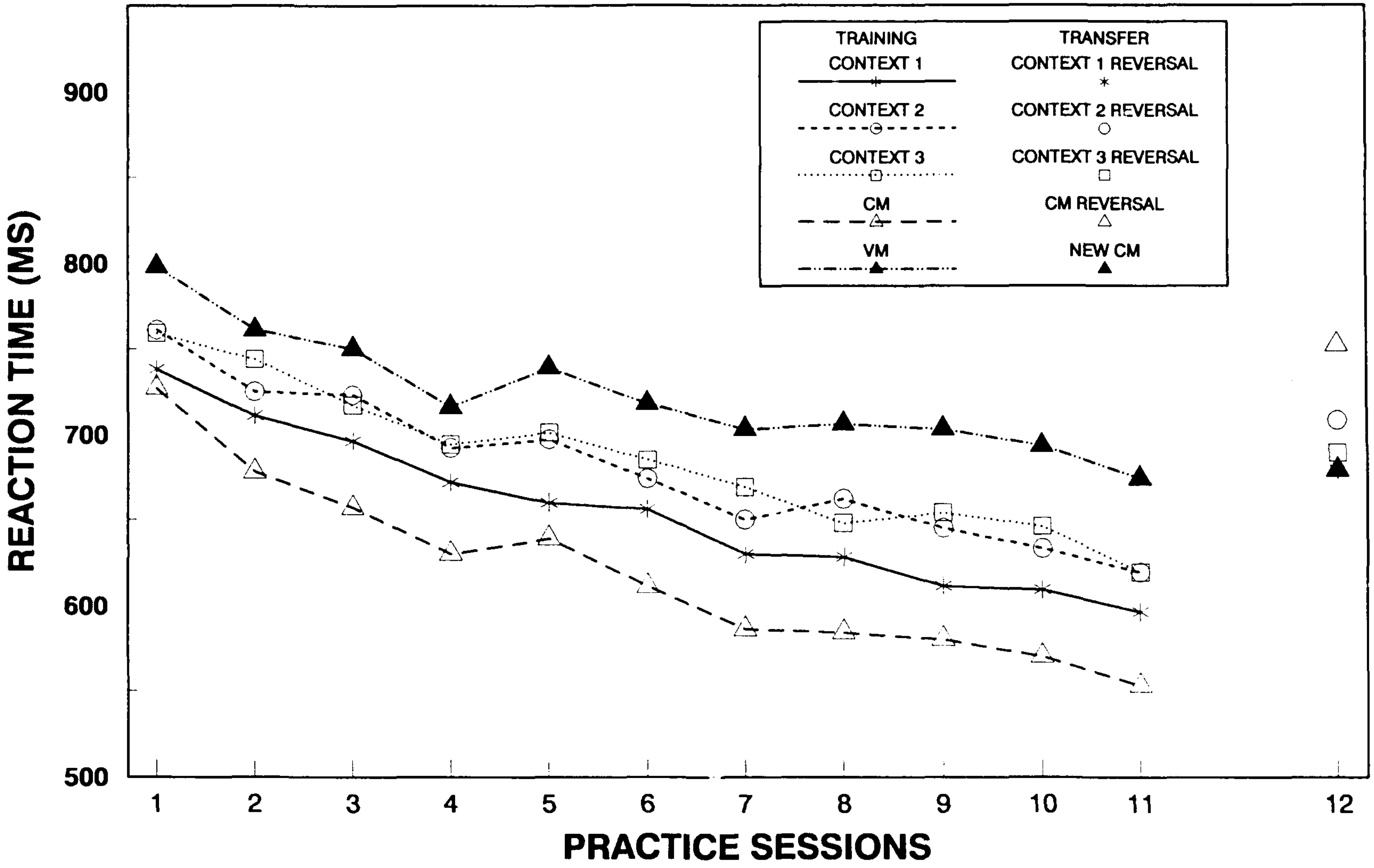


Figure 17. Reaction Time for Each Search Condition Plotted as a Function of Practice Session, for the Cycle 10 Condition.

rate yielded a non-significant effect, *F*(4,28) = 2.17, *p* = .0982.

To test the effects of transferring subjects to the reversal conditions separate comparisons were made between final level training RT and transfer RT for each condition (the difference scores are presented in the second column of Table 9). The comparisons were significant for Context 1, *F*(1,63) = 7.55, *p* < .0078, Context 2, *F*(1,63) = 8.6, *p* < .0047, and Context 3, *F*(1,63) = 5.42, *p* < .0232. The difference between Training RT and Transfer RT for the CM condition was also significant, *F*(1,63) = 42.89, *p* < .0001. The New CM condition was not significantly faster than the previous VM condition, *F*(1,63) < 1.

Discussion: Cycle Condition 10

The training data from the Cycle 10 condition correspond to our predictions: Performance in the Context conditions was superior to the VM condition but not as good as the CM condition. This result suggests that 10 trials were also sufficient to allow a temporary biasing of the salience of target and distractor items.

Results: Cycle Condition 5

Training data. RT for correct trials from both the training (Sessions 1 to 11) and transfer (Session 12) phases of the experiment are shown in Figure 18 for the Cycle 5 condition. A Training Condition (Context 1, Context 2, Context 3, CM, VM) x Practice (Sessions 1 through 11) ANOVA showed that the main effects of Training Condition, *F*(4,28) = 9.79, *p* < .0001, and Practice, *F*(10,70) = 10.25, *p* < .0001, and the interaction between Training Condition and Practice, *F*(40,280) = 1.59, *p* < .0177, were statistically significant. Multiple comparisons were conducted among training conditions for performance at the end of training (i.e., final 200 trials per condition). The CM condition differed from VM, *F*(1,28) = 55.12, *p* < .0001, and the CM condition was significantly faster than all the Context conditions, *F*(1,28) = 15.33, *p* < .0005, *F*(1,28) = 30.76, *p* <

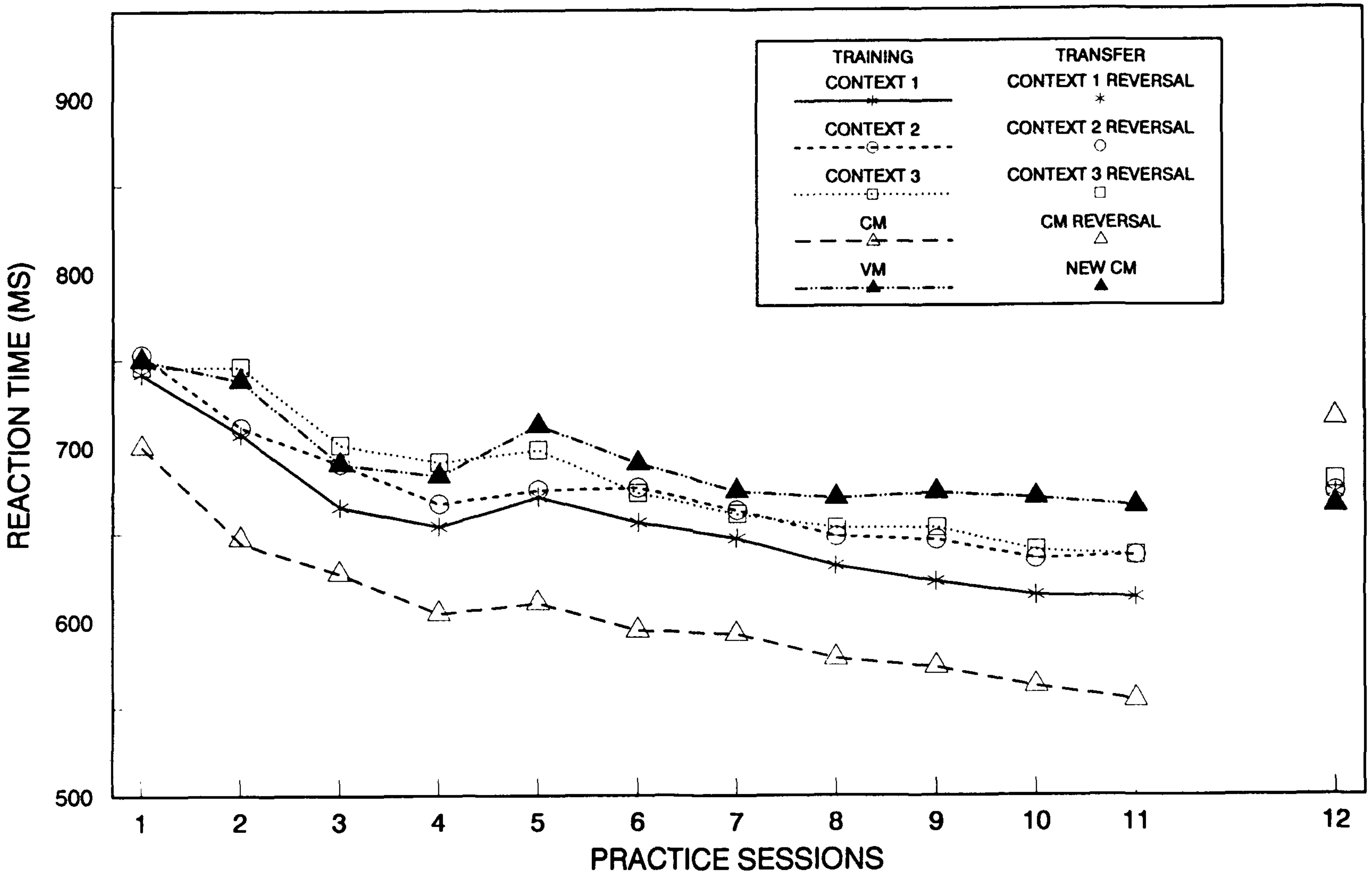


Figure 18. Reaction Time for Each Search Condition Plotted as a Function of Practice Session, for the Cycle 5 Condition.

.0001, and $F(1,28) = 31.04$, $p < .0001$, for comparisons with Context 1, Context 2, and Context 3, respectively. VM was significantly slower than only the Context 1 condition, $F(1,28) = 12.32$, $p < .0015$. None of the Context conditions differed significantly from each other in terms of RT.

Accuracy was 98 percent in the CM condition, 95 percent in the VM condition, and 95 percent across all of the Context conditions.

Transfer data. A one-way ANOVA conducted on the RT data to test the effect of Transfer Condition (Context 1 Reversal, Context 2 Reversal, Context 3 Reversal, CM Reversal, New CM) yielded a significant effect, $F(4,28) = 3.24$, $p < .0265$. At transfer the Context conditions did not differ among themselves and the New CM condition was not significantly different from any of the Context conditions. However, all conditions were significantly different from the CM Reversal, as shown by a Newman-Keuls comparison of RT means.

A similar analysis conducted on the accuracy data also yielded a significant effect, $F(4,28) = 2.84$, $p < .0428$. The New CM condition was statistically more accurate than the CM Reversal condition and Context Reversal 1, $F(1,28) = 5.14$, $p < .0313$, and $F(1,28) = 4.25$, $p < .0487$, respectively. Accuracy was 93 percent for the CM Reversal condition and 96 percent for the New CM condition. Accuracies were 96, 93, and 95 percent for the Context Reversals 1, 2, and 3, respectively.

To test the effects of transferring subjects to the reversal conditions, separate comparisons were made between final level training RT and transfer RT for each condition (the difference scores are presented in the third column of Table 9). The comparisons were significant for Context 1 only $F(1,63) = 5.54$, $p < .0217$. Context 2, $F(1,63) = 2.06$, $p = .156$, and Context 3, $F(1,63) = 2.90$, $p = .0936$ were not significantly affected by reversal. Reversing the CM target and distractors significantly slowed reaction time, $F(1,63)$

= 42.89, p < .0001. The mean RTs for New CM condition and the previous VM condition were equal; thus, there was obviously not a significant difference.

Discussion: Cycle Condition 5

The training data from the Cycle 5 condition correspond only partially to the results of the Cycle 50 and Cycle 10 conditions: Only performance in the Context 1 condition was superior to that in the VM condition. CM performance was faster than in all three of the Context conditions which did not significantly differ from each other. However, the fact that only the Context 1 condition was better than VM suggests that five trials may not be sufficient to allow salience-biasing of all targets and distractors when multiple context conditions are being trained. These results further suggest that there may be some benefit for the first context condition encountered in a series.

It is important to note that all subjects performed best in their "Context 1" condition (that is, the first context condition encountered). A strength interpretation of this finding (Schneider & Detweiler, 1987; Shiffrin & Czerwinski, 1988) would suggest that not only is a temporary biasing occurring but also target and distractor strengthening is occurring. With only five repetitions the gain produced by target detection for the first context condition is never overcome by the other conditions. This would be predicted if target learning is faster than distractor inhibition. Such a prediction is substantiated by simulation data (Schneider and Detweiler, 1987). Further experimentation is required to address this important issue.

Results: Cycle Condition 1

Training data. RT for correct trials from both the training (Sessions 1 to 11) and transfer (Session 12) phases of the experiment are shown in Figure 19 for the Cycle 1 condition. A Training Condition (Context 1, Context 2, Context 3, CM, VM) x Practice (Sessions 1 through 11) ANOVA was conducted. The main effects of Training Condition,

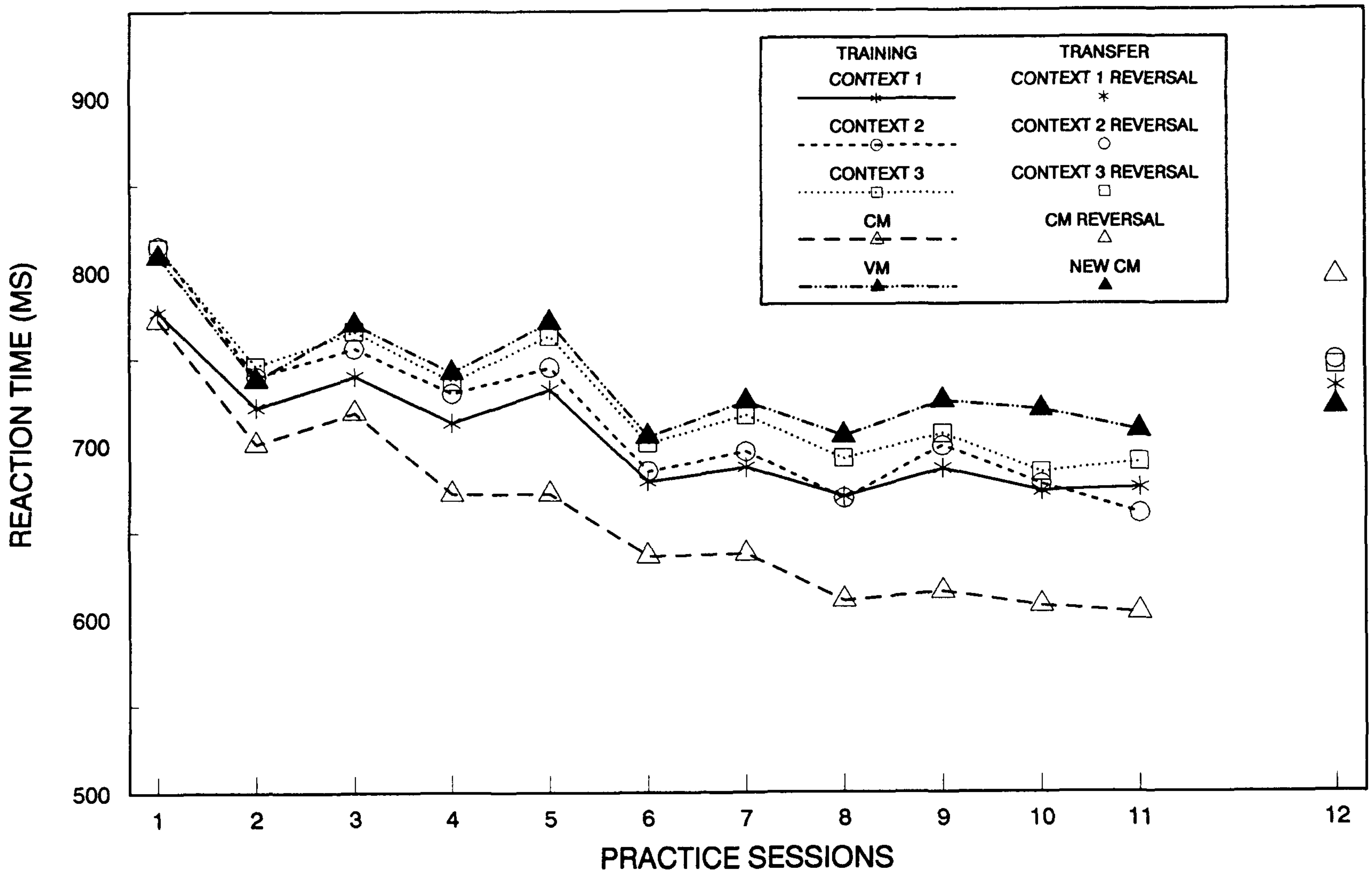


Figure 19. Reaction Time for Each Search Condition Plotted as a Function of Practice Session, for the Cycle 1 Condition.

$F(4,28) = 3.87$, $p < .0126$, and Practice, $F(10,70) = 11.32$, $p < .0001$ were significant as was the interaction between Training Condition and Practice, $F(40,280) = 2.22$, $p < .0001$. Multiple comparisons were conducted among training conditions for performance at the end of training (i.e., final 200 trials per condition). The CM condition differed from VM, $F(1,28) = 23.52$, $p < .0001$, and the CM condition was significantly faster than all of the Context conditions, $F(1,28) = 10.99$, $p < .0025$, $F(1,28) = 7.02$, $p < .0131$, and $F(1,28) = 16.21$, $p < .0004$, for comparisons with Context 1, Context 2, and Context 3, respectively. In addition, VM was significantly slower than only the Context 2 condition, $F(1,28) = 4.84$, $p < .0362$. None of the Context conditions differed significantly from each other in terms of performance.

Accuracy was 98 percent in the CM condition, 94 percent in the VM condition and 95 percent across all the context conditions.

Transfer data. A one-way ANOVA conducted to test the effect of Transfer Condition (Context 1 Reversal, Context 2 Reversal, Context 3 Reversal, CM Reversal, New CM) was significant, $F(4,28) = 3.00$, $p < .0353$. RT in the New CM condition was significantly faster than the CM Reversal condition $F(1,28) = 10.32$, $p < .0033$. At transfer the Context conditions did not differ among themselves and the New CM condition was not significantly different from any of the Context conditions. However, Context Reversals 1, 2 and 3 were all significantly different from the CM Reversal, $F(1,28)= 7.28$, $p < .0117$, $F(1,28) = 4.24$, $p < .049$, and $F(1,28) = 4.87$, $p < .0358$, respectively.

The main effect of Transfer condition was also significant for the accuracy scores, $F(4,28) = 4.58$, $p < .0057$. The New CM condition was statistically more accurate than the CM Reversal condition $F(1,28) = 12.98$, $p < .0012$, and the Context Reversals 1, 2, and 3 [$F(1,28) = 6.44$, $p < .0170$, $F(1,28) = 4.53$, $p < .0422$, and $F(1,28) = 14.18$, $p <$

.0008, respectively]. Accuracy was 92 percent for the CM Reversal condition, 97 percent for the New CM condition, 96, 95, and 98 percent for Context Reversals 1, 2, and 3; respectively.

To test the effects of transferring subjects to the reversal conditions, separate comparisons were made between final level training RT and transfer RT for each condition (the difference scores are presented in the last column of Table 9). The comparisons were significant for Context 1, $\underline{F}(1,63) = 5.88$, $\underline{p} < .0182$, Context 2, $\underline{F}(1,63) = 13.30$, $\underline{p} < .0005$, and Context 3, $\underline{F}(1,63) = 5.08$, $\underline{p} < .0278$. Reversing the CM target and distractors significantly slowed RT, $\underline{F}(1,63) = 64.73$, $\underline{p} < .0001$. The New CM condition was not significantly faster than the previous VM, $\underline{F}(1,63) < 1$.

Discussion: Cycle Condition 1

The Cycle 1 condition data present a qualitatively different pattern for the context conditions when compared with the other cycle conditions. Also, overall, all conditions except VM were slowed relative to the other cycle conditions (see below). The present data suggest that when context is cycled every trial the amount of exposure is insufficient for benefits to accrue. This finding is not surprising if one assumes that context does not immediately affect performance. A strength based interpretation also would predict the present findings. That is, with context cycling every trial, a stimulus category occurs as often as a target as it occurs as a distractor; hence, its strength is incremented and decremented across trials. Without repeated exposures as a target, a given context target set has no opportunity to accrue strength beyond that found normally for inconsistent or partially inconsistent conditions. In the Cycle 5 condition, there was an orderly relationship among the performance levels as a function of when in training a context condition was first encountered. However, in the present condition such an orderly effect was not present. Subjects' performance in the context conditions was not a function of context presentation order; hence, it seems that the differences between Context 2 and VM seems likely to be due to random variation and not a true effect.

Results: Cycle Comparisons

RTs for correct trials from both the training (Sessions 1 to 11) and transfer (Session 12) phases of the experiment are shown in Figure 20 for all four Cycle conditions. A Cycle Condition (cycles 1, 5, 10, and 50) x Search Condition (Context 1, Context 2, Context 3, CM, and VM) x Session (Sessions 1 through 11) ANOVA was conducted on the RT training data. The main effects of Search Condition, $F(4,112) = 35.2$, $p < .0001$, and Session, $F(10,280) = 67.71$, $p < .0001$, were significant. The two-way interactions of Session x Cycle Condition, $F(30,280) = 1.88$, $p < .0047$, and Session x Search Condition, $F(40,1120) = 3.7$, $p < .0001$, were also significant as was the third-order interaction Session x Search Condition x Cycle Condition, $F(120,1120) = 1.3$, $p < .0219$.

A comparison of the Cycle conditions, as presented in Figure 20, suggested that the differentiation between the context conditions and the CM and VM conditions occurred very early in training for the Cycle 50 and Cycle 10 conditions. However, this did not appear to be the case for the Cycle 5 and Cycle 1 conditions.

General Discussion

The present data are important from both a basic and applications-oriented perspective. In summary, the the following main findings can be derived from this experimental series.

First, all CM conditions improved to an asymptotic performance level superior to any context or VM performance level regardless of cycle condition. However, CM performance in the Cycle 1 condition was slower (although nonsignificant) than CM performance in any other cycle condition.

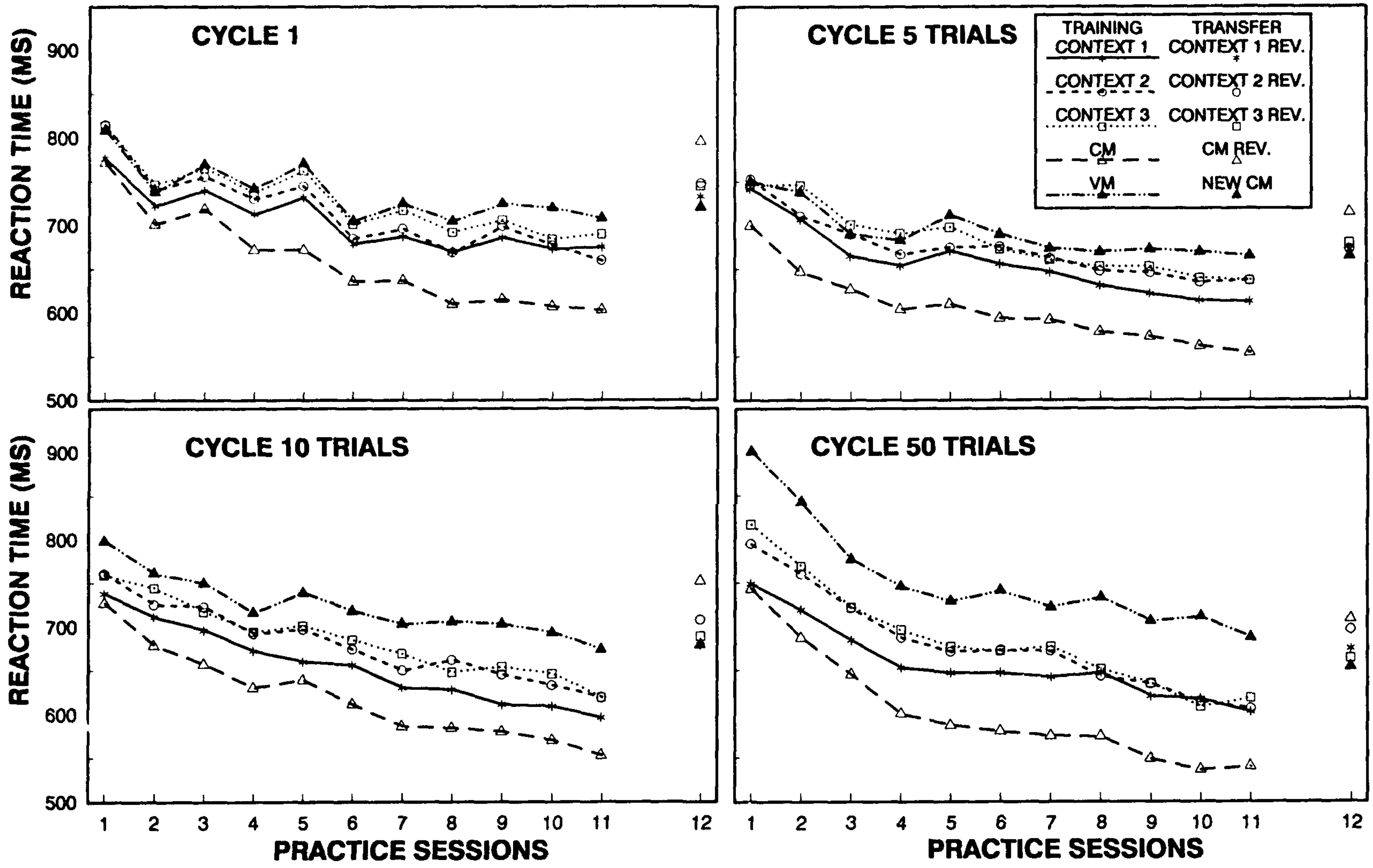


Figure 20. Reaction Time for Each Search Condition, Plotted as a Function of Practice Session, for Each Cycle Condition.

