SEARCH AND RETRIEVAL PROCESSES IN LONG TERM MEMORY

BY

RICHARD M. SHIFFRIN

TECHNICAL REPORT NO. 137

August 15, 1968

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

TECHNICAL REPORTS

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

(For reports no. 1 - 44, see Technical Report no. 125.)

50 R. C. Atkinson and R. C. Calfee. Mathematical learning theory. January 2, 1963. (In B. B. Wolman (Ed.), Scientific Psychology. New York: Basic Books, Inc., 1965. Pp. 254-275)

51 P. Suppes, E. Crothers, and R. Weir. Application of mathematical learning theory and linguistic analysis to vowel phoneme matching in Russian words. December 28, 1962.

52 R. C. Atkinson, R. Calfee, G. Sommer, W. Jeffrey and R. Shoemaker. A test of three models for stimulus compounding with children. January 29, 1963. (J. exp. Psychol., 1964, 67, 52-58)

53 E. Crothers. General Markov models for learning with inter-trial forgetting. April 8, 1963.

54 J. L. Myers and R. C. Atkinson. Choice behavior and reward structure. May 24, 1963. (Journal math. Psychol., 1964, 1, 170-203)

55 R. E. Robinson. A set-theoretical approach to empirical meaningfulness of measurement statements. June 10, 1963.

56 E. Crothers, R. Weir and P. Palmer. The role of transcription in the learning of the orthographic representations of Russian sounds. June 17,. 1963.

57 P. Suppes. Problems of optimization in learning a list of simple items. July 22, 1963. (In Maynard W. Shelly, II and Glenn L. Bryan (Eds.), Human Judgments and Optimality. New York: Wiley. 1964. Pp. 116-126)

58 R. C. Atkinson and E. J. Crothers. Theoretical note: all-or-none learning and intertrial forgetting. July 24, 1963.

59 R. C. Calfee. Long-term behavior of rats under probabilistic reinforcement schedules. October 1, 1963.

60 R. C. Atkinson and E. J. Crothers. Tests of acquisition and retention, axioms for paired-associate learning. October 25, 1963. (A comparison of paired-associate learning models having different acquisition and retention axioms, J. math. Psychol., 1964, 1, 285-315)

61 W. J. McGill and J. Gibbon. The general-gamma distribution and reaction times. November 20, 1963. (J. math. Psychol., 1965, 2, 1-18)

62 M. F. Norman. Incremental learning on random trials. December 9, 1963. (J. math. Psychol., 1964, 1, 336-351)

63 P. Suppes. The development of mathematical concepts in children. February 25, 1964. (On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 1965, 30, 60-96)

64 P. Suppes. Mathematical concept formation in children. April 10, 1964. (Amer. Psychologist, 1966, 21, 139-150)

65 R. C. Calfee, R. C. Atkinson, and T. Shelton, Jr. Mathematical models for verbal learning. August 21, 1964. (In N. Wiener and J. P. Schoda (Eds.), Cybernetics of the Nervous System: Progress in Brain Research. Amsterdam, The Netherlands: Elsevier Publishing Co., 1965. Pp. 333-349)

66 L. Keller, M. Cole, C. J. Burke, and W. K. Estes. Paired associate learning with differential rewards. August 20, 1964. (Reward and information values of trial outcomes in paired associate learning. (Psychol. Monogr., 1965, 79, 1-21)

67 M. F. Norman. A probabilistic model for free-responding. December 14, 1964.

68 W. K. Estes and H. A. Taylor. Visual detection in relation to display size and redundancy of critical elements. January 25, 1965, Revised 7-1-65. (Perception and Psychophysics, 1966, 1, 9-16)

69 P. Suppes and J. Donio. Foundations of stimulus-sampling theory for continuous-time processes. February 9, 1965. (J. math. Psychol., 1967, 4, 202-225)

70 R. C. Atkinson and R. A. Kinchla. A learning model for forced-choice detection experiments. February 10, 1965. (Br. J. math stat. Psychol., 1965, 18, 184-206)

71 E. J. Crothers. Presentation orders for items from different categories. March 10, 1965.

72 P. Suppes, G. Groen, and M. Schlag-Rey. Some models for response latency in paired-associates learning. May 5, 1965. (J. math. Psychol., 1966, 3, 99-128)

73 M. V. Levine. The generalization function in the probability learning experiment. June 3, 1965.

74 D. Hansen and T. S. Rodgers. An exploration of psycholinguistic units in initial reading. July 6, 1965.

75 B. C. Arnold. A correlated urn-scheme for a continuum of responses. July 20, 1965.

76 C. Izawa and W. K. Estes. Reinforcement-test sequences in paired-associate learning. August 1, 1965. (Psychol. Reports, 1966, 18, 879-919)

77 S. L. Blehart. Pattern discrimination learning with Rhesus monkeys. September 1, 1965. (Psychol. Reports, 1966, 19, 311-324)

78 J. L. Phillips and R. C. Atkinson. The effects of display size on short-term memory. August 31, 1965.

79 R. C. Atkinson and R. M. Shiffrin. Mathematical models for memory and learning. September 20, 1965.

80 P. Suppes. The psychological foundations of mathematics. October 25, 1965. (Colloques Internationaux du Centre National de la Recherche Scientifique. Editions du Centre National de la Recherche Scientifique. Paris: 1967. Pp. 213-242)

81 P. Suppes. Computer-assisted instruction in the schools: potentialities, problems, prospects. October 29, 1965.

82 R. A. Kinchla, J. Townsend, J. Yellott, Jr., and R. C. Atkinson. Influence of correlated visual cues on auditory signal detection. November 2, 1965. (Perception and Psychophysics, 1966, 1, 67-73)

83 P. Suppes, M. Jerman, and G. Groen. Arithmetic drills and review on a computer-based teletype. November 5, 1965. (Arithmetic Teacher, April 1966, 303-309.

84 P. Suppes and L. Hyman. Concept learning with non-verbal geometrical stimuli. November 15, 1968.

85 P. Holland. A variation on the minimum chi-square test. (J. math. Psychol., 1967, 3, 377-413).

86 P. Suppes. Accelerated program in elementary-school mathematics -- the second year. November 22, 1965. (Psychology in the Schools, 1966, 3, 294-307)

87 P. Lorenzen and F. Binford. Logic as a dialogical game. November 29, 1965.

88 L. Keller, W. J. Thomson, J. R. Tweedy, and R. C. Atkinson. The effects of reinforcement interval on the acquisition of paired-associate responses. December 10, 1965. (J. exp. Psychol., 1967, 73, 268-277)

89 J. I. Yellott, Jr. Some effects on noncontingent success in human probability learning. December 15, 1965.

90 P. Suppes and G. Groen. Some counting models for first-grade performance data on simple addition facts. January 14, 1966. (In J. M. Scandura (Ed.), Research in Mathematics Education. Washington, D. C.: NCTM, 1967. Pp. 35-43.

(Continued on inside back cover)

SEARCH AND RETRIEVAL PROCESSES IN LONG-TERM MEMORY

by

Richard M. Shiffrin

TECHNICAL REPORT NO. 137

August 15, 1968

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

PRECEDING PAGE BLANK NOT FILMED.

ACKNOWLEDGMENTS

I am indebted to Professor Richard C. Atkinson, chairman of my dissertation committee, for his frequent advice and encouragement during the course of this dissertation. Thanks are due to Professor William K. Estes and Professor Gordon H. Bower, who also served on my dissertation committee, and to Professor Walter Kintsch, for their comments at various stages of this paper. Michael S. Humphreys was instrumental in the design of Experiment I, and has engaged in valuable conversations with the writer. Helen Butler and her staff have been invaluable in the final preparation of the paper, particularly Maria Jedd who painstakingly prepared the many figures. Thanks are due to Carolyn Young who typed the major portion of the final manuscript. Finally, thanks are especially due to my wife Susan, for typing numerous rough drafts and a portion of the final draft, for providing cheerful encouragement, and for relieving me of many concerns extraneous to this project.

The research reported here was accomplished while I held a National Science Foundation Terminal Graduate Fellowship. Support for the research was provided, in part, by the National Aeronautics and Space Administration Grant NGR-05-020-244.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

List of Tables

List of Figures

INTRODUCTION

In the past fifteen years, there has been an increasing interest in theories of human memory that consider storage and retrieval to be probabilistic processes that may vary randomly from one moment to the next. These theories for the most part can be regarded as variants of Stimulus Sampling Theory (Estes, 1959; Atkinson and Estes, 1963), and stimulus fluctuation theory (Estes, 1955a,b). A fairly large number of memory variables have been analyzed by quantitative, mathematical models within this framework. Heretofore these models have tended to be quite restrictive, their range of application being limited to a small number of variations within simple situations. In addition, these models have been concerned primarily with the memory acquisition process rather than the memory loss process. This report attempts to extend this earlier work by introducing a theory which can deal quantitatively and simultaneously with many of the variables previously examined individually, and which will deal as extensively with forgetting as learning. The theory is formulated in the spirit of Stimulus Sampling Theory, but due to the complexity of the data examined, is not a direct extension of the earlier models which have largely taken the mathematical form of multi-state Markov models.

The theory is conceived of as a quantitative alternative to primarily qualitative theories such as "two-factor theory" (Postman, 1961), although the variables dealt with in the two cases do not entirely overlap. The direct antecedents of the present work are the theoretical papers of Atkinson and Shiffrin (1965, 1968) and Shiffrin and Atkinson

(1968). As a result, the theory is primarily concerned with an elaboration of a complex search and retrieval process from long-term memory.

Chapter I of the present report outlines the general framework of the theory. Chapter II describes and presents the results of two experiments designed to provide a wide range of data to test a quantitative version of the overall framework. The first experiment is concerned with the probabilistic nature of retrieval, and forgetting of individual items. The second experiment is concerned with intrusion phenomena in responding, and with interference phenomena following the altering of the response assigned with a stimulus. A number of other variables which are examined will be described in the text. Chapter III presents a specific quantitative model based on the theory of Chapter I, and applies it to the results of the two experiments.

CHAPTER I

A THEORY OF STORAGE AND RETRIEVAL IN LONG-TERM MEMORY

This chapter begins with a brief survey of the human memory system, largely following the format of Atkinson and Shiffrin (1965, 1968). The report will then turn to a detailed discussion of a theory of storage and retrieval for long-term memory. Although the system is meant to be quite general, the theory will be described as it applies to a continuous paired-associate learning task. Such a task consists of a series of anticipation trials. On each trial a stimulus is presented for test and then paired with a response for study. The task is called continuous because new stimuli are continually being introduced at randomly spaced intervals. The theory is described in relation to this task because it is the one utilized in the experiments described in Chapter II.

The Memory System

It has proved of value (Atkinson and Shiffrin, 1968) to dichotomize memory processes on a dimension of subject control. Thus, on the one hand, there are "structural processes" which are permanent, unvarying features of the memory system, features which may not be modified at the will of the subject. On the other hand are "control processes" which are selected, constructed, and used at the option of the subject, and may vary greatly from one task to another. This distinction was set forth in great detail in the report cited, and will not be belabored here. In the remaining portions of this chapter it will be clear that most of the processes discussed, from storage mechanisms to search schemes, are under subject control to one degree or another. Except where special emphasis is required, the distinction between structural and control processes will not be stated explicitly.

The three major components of the memory system are the "sensory register," the "short-term store" (STS), and the "long-term store" (LTS). The sensory register accepts incoming sensory information and holds it very briefly while it is given minimal processing and then transferred to STS. If a large amount of information is presented quickly, then only a portion of this information can be transmitted to STS, and the precise characteristics of the sensory register will become quite important. In the experiments to be considered in this report, however, the presentation rates are slow enough, and the information quantities are small enough, that the information presented can be assumed to transit the sensory register and enter STS essentially intact. In the following, then, discussion of the sensory register will be omitted.

The short-term store is the subject's working memory; it is used for the momentary holding of information utilized by control processes such as the storage mechanisms and search schemes. Information will decay and be lost from this store within about 30 seconds or less if unattended, but may be maintained there indefinitely by rehearsal. In some situations, such as those discussed in Section 4 of Atkinson and Shiffrin (1968), the primary function of STS is one of memory -- that is, information will be maintained there via rehearsal from the time of presentation until the moment of test. The situations in which STS assumes this function are ones in which the study-test intervals are short, interference is high, and long-term learning is difficult. In other situations, such as the ones examined in this report, the memory function of STS is utilized in a different manner; STS is used for the temporary holding of information needed for long-term processing. Thus information needed for coding and search schemes is temporarily stored in STS. Although STS is utilized for the transient handling of information, it is not utilized for maintenance of the information until the moment of test.

The long-term store is a permanent repository for information. It will be assumed that information once stored is never thereafter lost or eliminated from LTS, but the subject's ability to retrieve this information will vary considerably with such variables as time and the amount of intervening, interfering material. The interaction between STS and LTS, in terms of the mechanisms and stages of storage and retrieval, is the main concern of this chapter. We turn to these considerations directly.

Storage and Retrieval

The discussion here follows the terminology of Shiffrin and Atkinson (1968). Storage refers to the set of processes by which information initially placed in STS is examined, altered, coded, and permanently placed in LTS. Retrieval refers to the inverse operations by which desired information is sought for, recovered, and emitted at test. It is convenient to subdivide both storage and retrieval into three components. The components of storage are "transfer," "placement," and "image-production." The transfer mechanism includes those control processes by which the subject decides what to store, when to store, and how to store information in LTS. The placement mechanism determined the LTS location in which an ensemble of information under consideration will be stored. Image-production is the process by which a portion of the information ensemble presented for storage will achieve permanent status in LTS. The components of retrieval are "search," "recovery," and "response-generation." Search is the mechanism by which an image is located in memory. Recovery is the mechanism by which some or all of the information in a stored image is recovered and made available to the short-term store. Response generation consists of the processes by which the subject translates recovered information into a specific response.

Before detailing the above processes, there are several general comments to be made about LTS as a whole. First, the use of the term "location" is not meant to imply necessarily a specific cortical area; rather, an LTS location is a psychological construct used to denote closeness of storage. The closer the location of two stored images, the more likely the examination of one will occur jointly with the examination of the other. Thus to say an image is stored in a single LTS location is to imply that the information in the image will tend to be recovered together. Second, a number of different terms will be used to denote an ensemble of information stored in some LTS location: ensemble of information, image, and code will be used interchangeably.

Finally, the structure of LTS may be clarified by an analogy with computer memories. A location-addressable memory is the normal computer memory; if the system is given a memory location, it will return with the contents of that location. A content-addressable memory is constructed so that the system may be given the contents of a word and will return with all the memory locations containing those contents. A location-addressable memory must be programmed before this is possible: an exhaustive search is made of all memory locations and the locations of all matches recorded. There are two primary methods for construction of content-addressable memories. In one, a fast parallel search is made of all locations simultaneously, with a buffer recording the locations of matches. In the other, the contents themselves contain the information necessary to identify the location where those contents are stored. This latter possibility can occur if the information is originally stored in accord with some precise plan based on the contents, as in some form of library shelving system. When followed at test, this storage plan will lead to the appropriate storage location. For example, a library with a shelving system based on the contents of books would store a book on the waterproofing techniques for twelfth century Egyptian rivercraft in a very precise location. When a user later desires a book with these contents, the librarian simply follows the shelving plan used for storage and directly reaches the storage location. This type of memory will be termed self-addressing. The point of view adopted in this report is that LTS is largely a self-addressing memory. That is, to a fair degree of accuracy, presented information will lead at once to a number of restricted locations where that information is likely to be stored. To give this discussion concrete form consider an experiment in which a series of consonant trigrams are presented and the subject's task is to tell whether each one has been presented previously or not. Suppose JFK is presented. In a location-addressable memory an exhaustive search would be carried out comparing JFK with each stored code. In a content-addressable memory of the first type, a parallel search is carried out which gives the locations of codes containing JFK. We assume, however, that LTS is self-addressing; hence a search is at once made of those locations where JFK is momentarily most likely to be stored. These locations are defined by a number of fairly restricted areas. The long-term store is assumed to be only partially self-addressing in that a search must next be initiated within each probable area to determine whether the desired information is indeed present. We now turn to a detailed discussion of storage and retrieval.

Storage

It is convenient to discuss the three components of the storage process in an order opposite to that normally obtaining. Thus we consider first the image-production mechanism. Image-production refers to the process by which some portion of an ensemble of information directed to some LTS location is permanently fixed there. The subject can control this mechanism in two primary ways. In the first, the subject may control the number of presentations of the information ensemble, more repetitions resulting in a larger proportion of information stored in the final image. In the second, the duration of the period of presentation may be controlled by the subject -- the longer the period during which the information resides in STS, the larger the proportion of information stored. Apart from these means, image production is beyond the control of the subject. In many applications it will simply be assumed that a random proportion of the presented information will be permanently stored.

No distinction will be made in this report between the quality and quantity of stored information; rather each image, or portion of an image, will be described by a strength measure which lumps both quality and quantity. The strength of an image will be a number between 0 and ∞, the higher the number the greater the strength. In the paired-associate situation, it is necessary to consider three strength measures, one describing stimulus related information, one describing response related information, and one describing stimulus-response associative information. This varied information may or may not be stored in the same LTS location. Specifically, it will be assumed that the stimulus information stored will have a strength distribution $F_s(I)$, the response information will have a strength distribution $F_r(I)$, and the associative information will have a strength distribution $F_a(I)$. (It should be apparent that these measures may be partially independent from each other. For a given stimulus-response pair, the subject may store information solely concerned with the stimulus, solely concerned with the response, or partially concerned with their association; these measures may even be stored in separate locations.) The form of the three distributions above will vary according to the experimental task and the techniques of storage adopted by the subject, but in general will have some spread. For example, a "good" stimulus-response pair is one that will typically result in a larger amount of stored information than a "bad" pair.

The placement process determines where information shall be stored. As pointed out previously, LTS is assumed to be largely a self-addressing memory; hence the information stored will partially direct itself to its own storage location. Thus a visual image of a cowboy will be stored in the appropriate region of the visual area of LTS. From a different point of view, it may be seen that placement will be determined by the form of the code adopted by the subject. A visual code will result in a different storage location than an auditory code. A mediator may establish its own storage location; for example, the pair QWZ - 64 may be stored via use of the mediator "the 64,000 dollar question," and the location used may be in the "television-quiz-show" region of LTS. In a paired-associate task, (when inter-pair organizational schemes are not feasible, as in continuous paradigms), the placement method yielding the best performance is one in which the location of storage is as unique as possible while simultaneously being recoverable at test. Since the stimulus is presented at test, it is most efficient to store in a location determined by stimulus information. Experiments demonstrating the relative efficacy of, say, visual imagery instructions as opposed to no instructions, demonstrate that subjects are not often aware of the most effective placement techniques to be utilized. Considerable subject differences are often found in long-term memory experiments for this reason.

The transfer process consists of subject decisions and strategies detailing what to store, when to store, and how to store information currently available in STS. It is a rather important process in most experiments because of the high degree of control that the subject exerts over it. When to store is the first decision that must be made. Consider a new paired-associate that has not been seen previously; the subject must decide whether to attempt to encode this pair. If the study time is long enough, and if the presented information is simple enough, then a coding attempt may always be made. In most experiments, however, these conditions are not met, and the subject will not find it feasible to attempt to encode every item. In this event, the decision to encode will be based upon momentary factors such as the expected ease of encoding, the time available for encoding, the importance of the item, the extent to which the item fits into previously utilized storage schemata, and so forth. In continuous experiments with homogenous items, these factors will vary randomly from trial to trial and we may assume that α, the probability of attempting to store a new item, is a parameter of a random process, and identical for each new item presented. The same holds for a previously presented item about which no information can currently be retrieved from LTS. In this latter case, however, the image stored will be in a different location than the unretrievable previous image; thus an item may have two or more codes stored in LTS over a period of reinforcements. At a subsequent test the information in each of these codes will have some chance of retrieval. If an item is currently retrievable from LTS when presented for study, then the subject has several options. When sufficient time is available for study, the subject may decide to store a new code in a new location. With less time available, information may merely be added to the current code. In complex tasks with short study periods the subject may be satisfied with simply tagging the current code with temporal information that will update it to the present time.

When a stimulus that has previously been presented with one response, called R1, is presented for study with a new response, called R2, several mechanisms may come into play. Either instructional set or individual initiative may lead a subject to add the information encoding the R2 response to the code for the R1 response (if this code is present in LTS and currently retrievable); this mechanism can be called "linking" or "mediating." Mediating is especially useful if a future test will require that both the R1 and R2 responses be given. In other situations, especially those where the subject is instructed to "forget" the R1 pairing when the R2 pairing is presented, the R2 pairing may be coded in independent fashion and stored in a new location. As was the case for a new item, it is assumed that the probability of attempting to code is a parameter α_0, which may be different than α. Note that there is no assurance that α or α_0 will not change from one reinforcement to the next. Especially in list structured experiments, there may be increasing incentive for coding unretrievable items as learning proceeds. However, in the continuous tasks we shall be discussing, it is not unreasonable to expect this probability to remain constant over successive reinforcements.