Second, although VM asymptotic performance was the same across cycle conditions, the amount of VM improvement was a direct function of cycle condition. We found no performance improvement for VM in the Cycle 1 condition, minimal improvement in Cycle 5, moderate improvement in Cycle 10, and considerable (relatively speaking) improvement in Cycle 50. This finding has never been documented before and is important for at least two reasons: (a) It may allow an understanding of why VM performance improvement is seen in some experiments and not others; and (b) it suggests that the amount of improvement is not due to stimulus related factors in VM training.

Third, the context effect seems to be dependent on how the context is cycled. The differentiation between context conditions and VM is related to the cycle condition, with context in the Cycle 50 condition showing the strongest and earliest differentiation from VM. Context performance in the Cycle 1 condition is the least differentiated from VM.

These data suggest that when the training developer can isolate pure CM components for training, then factors such as how the training is cycled with other conditions is of less importance than when they are training a less than totally consistent condition. When conditions are less than totally consistent, how the training is packaged may be crucial for predicting performance as a function of practice.

VII. EXPERIMENTAL SERIES 6: LEARNING AND PERFORMANCE RETENTION IN A HIGH-PERFORMANCE-SKILL-BASED, PROBLEM-SOLVING TASK

Introduction

The purpose of this section is to describe a complex, battle management analog task developed to facilitate further investigation of real-world application of automatic/controlled processing principles. The present task was designed as a test-bed for issues of training design, component information coordination, effects of part-whole task sequencing, complex performance under speed stress, retention of component/whole task (as a function of type of training), etc. However, to use the task to accomplish these goals, task performance must demonstrate characteristics of high-performance skill in both acquisition and asymptotic performance. Hence, the major purpose of the present investigation was to document the validity of our task as a true high-performance-skills-dependent task.

The present two experiments involve examining characteristics of subjects' performance in a relatively complex "strategic planning" task. Through pilot testing we have developed what will be referred to as a "dispatching" task. This task was chosen because it allows manipulation and examination of important information-processing components found in most complex tasks (e.g., see Fisk et al., 1987; Kyllonen & Woltz, 1989; Salthouse & Somberg, 1982). The information processing components that are assessed are (a) visual search, (b) memory scanning, (c) working memory (and effect of varying memory loads), (d) decision making, and (e) response selection/execution. This present class of tasks provides a rich converging data set for the understanding of automatic and controlled processing from a training and retention perspective.

The present task is conceptually similar to tasks performed by a Fighter Duty Officer. The task has several procedural components, requires learning a substantial amount of declarative knowledge, and is very heavily rule-based. In addition, the task has both memory and visual search components. Although the task is conceptually simple, the subject must choose the optimum "driver" for a given "delivery," and the subject must learn rules associated with how to determine load level, load type, and delivery location characteristics. In addition, the subject must learn to associate 27 drivers to given "license classes" (license classification determines who can carry out the mission). The software is set up to allow varying degrees of access to help screens and feedback. Our expert system scenario generator, allows the generation of scenarios with varying degrees of consistency and allows for the "loss" of personnel, equipment, and delivery locations. For follow-on experiments we also have the capability to time-stress the decision-making process, as well as add subsidiary tasks to increase mental workload. The participants serve as "dispatchers" and, for each trial, they receive an "order" for a specific amount of a specific cargo to be delivered by a specific vehicle. A visual display of the name of the desired cargo, its weight, the cargo's destination, and the required vehicle is presented in a two-by-two matrix in the center of the computer display. Based on the order, the subject's task is to first determine the range of possible operators whose license qualifies him/her to deliver the cargo and retain those names in memory. The subject then presses the space bar and is presented with four driver names (displayed in a two-by-two matrix) and must quickly determine which driver is the optimum driver. For this aspect of the task, accuracy and response time are the dependent variables.

Subjects have access to extensive help screens via single key presses. The help screens provide all needed declarative information and rule-based knowledge needed to perform the task and can serve as external memory aids. All keystrokes are stored such that how the subject traverses through help is recorded. In addition, time spent in each help screen is recorded. This serves as an on-line nonintrusive "protocol" analysis.

The present task requires memory scanning (subjects must hold a derived list of potential drivers in memory) and across trials the number of potential drivers (hence memory load) is manipulated, allowing data converging with the standard memory search studies. Subjects must learn rules associated with performing the task; hence, rule-based learning (necessary for most complex-skill-based tasks) can be assessed. Subjects must decide when and how to optimally access help screens (a decision component) and also scan a display to locate the optimum driver (corresponding to standard visual search tasks). The task component selection is based on an information-processing task-analytic methodology developed to isolate trainable information-processing components across a range of real-world complex tasks.

To examine the important issue of skill retention, subjects were called back 6 weeks subsequent to final practice for evaluation of performance retention. (Six weeks was chosen based on retention characteristics across 1-year retention intervals; see Appendix A). We examined retention from a global task performance perspective and determined what components deteriorated with disuse over the retention interval.

Experiment 1 - Method

Subjects. Five undergraduates, four males and one female, from the Georgia Institute of Technology served as subjects and received $5.00 per hour for their participation. Four subjects had participated in a previous experiment conducted in the Human Attention and Performance Laboratory; the other subject was a senior psychology major but had not participated in any previous experiments in the Human Attention and Performance Laboratory. Subjects were tested for both near and far visual acuity and had at least 20/30 and 20/40 vision, respectively.

Experimental Task. The experimental task was a dispatching task presented by microcomputer, in which subjects served as dispatchers. The dispatcher received an order, initiating a trial, for a specific amount (in kilograms) of a particular cargo to be delivered by a specific vehicle. Each order was initiated when the subject pressed the space bar. A visual display of the name of the cargo to be delivered, the cargo's weight, the name of its destination, and the vehicle to be used for delivery was presented in a two by two matrix in the center of the computer screen. The dispatcher's task was to determine the potential range of operators whose licenses would qualify them to deliver the cargo, based on the particular order.

Stimuli. The stimuli which comprised the basic elements of the experimental task belong to six categories (a) cargo, (b) weight, (c) destination, (d) distance, (e) vehicle, and (f) operator license. The design of the experimental task determined these categories. We chose to use the metric system (kilograms and kilometers) to describe the weights and distances used in the task.

The names associated with each of the categories were derived using two different techniques. The first technique was employed in constructing the "vehicle" category. Using this method, a system was constructed in which vehicle names were definable along one or more dimensions, according to a set of logical rules. A thorough understanding of the rule(s) was required in order to learn the vehicle names. Learning the arbitrary assignment of a specific name to the category to which it belonged (i.e., rote memorization) was not required by use of this technique.

The second technique was employed in constructing the destination and operator license categories. Using this method, both the operator names associated with each license category and the company names associated with each destination category were assigned in a wholly arbitrary manner, without reliance on an underlying set of logical rules. Thus, learning the operator and company names required the rote memorization of specific names, along with their associated categories.

Selection of the company names was based on perusal of the yellow pages of the Atlanta metropolitan area phone directory, from which names of actual businesses were drawn. Our principal goal was to minimize any prior associations; that is, destination names were selected so as not to imply any particular enterprise or company with which subjects might be familiar. Thus, the selection criterion was that the names of the businesses had to be nondescript. After selection of a name from the phone directory, the name was modified by changing its "suffix" to one of the following: Co., Inc., Corp., Ltd., Assoc., Industries, Products, Enterprises, Systems, or Technology. The result was a generic, all-purpose business name (e.g., Ajax Inc.).

To select the names of human operators to be associated with different license types, the Battig and Montague category norms (1969) were employed. Again, selection criteria were based on an effort to minimize subjects' prior associations or familiarity with operator names (e.g., "my best friend, Tom"; "my mother, Alice"). First, we constructed a list of names that were rated lowest in prototypicality, were a maximum of seven letters in length, and were visually distinct was constructed. Four graduate students in psychology, to whom the experimental task was described, were asked to eliminate any names they considered unusual, confusable, or unisex.

Because the different subcategories of cargo to be used in the experimental task were, arguably, natural subcategories (general purpose, liquid, hazardous), we selected cargo names that would be easily, if not naturally, associated with each subcategory. All categories, subcategories, and exemplars are listed within the description of the task presented in Appendix E.

Equipment. Epson Equity I+ microcomputers equipped with Epson MBM-2095 monochrome monitors (green phosphor, 50-Hz refresh rate) and Epson multimode graphics adapters were used to present the task. The microcomputers were programmed with Turbo Pascal version 5.0 to generate files containing task "orders" (see below), present the experimental task, record response behavior, and perform descriptive data analysis. A Heath model AD-1309 white/pink noise generator was used to generate pink noise, which was fed into a Realistic model SA-150 integrated stereo amplifier and output through speakers at a sound level of approximately 55dB A. In this manner external sounds were masked.

Procedure and Design. The procedure for the training phase was as follows. Upon their arrival, subjects were given extensive written instructions for performing the task. These instructions are included in Appendix E. After the subjects read the instructions, the experimenter explained that he would remain in the room with the subject and would ask questions regarding task behavior, as well as answer questions.

Subjects were given a form on which to record their own response latency and accuracy performance by block, across each session. They were also given pen and paper to record any comments they might have. These comments are included in Appendix F. Also, periodically, subjects were asked to record their strategies for performing the task. When subjects were finished reading the instructions, the experimenter removed the instructions. However, they were allowed to review the instructions between blocks and at the end of the session; all did so during the first session. All subjects also reviewed the instructions prior to the beginning of Session 2. Prior to the beginning of Session 3 only two subjects examined the instructions briefly.

The experiment was divided into discrete trials, blocks, and sessions. There were a total of 10 Sessions. Sessions 1 through 4 contained two blocks; Sessions 5 through 9 contained three blocks; and Session 10 contained four blocks. Thus, there were 27 blocks. Also, there were 36 trials per block, for a total of 972 trials. Each trial represents an "order." As described previously, a software program generated the files containing these orders. The sequence of presentation was random and an identical sequence was used for all subjects.

As described previously, the dispatcher's task was to select the range of all possible operators qualified (i.e., licensed appropriately) to deliver a particular type of cargo. Extensive help (in the form of text screens describing cargos, vehicles, and destination points, along with the different license types associated with operators) was provided to assist subjects in selecting the operators. The help menu was accessed by pressing the 'H' key and selecting the desired help. Help was available only while the subject was studying the order. When the subject was ready to proceed to the screen which contained the names of the available operators, he or she could no longer access help.

When the subject was finished studying information pertaining to the order, he or she pressed the space bar; orientation points (four '+' symbols arranged in a two-by-two matrix with the 'o' symbol centered horizontally and vertically between the '+' symbols) then were displayed for 500 ms. Immediately following the display of the orientation points, four names were displayed in the same two by two matrix. All names were operator names. The dispatcher's task was to select the operator who had the lowest or minimal level of license but was still qualified to deliver the cargo. Thus, there were trials in which more than one operator was qualified to deliver the cargo. There was always at least one qualified operator, but never more than one "optimal" operator. Subjects selected their choice by pressing the '7', '9', '1', or '3' keys of the numeric keypad. These keys represented the top left, top right, bottom left, and bottom right corners of the two-by-two matrix and were labeled 'TL', 'TR', 'BL', and 'BR', respectively.

On correct trials, subjects received feedback informing them that their choice was correct. On incorrect trials, they were told that their choice was incorrect and given the name of the correct operator. At the end of each block, subjects were given their mean response time in milliseconds and their accuracy in terms of percentage of correct responses.

Data Collection. All keystrokes were captured and stored by the computer program. Hence, a complete record of each subject's use of help was recorded. Also, the time between each keystroke was stored such that it was possible to determine the amount of time spent in each help screen, in the study screen, etc. Finally, each subject's decision accuracy (accuracy for choosing the optimal operator in the decision screen), as well as the decision latency on each trial, was recorded (see Appendix G for a more detailed account of data collection).

Experiment 1 - Training Results

All indices of task performance improved dramatically across the 27 blocks of training. For group data, accuracy increased and total study time (time studying the work order screen plus time in help screens), study time (time studying work orders), and help time decreased according to a typical power function:

$$y = ax^b \quad \text{where}$$

'y' represents the index of performance (e.g., percent correct), 'a' represents performance at Block 1, 'x' represents the block number, and 'b' represents the rate of improvement. Most individual data correspond also to this power function. An additional indication of the development of proficiency was the reduction in variance of the various indices of performance across blocks, reflected in standard deviations.

Decision Latency. Improvement in mean decision latency did not follow the power function typical of most training situations. However, the reader is reminded that the scenarios were generated using a random process; therefore, level of difficulty varied across blocks. Mean decision latencies declined from 8.16 seconds (sec) at Block 1 to 2.99 sec at Block 27, with standard deviations of 6.99 sec and 2.16 sec, respectively. Unless specified otherwise, times reported are for all trials. Error trial times tended to be slower. Table 10 presents decision latency as a function of block number.

Accuracy. Accuracy performance improved in a manner more typical of training situations. Mean accuracy rose from 67.22 percent correct at Block 1 to 98.89 percent correct at Block 27, with standard deviations of 12.33 percent and 1.52 percent, respectively (see Table 11). The accuracy data are represented by the following equation:

$$y = 69.66x^{0.107}$$

This fit accounts for 90.4 percent of the variance.

Total Study Time. Initially, participants spent a great deal of time examining all available help information. As described previously, total study time consists of study time and help time. Mean total study time declined from 70.15 sec at Block 1 to 2.92 sec at Block 27, with standard deviations of 69.05 sec and 2.18 sec, respectively. Table 12 presents total study time as a function of block number. Mean total study time is represented by the following equation:

$$y = 63.963x^{-0.930}$$

This fit accounts for 98.1 percent of the variance.

Study Time. Mean study time declined from 18.92 sec at Block 1 to 2.73 sec at Block 2,7 with standard deviations of 18.37 sec and 1.94 sec, respectively. Table 13 presents

Table 10. Decision Latency (Seconds) as a Function of Block

Block	Mean	SD
1	8.16	6.99
2	6.19	5.74
3	7.30	6.47
4	6.71	6.30
5	6.39	6.04
6	6.60	6.91
7	5.44	5.69
8	5.38	4.34
9	5.42	5.70
10	5.60	6.56
11	4.65	4.82
12	6.08	6.26
13	5.95	5.83
14	4.19	3.96
15	4.87	4.78
16	3.86	3.85
17	3.94	3.12
18	3.52	3.15
19	5.29	4.45
20	3.92	2.84
21	3.90	3.17
22	3.27	2.15
23	4.10	3.11
24	3.48	3.08
25	4.06	3.99
26	4.64	3.56
27	2.99	2.16

Table 11. Percent Correct as a Function of Block

Block	Mean	SD
1	67.22	12.33
2	73.89	18.07
3	80.55	09.21
4	81.67	12.51
5	79.17	15.13
6	90.00	10.13
7	85.56	09.09
8	90.00	08.91
9	90.55	09.55
10	89.44	11.52
11	92.78	07.24
12	92.78	07.50
13	91.66	06.80
14	86.09	09.86
15	88.90	09.62
16	93.89	05.69
17	95.00	03.04
18	96.11	03.17
19	90.56	09.94
20	96.66	03.05
21	97.22	03.40
22	96.11	02.49
23	98.33	02.49
24	97.22	03.93
25	97.22	03.40
26	98.89	01.52
27	98.89	01.52

Table 12. Total Study Time (Second) as a Function of Block

Block	Mean	SD
1	70.15	69.05
2	29.99	22.06
3	28.48	29.04
4	16.75	15.48
5	15.03	13.28
6	10.90	09.49
7	10.49	09.34
8	07.97	08.23
9	08.79	13.57
10	06.84	07.60
11	05.74	05.39
12	07.15	07.63
13	06.48	07.14
14	04.99	04.43
15	05.06	04.44
16	04.28	03.58
17	04.01	03.94
18	04.76	04.44
19	04.53	04.08
20	03.59	03.30
21	04.14	03.92
22	03.23	02.84
23	03.86	03.18
24	03.70	02.91
25	03.70	03.08
26	03.11	02.55
27	02.92	02.18

Table 13. Study Time (Second) as a Function of Block

Block	Mean	SD
1	18.92	18.37
2	10.08	07.68
3	09.65	08.96
4	07.70	05.35
5	06.25	03.86
6	06.50	04.25
7	06.47	04.32
8	05.49	04.57
9	04.88	04.00
10	05.47	05.27
11	04.67	03.94
12	05.27	04.79
13	05.20	04.62
14	04.57	04.06
15	04.31	03.16
16	03.71	02.43
17	03.82	03.76
18	04.12	03.50
19	04.06	03.04
20	03.42	02.93
21	03.64	03.11
22	02.97	02.37
23	03.60	02.72
24	03.39	02.53
25	03.45	02.74
26	02.95	02.40
27	02.73	01.94

study time as a function of block number. Mean study time is represented by the following equation:

$$y = 16.378x^{-0.506}$$

This fit accounts for 95.8 percent of the variance.

Help Time. Mean help times declined from 51.24 sec at Block 1 to 0.18 sec at Block 27, with standard deviations of 54.92 sec and 0.62 sec, respectively. Table 14 presents help time as a function of block number. Mean help time is represented by the following equation:

$$y = 108.753x^{-1.886}$$

This fit accounts for 92.9 percent of the variance.

Help times were partitioned further into the mean time spent in each individual screen. Initially, participants engaged in general exploratory behavior, examining all available help screens. Quickly, however, they reduced their help needs to four screens: weight information, license categories information, destination names, and operator names. By Block 6 these were further reduced to destination and operator names. Finally, by Block 20, access to any help screen was trivial.

At Block 1, mean time spent studying the operator names was 17.84 sec, with a standard deviation of 20.96 sec. By Block 10, access of this help screen was modest (M=0.25 sec and SD=1.02 sec) and by Block 20 had all but disappeared (M=0.10 sec and SD=0.13 sec). Table 15 presents mean time spent studying operator names as a function of block number. At Block 1, mean time spent studying destination names was 6.70 sec, with a standard deviation of 7.88 sec. By Block 14, access of this help screen was modest (M=0.27 sec SD=0.95 sec) and by Block 20 was negligible (M=0.08 sec and SD=0.36 sec). Table 16 presents mean time spent studying destination names as a function of block number.

Experiment 1 - Discussion

In this experiment we examined skill acquisition in a cognitive task. The task was designed such that we could

Table 14. Help Time (Second) as a Function of Block

Block	Mean	SD
1	51.24	54.92
2	19.91	18.08
3	18.83	24.75
4	09.05	12.77
5	08.78	11.16
6	04.40	07.61
7	04.02	07.54
8	02.48	05.29
9	03.91	11.66
10	01.38	03.57
11	01.06	02.34
12	01.88	04.76
13	01.28	04.33
14	00.42	01.35
15	00.74	02.16
16	00.57	02.11
17	00.19	00.83
18	00.64	01.73
19	00.47	01.78
20	00.17	00.75
21	00.50	01.36
22	00.26	00.96
23	00.26	00.87
24	00.31	00.94
25	00.25	00.88
26	00.17	00.59
27	00.18	00.62

Table 15. Operator Names Screen Time (Second) as a Function of Block

Block	Mean	SD
1	17.84	20.96
2	07.90	10.69
3	09.79	18.99
4	05.12	09.95
5	04.26	07.95
6	01.85	05.29
7	01.51	04.53
8	00.72	02.00
9	02.35	10.19
10	00.25	01.02
11	00.08	00.47
12	00.54	03.33
13	00.20	01.68
14	00.01	00.15
15	00.15	01.27
16	00.16	01.20
17	00.05	00.44
18	00.04	00.28
19	00.09	00.92
20	00.01	00.13
21	00.02	00.17
22	00.00	00.00
23	00.01	00.12
24	00.00	00.00
25	00.02	00.25
26	00.00	00.00
27	00.00	00.00

Table 16. Destination Names Screen Time (Second) as a Function of Block

Block	Mean	SD
1	6.70	7.88
2	3.74	4.77
3	3.16	5.53
4	1.15	2.05
5	1.74	2.67
6	0.92	2.00
7	0.99	2.12
8	0.54	1.72
9	0.63	1.57
10	0.52	1.91
11	0.55	1.38
12	0.57	1.34
13	0.72	2.31
14	0.27	0.95
15	0.31	0.81
16	0.23	0.74
17	0.06	0.29
18	0.37	1.19
19	0.24	0.95
20	0.08	0.36
21	0.26	0.78
22	0.15	0.57
23	0.14	0.50
24	0.11	0.44
25	0.05	0.25
26	0.06	0.40
27	0.07	0.28

partition it into different components. Thus, we could examine performance on each of these components. Also, as in the "real world," these different components had varying degrees of consistency. The rules governing the task were consistent. The names associated with the different categories (e.g., operators and destinations) remained consistent throughout the experiment. On the other hand, the target and distractor names were inconsistent. The same order scenario could be present on multiple trials, yet the operator who was the target on trial x might well have been one of the distractors on trial y and one of the distractors on trial x could turn out to be the target on trial y.

To become proficient at the task subjects had to sufficiently encode a moderately large set of associations such that they could be readily retrieved (e.g., operator names with license categories). Also, they not only had to understand the rules, but had to refine their understanding of the rules and the situations under which those rules applied. For example, to determine the minimum level operator license required to perform the task requires two pieces of information: the vehicle type (the license is associated with the vehicle, not with cargo or weight per se) and the distance class of the destination. Three subjects did get stuck on this problem. Early in the experiment, they mentioned to the experimenter that they could not understand some of their errors; they believed there was a bug in the computer program. When this occurred the experimenter then went over one of the suspect trials step-by-step with each subject and explained why the target was the optimal choice. Subjects stated that they had been ignoring the information provided by the vehicle. Finally, both rules and names had to be internalized to attain maximum performance. Operationally, we assume that this has occurred when no help is accessed/used.

We found that in the consistent components of the task, performance improvement followed a power function which had been previously referred to as the ubiquitous "law" of skill acquisition (Newell &Rosenbloom, 1981). This was seen in help time, study time, and total study time. Across training, we also found considerable reduction in variance in these performance measures. Both of these results are consistent with previous findings in the automaticity literature (cf. Kanfer &Ackerman, 1989). We also found that both decision latency and its variance declined across performance. However, improvement followed no predictable pattern. As mentioned previously, this component of the task was inconsistent and this finding is typical of performance in variably mapped conditions found in the automaticity literature. It is interesting to recall that improvement in accuracy is not described well by a power function. At first this might seem inconsistent with what we said above about improvement in decision latency. However, these results are not at odds with findings in the literature. We believe that this pattern of improvement is related to performance in the consistent components of the task. Early in training, subjects are still learning rules and associations. Understanding and memory are imperfect. This is reflected in poor accuracy scores. As their understanding and memory improve, so does their accuracy. Furthermore, in variably mapped tasks subjects are able to maintain high levels of accuracy, albeit at a cost in reaction time.