Each of the components of the storage process are accomplished by the subject via one action: the generation and maintenance in STS of the information intended for storage. It is assumed that information is transferred to LTS from STS during the period that the information resides in STS.*

*Throughout this paper, transfer of information is not meant to imply that the information is removed from one location and placed in another. Rather, transfer implies the copying of information from a location without affecting it in any way.

Retrieval

When a test occurs the subject will first search STS and then LTS for the desired information. The STS search is assumed to be a relatively fast and accurate process compared with the LTS search. In the following, we shall consider only the case where the desired information is not found in STS, and the retrieval process will be considered solely as it applies to LTS. LTS retrieval is assumed to take place as follows. The search process generates an image to be examined. The recovery process makes some of the information contained in this image available to STS. Finally, response-production consists of decisions concerning whether to output a response found, whether to cease searching, or whether to continue the search by examining another image. The search continues until it terminates of its own accord, or until an external time limit of the experimental procedure has expired. Retrieval is best described as a rather complex sequential search scheme.

Search. Because memory is assumed to be partially self-addressing, a stimulus presented for test will at once lead to a number of likely LTS locations where information about that stimulus may be stored. In certain cases the stimulus will have some characteristic so salient that a storage location is defined uniquely and precisely. This location will then be examined. If the experiment is such that certain stimuli presented for test may be new (not presented previously), and if no stored information is found in the location indicated, the subject may decide that the stimulus is new, and cease further search. There will be a bias mechanism determining how much information must be present for the search to continue. In most cases, the information required will be extremely minimal, since the coded image itself may be stored in a location other than the one indicated by the salient stimulus characteristic.

Regardless of the salience of the stimulus characteristics, the images or codes examined will initially be determined by stimulus information $[F_s(I)]$. That is, the locations in memory to be examined will be roughly indicated by information contained in the stimulus presented. Within the regions thus indicated, an image will be chosen for examination partly on the basis of recency (temporal information stored), partly on the basis of its strength, and partly on the basis of chance. Once the search has begun successive images examined will depend not only upon stimulus information, but also upon associative information recovered during the search. In a continuous paired-associate task the conception of the search may be simplified somewhat, as illustrated in Figure I-1. We first define a "subset" of codes in LTS which will eventually be examined if the search does not terminate via a response recovery and output. This subset will be termed the "examination-subset." It is then possible to consider the order of search through this subset. Figure I-1 portrays this process. The stimulus of the paired-associate labeled number 18, on the far left, has just been presented for test, on trial 70. The second row from the bottom in the Figure gives the sequence of presentations preceding this test. The third row from the bottom gives the images stored in LTS for each item presented, where the height of the bar gives the strength of the code stored (lumping stimulus, associative, and response information.) The fourth row from the bottom gives those codes that are in the examination-

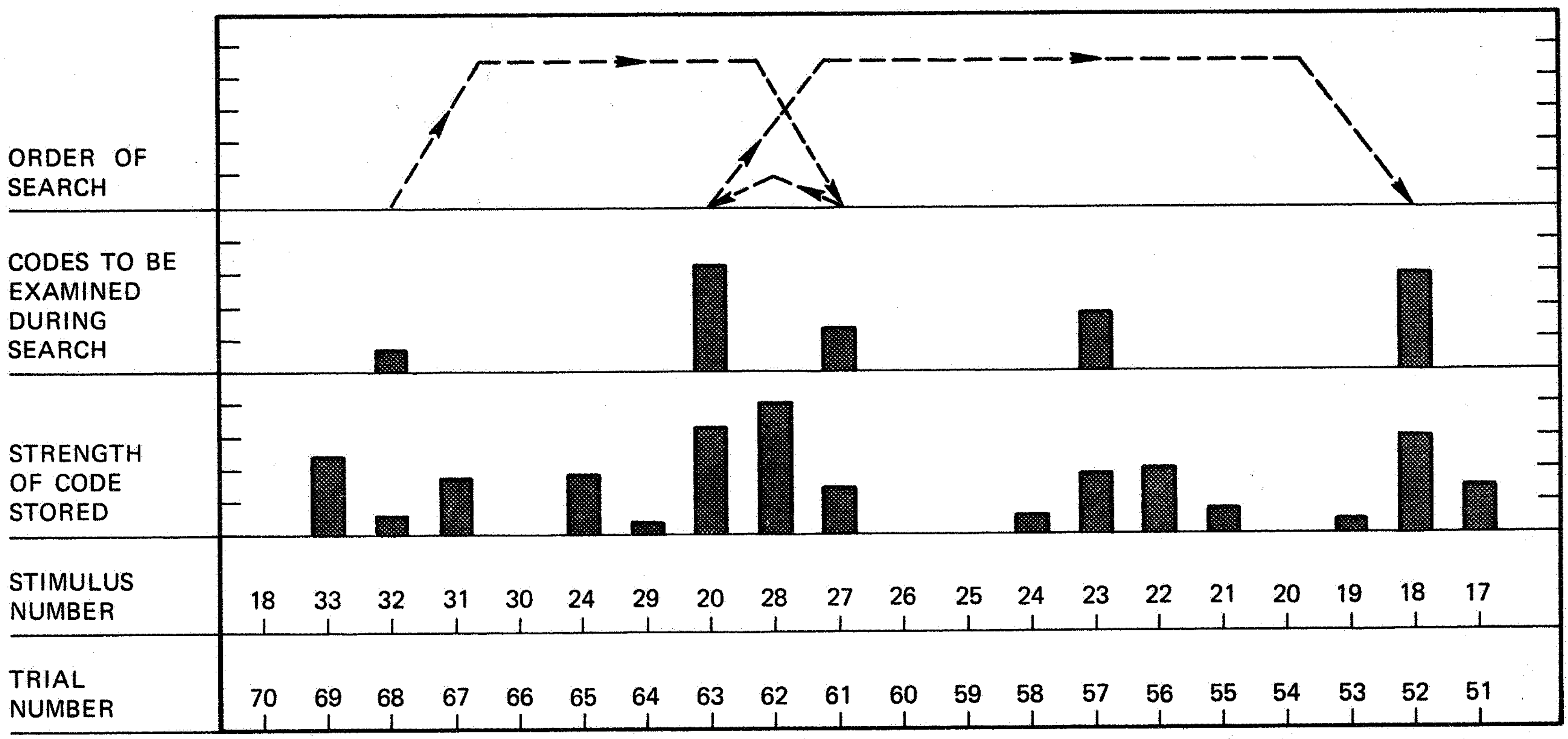

Figure I-1. An LTS Search in a Continuous Memory Task.

subset. The arrows on the top of the Figure give the order of search through the subset. Thus item 32 was first examined and rejected, then item 27, then item 20. Finally, the code for item 18 was examined, the response coded there was recovered and accepted, and the search ended with a correct response. Note that item 23 was not examined because the search terminated.

In continuous tasks it may be assumed generally that the order of search through the subset of codes is a function both of the "age" and strength of the codes involved, where age is related to the number of items that have intervened between storage of a code and the present test. It seems clear that temporal information must be an important determiner of search order. In free recall tasks, for example, successive series of items are presented to the subject. Following each series, the subject attempts to output the members of the series. The important finding for present purposes is that intrusions from one series in the responses for a following series are extremely rare; apparently subjects can order their search temporally so that only the members of the most recent list are examined during retrieval. The question of the degree to which search order depends upon temporal factors will be examined in Chapters II and III, and will not be discussed here.

There are several factors which help determine which codes will be in the examination-subset. Denote the image which encodes the pair currently being tested as a <u>c-code</u>. A <u>c-code</u> should have a higher probability of being in this subset the higher its strength (primarily the amount of its stimulus information). Other images, denoted <u>i-codes</u>, should have a probability of being in the subset which is greater, the greater the degree of generalization between its stimulus information and the stimulus being tested. In general, however, i-codes will have a much smaller probability of being in the subset than a c-code of equal strength. As a result, the total number of codes making up the subset of codes to be examined may be fairly small.

Recovery. Recovery refers to the extraction of information from the image under examination. The recovery of a desired complex of information, if this information is actually encoded in the image under examination, should be a monotonic function of the strength of the image. A number of decisions are dependent upon the outcome of the recovery process. Stimulus information recovered is largely responsible for accepting or rejecting the image as containing the desired response. That is, regardless of response information recovered, if the stimulus information is discrepant with the stimulus being tested, then the search will skip by this image and continue elsewhere. Response information recovered allows the subject to emit the encoded response. Associative information recovered will often serve the purpose of directing the search to a different LTS location where an image encoding the response may be stored.

Response Generation. Following recovery of information from an image, a decision process must be utilized to decide whether to emit a response, and if so, what response. It will normally be the case that the stimulus information recovered from a c-code will be congruent with the stimulus being tested, and a decision will then be made to attempt to output the response if at all possible. Whether a response can be emitted will depend upon the response information recovered. In cases where the response set is well delineated, a criterion is assumed to be set which will monitor the sensitivity of the output process. If the criterion is set quite low, then many responses will be emitted, but they will often be wrong. If the criterion is set quite high, few responses will be given, but these will almost always be correct. For i-codes the probability of emitting a response will be considerably lower than for c-codes; this occurs because output may be suppressed when the recovered stimulus information does not match the stimulus being tested. Thus a response will be emitted after examination of an i-code considerably less often than after examination of a c-code. In some applications (as in Chapter III) the recovery and response generation processes will be lumped for simplicity into a single process. In this event the probability of output of the response encoded will be a function of the strength for c-codes. For i-codes the strength will be multiplied by a generalization parameter less than one; the resultant quantity will be termed the "effective strength" of the i-code. The probability of output will then be the same function as for c-codes, but the function will be based upon the effective strength of the i-code. This scheme will be discussed fully in Chapter III.

Search Termination. Depending upon the task, a variety of mechanisms help determine when the search ceases. If the test interval is quite short, then the search may continue until a response is output or time runs out. Furthermore, if the test interval is short, the subject may output the first likely response recovered in the search. When longer response periods are available, then the search might be allowed to continue until a number of likely responses are recovered; these responses will then be evaluated and a first choice chosen for output. When sufficient time is available, the subject may adopt one of a number of sophisticated termination schemes. These were discussed in Atkinson and Shiffrin (1965) and will not be discussed further here.