Experiment 2 - Method

Subjects. Four subjects who completed Experiment 1 completed the retention experiment as well. Subjects were not informed of the retention phase during the training study. Instead, they were contacted approximately 55 days following the final day of the training experiment and asked if they would be willing to return for a second study. These four subjects agreed to return; the other had graduated in the interim. The retention experiment was initiated 60 days following the final session of the training phase.

Experimental Task. The design of the experimental task was identical to that in Experiment 1.

Stimuli. A subset of the stimuli (Blocks 1 through 20) used in Experiment 1 was used. Furthermore, the order of presentation was identical to that used in Experiment 1. Consequently performance on, for example, Block 7 Experiment 1 could be compared directly with performance on Block 7 Experiment 2.

Procedure and Design. The procedure was almost identical to that in the training experiment. In this experiment subjects received more blocks per session: four each in the first and second sessions and six each in the third and fourth. Thus, subjects received 20 blocks, for a total of 720 trials.

Experiment 2 - Retention Results

As mentioned previously, the stimuli used in Experiments 1 and 2 are isomorphic. Subjects retained a substantial ability to perform the task; in some cases they were operating at over 80 percent of the end-of-training levels. By Block 7 (less than 2 hours of practice), all indices of performance indicate that subjects were operating at end-of-training levels. To facilitate understanding performance relative to a given point in training, another measure was added--mean performance savings. For decision latency, accuracy, total study time, and study time, a savings score was computed for each subject (with the averaged savings scores reported). This score was derived for each retention block (retention Block 1 through 20) relative to each training block. That is, the savings score for Block 1 at retention is relative to performance during training on Block 1. Similarly, the savings score on Block 20 at retention is calculated based on Block 20 retention performance relative to Blcck 20 training performance. For a given block of trials, each savings score was calculated by subtracting a given Experiment 2 performance score from the corresponding Experiment 1 performance score and dividing it by the Experiment 1 score. In the case of the accuracy measure, the Experiment 1 measure was subtracted from the Experiment 2 measure, thus maintaining a positive value.

In general, participants continued to show improvement according to all indices of performance. Both mean total study time and study time improved according to the power function described in Experiment 1. In several cases, such as accuracy and access of help screens, performance had reached ceiling.

Decision Latency. Mean decision latency declined from 5.89 seconds (sec) at Block 1 to 3.15 sec at Block 20, with standard deviations of 5.66 sec and 2.73 sec, respectively. Unless specified otherwise, times reported are for all trials. Error trial times tended to be slower. Tables 17 and 18 present mean decision latency and mean proportion of savings in decision latency as a function of block number, respectively. Decision latency savings were considerable, ranging from a mean of 0.390 (SD = .062) at Block 3 to a mean of 0.205 (SD = 0.036) at Block 14. Table 17 and Table 18 taken together show that performance improves over the retention interval and that retention performance as measured by decision latency was always significantly better (for equivalent training block number) than training performance. The fact that retention performance was always better than training performance (for equivalent block numbers) can be seen in Table 18 because, although the saving scores decreased with practice, they were always positive and significantly greater than zero.

Accuracy. At Block 1, mean accuracy was 81.95 percent with a standard deviation of 3.58. By Block 5, mean accuracy was 95.14 percent, with a standard deviation of 2.66. From this point on, accuracy remained above 90 percent correct. Tables 19 and 20 present mean accuracy and

Table 17. Decision Latency (Second) as a Function of Block

Block	Mean	SD
1	5.89	5.66
2	4.91	5.05
3	4.46	3.73
4	4.48	4.38
5	4.28	3.75
6	4.03	4.44
7	3.93	4.23
8	4.53	4.89
9	3.08	2.30
10	3.01	2.24
11	2.75	2.01
12	3.03	2.00
13	3.96	3.48
14	3.32	3.35
15	3.41	2.88
16	2.96	2.60
17	2.89	2.47
18	2.57	2.33
19	3.91	3.31
20	3.15	2.73

Table 18. Decision Latency Savings (Proportion Saved) as a Function of Block

Block	Mean	SD
1	0.303	0.260
2	0.275	0.196
3	0.390	0.062
4	0.317	0.192
5	0.292	0.123
6	0.366	0.072
7	0.278	0.082
8	0.212	0.158
9	0.300	0.235
10	0.353	0.189
11	0.309	0.211
12	0.354	0.279
13	0.273	0.144
14	0.205	0.036
15	0.286	0.106
16	0.230	0.072
17	0.260	0.107
18	0.304	0.160
19	0.271	0.143
20	0.207	0.158

Table 19. Percent Correct as a Function of Block

Block	Mean	SD
1	81.95	3.58
2	86.11	9.89
3	83.34	6.00
4	88.89	9.88
5	95.14	2.66
6	97.22	2.27
7	93.75	9.18
8	90.97	2.66
9	94.44	2.27
10	93.06	1.60
11	97.92	1.39
12	93.75	4.16
13	96.53	5.26
14	94.45	5.07
15	93.06	3.58
16	95.14	2.66
17	97.92	2.66
18	93.06	3.58
19	96.53	3.49
20	95.14	3.50

Table 20. Accuracy Savings (Proportion Saved) as a Function of Block

Block	Mean	SD
1	0.174	0.157
2	0.133	0.156
3	0.036	0.122
4	0.128	0.153
5	0.242	0.288
6	0.109	0.158
7	0.123	0.133
8	0.038	0.083
9	0.067	0.147
10	0.085	0.135
11	0.080	0.070
12	0.034	0.060
13	0.080	0.081
14	0.141	0.111
15	0.068	0.141
16	0.033	0.074
17	0.045	0.039
18	-0.021	0.049
19	0.102	0.086
20	-0.007	0.049

mean proportion of savings in accuracy as a function of block number, respectively. Accuracy savings were more modest than those for decision latency, ranging from a mean of 0.242 (SD = .288) at block five to a mean of -0.021 (SD = 0.049) at block 18. This is to be expected because in both experiments subjects eventually reached ceiling.

Total Study Time. Mean total study times declined from 12.11 sec at Block 1 to 2.06 sec at Block 20, with standard deviations of 14.54 sec and 1.78 sec, respectively. Tables 21 and 22 present mean total study time and mean proportion of savings in total study time as a function of block number, respectively. Mean total study time is represented by the following equation:

$$y = 9.260x^{-0.478}$$

This fit accounts for 89.0 percent of the variance. The greatest amount of savings was found for total study time, ranging from a mean of 0.811 (SD = 0.084) at Block 1 to a mean of 0.274 (SD = 0.373) at Block 19.

Study Time. Mean study times declined from 5.71 sec at Block 1 to 2.58 sec at Block 20, with standard deviations of 4.73 sec and 1.78 sec, respectively. Tables 23 and 24 present mean study time and mean proportion of savings in study time as a function of block number, respectively. Mean study time is represented by the following equation:

$$y = 11.920x^{-0.616}$$

This fit accounts for 85.6 percent of the variance. Like decision latency savings, study time the level of savings was considerable, ranging from a mean of 0.645 (SD = 0.143) at Block 1 to a mean of 0.239 (SD = 0.366) at Block 19.

Help Time. Upon their return, participants made efficient use of help. The only nontrivial access of help involved the operator names and destination names. Mean

Table 21. Total Study Time (Second) as a Function of Block

Block	Mean	SD
1	12.11	14.54
2	06.55	08.18
3	04.84	05.38
4	04.13	05.05
5	04.41	05.80
6	03.92	03.12
7	03.32	03.10
8	03.66	03.42
9	02.70	02.74
10	02.36	02.14
11	02.62	02.00
12	02.77	02.04
13	02.92	02.34
14	02.84	02.57
15	02.54	03.28
16	02.56	02.06
17	02.52	02.17
18	02.71	02.29
19	02.96	02.92
20	02.06	01.78

Table 22. Total Study Time Savings (Proportion Saved) as a Function of Block

Block	Mean	SD
1	0.811	0.084
2	0.722	0.078
3	0.732	0.135
4	0.639	0.200
5	0.688	0.060
6	0.605	0.082
7	0.624	0.119
8	0.483	0.161
9	0.566	0.254
10	0.514	0.268
11	0.449	0.215
12	0.456	0.292
13	0.363	0.340
14	0.352	0.289
15	0.475	0.151
16	0.370	0.162
17	0.322	0.268
18	0.375	0.294
19	0.274	0.373
20	0.393	0.382

Table 23. Study Time (Second) as a Function of Block

Block	Mean	SD
1	5.71	4.73
2	4.38	3.87
3	3.68	3.05
4	3.46	2.85
5	2.94	1.76
6	3.58	2.66
7	3.13	2.69
8	3.56	3.32
9	2.47	2.27
10	2.24	1.73
11	2.37	1.65
12	2.68	1.92
13	2.92	2.34
14	2.84	2.57
15	2.38	2.94
16	2.44	1.92
17	2.52	2.17
18	2.66	2.27
19	2.85	2.63
20	2.06	1.78

Table 24. Study Time Savings (Proportion Saved) as a Function of Block

Block	Mean	SD
1	0.645	0.143
2	0.407	0.268
3	0.491	0.098
4	0.439	0.074
5	0.526	0.133
6	0.357	0.188
7	0.432	0.085
8	0.245	0.276
9	0.386	0.241
10	0.443	0.260
11	0.383	0.237
12	0.351	0.266
13	0.334	0.295
14	0.331	0.297
15	0.437	0.131
16	0.309	0.174
17	0.288	0.287
18	0.298	0.309
19	0.239	0.366
20	0.360	0.418

help times declined from 6.40 sec at Block 1 to 0.00 sec at Block 20, with standard deviations of 11.68 sec and 0.00 sec, respectively. Table 25 presents mean help time as a function of block number.

At Block 1, mean time spent studying the operator names was 3.11 sec, with a standard deviation of 8.26 sec. At Block 7, there was no access of this help screen, and thereafter access was trivial. Table 26 presents mean time spent in operator names screen as a function of block number. No savings scores were computed for any of the help screens because frequently there were blocks where subjects did not access help. In these cases, the formula for computing savings is not meaningful.

At block 1, mean time spent studying destination names was 1.41 sec, with a standard deviation of 2.36 sec. By block 8 access of this help screen was modest ($M = 0.46$ sec and $SD = 300.67$ sec) and there was no access in Blocks 17 and 20. Table 27 presents mean time spent in destination names screen as a function of block number.

Experiment 2 - Discussion

In this experiment we investigated retention of a complex cognitive skill 60 days following initial training. Our experimental design provided us with a measure of retention; savings and our task provided us with the opportunity to examine retention at the component level. It is not surprising that we found performance had declined after 60 days; however, the decline was modest when overall task performance is considered. More important, it appears that the major locus of the decline can be isolated to certain task components/information-processing components. This is important because we can begin to analyze retention performance in a manner that will allow us to understand what is being retained and what is being lost.

The savings indices provided us with a metric of retention. On the first block at retention, savings for total study time, study time, accuracy, and decision latency

<u>Table 25</u>. Help Time (Second) as a Function of Block

Block	Mean	SD
1	6.40	11.68
2	2.17	06.67
3	1.16	03.31
4	0.67	03.38
5	1.47	04.95
6	0.37	01.06
7	0.19	00.94
8	0.01	00.54
9	0.23	01.22
10	0.12	00.69
11	0.24	00.79
12	0.09	00.43
13	0.19	00.80
14	0.09	00.46
15	0.16	01.25
16	0.12	00.59
17	0.00	00.00
18	0.05	00.29
19	0.12	00.68
20	0.00	00.00

Table 26. Operator Names Screen Time (Second) as a Function of Block

Block	Mean	SD
1	3.11	8.26
2	1.01	5.83
3	0.46	2.05
4	0.36	2.96
5	0.64	3.55
6	0.07	0.64
7	0.00	0.00
8	0.02	0.18
9	0.06	0.63
10	0.01	0.15
11	0.00	0.00
12	0.00	0.00
13	0.09	0.11
14	0.00	0.00
15	0.07	0.79
16	0.00	0.00
17	0.00	0.00
18	0.00	0.00
19	0.00	0.00
20	0.00	0.00

Table 27. Destination Names Screen Time (Second) as a Function of Block

Block	Mean	SD
1	1.41	2.36
2	0.60	1.48
3	0.32	0.83
4	0.17	0.68
5	0.32	1.00
6	0.13	0.41
7	0.12	0.66
8	0.05	0.30
9	0.08	0.40
10	0.04	0.23
11	0.15	0.56
12	0.05	0.24
13	0.10	0.43
14	0.05	0.26
15	0.05	0.31
16	0.08	0.42
17	0.00	0.00
18	0.03	0.17
19	0.08	0.46
20	0.00	0.00

were considerable (M = 0.811 and SD = 0.084, M = 0.645 and SD = 0.143, M = 0.174 and SD = 0.157, M = 0.303 and SD = 0.260, respectively). These findings indicate that subjects' retention of this skill was excellent. Furthermore, there were appreciable savings for three out of four of these indices throughout the entire experiment. As accuracy reached ceiling, savings, of course, became negligible at best and there were even two blocks where there were trivial losses. Clearly, the degree of consistency present in the overall task was such that retention performance was optimized, though not perfect.

One component of skill that declined appreciably was memory for specific names. Evidence for this decline is provided by subjects' use of help. Upon their return, subjects accessed most available help either trivially or not at all. Out of a total of 144 trials in Block 1 (four subjects times 36 trials), distance, cargo, weight, and vehicle categories help screens were each accessed only once (and never again in the entire experiment) and cargo names and vehicle names help screens were never accessed. Destination and license categories help screens were examined cursorily during Block 1; however, subjects made appreciable use of both the operator and destination names help screens. By Block 6 or 7, use of these screens had become trivial.

It appears that the subjects retained the structure of the task quite well. Two pieces of evidence provide support for this statement. First, initial accuracy was quite good, approximately 82 percent. It is doubtful that subjects would be able to achieve this level of accuracy if their knowledge of the rules governing the task had not remained solid. (Also, they did not expect to return and they were given no instructions.) Second, subjects made efficient use of help. That is, they avoided help that was superfluous; they knew where not to look. For example, they remembered that the weight information is unnecessary and even misleading; the vehicle information overrides it. Furthermore, because they did not access vehicle help, they must have recalled that all one needs to know about the vehicle is that if the first digit in the suffix is a 1 then the vehicle is light duty: if it is a 2, then the vehicle is medium duty; and if it is a 3, then the vehicle is heavy duty. The actual name is unimportant.

By Block 7 (less than 2 hours of practice), all indices of performance indicate that subjects were operating at end-of-training levels. It is interesting to note that by this point in the retention experiment the need to access both the operator and destination names from help had virtually disappeared. This seems to indicate that although initial access to the declarative information was reduced, restrengthening the access to the information required minimal retraining. It would appear that memory for names was the single most limiting factor in retention of skill in this task. This indicates that declarative knowledge decayed more relative to procedural knowledge.

Summary

In this experiment we examined the acquisition and retention of a cognitive skill in a complex task which consisted of a number of components with varying degrees of consistency. We set out to examine the validity/generalizability of previous findings from the automaticity literature to tasks with more ecological validity. We found that when components were consistent performance improved according to the power law (Newell & Rosenbloom, 1981) and variance was reduced. In the case of our only inconsistent component, overall performance improved and variance was reduced but the pattern of improvement was erratic, much like performance in a task with varied mapping between stimulus and response.

Retention performance was amazingly good. The quality of this performance is attributed to the degree of consistency present in the task at training and the persistence of the subjects' procedural knowledge.

We are currently working to replicate and extend these findings. An even more detailed analysis of the components of training and retention is our goal. We are currently in the process of refining our task to provide us with a tool to achieve this goal. We feel that investigations of training and retention in ecologically valid tasks are desperately needed. In fact, it could be argued that studying training without examining retention is like preparing a meal without tasting it.

VIII. AUGMENTED PROCESSING PRINCIPLES

One important outcome of the research program is the opportunity to specify what we refer to as processing principles. Such processing principles illustrate human performance guidelines that have been shown to be important for the development of "knowledge engineering" for understanding and developing training programs for complex operational tasks. Research conducted prior to AFHRL's investment in the understanding of the limits and extension of automatic/controlled processing theory to more mission-oriented tasks was well described by Fisk et al. (1987). Those principles of human performance can be summarized as follows:

Early Principles of Human Performance (from Fisk et al., 1987)

1. Performance improvements will occur only for situations where stimuli (or information) can be dealt with the same way from trial to trial.

2. The human operator is limited, not by the number of mental operations required, but by the number of inconsistent or novel cognitive (or psychomotor) operations.

3. To alleviate high workload situations, consistent task components must be identified and, once identified, training of those components should be given to develop automatic component processes.

4. Similar to number 3, to make performance reliable under environmental stressors (alcohol, fatigue, heat, noise, etc.), training should be conducted to develop automatic task components.

5. For tasks requiring sustained attention (vigilance), automatic target detection should be

developed prior to participating in the vigilance task; also, variably mapped information should not be presented in a continual and redundant pattern.

6. When preparing training programs, instructional designers should consider the nature of the underlying processing modes (automatic or controlled) in choosing part-task training strategies.

Based on the present work, as well as that described by Fisk et al., 1990 and other Air Force-sponsored research, we are now in a position to add to these human performance guidelines. The present augmented guidelines allow a more precise specification of human performance principles for determining performance limits and training program design for high-performance-skills training in complex, real-world tasks. Throughout this technical report we have presented data illustrating the following augmented human performance guidelines:

Augmented Processing Principles

1. Performance improvements will occur only for consistent elements of a task and the degree of improvement is directly related to the degree of consistency. [Section IV and Schneider & Fisk, 1982]

2. Performance is limited by the number of inconsistent cognitive operations; however, performance may also be limited by the type of task structure (e.g., memory versus visual versus hybrid memory/visual search). [Fisk & Rogers, in press]

3. Consistency need not be related to the individual stimulus level. Consistent relationships among stimuli, rules, and context should be identified

when considering part-task training strategies. [Section VI and Fisk & Lloyd, 1988; Fisk & Rogers, 1988; Fisk, Oransky, & Skedsvold, 1988; Myers & Fisk, 1987]

4. Global consistency can dominate performance improvement if lower-level consistency is absent. Instructional designers should locate, understand, and capitalize on global consistencies. [Section V and Fisk & Eboch, 1989; Fisk, Oransky, & Skedsvold, 1988]

5. Context affects performance in two major ways: (a) Contextual cues may be used to bias performance and <u>mimic</u> the effects of consistency; however, performance in this situation remains resource sensitive. (b) Contextual cues may activate automatic sequences of behavior. Context activation follows lawful temporal development. [Section VI and Fisk & Rogers, 1988]

6. Performance improvement occurs for lower-level, stimulus-based consistencies regardless of higher-order inconsistency. However, learning at the higher-order relational level is greatly attenuated by any degree of global inconsistency. [Section V and Fisk & Thigpen, 1988].

7. A direct relationship exists between amount of consistent practice and stimulus activation strength. However, the functional relationship is disrupted (i.e., <u>more</u> training is not necessarily better) when the to-be-learned stimuli can be unitized. Once a "superset" is developed, the activation of one element "strengthens" the other unitized elements. [Section II]

8. Disruption due to recombination of automatized task components is directly related to the "priority strength" of competing components. [Fisk et al., in press]

9. Part-task training can result in efficient associative learning, at least for semantic-based processing. Target strengthening (priority learning) benefits most from part-task training. [Section III]

10. Long-term retention of automatized task components is related to the type of task-specific processing: Memory access shows no decay for at least 1 year and visual search shows statistically nonsignificant (8 percent) decay after a year. Maximum decay (18 percent) is related to the coordination of component information, not component activation. [Appendix A]

IX. REFERENCES

Ackerman, P.L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. Intelligence, 10, 101-139.

Ackerman, P.L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of Experimental Psychology: General, 117, 288-318.

Adams, J.A. (1960). Part trainers. In G. Finch (Ed.), Educational and training media: A symposium (Publication 789). Washington, DC: National Academy of Science, National Research Council.

Adams, J.A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin, 101, 41-74.

Adams, J.A., & Hufford, L.E. (1961). Effects of programmed perceptual training on the learning of contact landing skills (NAVTRADEVCEN 247-3). Port Washington, NY: U.S. Naval Training Device Center.

Ammons, R.B., Ammons, C.H., & Morgan, R.L. (1956). Transfer of skill and decremental factors along the speed dimension in rotary pursuit. Perceptual and Motor Skills, 6, 43.

Anderson, J.R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Anderson, J.R. (1983). The architecture of cognition Cambridge, MA: Harvard University Press.

Bailey, J.S., Hughes, R.G., & Jones, W.E. (1980). Application of backward chaining to air-to-surface weapons delivery training (AFHRL-TR-79-63, AD085 610). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.

Battig, W.F., & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 80, (Whole).

Battiste, V. (1987). Part-task vs. whole-task training on a supervisory control task. In Proceedings of the Human Factors Society 31st Annual Meeting (pp. 1365-1369). Santa Monica, CA: Human Factors Society.

Briggs, G.E., & Brogden, W.J. (1954). The effect of component practice on performance of a lever-positioning skill. Journal of Experimental Psychology, 48, 375-380.

Briggs, G.E., & Naylor, J.C. (1962). The relative efficiency of several training methods as a function of transfer task complexity. Journal of Experimental Psychology, 64, 505-512.

Briggs, G.E., Naylor, J.C., & Fuchs, A.H. (1962). Whole versus part training as a function of task dimensions (NAVTRADEVCEN 950-2). Port Washington, NY: U.S. Naval Training Device Center.

Briggs, G.E., & Rockway, M.R. (1966). Learning and performance as a function of the percentage of pursuit components in a tracking display. Journal of Experimental Psychology, 71, 165-169.

Briggs, G.E., & Waters, L.K. (1958). Training and transfer as a function of component interaction. Journal of Experimental Psychology, 56, 492-500.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press.

Collen, A., Wickens, D.D., & Daniele, L. (1975). The interrelationship of taxonomic categories. Journal of Experimental Psychology: Human Learning and Memory, 1, 629-633.

Cream, B.W., Eggemeier, F.T., & Klein, G.A. (1978). A strategy for the development of training devices. Human Factors, 20, 145-158.

Dumais, S.T. (1979). Perceptual learning in automatic detection: Processes and mechanisms. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.

Duncan, J. (1986). Consistent and varied training in the theory of automatic and controlled information processing. Cognition, 23, 279-284.

Durso, F.T., Cooke, N.M., Breen, T.J., & Schvaneveldt, R.W. (1987). Is consistent mapping necessary for high speed search? Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 223-229.

Eberts, R., & Schneider, W. (1985). Internalizing the system dynamics for a second-order system. Human Factors, 27, 371-395.

Eggemeier, F.T., Fisk, A.D., Robbins, R., Lawless, M.T., & Spaeth, R. (1988). High-performance skills task analysis methodology: An automatic human information processing theory approach (Final Report AFHRL-TP-88-32, AD-B128 366). Wright-Patterson AFB, OH: Logistics and Human Factors Division, Air Force Human Resources Laboratory.

Feurzeig, W., & White, B.Y. (1983). Development of an articulate instructional system for teaching arithmetic procedures (BBN Report No. 5484). Cambridge, MA: BBN Laboratories.

Fisk, A.D., Ackerman, P.L., & Schneider, W. (1987). Automatic and controlled processing theory and its applications to human factors problems. In P.A. Hancock (Ed.), Human Factors Psychology (pp. 159-197). Amsterdam: North-Holland.

Fisk, A.D., & Eboch, M.M. (1989). An automatic/ controlled processing theory application to training component map reading skills. Applied Ergonomics, 20, 2-8.

Fisk, A.D., & Gallini, J.K. (1989). Training consistent components of tasks: Developing an instructional system based on automatic/controlled processing principles. Human Factors, 31, 453-463.

Fisk, A.D., Hodge, K.A., Lee, M.D., & Rogers, W.A. (1990). Automatic information processing and high-performance skills: Acquisition, transfer, and retention (AFHRL-TR-89-69, AD-A221 744). Wright-Patterson AFB, OH: Logistics and Human Factors Division, Air Force Human Resources Laboratory.

Fisk, A.D., Lee, M.D., & Rogers, W.A. (in press). Recombination of automatic processing components: The effects of transfer, reversal, and conflict situations. Human Factors.

Fisk, A.D., & Lloyd, S.J. (1988). The role of stimulus-to-rule consistency in learning rapid application of spatial rules. Human Factors, 30, 35-49.