Applications and Extensions

We shall next consider applications of the theory to a variety of manipulations which may be carried out in the context of a continuous paired-associate design. Primarily we shall discuss those variations which were actually employed in the experiments presented in Chapter II.

Recognition and Recall. In a recognition test, a specific item is presented and the subject must attempt to ascertain whether this item has been presented previously in the session or not. It has sometimes been assumed that use of such a test will eliminate search from the retrieval process, but this is not necessarily correct. Characteristics of the item presented will lead the subject to examine some restricted LTS region for relevant information. The more salient are these characteristics, the more restricted will be the region indicated, and the smaller will be the search needed to locate the desired information. In general, however, some search will be required. When a stimulus is presented in a recall test where the number of responses is large, a considerably more extensive search is required. This occurs because stimulus information alone is required for the recognition phase, but the response may be encoded in quite another LTS location than that indicated by any salient stimulus characteristics. In a continuous paired-associate task with recall tests, recognition is still an important process; for example, the subject may recognize that a stimulus presented for test is new and has not been previously presented; upon such a recognition, the search will cease. When the task is such that the subject may either refrain from responding or emit a response, then wrong responses actually emitted are called intrusions. Due to the recognition process, the intrusion rate for new items being tested may be considerably lower than that for previously presented items.

Ranking. The task may require the subject to rank a series of responses in the order of their perceived likelihood of being correct. When the retrieval scheme is such that the search ceases when the first likely response is recovered, then the response ranked first will often be correct. However, responses ranked after the first will be correct only to the degree expected by pure guessing. If on the other hand, enough time is available for several likely responses to be recovered and considered, then responses ranked after the first will be correct at an above chance level. The degree to which the rankings after the first will be above chance will depend upon the decision process used to choose between likely responses, and also the coding schemes used.

Second-Guessing. Second-guessing refers to a procedure in which the subject is told whether his first response is wrong; if it is wrong he is then allowed to make an additional response, called the second-guess. First consider the case where a search procedure is used that would not result in an above chance ranking effect, i.e., the first likely response recovered in the search is output. When informed of an incorrect response, the subject will initiate another search of LTS. Performance on the second-guess will be partly determined by the degree of dependence of the second search upon the original search. If the second search is completely dependent, both in terms of the items making up the examination subset and also the order of search, then a correct second-guess can be made only in those instances where the wrong first response was an intrusion emitted before the c-code was examined in the original search. In these instances, the second search may continue beyond the point of the intrusion and thereafter result in a correct recovery. On the other hand, if the searches are completely independent, then correct recoveries can be made during the second search in cases where the c-code was present in LTS but not in the examination subset during the original search. In this event, the c-code might be in the examination subset during the second search. These considerations are complicated slightly if the original search was of the type which recovers several likely response alternatives, ranks them, and outputs the most likely. In this case, it is possible for the subject to forego a second search entirely and simply give the response ranked second most likely during the original search. If a second search is nevertheless engaged in, then the final response given must be the result of a decision process involving all the likely response alternatives recovered during both searches.

Regardless of the form of the second-guess search, there is no guarantee that the parameters of this search will be the same as on the original search. In particular, it would be natural for the subject to lower his criterion for output of recovered responses, since the original error indicates that the state of knowledge regarding the correct answer may be quite weak.

Interference Phenomena. Interference refers to a paradigm in which the first response paired with a stimulus (R1) is changed to a different response (R2); a subsequent test for R1 is called a retroactive interference condition, while a subsequent test for R2 is called a proactive interference condition. Although considerable work on interference phenomena has taken place within designs employing repeated presentations of whole lists of paired-associates, it is currently uncertain what form these phenomena will take in a continuous task. This entire question will be discussed more fully in subsequent chapters of this report. For the present we should merely like to point out that the theory can predict either proactive or retroactive interference effects. That is, learning of the R1 response may hinder recall of the R2 response, or vice versa. The predictions will depend upon the precise form of the assumptions regarding order of search and the addition of information to codes currently stored in LTS. For example, if search order is strictly temporal and proceeds starting with the most recent item, and if the original response code is older than the new response code, then no proactive effect will be expected. This prediction results from the following argument. In those cases where both the old and new codes for a stimulus are simultaneously in the examination subset, the new response code will always be examined prior to the older response code. Hence the probability correct will not be affected by the presence or absence of the older code.* On the other hand, a strong retroactive

*This is not quite true, but approximately so. Recovering the R1 response and emitting it will insure that an error is made. On the other hand, a different type of intrusion, or a pure guess, will be correct at the chance level. Thus the above argument is true when the chance level is zero, and is almost true when the chance level is quite low.

effect will be expected in this case, at least if the search terminates at the R2 code an appreciable proportion of the time.

To the degree that the strictly temporal search order assumption is relaxed, a proactive effect will be expected. However, if information is added to the R1 code that the response has been changed, then the search will bypass that code and continue; thus the proactive effect will be dependent on the information added to the R1 code when the response is changed. These same factors apply to retroactive interference. This discussion should make it clear that the theory has a good deal of freedom with regard to interference predictions. Experiment II in the next chapter examines proactive interference, and further discussion is reserved until that point.

Latencies. The recovery of a response from STS is assumed to be associated with a very short latency. The latency associated with a response recovery from LTS is assumed to be monotonically related to the number of codes examined before the response is given, the more codes examined, the slower the response. For the present discussion, components of response time associated with the decision processes involved in retrieval will be ignored. This rather simple conception of latencies leads to a large number of predictions. The latency of pure guesses should be quite long, since guesses occur only at the conclusion of an unsuccessful LTS search. The latency of intrusions will depend upon the order of search, but will probably be somewhat larger than correct response latencies. The latency of a correct response is expected to increase as the length of the period since the previous presentation increases, since a greater number of codes will tend to be examined prior to the c-code as this period increases. The correct response latency will be expected to decrease as the number of reinforcements increases, since the c-code will tend to be stronger, and codes of greater strength will tend to be examined earlier in the search. This list of predictions may be extended in a natural fashion to change-of-response conditions, and to second-guess conditions, but further discussion will be reserved until the latency data of Experiment II is examined.

CHAPTER II

THE EXPERIMENTS: DESIGN, PROCEDURE, AND RESULTS

The two experiments of the present study were designed to investigate various facets of search and retrieval from long-term memory, and to provide a source of quantitative data against which a specific version of the theory outlined in Chapter I could be tested. Although both experiments utilized a continuous paired-associate design, the differences between them were considerable and their procedures will be described separately. The experiments are referred to as continuous because a particular item may have had its first presentation on any trial of the experiment, appeared a few times at varying intervals, and then been discarded. Each trial of the experiments consisted of a test phase followed by a study phase. During the test phase a stimulus was presented alone and the subject was then tested in some detail concerning his knowledge of the correct response. During the study phase, the stimulus just tested was presented with a response to be remembered. In what follows, we use the term _lag_ to refer to the number of trials intervening between two successive presentations of a particular stimulus.

Experiment I

Design Justification. Experiment I was designed with several objectives in mind. A primary aim was the independent establishment of the imperfect-search characteristics of memory retrieval in the paired-associate situation. In order to accomplish this, a design was utilized which would separate two components of "second-guessing" performance: the partial-information component and the imperfect-search component.

A number of paired-associate experiments have shown that performance on a second response (following information that a first response was incorrect) may be well above chance level (Bower, 1967; Binford and Gettys, 1965); other experiments have shown that ranking of responses in their order of being correct can result in rankings beyond the first choice which are also above the chance level (Bower, 1967). These findings can be explained by either of two models: in the first, retrieval from memory results in recovery of partial information about more than one response; in the second, retrieval results in recovery of information about only one response; but if it's an error, a second search of memory results in recovery of new information about some other response. These models are separated in Experiment I by utilizing both rankings and second-guesses on each test trial.

The second major objective of Experiment I was the examination of changes in retrieval of individual items from memory, in a steady-state situation. Forgetting, particularly, needs extensive examination in a continuous task, since almost all the research on long-term forgetting has utilized a list-structure design. In such a design performance changes are measured for whole lists, and then inferred for individual items, but this inference lacks validation. For this reason, list structure is eliminated in Experiment I by using a continuous task: new items are continually being introduced, and old items eliminated.

A third objective of Experiment I was the demonstration that a class of previously used models for paired-associate learning suffered from certain deficiencies, deficiencies not present in the theory of Chapter I (henceforth called LTS theory). The design of Experiment I is similar to those used by Bjork (1966) and Rumelhart (1967). Each of these workers used a model to describe their data which has been called the GFT. The GFT model is basically a three state Markov model with a long term absorbing state (L). The probability that an item will be in L increases as the number of presentations of the item increases. Once an item enters L, a correct response will always be given and the item cannot thereafter leave L. Thus the GFT implies that the probability correct following a given sequence of reinforcements cannot be lower than a certain minimum, regardless of the lag of the current test; the minimum is determined by the probability that the item is in the state L at the time of test, which is not affected by the previous lag. These predictions are quite at odds with LTS theory: as long as new items are continually being introduced, LTS theory predicts that the probability correct should decrease toward chance as the lag increases. It is not surprising that the Bjork data was handled well by the GFT, because the design used did not allow for the continual introduction of new items; rather the design basically utilized a list structure, so that all items late in the session had been presented many times before. In such a situation LTS theory predicts that all items will become permanently learned, much as if an absorbing state was present; the prediction is based on many factors, which are described in Shiffrin and Atkinson (1968). Thus either GFT or LTS theory will provide an adequate description of list-structured designs. The Rumelhart study, on the other hand, used a design in which new items are continually being introduced; nevertheless the GFT model fit the data quite adequately. We propose that the GFT model proved adequate only because the range of lags examined was quite restricted, never being larger than 32. It should be possible to demonstrate that the GFT model is inadequate if a large enough range of lags is examined. For example, if the probability correct at very long lags tends toward chance, then a model in which an appreciable number of items enter an absorbing state will not be appropriate. For these reasons, the range of lags examined in Experiment I is very large, ranging from 0 to about 225.

Design. A daily session for each subject consisted of a series of 440 trials, each made up of a test phase followed by a study phase. On each trial a stimulus, possibly one not presented previously, was chosen according to a prearranged schedule and presented for test. Following the test phase that same stimulus is presented with a correct response during the study phase. The sequence in which the stimuli are presented for test and study are the same for every subject and every session; Appendix 1 gives the actual sequence used. In the Appendix, the sequence of trials is given in terms of the stimulus number. For a given subject and session each stimulus number represents some randomly chosen stimulus (actually a consonant trigram). Thus the sequence of trials remained fixed, but the actual stimuli and responses were changed from session to session.