Fisk, A.D., McGee, N.D., & Giambra, L.M. (1988). The influence of age on consistent and varied semantic category search performance. Psychology and Aging, 3, 323-333.

Fisk, A.D., Oransky, N.A., & Skedsvold, P.R. (1988). Examination of the role of "higher-order" consistency in skill development. Human Factors, 30, 567-581.

Fisk, A.D., & Rogers, W.A. (1988). The role of situational context in the development of high-performance skills. Human Factors, 30, 703-712.

Fisk, A.D., & Rogers, W.A. (in press, June, 1991). Toward an understanding of age-related memory and visual search effects. Journal of Experimental Psychology: General.

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

Fisk, A.D., & Schneider, W. (1982). Type of task practice and time-sharing activities predicts deficits due to alcohol ingestion. In Proceedings of the Human Factors Society 26th Annual Meeting (pp. 926-930). Santa Monica, CA: Human Factors Society.

Fisk, A.D., & Schneider, W. (1983). Category and word search: Generalizing search principles to complex processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 177-195.

Fitts, P. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), Categories of human learning (pp. 243-285). New York: Academic Press.

Fitts, P., & Posner, M.I. (1967). Human performance. Belmont, CA: Brooks/Cole.

Flexman, R.E., Roscoe, S.N., Williams Jr., A.C., & Williges, B.H. (1972). Studies in pilot training: The anatomy of transfer. Aviation Research Monographs, 2.

Folds, D.J., Gerth, J.M., & Engelman, W.R. (1987) Enhancement of human performance in manual target acquisition (USAFSAM-TR-86-18). Atlanta, GA: Georgia Institute of Technology, Systems Engineering Laboratory.

Frederiksen, J.R., Warren, B., & Rosebery, A.(1985a). A componential approach to training reading skills: Part I. Perceptual units training. Cognition and Instruction, 2, 91-130.

Frederiksen, J.R., Warren, B., & Rosebery, A. (1985b). A componential approach to training reading skills: Part II. Decoding and use of context. Cognition and Instruction, 2, 271-338.

Frederiksen, J.R., & White, B.Y. (1989). An approach to training based upon principled task decomposition. Acta Psychologica, 71, 89-146.

Freedle, D.O., Zavala, A., & Fleishman, E.A. (1968). Studies of component-total relations: Order of components, total task practice, and total task predictability. Human Factors, 10, 33-40.

Gopher, D., & North, R.A. (1974). The measurement of capacity limitation through single and dual-task performance with individual adjustment of difficulty. Proceedings of the Human Factors Society 18th Annual Meeting (pp. 480-485). Santa Monica, CA: Human Factors Society.

Gordon, N.B. (1959). Learning a motor task under varied display conditions. Journal of Experimental Psychology, 57, 65-73.

Hancock, P.A. (1984). Environmental stressors. In J. S. Warm (Ed.), Sustained attention in human performance. New York: John Wiley.

Hancock, P.A., & Pierce, J.O. (1984). Toward an attentional theory of performance under stress: Evidence from studies of vigilance in heat and cold. In A. Mital (Ed.), Trends in ergonomics/human factors I (pp. 1-7). Amsterdam: North-Holland.

Hodge, K.A., & Fisk, A.D. (1989). Transfer of training as a function of semantic relatedness in a category search task. In Proceedings of the Human Factors Society 33rd Annual Meeting (pp. 1253-1257). Santa Monica, CA: Human Factors Society.

Jaeger, R.J., Agarwal, G.C., & Gottlieb, G.L. (1980). Predictor operator in pursuit and compensatory tracking. Human Factors, 22, 497-506.

Jennings, A.E., & Chiles, W.D. (1977). An investigation of time-sharing ability as a factor in complex performance. Human Factors, 19, 535-547.

Jensen, R.S. (1979). Prediction and quickening in perspective flight displays for curved landing approaches. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana, IL.

Kanfer, R., & Ackerman, P.L. (1989). Dynamics of skill acquisition: Building a bridge between abilities and motivation. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 5, pp. 83-134). Hillsdale, NJ: Erlbaum.

Klapp, S.T., Martin, Z.E., McMillan, G.G., & Brook, D.T. (1987). Whole-task and part-task training in dual motor tasks. In L. S. Mark, J. S. Warm, & R. L. Huston (Eds.), Ergonomics and Human Factors: Recent Research (pp. 125-130). Amsterdam: North-Holland.

Kristofferson, M.W. (1977). The effects of practice with one positive set in a memory scanning task can be completely transferred to a different positive set. Memory and Cognition, 5, 177-186.

Kyllonen, P.C., & Woltz, D.J. (1989). Role of cognitive factors in the acquisition of cognitive skill. In R. Kanfer, P.L. Ackerman, & R. Cudeck (Eds.), Abilities, motivation, and methodology (pp 239-280). NY: Freeman a& Co.

LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 6, 293-323.

Lintern, G., & Roscoe, S.N. (1980). Visual cue augmentation in contact flight simulation. In S. N. Roscoe (Ed.), Aviation psychology. Ames, IA: Iowa State University Press.

Lintern, G., Thomley, K., Nelson, B., & Roscoe, S.N. (1984). Content, variety and augmentation of simulated visual scenes for teaching air-to-ground attack (NAVTRAEQUIPCEN 81-C-0105-3). Orlando, FL: Naval Training Equipment Center.

Lintern, G., & Wickens, C.D. (1987). Attention theory as a basis for training research (ARL-87-2/NASA-87-3). Savoy, Illinois: University of Illinois, Institute of Aviation, Aviation Research Laboratory.

Logan, G.D. (1978). Attention in character classification: Evidence for the automaticity of component stages. Journal of Experimental Psychology: General, 107, 32-63.

Logan, G.D. (1979). On the use of concurrent memory load to measure attention and automaticity. Journal of Experimental Psychology: Human Perception and Performance, 5, 189-207.

Logan, G.D. (1985). Skill and automaticity: Relations, implications and future directions. Canadian Journal of Psychology, 39, 367-386.

Logan, G.D. (1988a). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Logan, G.D. (1988b). Automaticity, resources, and memory: Theoretical controversies and practical implications. Human Factors, 30, 583-598.

MacKay, D.G. (1982). The problem of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. Psychological Review, 89, 483-506.

Mane, A.M. (1984). Acquisition of perceptual-motor skill: Adaptive and part-whole training. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 522-526). Santa Monica, CA: Human Factors Society.

McGrath, J.J., & Harris, D.H. (Eds.) (1971). Adaptive training. Aviation Research Monographs (Vol. 1). Champaign, Illinois: University of Illinois, Aviation Research Laboratory.

McGuigan, F.J., & MacCaslin, E.F. (1955). Whole and part methods in learning a perceptual-motor skill. American Journal of Psychology, 68, 658-661.

Myers, G.L., & Fisk, A.D. (1987). Application of automatic and controlled processing theory to industrial training: The value of consistent component training. Human Factors, 29, 255-268.

Naylor, J.C., & Briggs, E.G. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. Journal of Experimental Psychology, 65, 217-224.

Neches, R., Langley, P., & Klahr, D. (1987). Learning, development, and production systems. In D. Klahr, P. Langley, & R. Neches (Eds.), Production system models of language and development. Cambridge, MA: MIT Press.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Nissen, M.J., & Bullemer, P. (1984). Attentional requirements of learning: Evidence from performance measures. Paper presented at the Psychonomic Society, San Antonio, TX.

Posner, M.I., & Snyder, C.R.R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), Information processing and cognition (pp. 55-85). Hillsdale, NJ: Erlbaum.

Poulton, E.C. (1974). Tracking Skill and Manual Control. New York: Academic Press.

Prinz, W. (1979). Locus of the effect of specific practice in continuous visual search. Perception and Psychophysics, 25, 137-142.

Rabbitt, P.M.A., Cumming, G., & Vyas, S.M. (1979). An analysis of visual search: Entropy and sequential effects. In S. Dornic (Ed.), Attention and Performance VI (pp. 363-386). Potomac, MD: Erlbaum.

Rieck, A.M., Ogden, G.D., & Anderson, N.S. (1980). An investigation of varying amounts of component-task practice on dual-task performance. Human Factors, 22, 373-384.

Rogers, W.A. (1989). Target and distractor learning in visual search: Age-related differences. Unpublished master's thesis, Georgia Institute of Technology, Atlanta, GA.

Roscoe, S.N., Saad, F., & Jensen, R.S. (1979). Analysis of intraserial transfer on curved landing approaches with pursuit and compensatory displays (Illiana-79-1). Champaign, Illinois: Illiana Aviation Sciences.

Roscoe, S.N., & Williges, B.H. (1980). Measurement of transfer of training. In S. N. Roscoe (Ed.), Aviation psychology. Ames, IA: Iowa State University Press.

Rumelhart, D.E., & McClelland, J.L. (1987). Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1). Cambridge, MA: MIT Press.

Salthouse, T.A., & Prill, K. (1983). Analysis of a perceptual skill. Journal of Experimental Psychology: Human Perception and Performance, 9, 607-621.

Salthouse, T.A., & Somberg, B.L. (1982). Skilled performance: Effects of adult age and experience on elementary processes. Journal of Experimental Psychology: General, 111, 176-207.

Schneider, W. (1985a). Training high-performance skills: Fallacies and guidelines. Human Factors, 27, 285-300.

Schneider, W. (1985b). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Martin (Eds.), Attention and Performance XI (pp. 475-492). Hillside, NJ: Erlbaum.

Schneider, W., & Detweiler, M. (1987). A connectionist/ control architecture for working memory. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 53-118), Volume 21. New York: Academic Press.

Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. Human Factors, 30, 539-566.

Schneider, W., Dumais, S.T., & Shiffrin, R.M. (1984). Automatic and control processing and attention. In R. Parasuraman, R. Davies, & J. Beatty (Eds.), Varieties of Attention (pp. 1-27). New York: Academic Press.

Schneider, W., & Fisk, A.D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. Perception and Psychophysics, 31, 160-166.

Schneider, W., & Fisk, A.D. (1984). Automatic category search and its transfer. Journal of Experimental Psychology: Learning, Memory and Cognition, 10, 1-15.

Schneider, W., & Mumme, M. (1987). A connectionist control architecture for attention, automaticity and the capturing of knowledge. Unpublished manuscript, University of Pittsburgh, Pittsburgh, PA.

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

Sheppard, D.J. (1984). Visual and part-task manipulations for teaching simulated carrier landings (NAVTRAEQUIPCEN 81-C-0105-9). Orlando, FL: Naval Training Equipment Center.

Shiffrin, R.M. (1988). Attention. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), Stevens' handbook of experimental psychology (pp. 739-811). New York: Wiley.

Shiffrin, R.M., & Czerwinski, M.P. (1988). A model of automatic attention attraction when mapping is partially consistent. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 562-569.

Shiffrin, R.M., & Dumais, S.T. (1981). The development of automatism. In J. R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 111-140). Hillsdale, NJ: Erlbaum.

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

Simon, C.W., & Roscoe, S.N. (1981). Application of a multifactor approach to training research (NAVTRAEQUIPCEN 78-C-0060-6). Orlando, FL: Naval Training Equipment Center.

Smode, A. (1958). Learning and performance in a tracking task under two levels of achievement information feedback. Journal of Experimental Psychology, 56, 297-304.

Sperling, G., Budiansky, J., Spivak, J.G., & Johnson, M.C. (1971). Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. Science, 174, 307-311.

Stammers, R.B. (1980). Part and whole practice for a tracking task: Effects of task variables and amount of practice. Perceptual and Motor Skills, 50, 203-210.

Thigpen, M.R., & Fisk, A.D. (1988, March). Top-down versus bottom-up stimulus consistency: Does level of processing interact with stimulus consistency? Presented at the Annual Meeting of the Southern Society for Philosophy and Psychology, Miami, FL.

Vidulich, M., Yeh, Y., & Schneider, W. (1983). Time-compressed components for air-intercept control skills. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 161-164). Santa Monica, CA: Human Factors Society.

Wechsler, D. (1981). Wechsler Adult Intelligence Scale: Revised. New York: Psychological Corporation.

Westra, D.P. (1982). Investigation of simulator design features for carrier landing: II. In-simulator transfer of training (NAVTRAEQUIPCEN 81-C-0105-1). Orlando, FL: Naval Training Equipment Center.

White, B.Y. (1981). Designing computer games to facilitate learning (AI-TR-619). Cambridge, MA: Massachusetts Institute of Technology, Artificial Intelligence Laboratory.

White, B.Y. (1984). Designing computer activities to help physics students understand Newton's laws of motion. Cognition and Instruction, 1, 69-108.

White, B.Y., & Frederiksen, J.R. (1985). QUEST: Qualitative understanding of electrical system troubleshooting. ACM SIGART Newsletter, 93, 34-37.

White, B.Y., & Frederiksen, J.R. (1986a). Progressions of qualitative models as a foundation for intelligent learning environments (BBN Report No. 6277). Cambridge, MA: BBN Laboratories.

White, B.Y., & Frederiksen, J.R. (1986b). Intelligent tutoring systems based upon qualitative model evolutions. In Proceedings of the Fifth National Conference on Artificial Intelligence, Philadelphia, PA.

Wightman, D.C. (1983). Part-task training strategies in simulated carrier landing final approach training (NAVTRAEQUIPCEN IH-347). Orlando, FL: Naval Training Equipment Center.

Wightman, D.C., & Lintern, G. (1985). Part-task training for tracking and manual control. Human Factors, 27, 267-284.

Wightman, D.C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final approach training. Human Factors, 29, 245-254.

APPENDIX A: RETENTION OF TRAINED PERFORMANCE IN CONSISTENT MAPPING SEARCH AFTER EXTENDED DELAY

A substantial amount of research has been conducted to investigate performance improvement with practice on consistently mapped search/detection tasks. In general, this work has shown that both the nature and extent of improvement are dependent on how consistently subjects can deal with a task (Schneider & Fisk, 1982). It is often found that, with consistent mapping practice, performance that initially appears dominated by relatively slow, effortful, and serial search processes seems to become dominated by fast, relatively effortless, and parallel search. Much has been written concerning the nature and mechanisms for such changes (e.g., see Anderson, 1982; Logan, 1988; Rosenbloom & Newell, 1986; Schneider, 1985; Schneider & Detweiler, 1987, 1988; Shiffrin & Schneider, 1977); unfortunately, little is known about the retention of learning associated with such performance improvements.

The study of the retention of learned material has had a prominent place in psychology from its earliest days (e.g., Ebbinghaus, 1885/1964) and continues to be important for psychological theory development in areas such as memory and human performance (e.g., see Bahrick, 1979, 1984; Kolers, 1976; Salasoo, Shiffrin, & Feustel, 1985), as well as instructional systems design (e.g., see Hagman & Rose, 1983; Johnson, 1981; Mengelkoch, Adams, & Gainer, 1971) and the analysis of individual differences (Gentile, Monaco, Iheozor-Ejiofor, Ndu, & Ogbonaya, 1982; Kyllonen & Tirre, 1988; Shuell & Keppel, 1970; Underwood, 1954).

Bahrick (1979, 1983, 1984; Bahrick, Bahrick, & Wittlinger, 1975) has collected a compendium of data on the very-long-term retention of various types of information. His results serve as an example of how the study of retention characteristics can be important for a more complete understanding of human performance and learning. Those data have made a fundamental contribution to the understanding of human memory. Bahrick examined what might be considered relatively "permanent knowledge." His research suggests that, although a portion of what we might think of as relatively permanent knowledge remains accessible only if used periodically, portions of that learned information attain what Bahrick calls "permastore" status. Bahrick's data point to the importance of the quality and extent of training at the time of initial learning. For example, his Spanish language retention data (Bahrick, 1984) demonstrated that the students who received low levels of training retained little knowledge of Spanish language whereas more extensive training led to up to 70 percent retention after 25 years or more. His data also showed the classic effect that initial training level predicts retention level; that is, after about 5 years, forgetting had reached a plateau but students receiving "A" grades in original coursework reached a higher retention level than those receiving Bs, etc. This effect is well documented in the retention literature (e.g., see Farr, 1987) for studies using shorter retention intervals.

Not all improvements in information processing gained via practice are retained. Salasoo et al. (1985) examined the development and long-term retention of two separable memory factors that facilitate the detection of letter strings. In their experiments they investigated the repetition effect (prior occurrence of an item speeds later identification of that item) and the development of associatively connected memory codes. Salasoo et al. demonstrated that repeated presentations of a nonword letter string led to "codification" (the unitization of a memory code that can be automatically activated even by fragments of the nonword string). Such codification eliminated the word superiority effect and repetition effects were present for both words and nonwords. Testing 1 year later revealed that codification was still intact but the repetition advantage had vanished both for the trained words and trained nonwords. These results suggest that certain memorial processes may be more resistant to decay than other processes, a point we will return to in the general discussion of our data.

There is further evidence that components of skilled performance may be retained at different levels across a retention interval. Kolers (1976) examined subjects' ability to read typographically inverted text approximately 1 year after they were trained to read that unfamiliar typography. Kolers found that subjects retained some of the previously trained ability to read the inverted text; text read for the first time during the retention test was read more quickly than approximately the 40th page of text (out of 160 total training pages) read during training. Furthermore, Kolers found that text which had been read the year before was read faster during the retention test than was the new text. Although a decrement in speed of reading the inverted text occurred after 1 year, these data suggested to Kolers that pattern-analyzing operations directed at the lexical objects were retained as well as, if not better than, semantic information.

It may not be surprising that some information or knowledge is retained for extended time periods whereas other information decays relatively quickly. However, an understanding of the characteristics of performance retention, within a given learning domain, may be valuable for understanding the structure of learning within that domain. Therefore, in this paper we focus on the retention of search/detection performance. Our goal was to examine and document the retention characteristics of memory, visual, and hybrid memory/visual search after subjects had received extensive consistent mapping practice.

The results from previous research examining the durability of performance improvement in search/detection tasks are somewhat equivocal. For example, Healy, Fendrich, and Proctor (1990) recently reported the extreme durability of performance in a pure visual search task (subjects searched for a single character in a display of 2, 4, or 16 characters). The subjects in their experiment demonstrated no forgetting of the detection skill even after a 1-month retention interval (with some evidence of retention beyond 6 months). In contrast, Rabbitt, Cumming, and Vyas (1979) found significant performance decay in a hybrid memory/visual search task (memory-set and display size both greater than one) after a six-week retention interval. Although these studies seem to be contradictory, it is important to note that the Healy et al. study utilized a pure visual search task and the Rabbitt et al. results are based on a hybrid memory/visual search task. There are good reasons to believe that memory, visual, and hybrid memory/visual search tasks are dominated by related but distinct processing mechanisms (see Fisk & Rogers, 1990, for a review); hence, in the present series of studies we systematically examined retention of performance in each of these classes of search tasks. This investigation allows more precise prediction of retention characteristics within the major classes of search detection tasks.

In the first two experiments, we examined retention of detection performance in memory scanning (Experiment 1) and in visual search (Experiment 2) approximately 1 month after training. In the first experiment, memory-set size varied from one to three items and display size was held constant at one item; thus, retention of pure memory search was assessed. This experiment examined the retention of associative learning and direct access (Logan, 1988) to that "codified" information. Experiment 2 examined the retention of what might be called perceptual tuning. In that experiment we utilized a multiple-frame task (Schneider & Shiffrin, 1977; Sperling, Budiansky, Spivak, & Johnson, 1971) with an adaptive training procedure to examine performance on a task that encouraged processing at the word or "word feature" level but not at the semantic category level. In this experiment, subjects searched for words from a single semantic category in rapidly displayed "frames" of multiple-distractor words.

In the third experiment we examined performance 1, 30, 90, 180, and 365 days following training on a hybrid memory/visual, semantic-category search task. Subjects received differing amounts of consistent practice across the categories used in the experiment. We also tested the subjects' ability to detect untrained words from the trained categories. Thus, we examined retention at intervals up to 1 year after practice as a function of amount of consistent mapping practice. In addition, we assessed the degree of "category" activation (performance on untrained words from the trained categories) at each retention interval.

Experiment 1 - Memory Scanning

Salasoo et al. (1985) have demonstrated that elements chunked together (or codified) as the result of training remain unitized even after some delay without practice. We tested the decay resistance of unitization using a different class of tasks to evaluate the generality of the Salasoo et al. finding. In this first experiment, we tested the effects of a 32-day retention interval on performance in a consistently mapped, memory-search task. This task was chosen because one aspect of consistent memory search seems to be the unitization of the memory-set elements; that is, the memory-set elements tend to become associatively connected as a function of practice (Schneider, 1985; Schneider & Detweiler, 1987; Schneider & Shiffrin, 1985; Shiffrin & Schneider, 1977). Associative learning allows the categorization (unitization) of the memory set; thus, working memory load is reduced and a more efficient search develops such that the entire memory set may be compared with the display elements in a single operation. For this efficient search to be used, however, the memory set must be well learned such that activation of one element in the memory set associatively activates other memory-set nodes in memory. If performance improvement in consistent mapping memory search is due, in a large part, to such associative learning, given the Salasoo et al. findings, we predicted little performance decrement after the 32-day retention interval.

Method

Participants. Fourteen right-handed volunteers, eight males and six females, were recruited from introductory psychology classes at the Georgia Institute of Technology. One male and two females failed to return for the retention phase; so, the data are presented for the remaining 11 subjects. Participants were tested for visual acuity of at least 20/30 (uncorrected or corrected) and near vision of at least 20/40. Participants received a combination of research credits and money.

Equipment. Epson Equity I+ microcomputers equipped with Epson MBM-2095-E monochrome monitors (green phosphor, 50-Hz refresh rate) with Epson multimode graphics adapters were programmed to present the task and collect data. The microcomputers were programmed with commercial software (Psychological Software Tools' Microcomputer Experimental Language) to present and time stimulus displays and to record responses. The '4' and '5' keys on the numeric keypad were labeled with a 'Y' and an 'N' corresponding to "yes" and "no," respectively. To mask external sounds, the task was performed within booths constructed of sound-deadening materials and pink noise was played at a sound level of approximately 55 dB(A).

Stimuli. Fourteen taxonomic categories, with exemplars selected from the Battig and Montague (1969) category norms, were used in the experiment. The categories were ALCOHOLIC BEVERAGES, ARTICLES OF CLOTHING, BUILDING PARTS, COUNTRIES, EARTH FORMATIONS, FLOWERS, FOUR-FOOTED ANIMALS, HUMAN BODY PARTS, MUSICAL INSTRUMENTS, OCCUPATIONS, RELATIVES, UNITS OF TIME, VEGETABLES, AND WEAPONS. Six words were chosen from each category according to four criteria: visual distinctiveness (assessed via pilot testing), semantic distinctiveness (Collen, Wickens, & Daniele, 1975), length (between four and seven letters), and high prototypicality (according to Battig and Montague).

All words were presented in uppercase. Participants were seated approximately 48 cm from the display. At that viewing distance, the average letter subtended 0.38 degree in width and 0.47 degree in height. Within a word, interletter separation was 0.19 degrees.

Design. The experiment consisted of two phases, training and retention. All manipulations in both training and retention were manipulated within-subject and within-block. In the training phase, there were two factors of interest: trial type (target present versus target absent) and memory-set size (1, 2 or 3 category labels). Probe size was constant at one exemplar. Each participant was trained on exemplars from three target categories and six distractor categories. All trials were consistently mapped. Assignment of categories to participants was counterbalanced by a partial Latin-square.

Each session consisted of 19 blocks of trials (42 trials per block). Subjects completed 10 sessions of training, for a total of 7,980 trials--half of which were target-present trials and half, target-absent. The retention phase consisted of one session (identical to a training session) 32 days following training.

Procedure. Each trial proceeded as follows. The memory set (one, two or three category labels) was displayed in the left center of the VDT screen at the beginning of each trial. Participants could study the memory set for up to 20 sec. To begin each trial participants pressed the space bar. An orientation display consisting of three '+' signs was presented for 500 ms in the same location as the display set to allow the participant to orient his or her gaze. Then the display set, consisting of either one target exemplar or one distractor exemplar was presented. The participant's task was to decide as quickly as possible whether a target was or was not present and press the '4' key (labeled Y) for target present or the '5' key (labeled N) for target absent.

Participants received the following performance feedback: After each correct trial, the participant's reaction time (RT) was displayed. After each incorrect trial, an error tone was sounded (for 500 ms) and the correct response displayed (for 1 sec). Following each block of trials, the participant was given his or her average correct trial RT and percent correct for that block. Participants were instructed to maintain an accuracy rate of 95 percent or better while responding as quickly as possible. If accuracy fell below 90 percent for any block, the computer instructed the subject to respond more carefully.