A particular stimulus could be presented for a maximum of eight trials (eight reinforcements), at varying lags. Table II-1 gives the sequence of lags associated with each "item-type," where a stimulus of item-type *i* is presented at successive lags according to the *i*th row of the table. The first column in Table II-1 gives the item-type. The next seven columns give the successive lags at which items of each type

TABLE II - 1

SEQUENCE OF LAGS FOR ITEM-TYPES
OF EXPERIMENT I

Item-type	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Number of Sequences
1	1	1	1	1	1	16	100	6
2	1	1	1	1	1	100	100	2
3	6	6	6	6	6	16	100	6
4	6	6	6	6	6	100	100	3
5	10	10	10	10	10	16	100	7
6	10	10	10	10	10	100	100	4
7	25	25	25	25	25	25	25	7
8	50	50	50	50	50	50	50	4
9	100	100	100	100				8
10	0	100	100	100	100			4
11	1	100	100	100	100			4
12	10	100	100	100	100			5
13	~225							6

are presented. The final column gives the number of stimuli of each item-type that are presented during each experimental session. As indicated in the table, the lags vary from 0 to about 225. The different stimuli of a given item-type are given first presentations which are spaced fairly evenly throughout each experimental session; the exact presentation schedule is presented in Appendix 1.

Four responses are used in Experiment I. When a stimulus is presented for test the subject responds by ranking the four responses in the order of their likelihood of being correct, using a random ranking if he does not know the correct answer. If the response ranked first is incorrect, then the subject is informed of this fact and he proceeds to rerank the three remaining alternatives, not necessarily in the same order as on the first ranking, and again guessing if the answer is not known. In order to make subsequent discussions clear, we adopt the following terminology. The subject's first four responses on a test trial are referred to as the "ranking." The second group of three responses (when given by the subject) is referred to as the "reranking." There is a further breakdown depending on the order of response. Thus the first response given on the test trial is called the "first-ranking," the second is called the "second-ranking," etc. The first response of the reranking (when the subject engages in reranking) is termed the "first-reranking" and so forth. It should be noted that the ranking responses in this experiment are akin to the responses given in the typical ranking experiment in the literature. Similarly, the first-ranking and first-reranking responses in this experiment are akin to the responses given in the typical second-guessing experiment.

Subjects. The subjects were ten students from Stanford University who received \$2.00 an hour for their services. Each subject participated in a minimum of 8 and a maximum of 11 experimental sessions. The sessions were conducted on weekday evenings and took approximately 1-1/4 hours each. The subjects were procured without regard for sex through the student employment service.

Apparatus. The experiment was conducted in the Computer-Based Learning Laboratory at Stanford University. The control functions were performed by computer programs running in a modified PDP-1 computer manufactured by the Digital Equipment Corporation, and under control of a time-sharing system. The subject was seated at a cathode-ray-tube display terminal; there were five terminals each located in a separate 7 × 8-ft. sound-shielded, airconditioned room. Stimuli and other information were displayed on the face of the cathode ray tube (CRT); responses were made on an electric typewriter keyboard located immediately below the lower edge of the CRT.

Stimuli and Responses. The stimuli were 990 consonant trigrams (CCC's) made up of all possible 3 letter permutations of the following consonants: B,D,F,G,J,K,P,Q,X,W, and Z. Thus a typical stimulus was JXQ. Ninety stimuli were randomly selected for use during each session, with the restriction that any stimulus used in a session could not be used in any succeeding session for that subject. Thus a subject could not take part in more than 11 sessions.

Four responses were used: the numbers 1,2,3, and 4. Thus the guessing probability of a correct first-ranking was 1/4 and the guessing probability of a correct first-reranking was 1/3.

Instructions

When a subject arrived for the first session he was given a sheet of instructions to read, as follows:

> "This is an experiment to test your memory. You will be sitting in a soundproof booth facing a T.V. screen with a typewriter keyboard below it. Each day take the same booth as the previous day. To start the session, type the semicolon (;). The experiment will then begin.
>
> You will be required to remember the response members of a number of paired-associates, each consisting of a nonsense-syllable paired with a number as a response. The responses will always be either 1,2,3, or 4. Each paired-associate will be presented a number of times during a session and you should try to learn it. Each trial will consist of a test followed by a study. On a test, the word "test" appears on the top of the screen, and then below it appears a nonsense-syllable. Below the syllable will appear the term "rank answers." You will try to remember the response paired with the syllable presented for test. To respond, type the number you think most likely to be the correct response; then type the second most likely number; then the third most likely, then the least likely. That is, you will rank the responses 1-4 in order of their likelihood of being correct. As you type these 4 responses, they will appear on the screen, your first choice being on the left. If you are satisfied with your answers, then type a carriage-return (CR). If not satisfied at any point, and you wish to change your ranking, type E and the screen will clear and you may type in a new ranking. If you make a typing mistake, the screen will clear your responses at once: in this case, type them in again.
>
> When you rank the responses and type a carriage-return, the computer will check to see whether your first ranked response was correct. If it was correct, you will go on to a study trial on the syllable you were just tested on. If your first rank was incorrect, then you will get one more chance: the words "wrong. rerank answers" will appear on the screen. You will then rerank the three remaining answers in the order of their likelihood of being correct. That is, the first number typed is the first choice, etc. These "reranks" do not have to correspond to the first re-rankings. If your first ranking was incorrect, search your memory again, and then make your best possible choices. As you type in your reranks they will appear on the screen. If you are satisfied with your three choices, then type a carriage return and the test trial will be terminated. The syllable you were tested on will then be presented with the

correct response for 2 seconds of study. Then after a short delay, the next test trial will begin.

Take the time you need to respond during test trials, but attempt to respond as quickly as possible without lowering your performance.

Your task is to learn and remember as many pairings as possible and to demonstrate this learning during the test phases of the trials. Feel free to use any codes or mnemonics you can devise in order to learn the pairs.

The way the experiment is being run, a syllable will first be presented for test on a trial, and then for study. Thus, especially at the start of a session, you will be tested on syllables whose response you have not yet seen. In this case, simply rank the responses randomly, i.e., guess. When guessing, do not always type in the answers in the same way - try to guess randomly. Furthermore, even if you feel you know the answer, do not always type in the remaining answers in the same order. Try to type these answers randomly also. Any questions? The experimenter will now review these instructions with you verbally."

The experimenter reviewed the instructions with the subjects and then introduced them to the computer and its operation. The entire first session was used to familiarize the subject with the apparatus and instructions, and to give him practice at the task.

Procedure

Each session consisted of a sequence of 439 trials, a trial being defined as a test followed by a study. Each trial involved a fixed series of events. (1) The word TEST appeared on the upper face of the CRT. Beneath the word TEST a specifically determined member of the stimulus set appeared, the stimulus member indicated by the presentation schedule given in Appendix 1. Below the stimulus appeared the words RANK ANSWERS. The subject then ranked the four responses by typing them in order on the keyboard, the most probably correct answer first, and so forth. The answers appeared on the CRT as they were typed. After ranking the four responses the subject typed a carriage-return and the rankings were evaluated by the computer. Previous to this point, the subject could begin his rankings anew by typing E. If the first-ranked response was wrong (even for stimuli never seen before) then the words WRONG. RERANK ANSWERS appeared on the CRT below the original rankings, which remained on the CRT. The subject then reranked the three remaining answers under the same conditions that pertained to the original rankings. The rankings and rerankings were self-paced, but instructions were used which insured that the subject took about 6-7 seconds for responding, on the average. (2) The CRT was cleared and a blank screen appeared for 1/4 second. (3) The word STUDY appeared at the top of the CRT. Beneath the word STUDY appeared the stimulus just tested along with the correct response. The correct pairing remained on the CRT for 2 seconds. (4) The CRT was blanked for 3/4 seconds. Then the next trial began. As indicated above, a complete trial took about 10 seconds or less and thus a session lasted about 1 hour and 15 minutes.

At the start of each session, the computer randomly assigned each subject 90 stimuli he had not seen in previous sessions. Each stimulus was then randomly assigned one of the four responses as the correct pairing to be used throughout that session. It should be noted again that the sequence of trials was the same for every subject-session, but the actual stimuli and responses differed. The first 12 trials of each session consisted of 10 filler items; these appeared seldom thereafter. From the 13th trial on, almost all trials were instances of one or another of the 13 item-types listed in Table II-1. These item-types were spaced roughly uniformly through the remaining 427 trials.

Altogether 83 subject-sessions of data were collected following the initial practice session. Because of computer stoppage or other extraneous reasons, only 58 sessions were entirely completed, but the remaining sessions were at worst within 10 or 20 trials of completion. The data collected on each trial consisted of the stimulus tested and its correct response, and the rankings and rerankings given by the subject. Latencies were not recorded. At the conclusion of the experiment, each subject filled out a written questionnaire.

Results of Experiment I

Table II-2 presents the summary results for each of the 10 subjects in the experiment. Tabled is the probability of a correct first-ranking lumped over all trials and sessions. The results are listed in order of increasing probability correct. It is evident that there are appreciable subject differences in overall ability in this task. Nevertheless, in order to gain precision of estimates, the remaining data are presented in a form lumped over all subjects. This should not overly distort the observed effects, since a consideration of the data to follow, where the number of observations permitted a subject by subject breakdown, consistently showed that the same qualitative effects hold for individuals as for the average data. Possible selection effects introduced by averaging will be discussed in Chapter III.

Table II-3 gives the probability of a correct first-ranking over successive days of the experiment (the practice session is not included). It is clear that no trend over days is present in the table. Apparently, proactive interference from session to session was minimal. The data to follow will be lumped over all sessions, excluding the practice session.