Results and Discussion

During training, mean RT decreased from 542 ms after the first session to 410 ms in the last session of practice. There were significant main effects of Practice, $F(9,90) = 54.15$, Memory-set size, $F(2,20) = 24.55$, and Trial Type (Positive vs. Negative), $F(1,10) = 8.04$ (unless otherwise indicated, alpha level was set at .05). The main effects notwithstanding, there was a significant interaction between Memory-set Size and Sessions, $F(18,180) = 12.53$. The interaction indicates that, as practice proceeded, memory set size had less of an influence on performance. An examination of comparison slope estimates (the slope of the line relating RT to number of memory comparisons) provides more evidence that training led to proficient performance. After Session 1 the slope estimates for target present and absent conditions were 28 ms and 20 ms per comparison, respectively. By Session 8, slope estimates in the target present and absent condition had stabilized at less than 4 ms. Accuracy also improved with practice, $F(9,90) = 5.59$, changing from 92 percent in Session 1 to 96 percent in

Session 10. No other effect of accuracy reached significance.

The central issue pertains to retention of the improved memory search performance: What happened after 32 days without practice? Reaction time performance at the last day of training and 32 days following training can be compared in Figure A-1. Clearly, the decline in performance was negligible (1.3 percent in the target absent condition and 1.1 percent in the target present condition). The comparison between the last session of training and the retention test revealed a main effect of Trial Type, $F(1, 10) = 5.79$ only. The effect of retention interval was not significant, $F < 1$. The other important comparisons, for the conceptual purpose of the experiment, involved the potential interactions with retention interval. None of those interactions reached significance: Trial Type X Time ($F < 1$), Memory-set Size X Time ($F(2,20) = 1.34$), and Trial Type X Memory-set Size X Time ($F < 1$).

The present data clearly indicate that what was gained during CM memory search practice did not decline within a retention interval of 32 days. Response speed was retained, as well as the elimination of the set size effect (i.e., scanning memory across three categories for a match was as fast as scanning for one). This finding supports the stability of associative learning that occurs during CM training. Associative learning results in the unitization or "codification" of the memory-set elements such that all elements need not be individually activated and compared during search (for a review see Schneider & Shiffrin, 1985). Hence, the learning related to CM memory search seems resistant to decay, at least for 32 days. This is consistent with, but extends memory retention findings by, Salasoo et al. (1985) and Bahrick (1984).

Experiment 2 - Visual Search

In the next experiment, we tested another group of participants to examine the effects of retention on a

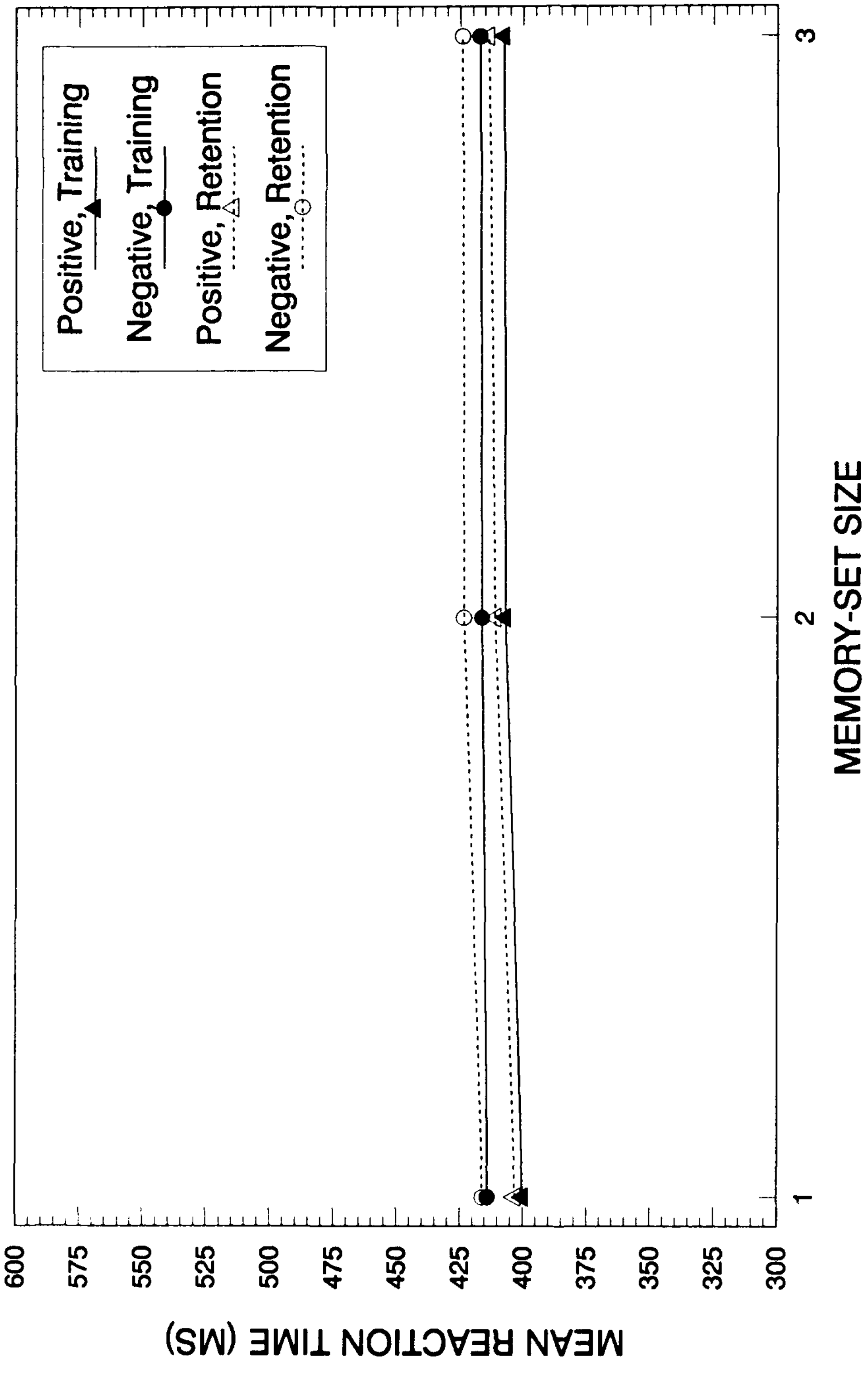


Figure A-1. Mean RTs (Correct Trials Only) for Last Session of Practice (Solid Symbols) and Performance 32 Days Subsequent to Practice (Open Symbols).

relatively pure visual search task. For this task, memory-set size was held constant at one and display-set size was held constant at three. Differences in retention performance between memory scanning and visual search might be predicted because, although similarities exist between memory scanning and visual search, these processes appear to involve different processing mechanisms (e.g., see Fisher, Duffy, Young, & Pollatsek, 1988; Flach, 1986; Hoffman, 1978, 1979; Logan, 1988; Schneider, 1985). Pure visual search benefits most from an ability to differentiate (i.e., filter) targets from distractors, whereas memory scanning is enhanced most from an ability to associate the elements in a target set into a single equivalence class. Examination of Kolers' (1976) data suggested that the precision or "perceptual tuning" that occurs with CM practice in visual search may decline with disuse. Although Kolers reports good retention when measured in savings scores, the decline in performance on his complex, inverted-text reading task was approximately 40 percent when the first retention page is compared with the last training page reading time. (It is important to note that the first page reading time at retention was four times faster than the first page reading time in training, approximately 4 seconds compared with approximately 16 seconds). Hence, although Kolers seems correct to have argued that specific pattern analyzing operations can be retained, the perceptual tuning that occurred with practice seemed to decay when not used. Other data suggest that performance improvement in a CM visual search task may be resistant to decay. Healy, Fendrich, and Proctor (1990) provided subjects with CM training for zero to four sessions. The subjects' task was to detect the letter 'H' within displays of 2, 4, or 16 characters. Healy et al reported a statistically significant display-size effect even after four practice sessions; however, their subjects' performance was no different after the 1 month retention interval when compared with performance after the last practice session. The difference in the findings of Kolers and of Healy et al. could be due to the type of task subjects were required to perform. Kolers' task was rather complex and was not a relatively pure visual search task. The task used by Healy et al. was comparatively rather simple and a relatively pure visual search task. In the next experiment we examined the retention of performance improvement in a demanding task but a task where visual search skill clearly dominated task performance.

Method

Subjects. Ten right-handed volunteers (five males) received a combination of research credits and money as compensation for participation in the experiment. Participants were tested for visual acuity of at least 20/30 (uncorrected or corrected) and near vision of at least 20/40.

Equipment. All equipment was the same as described in Experiment 1.

Design. All manipulations were within-subject. The study was divided into three phases: training, transfer, and retention test. Training consisted of one orientation session and 14 training sessions. During the orientation session we obtained demographic and health information, tested visual acuity, and instructed participants on how to perform the task. In addition, participants ran through an abbreviated session--seven blocks of trials with 30 trials per block for a total of 210 trials. The actual training sessions consisted of 14 blocks of trials per session (30 trials per block), for a total of 5,880 trials. An average of 20 percent of all trials were negative (target absent). (Negative trials were included to ensure that subjects attended to all display locations throughout the trial.) In any block, five, six or seven negative trials could be presented. The exact number for any particular block was permuted with the restriction that the mean number of negative trials per block was six.

There were two transfer sessions consisting of 11 blocks per session. Five conditions were manipulated across blocks, with two blocks of each condition per session: (a) Trained/Trained (TT)- the same category and exemplars on which an individual had previously trained; (b) Trained/Untrained (TU)- six new exemplars from the same category on which a participant had previously trained: (c) Highly-Related (HR)- six exemplars from a category which was highly semantically related (Collen et al., 1975) to the category on which a participant trained; (d) Moderately Related (MR)- six exemplars from a category moderately semantically related to the category on which a participant trained; and (e) Unrelated (UR)- six exemplars from a category unrelated to any other category used in either training or transfer. The five conditions were manipulated between blocks of trials and order of presentation was counterbalanced across participants. In addition, each transfer session for all participants began with one TT block as a "priming" situation. The retention testing occurred 30 days following the last transfer session.

Stimuli. The criterion used to select the categories and exemplars for training was the same as described in Experiment 1. The selection of categories for transfer (and retention) was also the same as that used in Experiment 1 with the constraint that the transfer categories were highly, moderately, or unrelated to the trained exemplars. During training, participants searched for target words (eight exemplars from a single category) against a background of distractor words (exemplars from six categories semantically unrelated to the target categories). During transfer, four new target categories were presented (six exemplars per category), as well as six new exemplars from the category on which participants trained. Also, to minimize confounding of target learning with distractor learning (Dumais, 1979; Fisk & Rogers, 1990; Kristofferson, 1977; Rogers, 1989), 48 exemplars from six new distractor categories were used during the transfer sessions.

Procedure. To test performance at the limits of each individual's visual search capacity, we developed an adaptive version of the "multiple-frame" detection task for the training phase of this experiment. This task was based upon multiple-frame tasks reported in the visual search/detection literature (e.g., Schneider & Shiffrin, 1977; Sperling et al., 1971). However, in our version of the task, frame time (the time from the onset of one display until the onset of the next display) was determined by each subject's individual accuracy.

All participants began the experiment at the same "speed," with frame time equal to 850 ms. If a participant's accuracy on any block was equal to or better than 86 percent correct (26 or more correct out of a total of 30 trials), frame time on the next block was decreased by 25 ms. If accuracy fell below 76 percent (23 or fewer correct), frame time on the next block was increased by 25 ms: otherwise frame time remained the same. Results from pilot testing indicated that this allowed accuracy to stabilize around 80 percent correct. Frame times for an individual's transfer sessions were derived using his or her mean frame time for the final two training sessions. Thus, frame time was held constant during transfer and retention phases with accuracy being the dependent measure for those sessions.

A representation of a single, multiple-frame trial is provided in Figure A-2. At the beginning of each trial, participants studied a memory set (a single semantic category) for a maximum of 20 sec. The subject initiated presentation of the frames by pressing the space bar. Each "frame" consisted of two displays presented sequentially. The first display of each frame consisted of three words displayed in a column. The second display of the frame

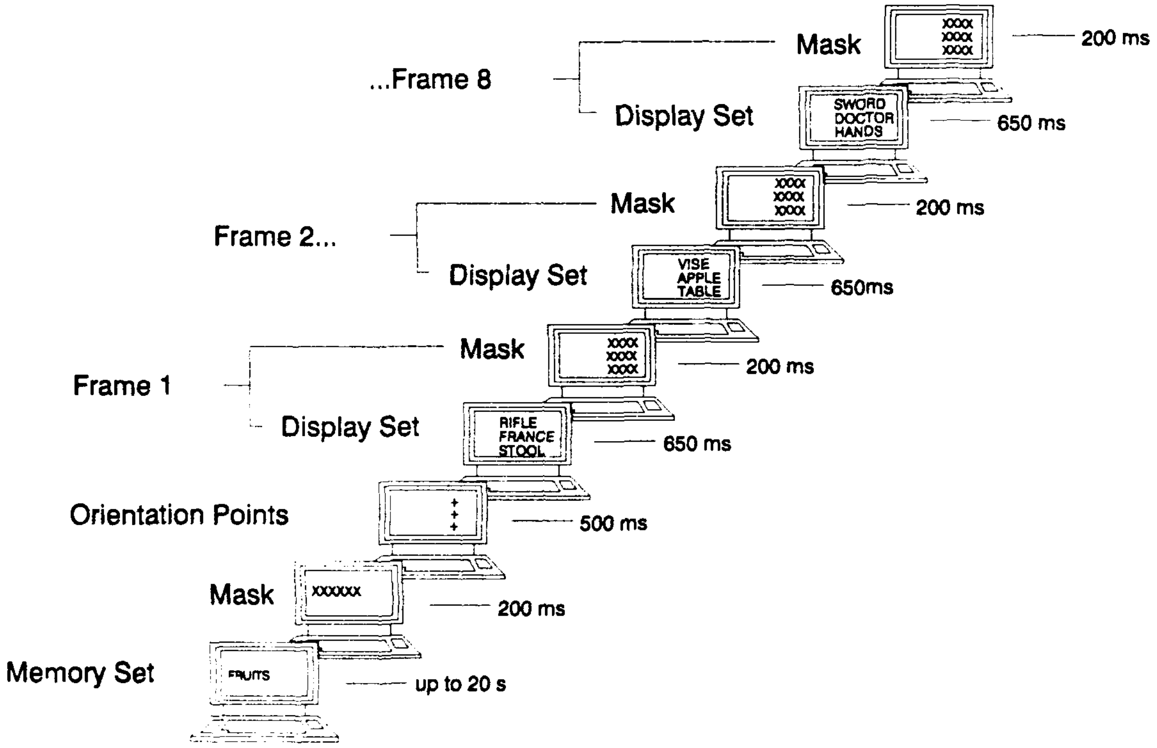


Figure A-2. A Representation of a Trial in the Multiple Frame Procedure. In this representation, frames 3 through 7 are omitted. The target, "APPLE", appears in the middle position on frame number 2.

contained a visual mask consisting of three rows of X's to prevent continued processing of the display set.

In this study, eight frames per trial were used. Each sequence of frames was presented following a 500 ms display of focus points (three "plus" signs (+) displayed in a column where the exemplars were to be displayed). Frame time was measured from the onset of display of one frame to the onset of the next frame (a zero interframe interval). While presentation time for the display set varied across blocks as a function of an individual's accuracy, presentation time of the visual mask remained constant at 200 ms.

Participants searched through 24 exemplars (eight frames x three exemplars per frame) to find a target. There were two kinds of trials: target present (positive trials) and target absent (negative trials). On positive trials one exemplar from the target category appeared in only one frame. The target could appear in Frames 2 through 7 (never Frame 1 or 8) in either the top, middle, or bottom position on the display. Both frame number and vertical position were selected randomly. If the trial was positive, the correct response was to press a key labeled T, M or B (corresponding to the 7, 4 or 1 keys on the numeric keypad) depending on the vertical location of the target exemplar. If the trial was negative, the correct response was to press a key labeled N (corresponding to the 5 key on the numeric keypad).

Participants could respond at any point during presentation of the frames and for up to 4 seconds after the final frame. Following the response, the display was cleared and feedback for that trial was presented. After each trial, participants received correlated visual and auditory feedback about their response. On correct responses the word "CORRECT!" was displayed. If the participant "missed" the target, then the message "ERROR, _exemplar_ was presented in _position_" (where _exemplar_ was the actual target word and position was the actual vertical position of the target for that trial) was displayed at the target location, simultaneously with presentation of a 1,200-Hz tone. If the participant "false-alarmed," then the microcomputer displayed "ERROR, there was no target present" in the right center of the screen, simultaneously with presentation of a 100-Hz tone. If the participant made an "error of position," then the microcomputer displayed "ERROR, exemplar was present in position" at the target location, simultaneously with presentation of a 500 Hz tone.

At the end of each block, participants received feedback and had an opportunity to take a break (and were encouraged to do so). First, information about performance on the just-completed block was displayed for 7 seconds. Then, cumulative feedback representing the individual's performance on each block was displayed. When a participant finished viewing the feedback screen he or she pressed the space bar to initiate the next block of trials.

Results and Discussion

Training. Subjects improved in this task in a manner similar to other consistent mapping training procedures. Frame times decreased from the initial 850 ms to an average of 165 ms by Session 15. The improvement in search performance, measured by decreasing frame time, was significant, $F(14,124) = 208.32$.

Transfer and Retention. The transfer data are shown in Figure A-3. Accuracy data from the transfer sessions (Sessions 16 and 17) were aggregated and analyzed with a one-way, within-subjects analysis of variance. There was a significant effect of transfer condition, $F(4,45) = 18.54$. A Newman-Keuls test revealed that performance in the T/T condition was superior to all other conditions and the T/U condition was more accurate than in both the MR and UR conditions. Performance in the HR condition was more accurate than for the UR condition. There were no

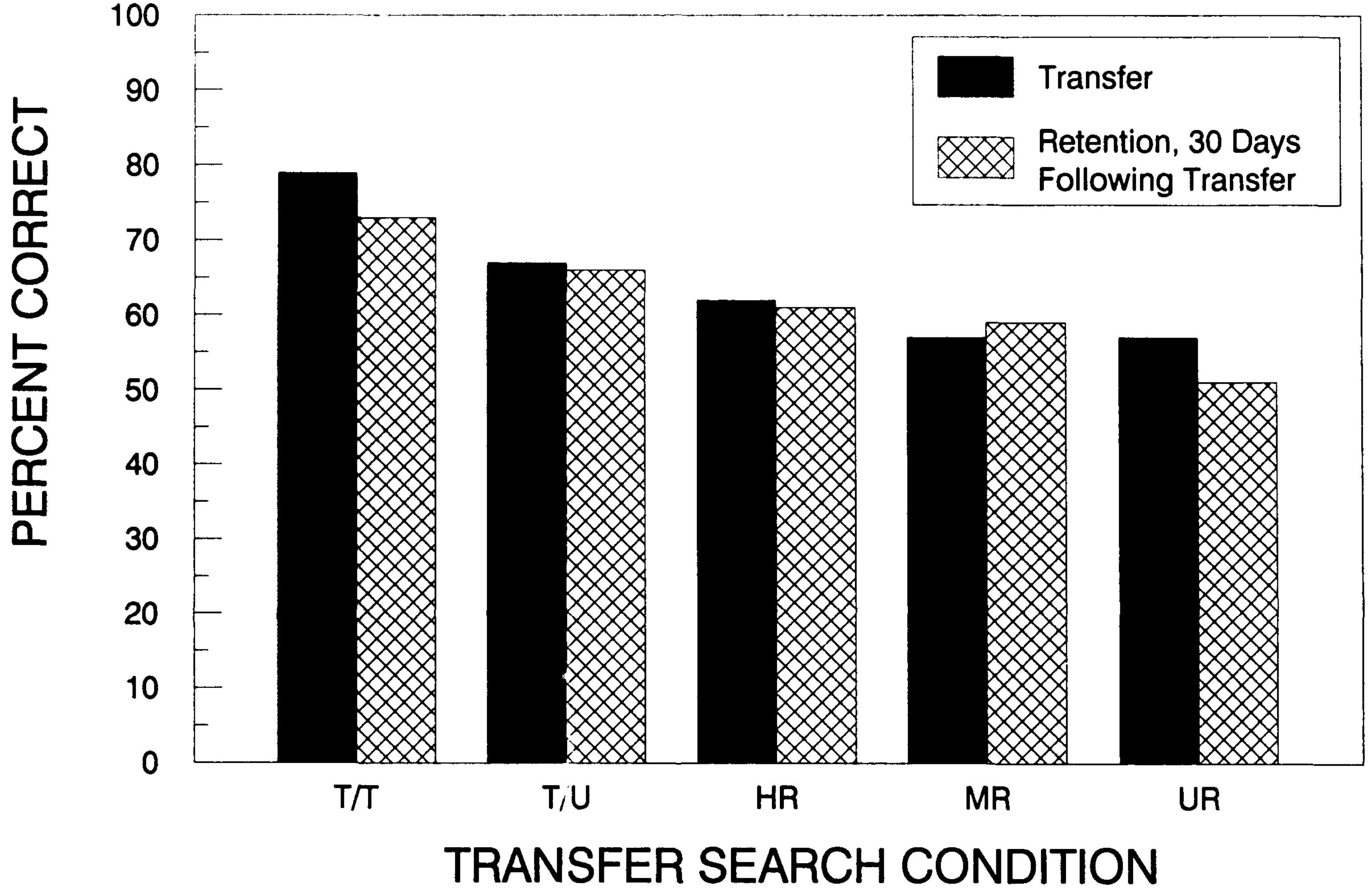


Figure A-3. Mean Accuracies are Plotted as a Function of Search Conditions. The graph compares performances at transfer with performance 30 days later. Please see text for description of X-axis labels.

significant differences between T/U and HR, HR and MR, or MR and UR.

Retention accuracy, 30 days following the final transfer session, is also shown in Figure A-3. An examination of Figure A-3 shows that there was very little decay in performance across the conditions. The difference in accuracy between transfer and retention was 7, 2, 3, 0, and 3 percentage points for the TT, TU, HR, MR, and UR conditions, respectively. A Search Condition X Transfer versus Retention (i.e., Session) ANOVA revealed a main effect of Search Condition, $F(4, 36) = 38.99$; however, there was no effect of Transfer versus Retention (no session effect) and no Search Condition X Session interaction, Fs < 1 in both cases. This ANOVA would suggest that there was no decay in performance, although this is somewhat misleading. The TT condition did show the most decay (in terms of difference score) and when individual comparisons are made between Transfer and Retention performance for each condition, only the TT condition produced minimal but significant decay, $F(1,9) = 6.01$. None of the other comparisons reached significance.

Discussion

The present results provide some support for the position that perceptual tuning does decay over a 1 month retention interval and seem to support our interpretation of Kolers' (1976) retention data. The performance decay observed in our experiment, when contrasted with data collected by Healy et al. (1990), suggest that only when a sensitive test of "perceptual tuning" is used will performance decay effect be observed. However, although a statistically significant decay was found for the explicitly trained stimuli, that decay was modest. As such, these results lend some support to the Healy et al. suggestion of "remarkable durability of the perceptual skill."

Experiment 3 - Hybrid Memory/Visual Search

Rabbitt, Cumming, and Vyas (1979) found significant decay in performance when subjects were tested 6 weeks subsequent to CM training. The task used by Rabbitt et al. was a hybrid memory/visual search task. In their task subjects searched a display of nine letters for any one of five memory-set elements (hence, subjects were required to search both memory and the display). The task used by Rabbitt et al. was more complex than the task used by Healy et al. (1990) in terms of information processing components (Schneider & Shiffrin, 1977). The decay found by Rabbitt et al. was greater than that observed in our Experiment 2. Given the lack of performance decay in our memory search experiment (Experiment 1), the modest decay in our pure visual search experiment (Experiment 2) and the lack of decay found in the Healy et al. visual search task, it is important to examine retention performance in a hybrid memory/visual search task. Hence, in the final experiment we examined the decay characteristics in a task similar to that used by Rabbit et al. but with stimuli consistent with our first two experiments. We manipulated memory-set size so that within the experiment we could simultaneously examine pure visual search (Memory-set size 1 and Display size 3) corresponding to the Healy et al. experiment as well as hybrid memory/visual search (memory-set size greater than 1, display size 3) corresponding to the Rabbitt et al. experiment. We also examined performance stability beyond the 30-day retention interval by also testing subjects at intervals of 90, 180, and 365 days.

Method

Participants. Twelve volunteers (mean age 25.8 years, six males, six females) completed the experiment. Ten were graduate students in psychology at the Georgia Institute of Technology and two were undergraduates. Participants were tested for corrected or uncorrected far vision of at least

20/30 and near vision of at least 20/40 and were paid for participation.

Equipment. The equipment was the same as described in Experiment 1 except that the '7', '4' and '1' keys on the numeric keypad were labeled 'T', 'M' and 'B' respectively, to indicate top, middle and bottom (mapping to target positions on the display).

Design. The experiment consisted of three phases: training, transfer, and retention. In each phase, all manipulations were within-subject and within-block. In the training phase, there were two factors of interest: search condition and memory-set size. Display-set size was constant at three. There were four search conditions (a) high amount of CM training (CM High, 4320 trials); (b) moderate amount of CM training (CM Moderate, 2160 trials); (c) low amount of CM training (CM Low, 720 trials); and (d) VM training (VM, 720 trials). Memory-set size varied from one to three items. There was a target exemplar present on every trial. There were three "target" categories associated with each CM condition. Six categories were used in the VM condition: Exemplars from these served as both targets and distractors. The six categories associated with the VM condition also served as "distractor" categories for CM conditions. Assignment of categories to participants was counterbalanced by a partial Latin-square. There were 12 sessions lasting an average of 40 minutes each. There were 20 blocks per session and 33 trials per block.

During transfer and retention, a new variable was added: exemplar type (trained versus untrained exemplars from the trained categories). In the untrained exemplar conditions, four new exemplars were added to each of the trained categories. There were four retention intervals: 30, 90, 180, and 365 days following training. During the single transfer session and for each retention test session, the participants received 480 trials (60 per condition).

Prior to each retention session, participants received six short blocks of "response" practice. This practice took approximately 15 minutes and was provided to allow participants to orient to the experimental environment and task (e.g., practice which keys to press). Categories and exemplars were semantically unrelated to those on which participants trained and to those on which they were tested during retention.

Procedure. Each trial proceeded as follows. The memory set (one, two or three category labels) was displayed in the left center of the VDT screen at the beginning of each trial. Participants could study the memory set for up to 20 sec. To view the display set, participants pressed the space bar. An orientation display consisting of three '+' signs was presented for 500 ms in the same location as the display set to allow the participant to focus his or her gaze. Then the display set, consisting of three words in a column, was presented. The participant's task was to identify the target (i.e., an exemplar from one of the categories in the memory set) and to indicate its location (top, middle or bottom) by pressing the corresponding key (labeled 'T', 'M' or 'B') on the keyboard. Participants were allowed a maximum of 6 sec to enter their responses. Participants received performance feedback as described in Experiment 1.

Stimuli. Fifteen semantically unrelated, taxonomic category labels (Collen et al., 1975) from the Battig and Montague (1969) category norms were used as memory set items in the training, transfer, and retention phases of the experiment. Six exemplars from each category were used during training and four new exemplars were introduced during the transfer and retention phases. Exemplars were selected according to the criteria described in Experiment 1.

Results

Training Results. An important question concerns whether there were differences in CM performance due to the differential amounts of training. Comparison of RT means (correct trials only) from Session 12 revealed that the CM High condition was faster (M = 650 ms) than CM Low RT [M = 727 ms, F(1, 11) = 6.61; CM Moderate was faster (673 ms) than CM Low [F(1,11) = 5.59; and CM Low was faster than VM [M = 1001 ms, F(1, 11) = 63.09. RT performance in the CM High condition was slightly faster than CM Moderate but the difference was not significant [F(1, 11) = 3.10. In general, then, performance was positively related to amount of CM training. Accuracies were stable across practice averaging 96 percent, 95 percent, and 95 percent for CM High, Medium, and Low training, respectively. Accuracy in the VM condition (84 percent) was lower than any CM condition.

An examination of comparison slope estimates provides additional evidence that increased CM training led to superior performance. These estimates describe the function that relates RT to the number of comparisons required to make the correct decision (the product of the number of items to be held in memory and the number of items to be searched in the display set). At Session 12, the comparison slope estimates for CM High, CM Moderate, CM Low and VM were 6.2 ms, 11 ms, 16.6 ms and 53.9 ms, respectively.

Retention Results: Trained Exemplars. Mean reaction times as a function of retention interval (for all conditions and collapsed across memory-set size) are presented in Figure A-4. Critical data for this investigation involve the pattern of RT performance decay for trained exemplars across search conditions and retention intervals. A comparison of mean RTs across search conditions revealed that, at Day 1 (Performance for the session one day following training is denoted as Day 1), CM High performance was faster than CM Low, F(1,11) = 5.06; CM

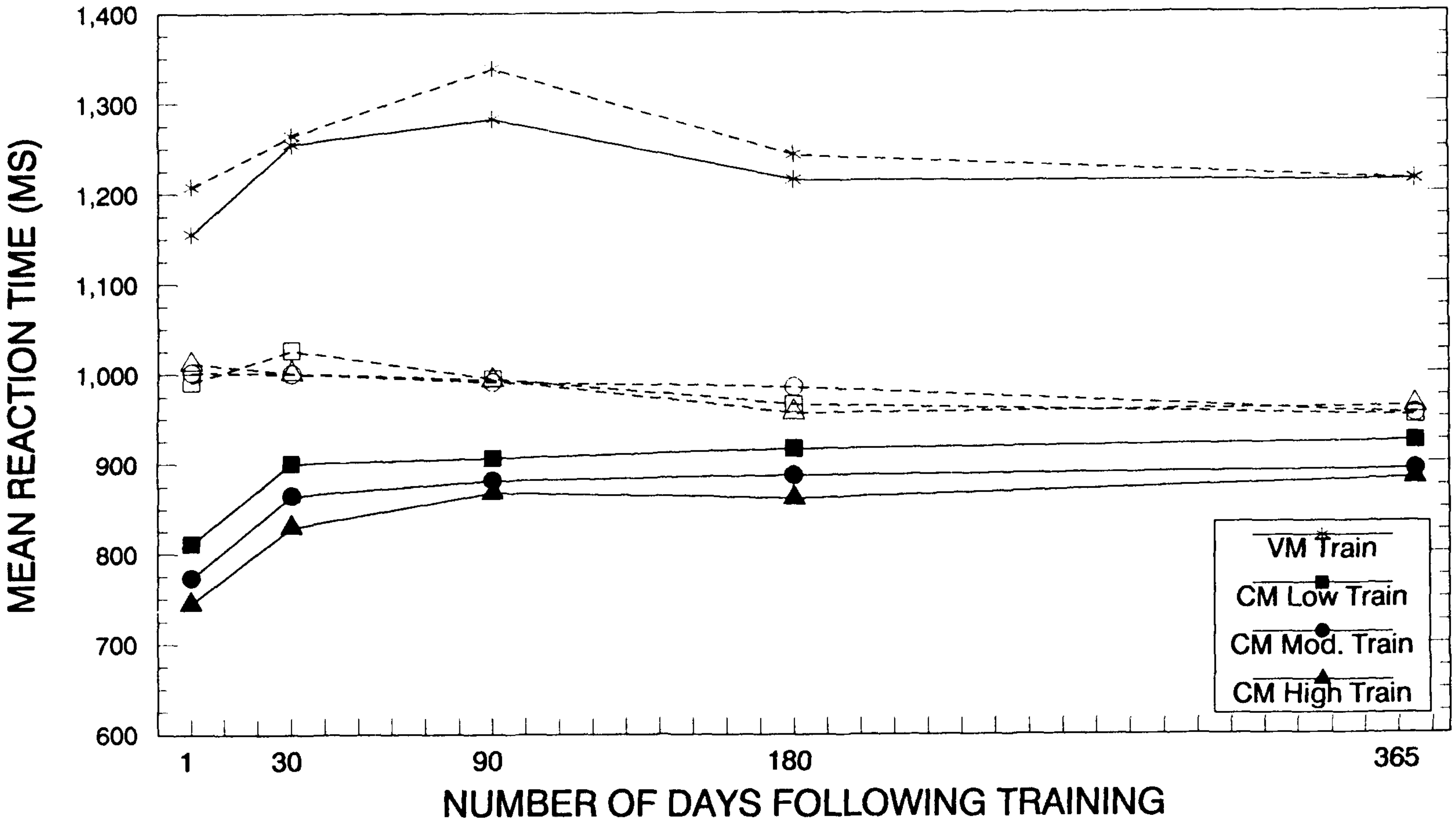


Figure A-4. Mean RTs (Correct Trials) are Plotted as a Function of Days Following Training. Performance on the trained exemplars is represented by the solid lines and performance on the untrained exemplars from the trained categories is represented by the dashed lines.

Moderate was faster than CM Low, F(1,11) = 5.59; and CM Low was faster than VM F(1,11) = 63.09.

Within the first 30 days following training, performance had declined 15 percent, 14 percent, and 12 percent for the CM High, Moderate, and Low training conditions, respectively. Following Day 1 there were no statistically significant differences between the CM High, Moderate or Low conditions. However, all CM conditions remained superior to VM across all retention sessions. Comparison of CM RTs across retention intervals revealed that performance in the CM High condition at Day 1 was faster than that in CM High at Day 30, F(1, 11) = 45.89. From Day 30 on, however, performance in the CM conditions did not vary significantly from one retention interval to the next. CM High performance at Day 30 was not significantly different from performance in CM High at Day 90, F(1,11) = 3.33; CM High at Day 90 is not significantly different from CM High at day 180, F < 1. Finally, CM High at Day 180 does not differ from performance on Day 365, F < 1. This pattern holds true for the CM Moderate and CM Low conditions. VM performance was erratic: from Day 1 through Day 90 performance was slower than baseline and from Day 90 to Day 365 performance was not different from Day 1.

Decay as a Function of Comparison Load. Given the significant decline in performance in all three CM conditions for the trained CM stimuli, we examined retention performance as a function of memory-set size across the CM conditions. The interaction between memory-set size and retention interval was significant, F(8,88) = 2.41. The three-way interaction among memory-set size, session, and amount of CM training was not significant, F < 1. The source of the significant interaction was due to the larger increase in performance decay as memory-set size (and hence comparison load) increased. For memory-set size one (pure visual search) there was a nonsignificant 8-percent decrease in performance, F(11,33) = 1.68. Search performance for memory-set size two and memory-set size three (hybrid memory/visual search) declined 13 percent and 18 percent, respectively; both of these declines were significant, F(11,33) = 2.74 and F(11,33) = 4.09; respectively. These results appear to reconcile the apparent discrepancy between the Rabbitt et al. (1979) and the Healy et al. (1990) retention data. The former researchers found significant performance decay and the latter investigators reported no decline in performance. The present results show the same pattern; that is, the significant decline after the 30-day retention interval is localized in our hybrid memory/visual search conditions (Rabbitt et al. paradigm) and not in the pure visual search condition (Healy et al. paradigm). This performance pattern occurred using the same stimulus material at the same retention interval which rules out several alternative explanations.

Retention Results for Transfer Conditions. RT performance for the untrained exemplars from the trained CM categories was marked by stability. Performance on the untrained exemplars from the trained VM condition shadowed the trained VM exemplars (i.e., performance was not stable). There are no statistically significant differences between the CM High, CM Moderate or CM Low transfer conditions at any retention interval. All CM transfer conditions were superior to both VM trained and VM transfer conditions across all retention intervals.

Retention Results: Accuracy Data. An overall ANOVA was performed on the accuracy data. The main effects of search condition F(3,33) = 18.99, memory-set size F(2,22) = 49.54, and training, F(1, 11) = 50.62 were significant. A Newman-Keuls test revealed no differences among CM conditions but the VM conditions were less accurate than any CM condition. There was no effect of retention interval F(3,44) = 1.92, indicating that accuracy across retention intervals was quite stable.

Discussion

There are four critical results from this experiment: (a) detection of both trained and untrained exemplars from the trained CM categories was superior to the VM conditions at all retention intervals; (b) trained CM conditions exhibited the greatest decrement in performance within 30 days following training, but after this initial decline, CM performance remained relatively stable; (c) the CM decline was largely due to performance in the hybrid memory/visual search conditions; and (d) the original ordering of performance levels produced by differential amounts of training was maintained at each retention interval, although the statistically significant differences among the trained CM conditions disappeared within 30 days.

The decline in performance on the CM trained exemplars notwithstanding, the present data suggest the remarkable stability of CM performance superiority relative to VM performance. The fact that CM performance remained superior to VM performance throughout the entire retention interval should not be lost in the discussions of performance decay over time.

The superiority of the untrained elements from the trained categories (the CM transfer conditions) to VM performance over the entire retention interval and the lack of decay in those CM conditions lend converging support to the findings of Experiment 1. In Experiment 1, we found no decay in CM-trained memory search. We interpret these data as suggesting the extreme stability of automatic access of well-trained, associatively connected semantic memory. The memory access data support previous investigations of the stability of codification, unification, or chunking (Salasoo et al., 1985).

Perhaps the most interesting finding from Experiment 3 is the decay in CM performance as a function of type of search (i.e., pure visual search versus hybrid memory/visual search). We found a nonsignificant decay in performance when we examined pure visual search, which replicates the Healy et al. (1990) experimental results. It is Important to note that when we examined the hybrid memory/visual search conditions, which conceptually replicate the Rabbitt et al. (1979) experimental design, we find significant decay in performance. These findings must be tempered somewhat in light of the Experiment 2 results which did show a small, but statistically significant decline in visual search performance. Clearly, situations can be created that will result in performance decay in visual search across retention intervals; however, those situations seem to be related to the need for extremely fine perceptual tuning.

The pattern of results demonstrated across the three experiments perhaps may be interpreted best within the context of a componential analysis of the processes underlying the complex hybrid memory/visual search task used in Experiment 3. The results of Experiment 1 reveal that access to automatized semantic memory search processes is not disrupted significantly (less than 2 percent) by an initial retention interval of 32 days. Further, a similar stability of component processes was revealed in Experiment 2, using a visual search paradigm. A performance decrement of less than 8 percent was demonstrated, a decrement which, although statistically significant, is considerably less than the large diminution in performance produced by aggregation of the two task components in the hybrid paradigm of Experiment 3 (18 percent decline for Memory-set size three, Display size three). The decline in retention performance yielded in the hybrid visual/memory search task cannot be solely attributable to the demonstrated decline in the visual search component nor to that demonstrated by the memory search component. Apparently an additional degree of complexity is present in the hybrid task, a complexity that is absent in either of the individual components.

In the hybrid memory/visual search task, an increasing level of integration of the mechanisms associated with visual and memory search components may be required (Logan, 1985; Schneider & Shiffrin, 1977). With sufficient CM training, the integration between automatic and controlled processes is facilitated (Logan, 1978; Schneider & Detweiler, 1988). However, it is possible that periods of inactivity produce an increasing demand upon the integrative mechanism associated with the control structure; hence, the substantial decline in performance. Models in which memory is accessed by the spreading of a limited amount of activation--a model such as ACT*--may produce a superadditive interaction between the difficulty of individual accesses and the number of accesses required. If this were the case then undetectable small main effects could combine to become detectable. Our present data cannot rule out this possibility; however, if difficulty (and not complexity) were the source of the hybrid memory/visual search results found in Experiment 3, then we would not expect the same pattern of data for our pure visual search results seen between Experiments 2 and 3 or between Experiment 3 (pure visual search) and the Healy et al., (1990) findings.

Given that the decline in performance stabilizes at approximately 30 days following training, it should be possible to predict longer-term performance decrements based upon performance at the 30-day mark. This predictive capability would be valuable for gauging performance levels across different time spans in a variety of tasks which draw upon both visual and memory search components. The basis for many skilled activities (e.g., in cardiopulmonary resuscitation) is to provide training on tasks that remain unused except in emergencies. Identification of the trade-off among amount of training, initial level of performance following training, and level of performance after various periods of delay without practice will allow a more precise assessment of "skill readiness." The present data may also serve to elucidate understanding of the effects of time without practice on skilled performance, an understanding that is essential to any effort to predict performance after a period of inactivity or establish which skill components to emphasize during training or instruction.

References

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. Journal of Experimental Psychology: General, 108, 296-308.

Bahrick, H. P. (1983). The cognitive map of a city -- 50 years of learning and memory. In G. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory, Vol 17 (pp 125-163). New York: Academic Press.

Bahrick, H. P. (1984). Semantic memory in permastore: Fifty years of memory for Spanish learned in school. Journal of Experimental Psychology: General, 113, 1-29.

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memories for names and faces: A cross-sectional approach. Journal of Experimental Psychology: General, 104, 54-75.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 80.

Collen, A., Wickens, D. D., & Daniele, L. (1975). The interrelationship of taxonomic categories. Journal of Experimental Psychology: Human Learning and Memory, 1, 629-633.

Dumais, S. T. (1979). Perceptual learning in automatic detection: Processes and mechanisms. Unpublished doctoral dissertation, Indiana University. Bloomington, IN.

Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology. (Translation by H. A. Rugers & C. E. Bussenius). New York: Dover. (Original work 1885, original translation 1913.)

Farr, M. J. (1987). The long-term retention of knowledge and skills: A cognitive and instructional perspective. New York: Springer-Verlag.

Appendix A (continued)

Fisher, D. L., Duffy, S. A. Young, C. & Pollatsek, A. (1988). Understanding the central processing limit in consistent-mapping visual search tasks. Journal of Experimental Psychology: Human Perception and Performance, 14, 253-266.

Fisk, A. D., & Rogers, W. A. (1990). Toward an understanding of age-related memory and visual search effects. Manuscript submitted for publication.

Flach, J. M. (1986). Within-set discriminations in a consistent mapping search task. Perception & Psychophysics, 39, 397-408.

Gentile, J. R., Monaco, N., Iheozor-Ejiofor, I. E., Ndu, A. N., & Ogbonaya, P. K. (1982). Retention by "fast" and "slow" learners. Intelligence, 6, 125-138.

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. Human Factors, 25, 199-213.

Healy, A. F., Fendrich, D. W., & Proctor, J. D. (1990). Acquisition and retention of a letter-detection skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 16, 270-281.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. Perception & Psychophysics, 23, 1-11.

Hoffman, J. E. (1979). A two-stage model of visual search. Perception & Psychophysics, 25, 319-327.

Johnson, S. L. (1981). Effect of training device on retention and transfer of a procedural task. Human Factors, 23, 257-272.

Kolers, P. A. (1976). Reading a year later. Journal of Experimental Psychology: Learning, Memory, and Cognition, 5, 554-565.

Kristofferson, M. W. (1977). The effects of practice with one positive set in memory scanning can be completely transferred to a different positive set. Memory & Cognition, 5, 177-186.

Kyllonen, P. C., & Tirre, W. C. (1988). Individual differences in associative learning and forgetting. Intelligence, 12, 393-421.

Logan, G. D. (1978). Attention in character classification: Evidence for the automaticity of component stages. Journal of Experimental Psychology: General, 107, 32-63.

Appendix A (continued)

Logan, G. D. (1985). Executive control of thought and action. Acta Psychologica, 60, 193-210.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Mengelkoch, R. F., Adams, J. A., and Gainer, C. A. (1971). The forgetting of instrument flying skills. Human Factors, 13, 397-405.

Rabbitt, P. M. A., Cumming, G., & Vyas, S. (1979). Improvement, learning, and retention of skill at visual search. Quarterly Journal of Experimental Psychology, 31, 441-459.

Rogers, W. A. (1989). Target and distractor learning in visual search: Age-related differences. Unpublished Master's thesis, Georgia Institute of Technology, Atlanta, GA.

Rosenbloom, P. S., & Newell, A. (1986). The chunking goal hierarchies: A generalized model of practice. In R. S. Michalski, J. G Carbonell, & T. M. Mitchell (Eds.), Machine learning: An artificial intelligence approach (Vol 2, pp. 247-288). Los Altos, CA: Morgan Kaufmann.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. Journal of Experimental Psychology: General, 114, 50-77.

Schneider, W. (1985). Toward a model of attention and the development of automaticity. In M. I. Posner & O. S. Martin (Eds.), Attention & Performance Volume XI (pp 475-492). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schneider, W., & Detweiler, M. (1987). A connectionist/ control architecture for working memory. In G. H. Bower (Ed.), The psychology of learning and motivation, Volume 21. New York: Academic Press.

Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. Human Factors, 30, 539-566.

Schneider, W., & Fisk, A.D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. Perception and Psychophysics, 31, 160-166.

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

Appendix A (continued)

Schneider, W., & Shiffrin, R. M. (1985). Categorization (restructuring) and automatism: Two separable factors. Psychological Review, 92, 424-428.

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

Shuell, T. J., & Keppel, G. (1970). Learning ability and retention. Journal of Experimental Psychology, 61, 59-65.

Sperling, G., Budiansky, J., Spivak, J. G., & Johnson, M. C. (1971). Extremely rapid visual search: The maximum rate of scanning letter for the presence of a numeral. Science, 174, 307-311.

Underwood, B. J. (1954). Speed of learning and amount retained: A consideration of methodology. Psychological Bulletin, 51, 276-282.

APPENDIX B: CATEGORIES AND EXEMPLARS USED IN EXPERIMENT 1 AND EXPERIMENT 2
(EXPERIMENTAL SERIES 2)

Target Exemplars:

FRUITS
APPLE
ORANGE
LEMON
PEACH
GRAPE
CHERRY

OCCUPATIONS
DOCTOR
FARMER
JUDGE
CLERK
LAWYER
DENTIST

BODY PARTS
ANKLE
LIVER
HEART
MOUTH
NOSE
HEAD

COUNTRIES
ITALY
FRANCE
ENGLAND
JAPAN
SWEDEN
NORWAY

CLOTHING
SHIRT
PANTS
JACKET
BLOUSE
DRESS
SWEATER

MUSICAL INSTRUMENTS
TUBA
CELLO
TRUMPET
HARP
GUITAR
FLUTE

Distractor Exemplars:

READING
BOOK
NOVEL
PAPER
ARTICLE
LETTER
ESSAY

DWELLINGS
CABIN
TENT
SHACK
HOTEL
HOME
MANSION

FURNITURE
SOFA
DESK
TABLE
CHAIR
COUCH
LAMP

RELATIVES
AUNT
MOTHER
SISTER
COUSIN
NEPHEW
NIECE

TIME
HOUR
WEEK
YEAR
DECADE
CENTURY
SECOND

EARTH FORMS
CANYON
ISLAND
RIDGE
VALLEY
OCEAN
PLATEAU

VEHICLES
BOAT
AUTO
SHIP
TRUCK
TAXI
BICYCLE

WEAPONS
SWORD
PISTOL
KNIFE
BOMB
RIFLE
ARROW

Appendix B (continued)

CATEGORIES AND EXEMPLARS USED IN EXPERIMENT 3 AND EXPERIMENT 4: SET 1 (EXPERIMENTAL SERIES 2)

Target Exemplars:

FRUITS
APPLE
LIME
LEMON
PEACH
GRAPE
CHERRY

OCCUPATIONS
DOCTOR
FARMER
JUDGE
DENTIST
TEACHER
CHEMIST

BODY PARTS
ANKLE
HEART
CHEST
LUNGS
FINGER
STOMACH

COUNTRIES
ITALY
FRANCE
ENGLAND
SWEDEN
NORWAY
GERMANY

CLOTHING
PANTS
BLOUSE
DRESS
SWEATER
SOCKS
GLOVES

MUSICAL INSTRUMENTS
TUBA
TRUMPET
PIANO
GUITAR
FLUTE
ORGAN

Distractor Exemplars:

TOOLS
HAMMER
CROWBAR
CHISEL
WRENCH
PLIERS
DRILL
LATHE
SANDER

BUILDING PARTS
WINDOW
ATTIC
FLOOR
CEILING
STAIRS
CHIMNEY
CLOSET
CELLAR

VEHICLES
BOAT
AUTO
SHIP
TRUCK
TAXI
BICYCLE
TRAIN
TRACTOR

WEAPONS
SWORD
PISTOL
KNIFE
BOMB
RIFLE
ARROW
CANNON
SPEAR

METALS
IRON
COPPER
STEEL
GOLD
SILVER
BRASS
BRONZE
NICKEL

COLOR
GREEN
YELLOW
BLACK
PURPLE
WHITE
BROWN
VIOLET
INDIGO

Appendix B (continued)

CATEGORIES AND EXEMPLARS USED IN EXPERIMENT 3 AND EXPERIMENT 4: SET 2 (EXPERIMENTAL SERIES 2)

Target Exemplars:

FRUITS
APPLE
LIME
LEMON
PEACH
GRAPE
CHERRY

OCCUPATIONS
DOCTOR
FARMER
JUDGE
DENTIST
TEACHER
CHEMIST

BODY PARTS
ANKLE
HEART
CHEST
LUNGS
FINGER
STOMACH

COUNTRIES
ITALY
FRANCE
ENGLAND
SWEDEN
NORWAY
GERMANY

CLOTHING
PANTS
BLOUSE
DRESS
SWEATER
SOCKS
GLOVES

MUSICAL INSTRUMENTS
TUBA
TRUMPET
PIANO
GUITAR
FLUTE
ORGAN

Distractor Exemplars:

READING
BOOK
NOVEL
PAPER
ARTICLE
LETTER
ESSAY
JOURNAL
LEAFLET

DWELLINGS
CABIN
TENT
SHACK
HOTEL
HOME
MANSION
CASTLE
TRAILER

SPORTS
TENNIS
SOCCER
HOCKEY
GOLF
ARCHERY
SKIING
BOWLING
RUGBY

RELATIVES
AUNT
BROTHER
SISTER
COUSIN
NEPHEW
NIECE
UNCLE
WIFE

TIME
HOUR
WEEK
YEAR
DECADE
CENTURY
SECOND
MINUTE
MONTH

EARTH FORMS
CANYON
ISLAND
RIDGE
VALLEY
OCEAN
PLATEAU
CLIFF
DESERT

APPENDIX C: FREQUENCY DATA FOR EXPERIMENTAL SERIES 2

Table C-1. Frequency of a Target for each Training Condition and Target Position for Experiment 1, Session 7.