TABLE II - 2

MEAN PROBABILITY CORRECT
FOR SUBJECTS OF EXPERIMENT I

Subject Number	10	7	4	2	9	1	6	3	8	5
Probability of Correct First-ranking	.45	.47	.51	.52	.54	.56	.59	.68	.69	.77

TABLE II - 3

MEAN PROBABILITY CORRECT
FOR SUCCESSIVE DAYS OF EXPERIMENT I

Day Number	1	2	3	4	5	6	7	8	9	10
Probability of Correct First-ranking	.58	.55	.58	.62	.61	.55	.56	.63	.54	.60

TABLE II - 4

PROBABILITY CORRECT AS A FUNCTION OF THE AVERAGE STATE
OF KNOWLEDGE CONCERNING THE ITEMS
MAKING UP THE PRECEDING LAG

	Low K Group Pr(C)	High K Group Pr(C)
Lag 1, Reinforcement 1:	.70	.75
Lag 6, R1:	.54	.61
Lag 10, R1:	.54	.57
Lag 25, R1:	.43	.52
Lag 50, R1:	.35	.43
Lag 100, R1:	.31	.39
Lag 1, Reinforcement 2:	.85	.88
Lag 6, R2:	.70	.76
Lag 10, R2:	.67	.68
Lag 25, R2:	.54	.57
Lag 50, R2:	.37	.43
Lag 100, R2:	.47	.46

Ranking Performance vs. Second-Guessing Performance. As stated earlier, a number of previous experiments have found that responses ranked after the first choice are correct at an above chance level. A hypothesis which can explain this finding holds that the subject sometimes retrieves from memory information which indicates the possible correctness of two or more responses. The subject examines this ambiguous information and then produces his rankings as the result of some type of decision process. Thus the correct response is sometimes ranked second rather than first, and the above finding is observed. Other experiments in the literature demonstrate that second-guesses, after the subject is told the first-guess is wrong, can result in performance well above chance levels. The hypothesis proposed above can also be utilized to explain this result: the subject engages in implicit ranking on the first guess and gives the response implicitly ranked first; if he makes an error, he then outputs the response he had previously ranked second. It is possible, however, that a substantial portion of the second-guessing effect may be explained by an alternative hypothesis: the subject makes his first guess on the basis of information available at the time; upon knowledge of an error he then engages in an additional search of memory. This second search sometimes results in retrieval of information not previously available to the subject, information which may then be used to respond correctly. This hypothesis is quite different from the first in its emphasis of the essentially probabilistic nature of the memory retrieval process.

The present experiment provides a means of separating these hypotheses. The essential statistic examines those instances where the response ranked first is wrong, but where the response <u>reranked first</u> is not the response <u>ranked second</u>. For these instances, a probability of correct first-reranking above the level expected by chance guessing implies that the second hypothesis is operative in the experiment. A convenient way to begin an analysis of the data is presented in Figure II-1. On the abscissa is the probability of a correct first-ranking divided into successive intervals which are marked on the graph. These intervals start at .30 since no item-type had a probability of correct first-ranking on any test after the first reinforcement which was below .30. For each interval we consider all trials in the sequence of 440 on which the probability of correct first-ranking lies in the interval. For these trials we graph (1) the probability that the second-ranked answer is correct and (2) the probability that the first reranked answer is correct. Both probabilities are plotted conditional upon a first-ranking error; thus the chance level for both probabilities is .33. In what follows we will refer to the first-reranking as second-guessing.

From the upper curve in Figure II-1 it is evident that a substantial amount of correct second-guessing has taken place. On the other hand, the lower curve indicates that virtually no initial ranking effect took place. The probability of correct second-ranking is barely above the chance level, the mean for all trials except those on which new stimuli are presented being .352. This probability is significantly above chance since it is based upon approximately 7000 observations, but it is obvious that the magnitude of the ranking effect is small compared with that of second-guessing. This result suggests that the second hypothesis presented above is appropriate for this experiment. That is, since the

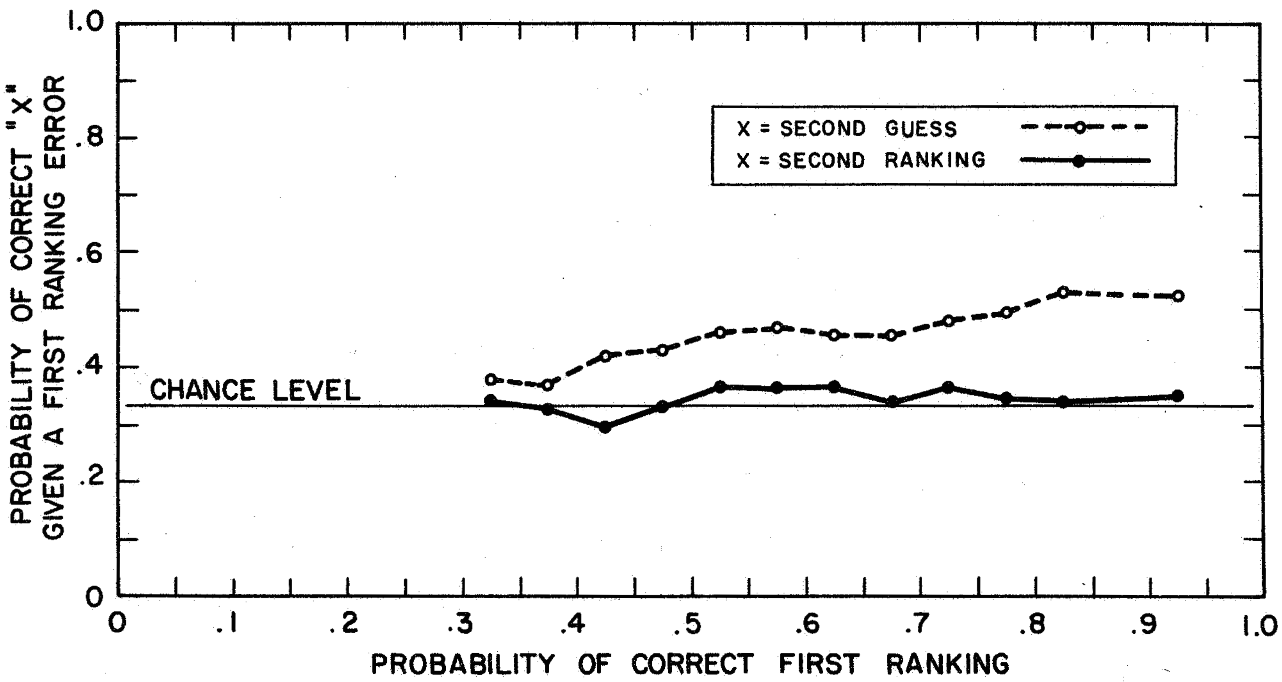

Figure II-1. Conditional Probabilities of Second-Guessing and Second-Ranking.

ranking effect was near chance, the majority of correct second-guesses were responses that were not ranked-second during initial ranking. Thus the subjects were utilizing information during second-guessing that was not utilized during initial ranking. A straightforward interpretation holds that after the error feedback a search was initiated which occasionally resulted in the correct response being found.*

It is most likely that the failure to find a large second-ranking effect was due to the instructions regarding response rate. Although responding was self-paced, the subjects were instructed to respond quickly enough to finish in an hour and a quarter, and had to respond rapidly as a result. Under these conditions, the subjects would be led to adopt a memory-search strategy which would output the first likely response alternative located in the search. If responding rates were lower, the subjects could adopt a strategy in which the memory-search continued until all likely alternatives could be recovered and evaluated. In this case a second-ranking effect would very likely result.

The failure to find a substantial ranking effect might lead us to expect that the reranking effect would also be minimal. This was indeed the case; rerankings after the first were correct with a conditional

---

*It conceivably could be argued that the subjects "knew" during their initial rankings the information they later used to second-guess, but nevertheless ignored it while making the rankings. This seems doubtful, especially if one takes the subjects own written comments into account: in several instances the subjects stated the second hypothesis almost verbatim on their final questionnaire. In any event, if the need arose, it is not difficult to formulate experiments to clear up this possible ambiguity, perhaps by giving positive payoffs for correct second-rankings.

probability of .498, almost exactly the level expected by chance. As a result, the remaining data analysis is considerably simplified. Only the first-ranking and first-reranking results will be considered and will be referred to as first-guessing and second-guessing respectively.

Learning and Forgetting. The title of this section should not be misconstrued: by learning and forgetting is meant only increases and decreases in retrieval. As indicated in Chapter I, our theoretical approach does not allow for the disappearance of stored information from memory, and the use of the term forgetting should not be taken to mean such.

In the following data the number of observations at each point may be found approximately by reference to Table II-1: for each item-type, multiply the entry in the column headed "NUMBER OF SEQUENCES" by 80, the approximate number of subject sessions. Figure II-2 presents the lag curves for first reinforcement items. The top panel presents the probability of a correct first-guess following an item's first reinforcement at a lag marked on the abscissa. The lower panel presents the probability of a correct second-guess conditionalized upon an error on the first guess. The observed data are plotted as open circles connected by dashed lines. The predictions are based on the model presented in Chapter III and may be ignored for the present. As might be expected in a continuous task, the lag curve decreases toward chance as the number of intervening items increases, albeit quite slowly. The chance level in the top panel is .25, and in the bottom panel is .33. The second-guessing curve is of interest because of its relatively small variance over the range of lags shown, and because of its maximum at about a lag of 10 or thereabout.

Discussion of the second-guessing data is reserved for the next chapter. The first-guess curve is most important because it demonstrates that the probability of a correct response tends toward chance as the lag increases. Thus the GFT model, or any model with a long-term absorbing state, will not provide an appropriate description of the data.

Figures II-3 and II-4 present the "learning" curves for each of the item-types in the experiment. The probability of a correct first-guess is plotted as a function of the number of presentations, for each item-type. The lag between successive presentations is listed in each graph as a small number placed between successive points on the predicted curve. In the two figures, the chance level is .25. Figures II-5 and II-6 present the same curves for second-guessing. These figures present the probability of a correct second-guess conditionalized upon a first-guess error; thus the chance level is .33. In each of these last four figures, all curves begin at the chance level, since on the first presentation the subject has not previously seen the item being tested. In Figure II-5 several observed points have been deleted from the Type 1 and Type 2 graphs. The number of observations at these points was below 30 (because the probability of a correct first-guess was so high).