Condition \ Target Position	Top	Middle	Bottom	No
PT2	106	108	108	100
PT3	104	108	107	101
WT6	105	107	107	101

Chi-Square$_6$ < 1, for Training Condition by Target Position

Table C-2. Frequency of a Target for each Training Condition and Target Position for Experiment 1, Session 7.

Condition \ Frame Number	2	3	4	5	6	7
PT2	71	72	69	70	71	69
PT3	70	72	70	70	72	66
WT6	69	71	70	69	70	71

Chi-Square$_{10}$ < 1, for Training Condition by Frame Number

Appendix C (continued)

Table C-3. Frequency of a Target for each Training Condition and Category for Experiment 1, Session 7.

Condition \ Frame Number	1	2	3	4	5	6
PT2	99	101	101	97	104	106
PT3	103	97	98	100	99	105
WT6	101	105	101	103	91	104

Chi-Square$_{10}$ = 1.5, p = .999 for Training Condition by Category

1 = Fruits
2 = Occupations
3 = Body Parts
4 = Countries
5 = Clothing
6 = Musical Instruments

Appendix C (continued)

Table C-4. Frequency of a Target for each Training Condition and Target Position for Experiment 1, Session 14.

Condition \ Target Position	Top	Middle	Bottom	No
PT2	107	108	106	104
PT3	105	103	107	99
WT6	105	107	107	103

Chi-Square$_6$ < 1, for Training Condition by Target Position

Table C-5. Frequency of a Target for each Training Condition and Target Position for Experiment 1, Session 14.

Condition \ Frame Number	2	3	4	5	6	7
PT2	69	71	72	71	71	71
PT3	70	70	67	71	70	66
WT6	70	72	70	70	71	69

Chi-Square$_{10}$ < 1, for Training Condition by Frame Number

Table C-6. Frequency of a Target for each Training Condition and Category for Experiment 1, Session 14.

Condition \ Frame Number	1	2	3	4	5	6
PT2	101	102	104	96	100	97
PT3	104	98	101	98	98	105
WT6	100	101	103	103	101	102

Chi-Square$_{10}$ < 1 for Training Condition by Category

1 = Fruits
2 = Occupations
3 = Body Parts
4 = Countries
5 = Clothing
6 = Musical Instruments

APPENDIX D: CERTAINTY SCALE DATA, EXPERIMENTAL SERIES 2

Table D-1. Frequency of Hits for each Training Condition as a Function of Frame Speed and Transfer Session for Experiment 1.

	PT2			PT3			WT6		
Frame Speed	180	220	260	180	220	260	180	220	260
<u>Transfer I</u> Scale Value									
5	231	277	301	222	273	294	259	301	317
4	48	37	33	24	28	26	34	30	22
3	18	12	15	21	12	9	5	5	4
2	2	0	0	0	1	1	0	0	1
1	0	0	0	0	0	0	0	0	0
<u>Transfer II</u> Scale Value									
5	258	288	296	255	267	305	267	294	339
4	44	37	33	41	43	23	49	45	26
3	16	24	25	7	14	6	16	10	5
2	1	0	0	0	0	0	0	0	1
1	0	0	0	0	1	0	0	0	0

5 = Absolutely Certain a Target Present
3 = Guess
1 = Absolutely Certain NO Target Present
(4 and 2 did not have labels, but represented a response between a guess and Absolute certainty)

<u>Appendix D</u> (continued)

Table D-2. Frequency of False Alarms for each Training Condition as a Function of Frame Speed and Transfer Session for Experiment 1.

	PT2			PT3			WT6		
Frame Speed	180	220	260	180	220	260	180	220	260
<u>Transfer I</u> Scale Value									
5	12	8	5	10	12	6	10	6	3
4	5	4	2	8	8	10	7	9	6
3	8	7	7	13	8	10	10	3	1
2	2	0	0	0	0	1	0	0	0
1	2	1	1	0	0	0	0	0	0
<u>Transfer II</u> Scale Value									
5	10	2	3	10	7	11	3	1	5
4	3	1	5	12	14	10	5	6	6
3	10	3	8	5	7	4	8	7	10
2	1	0	0	1	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0

5 = Absolutely Certain a Target Present
3 = Guess
1 = Absolutely Certain NO Target Present
(4 and 2 did not have labels, but represented a response between a guess and Absolute certainty)

Appendix D (continued)

Table D-3. Frequency of Correct Rejections for each Training Condition as a Function of Frame Speed and Transfer Session for Experiment 1.

	PT2			PT3			WT6		
Frame Speed	180	220	260	180	220	260	180	220	260
Transfer I Scale Value									
5	3	0	0	0	1	0	1	1	1
4	0	2	1	0	0	0	0	1	0
3	37	36	29	10	11	5	34	20	16
2	25	24	31	43	38	40	42	55	65
1	14	26	32	24	30	36	4	13	16
Transfer II Scale Value									
5	1	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	3	1	0
3	24	19	21	6	3	3	32	27	24
2	30	37	30	42	44	38	55	59	54
1	29	46	41	32	33	42	2	7	9

5 = Absolutely Certain a Target Present
3 = Guess
1 = Absolutely Certain NO Target Present
(4 and 2 did not have labels, but represented a response between a guess and Absolute certainty)

Table D-4. Frequency of Misses for each Training Condition as a Function of Frame Speed and Transfer Session for Experiment 1.

	PT2			PT3			WT6		
Frame Speed	180	220	260	180	220	260	180	220	260
Transfer I Scale Value									
5	23	11	13	34	15	13	22	20	16
4	11	5	6	14	15	11	21	14	11
3	64	56	35	42	24	14	35	23	21
2	23	19	14	47	43	37	54	34	36
1	12	15	15	28	21	27	2	5	4
Transfer II Scale Value									
5	17	15	15	23	16	18	17	5	7
4	9	1	4	26	15	10	6	6	8
3	33	34	22	21	12	6	33	19	16
2	19	10	7	38	33	32	43	52	27
1	35	23	30	21	31	32	1	1	3

5 = Absolutely Certain a Target Present
3 = Guess
1 = Absolutely Certain NO Target Present
(4 and 2 did not have labels, but represented a response between a guess and Absolute certainty)

APPENDIX E: INSTRUCTIONS FOR COMPLEX TASK (REPRODUCED EXACTLY AS SEEN BY THE SUBJECTS)

In this task you will perform the duties of a dispatcher. Your task is to select operators to deliver cargo to different destinations. You will receive the following information about an order: 1) the type of cargo to be delivered, 2) the weight of the cargo in kilograms (kg), 3) the vehicle which is available to transport the cargo, and 4) the destination to which the cargo is to be delivered. You must assign one operator (the optimal out of four choices) to deliver the cargo. All destinations, cargos, vehicles, operators, etc. are classified according to certain parameters. There is also a set of rules governing the decision-making process for selection of the optimal operator.

Now, let's explore the structure of the task in greater detail. First, we'll examine the classification scheme. There are six sets of classes (or categories, if you prefer):

1) cargo,
2) weight,
3) distance (to destination),
4) vehicle,
5) destination, and
6) operator license.

CARGO

There are three classes of cargo: **general purpose** (GP), **liquid** (LQ), and **hazardous** (HZ).

WEIGHT

There are three classes of cargo weight: **light** (L), **medium** (M), and **heavy** (H).

DISTANCE

There are three classes of distance to destination (**short range** (SR), **medium range** (MR), and **long range** (LR).

VEHICLES

There are nine classes of vehicles. Vehicles are divided into three principle classes based on the kind of cargo they can carry (general purpose, liquid, and hazardous). Each of these principle classes is divided further into three classes based upon weight rating (**light duty**, **medium duty**, and **heavy duty**).

DESTINATIONS

There are nine classes of destinations. Destinations are divided into three principle classes based upon the type of cargo which they receive (**general purpose**, **liquid**, or **hazardous**). Each of these principle classes is divided further into three classes based upon distance (**short**, **medium**, or **long**) from the shipping terminal.

OPERATOR LICENSES

There are nine classes of operator licenses. Licenses are divided into three principle classes based upon the distance the operator is permitted to transport cargo (**short**, **medium**, or **long range**) and the type of cargo to be delivered (**general purpose**, **liquid**, and **hazardous**). Also, each of these principle classes is subdivided into three more classes based upon the weight rating of the vehicle the operator is permitted to operate (**light duty**, **medium duty**, or **heavy duty**). The license classification system is a progressive one: an operator with a given license classification is permitted to do anything that an operator with a lower license classification can do (more about this later).

Appendix E (continued)

TABLES

The following tables present each set of classes followed by tables with actual operator names, vehicle names, destination names, etc., that belong to each class.

DISTANCE CLASSES

short	(SR)	0- 80	km
medium	(MR)	81-320	km
long	(LR)	321+	km

CARGO CLASSES

general purpose (GP)
liquid (LQ)
hazardous (HZ)

WEIGHT CLASSES

light	(L)	0- 1,500 kg
medium	(M)	1,501-10,000 kg
heavy	(H)	10,001+ kg

VEHICLE CLASSES

general purpose, light duty	(GP-LD)
general purpose, medium duty	(GP-MD)
general purpose, heavy duty	(GP-HD)
liquid, light duty	(LQ-LD)
liquid, medium duty	(LQ-MD)
liquid, heavy duty	(LQ-HD)
hazardous, light duty	(HZ-LD)
hazardous, medium duty	(HZ-MD)
hazardous, heavy duty	(HZ-HD)

Appendix E (continued)

DESTINATION CLASSES

general purpose, short range	(GP-SR)
general purpose, medium range	(GP-MR)
general purpose, long range	(GP-LR)
liquid, short range	(LQ-SR)
liquid, medium range	(LQ-MR)
liquid, long range	(LQ-LR)
hazardous, short range	(HZ-SR)
hazardous, medium range	(HZ-MR)
hazardous, long range	(HZ-LR)

LICENSE CLASSES

lowest	1.1:	general purpose, light duty, short range	(GP-LD-SR)
	1.2:	general purpose, medium duty, short range	(GP-MD-SR)
	1.3:	general purpose, heavy duty, short range	(GP-HD-SR)
	2.1:	liquid, light duty, medium range	(LQ-LD-MR)
	2.2:	liquid, medium duty, medium range	(LQ-MD-MR)
	2.3:	liquid, heavy duty, medium range	(LQ-HD-MR)
	3.1:	hazardous, light duty, long range	(HZ-LD-LR)
	3.2:	hazardous, medium duty, long range	(HZ-MD-LR)
highest	3.3:	hazardous, heavy duty, long range	(HZ-HD-LR)

CARGO

GP	LQ	HZ
lumber	water	mercury
books	milk	cobalt
clothes	whisky	asbestos

Appendix E (continued)

VEHICLES

GP-LD	GP-MD	GP-HD
Load Hog 1000	Load Hog 2000	Load Hog 3000
Freight King 100	Freight King 200	Freight King 300

LQ-LD	LQ-MD	LQ-HD
Tank King 1000	Tank King 2000	Tank King 3000
Route Master 100	Route Master 200	Route Master 300

HZ-LD	HZ-MD	HZ-HD
Haul Master 1000	Haul Master 2000	Haul Master 3000
Kargo King 100	Kargo King 200	Kargo King 300

DESTINATIONS

GP-SR	GP-MR	GP-LR
United Enterprises	Olympia Industries	Island Enterprises
Keystone Systems	Matrix Co.	Universal Systems
Paragon Inc.	Globe Products	Standard Corp.

LQ-SR	LQ-MR	LQ-LR
National Systems	Horizon Technology	Victory Corp.
Republic Enterprises	Acme Corp.	Ajax Industries
Phoenix Technology	Fidelity Systems	Excel Services

HZ-SR	HZ-MR	HZ-LR
Charter Systems	Marathon Corp.	Colonial Inc.
Federal Assoc.	Western Enterprises	Vulcan Assoc.
Triad Co.	Heritage Ltd.	Beta Corp.

OPERATORS

1.1: GP-LD-SR
Eloise
Julian
Gwen

1.2: GP-MD-SR
Bradley
Agatha
Conrad

1.3: GP-HD-SR
Eugene
Lester
Gina

2.1: LQ-LD-MR
Lolita
Rosalie
Barney

2.2: LQ-MD-MR
Valerie
Vance
Mable

2.3: LQ-HD-HR
Herbert
Vera
Adele

3.1: HZ-LD-LR
Nelson
Felix
Claude

3.2: HZ-MD-LR
Bernice
Troy
Olivia

3.3: HZ-HD-LR
Enid
Vincent
Stella

RULES

A set of rules governs the assignment of operators to deliveries. These rules follow.

VEHICLES

1. Any vehicle can travel any distance to deliver its cargo. There is no restriction of range for vehicles.

2. If a vehicle is classified as **"light duty"** (LD), then it can carry a maximum of 1,500 kilograms (kg).

3. If a vehicle is classified as **"medium duty"** (MD), then it can carry a minimum of 0 kg and a maximum of 10,000 kg.

4. If a vehicle is classified as **"heavy duty"** (HD), then it can carry a minimum of 0 kg and there is no maximum limitation.

5. If a vehicle is classified as **"general purpose"** (GP), then it can carry only cargo that is classified as general purpose.

6. If a vehicle is classified as **"liquid"** (LQ), then it can carry only cargo that is classified as liquid.

7. If a vehicle is classified as **"hazardous"** (HZ), then it can carry only cargo that is classified as hazardous.

DESTINATIONS

8. Any destination can receive any amount (i.e., weight) of cargo. There is no restriction for amount of cargo received by a destination.

9. If a destination is classified as **"general purpose"** (GP), then it can receive only cargo that is classified as general purpose.

10. If a destination is classified as **"liquid"** (LQ), then it can receive only cargo that is classified as liquid.

11. If a destination is classified as **"hazardous"** (HZ), then it can receive only cargo that is classified as hazardous.

12. If a destination is classified as **"short range"** (SR), then a vehicle must travel between 0 and 80 kilometers (km) to deliver its cargo.

13. If a destination is classified as **"medium range"** (MR), then a vehicle must travel between 81 and 320 km to deliver its cargo.

14. If a destination is classified as **"long range"** (LR), then a vehicle must travel more than 320 km to deliver its cargo.

LICENSES

General Purpose and Short Range

9. If an operator is classified 1.1, then he or she can operate:
1) vehicles which are classified **"general purpose"** and **"light duty"** (GP-LD)
and
2) can only deliver cargo to destinations which are classified **"short range"** (SR).

If 1.1, then vehicle = GP-LD and destination = SR.

10. If an operator is classified 1.2, then he or she can operate:
1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD) or **"medium duty"** (MD)
and
2) can only deliver cargo to destinations which are classified **"short range"** (SR).

If 1.2, then vehicle = GP-LD or GP-MD and destination = SR.

11. If an operator is classified 1.3, then he or she can operate:
1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (MD), or **"heavy duty"** (HD)
and
2) can only deliver cargo to destinations which are classified **"short range"** (SR).

If 1.3, then vehicle = GP-LD or GP-MD or GP-HD and destination = SR.

Liquid and Medium Range

12. If an operator is classified 2.1, then he or she can operate:
1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (GP-MD), or **"heavy duty"** (GP-HD)

or

2) vehicles which are classified **"liquid"** and **"light duty"** (LQ-LD)

and

3) can only deliver cargo to destinations which are classified either **"short range"** (SR) or **"medium range"** (MR).

If license = 2.1, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD and destination = SR or MR.

13. If an operator is classified 2.2, then he or she can operate:
1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (GP-MD), or **"heavy duty"** (GP-HD)

or

2) vehicles which are classified **"liquid"** and either **"light duty"** (LQ-LD), **"medium duty"** (LQ-MD)

and

3) can only deliver cargo to destinations which are classified either **"short range"** (SR) or **"medium range"** (MR).

If license = 2.2, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD and destination = SR or MR.

Liquid and Medium Range (continued)

14. If an operator is classified 2.3, then he or she can operate:
1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (GP-MD), or **"heavy duty"** (GP-HD)

or

2) vehicles which are classified **"liquid"** and either **"light duty"** (GP-LD), **"medium duty"** (LQ-MD), or **"heavy duty"** (LQ-HD)

and

3) can only deliver cargo to destinations which are classified either **"short range"** (SR) or **"medium range"** (MR)

If 2.3, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD and destination = SR or MR.

Hazardous and Long Range

15. If an operator is classified 3.1, then he or she can operate:
1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (GP-MD), or **"heavy duty"** (GP-HD)

or

2) vehicles which are classified **"liquid"** and either **"light duty"** (LQ-LD), **"medium duty"** (LQ-MD), or **"heavy duty"** (LQ-HD)

or

3) vehicles which are classified **"hazardous"** and **"light duty"** (HZ-LD)

and

4) can only deliver cargo to destinations which are classified either **"short range"** (SR) or **"medium range"** (MR) or **"long range"** (LR).

If 3.1, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD or HZ-LD and destination = SR or MR or LR.

Hazardous and Long Range (continued)

16. If an operator is classified 3.2, then he or she can operate: 1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (GP-MD), or **"heavy duty"** (GP-HD)

or

2) vehicles which are classified **"liquid"** and either **"light duty"** (GP-LD), **"medium duty"** (LQ-MD), or **"heavy duty"** (LQ-HD)

or

3) vehicles which are classified **"hazardous"** and either **"light duty"** (GP-LD), **"medium duty"** (LQ-MD)

and

4) can only deliver cargo to destinations which are classified either **"short range"** (SR) or **"medium range"** (MR) or **"long range"** (LR).

If 3.2, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD or HZ-LD or HZ-MD and destination = SR or MR or LR.

17. If an operator is classified 3.3, then he or she can operate: 1) vehicles which are classified **"general purpose"** and either **"light duty"** (GP-LD), **"medium duty"** (GP-MD), or **"heavy duty"** (GP-HD)

or

2) vehicles which are classified **"liquid"** and either **"light duty"** (LQ-LD), **"medium duty"** (LQ-MD), or **"heavy duty"** (LQ-HD)

or

3) vehicles which are classified **"hazardous"** and either **"light duty"** (HZ-LD), **"medium duty"** (HZ-MD), or **"heavy duty"** (HZ-HD)

and

4) can deliver cargo to destinations which are classified either **"short range"** (SR) or **"medium range"** (MR) or **"long range"** (LR).

If 3.3, then vehicle = GP-LD or GP-MD or GP-HD or LQ-LD or LQ-MD or LQ-HD or HZ-LD or HZ-MD or HZ-HD and destination = SR or MR or LR.

Dispatching Decision Rule

18. The operator with the lowest license classification who is qualified to operate the available vehicle is to be given the assignment. For example, Barney has a license classification of 2.1 and Olivia has a license classification of 3.2. If they are both qualified to do the job then Barney should be given the assignment. This is the rule that operates if one is attempting to minimize cost (i.e., send the operator who is paid the least).

THE TASK

Let's examine how all this comes together in the task. The experiment will be divided into X sessions of 2 blocks of 36 trials per block. You may take breaks between trials or between blocks. For each trial you will be presented with the following information in one computer display:

1) The name of the cargo to be delivered,
2) The weight of the cargo (in kilograms (kg)),
3) The name of the vehicle with which to deliver the cargo
4) The name of the destination to which the cargo is to be delivered.

This display is the 'study display'. You must study the information contained in this display and based on this information (and what you know about the structure and rules of the task) you must decide which operator (or operators) can make the delivery. While this is going to be extremely challenging it's not quite as bad as it might seem; we have provided on-line help. You access help by pressing the 'H' key. This brings up the help Main Menu. From this menu you can choose help on any of six topics:

1) distance,
2) cargo,
3) weight,
4) vehicle,
5) destination, and
6) license.

Some of these will have two levels of help (destination, for example). To choose a topic, simply press the key corresponding to the number of the help item (these number keys are located on the top row of the alphanumeric keypad). If there is a second level of help for the topic you select and you wish to view it, press the 'Page Down' key located in the upper right side of the keyboard. To leave any of the help screens press the 'Esc' key (this is the escape key) which is located in the top left corner of the key board). When you leave the help Main Menu you will return to the study display.

As soon as you have formulated a set of possible operators who can perform the task (The minimum number of possible operators for any delivery is three. Think about it.) press the spacebar and you will be presented with a display containing the names of four operators. There will always be four names. One, and only one, of these names will be the best answer (according to the 'decision dispatching rule'). The number of operators capable of performing the task will vary from one to four. Examine these names and make your decision as quickly as possible (without sacrificing accuracy). When you have made your decision press the key on the numeric keypad which corresponds to your choice.

APPENDIX F: COMMENTS FROM PARTICIPANTS IN DISPATCHING TASK (REPRODUCED EXACTLY AS PROVIDED BY THE SUBJECTS)

SUBJECT 1

Session 6

Strategy -- I look at what is given in this order:
1) type of cargo
2) truck # to determine L, M, or H Duty
3) destination

If type of cargo is hazardous I don't bother to look at the destination. After I get my answer in my head (type, duty, range) I either: 1) name the three people in that category or 2) I just think about the visual position of where the best person would be, then I either immediately see the right person or by the process of elimination find the best person.

Session 10

Rules -- There are three categories you need to keep track of: 1) cargo type 2) cargo weight 3) destination.

Drivers are to be assigned according to these categories. Some important rules must be followed:

1) A driver may not transport a type of cargo above which he or she is licensed for (general, liquid, or hazardous). e.g., a driver licensed for general may not transport hazardous.

2) A driver may not transport a weight above which he or she is (light, medium, or heavy) ** NOTE: A type of truck will also be given. The trucks are classified according to the weight they may carry. If the truck's possible weight that it can carry is above the actual weight of the cargo, this will override the weight of the cargo. You should substitute this weight (given in the truck name) when determining the driver.

3) A driver may not transport cargo to a longer distance (given by destination) than which he or she is licensed for (short, medium, long range).

The optimum driver must be used. If he or she is not available, the next (higher license) driver must be used.

If the original plan was to use the smiling faces & music through the whole experiment I think it would get obnoxious.

Appendix F (continued)

It was good for the first couple of blocks, but it might be distracting after a while.

The first day of instructions was overwhelmingly long.

I really enjoyed the task I thought it was neat!

On a scale of 1 (extremely easy) to 9 (extremely hard) she rated the task a 3.

SUBJECT 2

Session 1

Use spacebar instead of page down key.

Subject is studying the three names that fit and the three names in the next level up. If target falls within, OK, if not, going for the splatter.

Session 3

Found trials he thought were "incorrect". Showed him they were correct and he realized he was wrong to ignore the vehicle information.

Session 6

Strategy: First thing he looks at is the cargo. That tells him if he needs a 1, 2, or 3 for the target number. He checks the company, if it's one of the nine I need to recognize, the target number is changed upwards.

Company Lists

		S	M	L
	GP		X	X
cargo	LQ			X
	HZ			

only needs to know companies in categories marked with 'X'.