Several characteristics of these data should be noted at this time. First, as found by previous workers (Greeno, 1964; Peterson, Hillner, and Saltzman, 1962; Rumelhart, 1967), a distributed practice effect occurred. Consider item-types 10, 11, and 12 in Figure II-4. As the first lag was varied from 0 to 1 to 10, the probability correct after a subsequent lag of 100 rose from .37 to .44 to .49; i.e., the longer

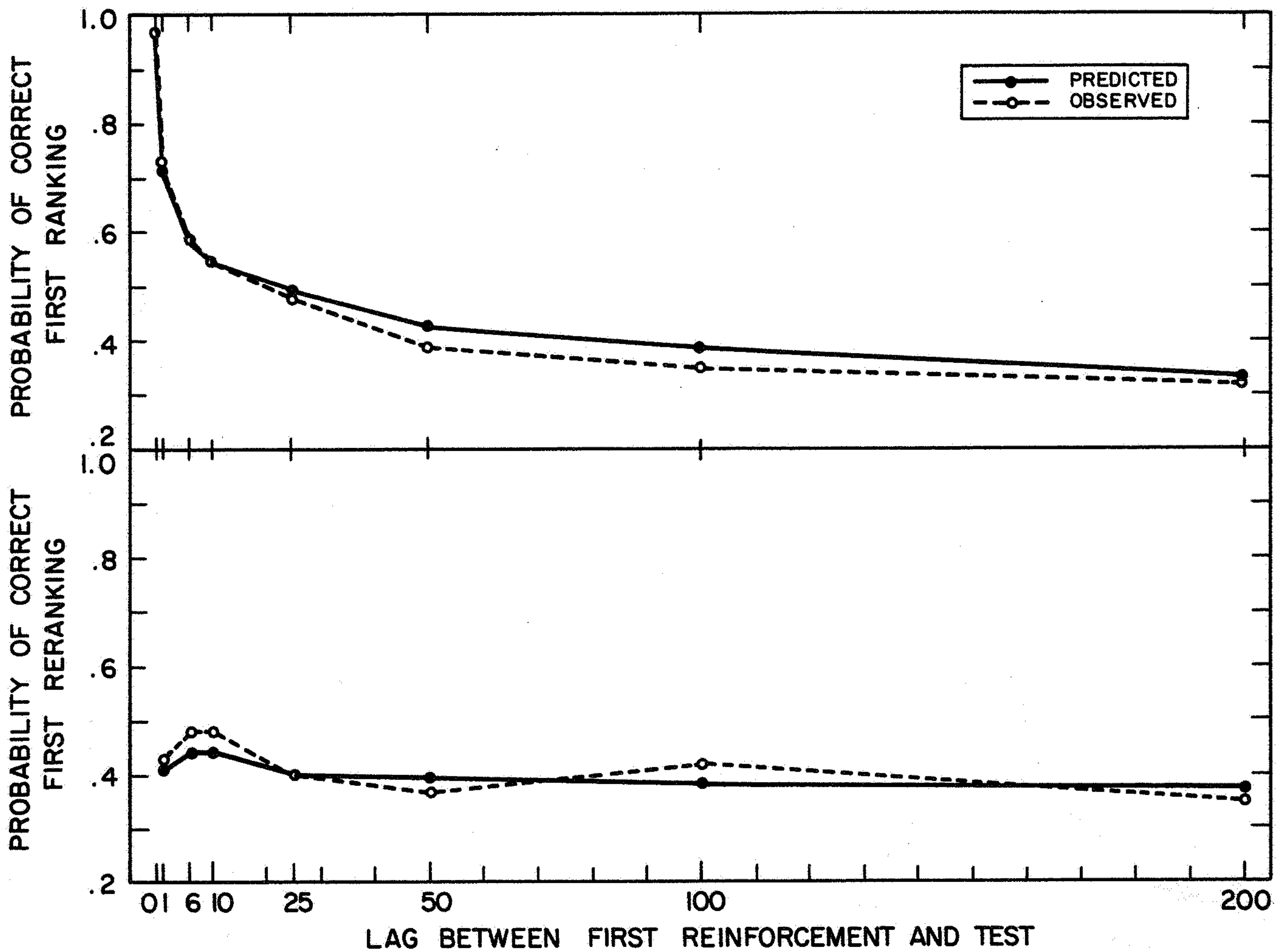

Figure II-2. Probability Correct as a Function of Lag, for Rankings and Rerankings.

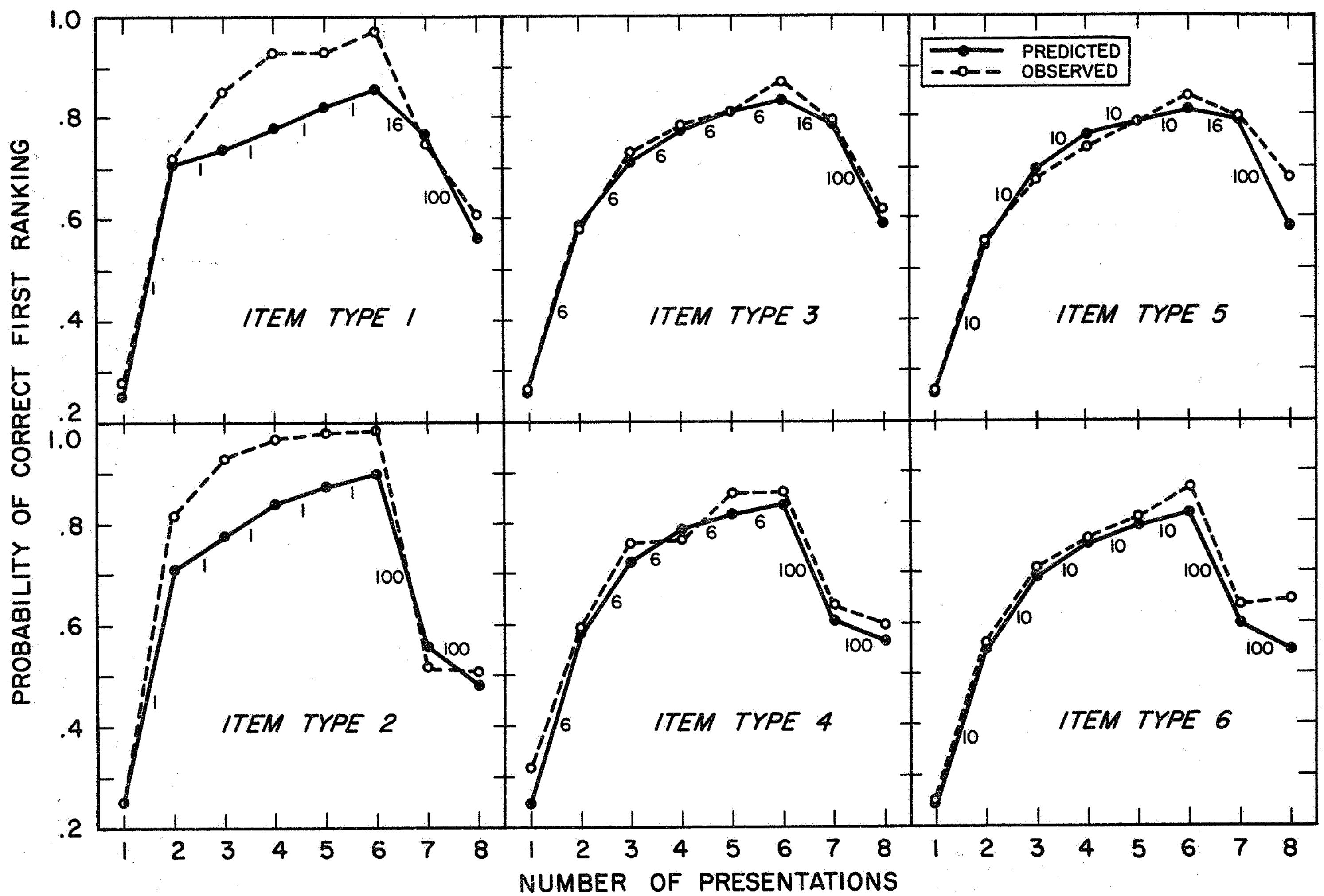

Figure II-3. Probability of Correct First Ranking as a Function of Number of Presentations, for Item-Types 1-6.

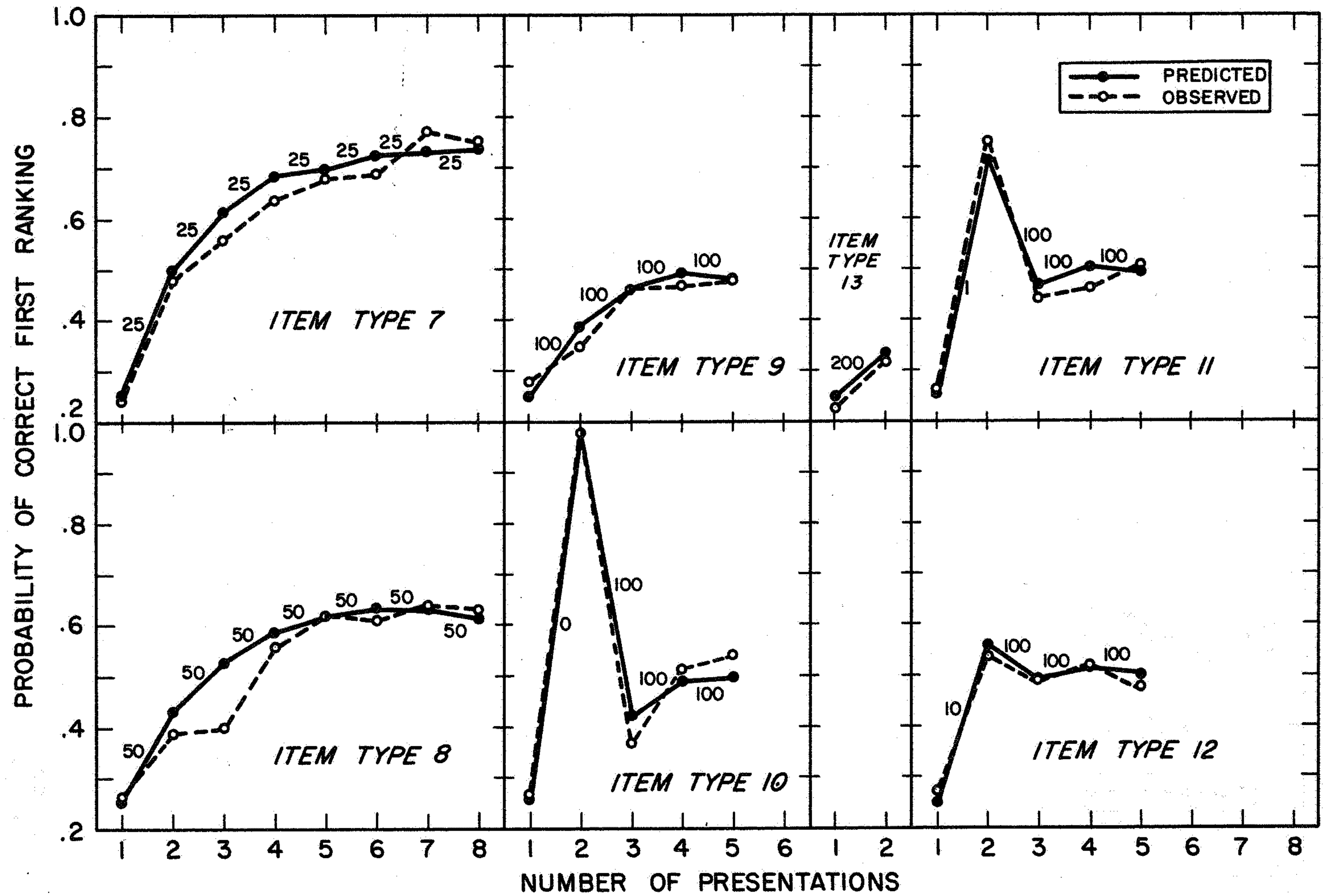

Figure II-4. Probability of Correct First Ranking as a Function of Number of Presentations, for Item-Types 7-13.