Then he checks the weight to find the second number. Last, he checks to see if the truck being used is greater than the cargo rating. This gives him the license type he needs. He used to not check the truck type and my scores reflected that. Often, the computer will show more than 1 person from the same license type. I know then that I can ignore them in this game, there can only be one "optimal" driver. That may be a flaw.

Session 10

Rules -- You are a dispatcher. You assign drivers to trucks delivering cargo to various locations. There are three types of cargo, in ascending order of difficulty: General Purpose, Liquid, and Hazardous. There are also three weight categories: Light, Medium, and Heavy. Added to that, there are three distance categories: Short, Medium, and Long.

Your job is to assign on of four possible drivers to a delivery. That driver must be the one who has the lowest qualifying license type. There are nine types of licenses, based upon two factors: Riskiest cargo/Longest Distance allowed, and heaviest load category. For instance, a driver with a 1.1 could only transport small loads of general purpose cargo short distances. A 1.2 would allow the driver to carry up to medium loads, but still only general purpose cargo for short distances. A 2.1 would allow the driver to haul liquids or general purpose cargo a short or medium distance, but still only light duty. Obviously, a 3.3 driver could carry anything, anywhere.

The job will have four factors: the cargo, the weight, the destination, and the truck carrying the cargo. Be wary the truck may have a heavier rating than the cargo needs! After you study the problem, using the help screens as necessary (they give info on destinations, cargos, drivers, et. al.) you will be given a choice of four drivers. There will only be one driver who fits the best: he/she may not be <u>ideal</u>, but will be the best out of those four.

How to toughen the task -- Time limits on blocks. No more than one driver for each license group. Have companies accept "lower" cargo types. Have dispatcher choose vehicle. Demerits for failure.

Strategy -- His strategy has changed quite a bit. It is essentially the same, except I don't even check the weight anymore. It's not necessary. Neither is the cargo, either, really...but I'd rather know nine cargos and nine "special case" companies than 27 companies.

Hmmm...actually, I would only need to memorize 18 companies...anything I didn't recognize would be class 3...too late now.

On a scale of 1 (extremely easy) to 9 (extremely hard) he rated the task a 2.

SUBJECT 3

Session 2

Problems

1) Hitting '0' to return from the Help Screen. Hitting 'Esc' I handle, no problem. It's odd, though expecting to hit 'Esc'...'Space' to get out (although I don't know why I do) and then having to hunt for the '0'.

2) A personal problem, so I don't guess this would really apply and it's certainly nothing that the program can be modified to account for. I think of the people's names as they appear in the matrix. As I get faster, the realization of where they are in the matrix translates instantly to the numeric keypad. Instead of hitting the person's name, I'll be hitting their slot on the matrix. This is all right if their name is in 1.2, 2.1, 2.3, 3.2 or 2.2. But otherwise, there's a strong chance my answer will be wrong. On this last block it dropped me from 100% accuracy to 94% accuracy.

Session 5

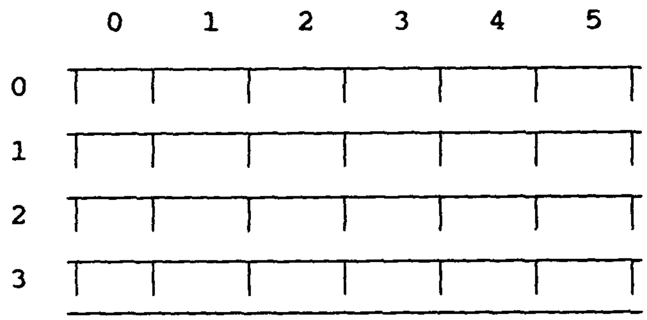
Right at first I didn't realize quite what the task was. I thought the names would be picked very close to the optimal drivers. I quickly realized that I would have to memorize the list. So I did. The names of the companies I (fortunately) never memorized. it took less time and conscious thought to depend on recognizing most of the ones that came through and checking the help screens on the rest.

I soon discovered that the only companies I would really need to memorize at all were the medium and long distance general product companies. Since all general product drivers can only go short distances, and all liquid drivers can only go short or medium distances, the length of the distance determined who could drive it...the rows of drivers that were eligible.

I scan the data in a clockwise circle from the top left. I "black out" the areas of a 3 X 3 matrix that contains the eligible drivers. I quickly scan the four available drivers to see if one of the ones in the optimal section are there. At the same time, I check to see if one from the next best column is there (as is often the case). If not, I pick the most likely, quickly check it against the rest, and enter my choice.

Session 10

Rules -- Your basic objective is to find the most efficient driver for a designated cargo. You are supplied with various data parameters which you must analyze in a minimum amount of time and which place certain boundaries on your choice of drivers. After viewing the parameters, you will select from four drivers, only one of whom will be the most efficient. It is important to note that you are choosing the most efficient driver available. Only the four you are offered are available. Imagine the drivers as having a certain ranking. After you have decided what the optimal ranking is for the given cargo, keep in mind that anyone of that rank or higher has the ability to carry the cargo (Actually, the ranking is a two dimensional ranking...imagine a grid:



A "higher ranking" would mean anyone in a row greater than or equal to the base row and in a column greater than or equal to the base column.)

The choice of drivers is based on three factors: 1) the type of material they can carry 2) the weight of the cargo they can carry and 3) the distance they can travel (assorted technical information.

Suggestions -- Mainly, I would suggest modification to the Help screen. Choosing the number; fine. Even hitting 'Esc' was fine (Of course, I am an ICS major, and well versed in instinctively grabbing for the 'Esc' key, so I imagine that could be a problem for others.) But having to hit '0' was not good. The space bar would have been ideal...except that you also use the space bar to get to the driver screen, and that could cause problems. I would suggest 'Esc' to get out of Help screen and 'Esc' to get out of the Help menu as well.

How to make the task harder -- 1) I liked the idea of disqualifying a driver for a certain amount number of trials after being chosen.

2) Avoid extremes. There were far too many 1A data parameters (i.e., low weight, short distances, general purpose) for one thing. And there were too many trials when you would have, say, three drivers from 1A or 3C, and another driver. This makes it very obvious which driver it

is, since it can't be one with another choice from the same area.

3) Possibly make the names of the trucks more important. If you make it so GP can be carried in LQ and HZ, and that LQ can be carried in HZ, and that LQ can be carried in HZ, it would make learning the names of the trucks more essential.

4) Have more similar names. Maybe it's just me, but I had a horrendous time with Eloise/Rosalie. For some reason I had difficulty keeping them separate.

On a scale of 1 (extremely easy) to 9 (extremely hard) he rated the task a 2.

SUBJECT 4

Session 1

names are too weird

too hard to get back to choices from list of names. it takes two moves...'Esc' and '0'.

don't know why I got something wrong -- was it my logic or was it remembering the order of names?

I had to go back to the rules to see the list of names in front of me to see why I got something wrong -- was I doing my figuring all wrong or was it remembering -- It was simply remembering the order wrong.

Session 2

lots of trials with 1.1. I think if I had the list of names on paper I could memorize them more quickly than on a computer screen. I've never had to memorize a screen and it is different than paper.

frustrating that there is no order to destination names like the vehicles (1000 = light, 2000 = medium and 3000 = heavy). How about all corps. are close, systems = medium, etc.

maybe change the color of different screens to take away monotony and help in memorizing.

my mean decision time I keep forgetting is being timed and I take my time.

if names were used that I could relate to then I could remember them better. I have no picture in my mind of Eloise or Gwen, etc.

Appendix F (continued)

Session 3

NOTE. Prior to running we discussed rules; particularly license and vehicle rules.

Now that I have a system that works, I never even think about the rules. for example: If hazardous just check the duty to tell 3.1, 3.2, or 3.3.

I have learned the name grid from the outside to middle. First, I learned 1.1 and 3.3. then 1.2 and 3.2, etc. Still have trouble with the middle.

Still frustrating to take three steps to get from name grid to choices.

BLOCK 2 My highest score yet 97%, starting to know that grid well and that makes me think my errors before were due to bad memory, not bad logic.

very much a system now, never think of rules

when it says "incorrect" maybe the name grid could pop up onto the screen with the correct answer highlighted instead of just gluing the name.

Session 4

BLOCK 1
Takes a while to get the memory back from yesterday

Give less names with SAME first letter. Easier to remember by first letter of name.

Enid is approximately equal to End and she's at the end -- easy to remember.

For the first time I was thinking TR and pressed BR by accident -- first time the mistake has been made by my hand not my brain.

BLOCK 2
So much easier when you're warmed up

I think I could be faster if there were no names just 1.3a 1.3b or 1.3c.

The only thing I still can't remember are those destinations -- the list is so random.

Session 5

BLOCK 1
This is my fifth day and I notice a definite increase in remembering the chart from the day before.

You might as well eliminate the weight of the cargo. I haven't looked at that since Monday.

Sometimes I hit the space bar for choices before I'm ready. That choice key should be far away. Space bar should send you directly to the grid of names.

BLOCKS 2 & 3
These are the only rules I ever think of: 1) If you drive HZ you can drive liquid and gen purpose.

2) If you drive long you can drive medium and short.

3) weight means nothing

4) If it is GP and MR use liquid medium range
If it is LQ and LR use hazardous long range

5) if it is going LR it must be done by a hazardous license

6) It doesn't matter where HZ is going just whether or not it is LD, MD, or HD

Session 6

make one special key to access the name grid.

how about flashing my decision time after each trial so I remember to try to be fast. When it just says correct or incorrect that becomes all I care about.

If I speed up I become slightly less accurate. From the experiment description I don't know whether you want fast and 92% or slow and 97%. Which is higher priority: speed, accuracy, or a combination of certain levels of each?

Session 7

I still don't know the destinations. They have no order.

when you hit the space bar to see choices maybe it could ask "are you sure"?

Session 8

Bring names closer together so I can read them all at once and be quicker. As it is I have to go from one name to the next and think. If they were closer I could take in the whole screen at once and decide quicker.

Actually, maybe that is just too hard to do for this experiment. I at least need to look at each one and think. The right one does not just jump out at me when looking at all four as a whole (BLOCK 2).

For some reason I thought of lumber as a liquid three times today.

Session 9

How about showing were the correct answer was on the four possible answers. Don't just say answer is Agatha. Highlight the name in the context of the other names.

Accuracy goes up with time spent before hitting space bar. How about telling me that time too.

Session 10

Rules -- cargo must be taken by capable driver. If a driver can drive far he can also drive close and medium. If he can drive HZ then he can drive LQ and GP. If he can drive HD then also light and medium. Must choose the best driver

Suggestions -- I don't like the four corners set-up. Would rather all in a row.

Give less examples in instructions. There were so many that I skipped a lot out of laziness. If there were fewer I would of concentrated on them more. No need to give every possibility

How to make task harder -- Only allow 5 seconds per help screen per trial.

On a scale of 1 (extremely easy) to 9 (extremely hard) he rated task a 2.

Updated strategy -- If GP 1) check range and type of truck
If LQ 1) check type of truck --> if it's long range it will jump out at you

If HZ just check truck for 1.3, 2.3, or 3.3.

SUBJECT 5

Session 6

Strategy -- On the first screen I look at the weight then the substance. Next, I look at the destination. If it's one I don't know then I use the help to look it up. Then, I use Help if I am not sure of the people around the weight/range. When, I go to the choice screen I usually visualize where the people are on the license screen and pick the appropriate choice. If one of the choices is in the exact category of a licenses division then I pick him without considering the other three. This is the same basic process I used since the beginning. As time went on, I used the help screens less.

NOTE. We looked at trials that the subject thought were program errors. In the process, he realized the importance of vehicular information.

Session 10

Rules -- To perform the task you must pick the lowest qualified driver for the task. Each driver is divided into license categories. The lowest category allows a driver to drive general purpose, light weight trucks a short distance. The next two license categories allow medium weight and then heavy weight. The next higher license allows a driver to drive a low weight liquid truck a short or medium distance. This also allows him to drive any previous license group trucks a short or medium distance. The next two license categories allow the driver to drive medium and then heavy liquids short or medium distances. The next category allows a driver to drive a hazardous material truck of light weight any distance. He can also drive a general purpose or liquid truck of any weight any distance. The next two categories allow him to also drive medium and the heavy weight hazardous materials.

To determine who is the lowest qualified driver for the task, three things must be examined. On the task screen there will be four categories to consider. These are a weight, type of cargo, a destination and a type of truck. The weight is not necessary for the decision. First, determine the type of cargo. Lumber, books, and clothes are all general purpose. Water, milk, and whisky are all liquid cargo. Mercury, cobalt, and asbestos are all hazardous materials. Next, look at the truck type. Any vehicle with a 100 or 1000 is light weight. A vehicle with a 200 or 2000 is medium weight. A vehicle with a 300 or 3000 is heavy. Next, look at the destination and determine if it is short, medium or long range. A help screen is provided during the task screen. If you do not remember a cargo, vehicle or destination type then reference it by pressing the 'H' key. The category wanted is selected by pressing the appropriate number. Also, the operators' names and qualifications are accessed by this. After this has been determined, then the appropriate name can be selected.

Suggestions -- In the help screens three things would be helpful. The first screens for destination and license are usually not necessary. Therefore the second half could be printed first and if the first was necessary the 'Page Down' could be used for it instead of for the more useful information. Also, a one key escape back to the task screen would speed up the process.

When an incorrect name is given as a choice, all of the name choices and the operator would know more about why he made the wrong choice.

The instructions could be given in a little less detail and in a different style.

The pink noise was probably more distracting than normal background noises.

How to make task harder -- Changing the license names, destinations or truck classifications would make task more difficult.

On a scale of 1 (extremely easy) to 9 (extremely hard) he rated the task a 4.

Updated strategy -- First, I look at the cargo type, then I determine the weight by looking for a 1, 2, or 3 in the truck name. Next, I determine the range. If I forget the destination type I use the Help to access it. If I feel uncertain about the operators, then I access Help. I look at the exact operator classification for the job and then the ones after it. I also review operators that I feel uncertain about, particularly frequently missed ones. When comparing operators on the assignment screen, I think about where they appear on the license screen and use that to determine target (if the answer is not obvious).

APPENDIX G: COMPLEX TASK USER'S MANUAL

Building the Screens

The first thing the experimenter must do is build the screens to be used by the program dispatch.exe. This involves creating ASCII text versions of the screens (any ASCII character, including the extended set, may be used), converting the ASCII files to binary files, and, finally, combining the various binary files into one large binary file which is actually used by complex3.com. The four files required for this process and their functions are as follows:

1. snapshot.com: This is a terminate-and-stay-resident (TSR) program used in conjunction with show.com to convert ASCII text files into binary files.

2. show.com: This program displays the ASCII text file so that a 'snapshot' of it may be taken.

3. looker.com: This program allows the experimenter to view the binary file to see how the actual screen will look.

4. diagcom.com: This program takes the various binary files and combines them into one large file called diagcom.dat, the file that is actually used by dispatch.exe.

First, create a subdirectory in which to do all this work and place the required files. The screen-building process begins by creating a series of ASCII text files corresponding to the help screens that will be available during the experiment. Although any DOS file name may be used, keep it simple and logical (e.g., 0.txt, 1.txt,...,n.txt).

Once the ASCII versions of the help screens are complete they must be converted to binary files. First, load snapshot.com by typing snapshot and then pressing the 'Enter' key. Second, display one of the ASCII files by typing show filename.ext, where filename.ext is the name of the ASCII file (e.g., 0.txt). Third, take the 'snapshot' by holding down the 'Ctrl' key and then pressing the 'break' key. The first time you take a snapshot it produces a binary file with the name diagram.0. Subsequent snapshots yield diagram.1, diagram.2, and so on.

These programs were written prior to this project and were originally used with PCs equipped with CGA adaptors. Consequently, taking the snapshot with the Epsons causes the machine to lock up and the PS/2s cannot be used at all. After each snapshot the computer must be re-booted and the process repeated until all screens are done (diagram.0 through diagram.n). If desired, one can view how the screen will actually appear to the subject by typing in the command looker diagram.x, where diagram.x is the binary file to view. It is important to keep track of which screen is associated with which diagram.x file because assignment of keys to their corresponding screens is based on this file name.

Now all the separate binary files (diagram.0 through diagram.n) must be combined into one large binary file that will be used by dispatch.exe. Type in the command diagcom and when prompted enter the number of files to combine minus one. For example, if there are 10 files to combine then enter 9. The resulting file, dispatch.dat, will contain all the help screens in binary format.

The SCREEN.DAT file is the inter-block information screen used to present additional information to the subject between blocks (e.g., if an operator is promoted or demoted). The first line will be the number of lines to read and then write to the screen. Note that this number may be zero. In this case there will be no info. If it is greater than zero then the info is presented and the program waits for the space bar to be pressed. Then a message informing which block is next is displayed and the routine waits for the space bar to be pressed. This message is presented whether or not an info screen is presented. For each block except the last there is a line in the file having the number of following info lines in the file to be displayed.

Building the Scenario (Stimuli) Files

The second thing the experimenter must do is build the files that contain the various scenarios that will be used as data by dispatch.exe. There is one scenario for every trial. The program generates blocks of trials where the experimenter specifies the number of trials per block and the number of blocks. This is an expert system type of program that has the dispatcher task rules built in and uses those rules to operate on data provided by the experimenter to generate its output. The four files required for this process and their functions are:

1. compgen.exe: This program uses the three data files to generate the stimuli file.

2. cg-class.dat: This program provides the categorization data. This is the default name and it can be changed.

3. cg-name.dat: This program provides the name data. This is the default name and it can be changed.

4. cg-block.dat: This program provides the trials per block and number of blocks. This is the default name and it can be changed.

To generate a stimuli file the experimenter must first assemble the three previously mentioned x.dat files. These must be in ASCII text format.

The cg-class.dat file lists each of the six different classes (or categories, as preferred). In order, these are distance, cargo, weight, vehicle, destination, and license. The first line contains the name of the class in upper case and each class is separated by a blank line. Within each description there is the name of the division (lower case) followed by the acronym for that division (upper case). In two instances, distance and weight, this is followed by the defining parameters for each division (weight ranges and distance ranges). In the case of licenses the numerical representation of each division precedes the name of the division.

The cg-name.dat file lists each name associated with each division of each class. The file is divided into four different classes. In order, they are cargo names, vehicle names, destination names, and operator names (actually, this is the license category). The name will appear in the program exactly as it appears in this ASCII text file (i.e., uppercase, lowercase, or mixed case). The number of names associated with any particular division of any particular category is flexible. In the January-April 1990 instantiation of this exercise, there are three names for each division of cargo (total of nine names), two names for each division of vehicles (total of 18 names), three names for each division of destinations (total of 27 names), and three names for each division of operators (total of 27 names).

In the cases of cargo and vehicles, each name is followed by the acronym associated with its division (acronym in uppercase).

In the case of destinations, each name is given a number (e.g., 1-27) and the names are ordered from general purpose, short range (GP-SR) to hazardous, long-range (HZ-LR). The number is followed by the name of the destination, which is followed by its division acronym (acronym in uppercase), which is followed by that destination's distance in kilometers.

In the case of operators, each name is again given a number and the names are ordered from general purpose, light duty, short range (GP-LD-SR) to hazardous, heavy duty, long range (HZ-HD-LR). The number is followed by the operator's name, which is followed by his or her license acronym, which is followed by the license division number (1.1-3.3).

The cg-block.dat file lists the number of trials per block and number of blocks. For example, if there are three blocks of 36 trials each, then the file would contain three lines with the number 36 on each line.

After these three files have been assembled, they should be saved; they will be used later to obtain frequency information about each block. Now, the stimuli file(s) may be generated. First, type the command compgen and press the 'Enter' key. Follow the prompts, and enter the names of the three x.dat files or press the 'Enter' key if the default file names are to be used. At the 'stimulus output name' prompt enter the name of the stimuli file to be created (sessnX.stm, where X is the session number, is desirable because the program will use this as the default). Once this is done, the program will prompt as to whether all information was entered correctly, if it was not, press the 'n' key and correct any mistake. If the information is correct, press the 'y' key and the program will execute. At the end, the program will display the number of blocks and trials per block that have been created.

Running the Experiment

To run the experiment the following four files must be in the subdirectory containing the program file (dispatch.exe): screen.dat, diagram.dat, fixtime.dat, and the stimuli file. Start the program by typing dispatch and pressing the 'Enter' key. Prompts then direct the following actions:

1. Type the subject number (1-99) and press the 'Enter' key.

2. Type the stimulus name or accept the default, sessnX.stm, where X is the session number (the program reads the X from the subject's data file) and press the 'Enter' key.

3. Type the number of trials per block (1-36) and press the 'Enter' key. Thirty-six is the default.

4. Type the number of blocks for the session (1-9) and press the 'Enter' key. Two is the default.

5. When prompted for the number of minutes for the session, press the 'Enter' key. This function is not operational.

6. If one or both of the stressor tasks is desired, type the 'y' key, followed by the 'Enter' key: otherwise, type the 'n' key and 'Enter'.

    a. If time to select the best operator name is to be limited, enter that time (in milliseconds).

b. If total time spent in the data/study screen and in help screens is to be limited, enter that time (in milliseconds).

7. There are three feedback options. When prompted for each, respond 'y' for "yes" or 'n' for "no," followed by 'Enter'.

a. Correct trial feedback

b. Block feedback

c. Help screen feedback (actually, "yes" lets the subject access help and "no" removes access).

8. The last prompt is for display adaptor type. The default (for Epsons) is monochrome ('m'). Color ('c') is the alternative (PS/2s).

Upon completion of these entries, there will be a prompt to verify their correctness. If they are correct, type 'y'; otherwise, type 'n'. Press the 'Enter' key when done.

If at any time the program must be stopped, there are two ways to accomplish this: Hold down the 'Ctrl' key and press the 'Break' key or reboot the computer. If the 'Ctrl-Break' combination is employed, then when the DOS prompt appears type the command fixtime and press the 'Enter' key. As an aside, the program will leave the time incorrect. Consequently, use the DOS time command to reset the clock.

Analyzeing the Data

In addition to the data files (results.#) to be analyzed, two files are used: results.com and comptime.com. First, a description of the raw data file is necessary. When a subject is tested for the first time, the program outputs a data file with the name results.#, where # is the subject's ID number. As long as this file is present in the subdirectory, data from subsequent sessions will be appended to it. The program also reads the most recent session number from this file and uses it to supply the default session number and stimuli file number at the beginning of the program.

Each raw data file has what are termed data lines and keystroke lines. There is a data line corresponding to each trial. Following each data line are a number of keystroke lines equal to the number of valid keystrokes performed during that trial less one (the target response keystroke is not represented because the information is contained in the data line).

Each data line begins with the '#' symbol as an identifier and is followed by numbers representing these 17 variables in the following order: subject's ID number, session number, trial number, block number, correct answer ('7'=top left, '9'=top right, '1'=bottom left, and '3'=bottom right), number of keys pressed during the trial, subject's choice ('7'=top left, '9'=top right, '1'=bottom left, and '3'=bottom right), whether the answer was correct (0=false, 1=true), whether the operator selected was qualifed to make the delivery (0=false, 1=true), the identification number of the operator in the top left position, the identification number of the operator in the top right position, the identification number of the operator in the bottom left position, the identification number of the operator in the bottom right position, response latency in milliseconds (ms), total amount of time spent in help (in ms), total amount of time spent studying the data screen (in ms), and the type of trial (1-27).

Each keystroke line is of the following form:

latency in ms '* ' keystroke ' *', where keystroke might be 'Esc' to represent the 'Escape' key or the actual key hit (including the space bar, which would be seen as '* *').

For the majority of statistical analyses, the following steps will be sufficient. The first step takes the raw data file (e.g., results.3) and writes the data lines minus the '#' symbol to a new file. This file is in a format acceptable to SAS. To begin, type in the command results and press 'Enter'. Then, at the first prompt, type in the name of file to process (e.g., results.3) and press the 'Enter' key, and at the second prompt, type the name of the file to which the output will be written (e.g., output.3). Upon completion, the program will display 'Execution is complete!' and the DOS prompt will return. It is best to take each of these files (one per subject) and, using Microsoft Word or any other text editor, concatenate them into one large file. This file may then be uploaded onto the mainframe to be analyzed.

The file comptime.exe is designed to present a view of time spent in each help screen. The results may also be written to the printer or disk. If output to disk, the data lines are similar to the raw data file but have variables representing the time spent in each particular help screen (from zero to who knows how many ms).

* U. S. GOVERNMENT PRINTING OFFICE: 1991--761-052/40035