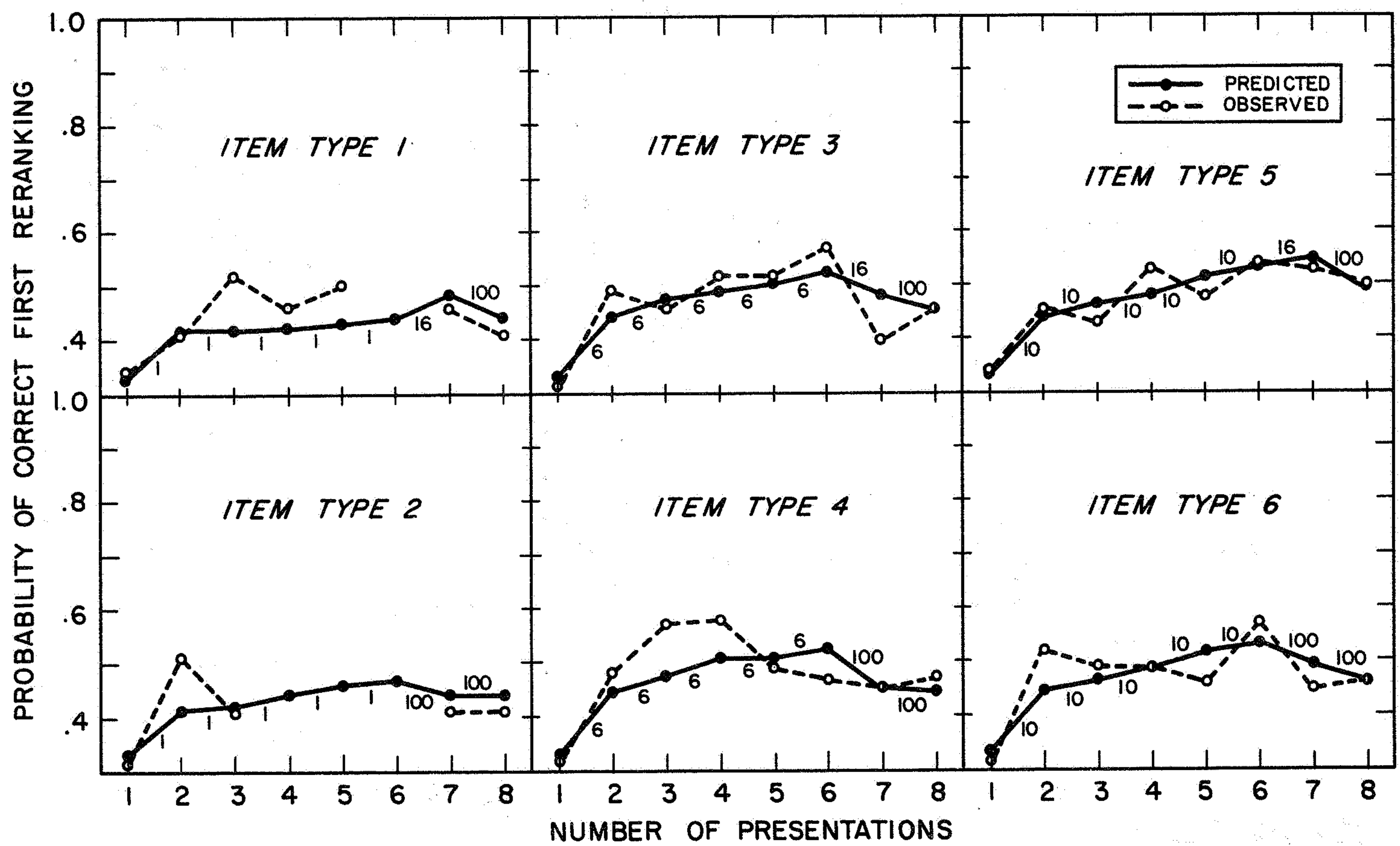

Figure II-5. Probability of Correct First-Reranking as a Function of Number of Presentations, for Item-Types 1-6.

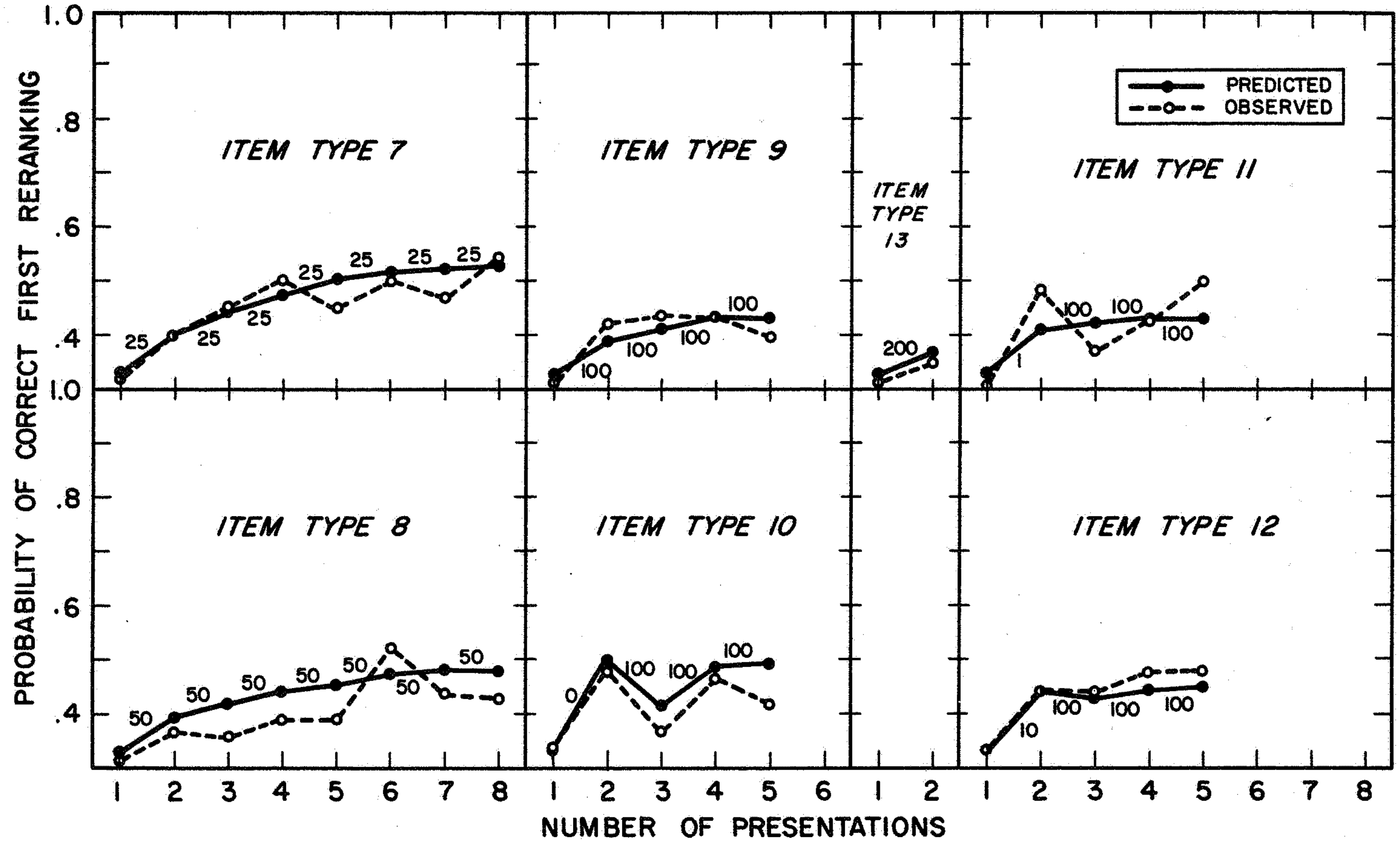

Figure II-6. Probability of Correct First-Reranking as a Function of Number of Presentations, for Item-Types 7-13.

the initial lag the better is performance after a long subsequent lag. A similar effect is seen in the graphs of item types 2, 4, and 6 in Figure II-3. Following five initial lags of either 1, 6, or 10, performance on two subsequent tests at lags of 100 rose from .52 to .62 to .65; i.e., performance is better at long lags the more spaced is the series of initial reinforcements.

It should be noted that item-types 9, 10, 11, and 12 seem to exhibit something like steady state characteristics; i.e., if reinforcements are given at lags of 100, performance seems to stabilize near the .50 level.* Item types 7 and 8 also seem to be approaching an asymptotic level of probability correct well below 1.0 (.75 and .63 respectively). These results further demonstrate that any model with a long term absorbing state which items enter an appreciable portion of the time will not provide an adequate description of the data. If the probability correct for an item in the absorbing state is p, then all curves at long lags should be asymptoting at p. This is not the case for these data even if p is allowed to be less than 1.0.

The Effects of Intervening Items. The lag curves above show that forgetting increases as the lag increases. It should be questioned whether it is the number of intervening items per se which determines the amount of forgetting. The theoretical position outlined in Chapter I implies that forgetting should, among other things, be a function of

*This result might lead to speculation that item-types 1-6, if given additional reinforcements at lags of 100, would exhibit a decrease in performance down toward the .50 level (which would be a strange sort of "learning," indeed).

the amount of new information stored during the intervening period. Therefore, the amount of forgetting should vary as a function of how well-known are the intervening items, if we accept the view that less new information is stored concerning well-known items. A similar expectation would hold if the degree of inter-stimulus interference were a determinant of forgetting; the greater the number of unknown stimuli that intervened, the greater the forgetting.* There are a number of experiments which bear on these points. Thompson (1967) demonstrated that a strong short-term effect exists in a situation where the subject adopts rehearsal as a predominant strategy; that is, a short series of extremely overlearned items following an item caused no forgetting, whereas an equal length series of unknown items caused dramatic decrements in performance. This short-term memory rehearsal effect should be differentiated, however, from the long-term memory retrieval effect proposed above; we shall return to this point shortly. Calfee and Atkinson (1965) proposed a trial-dependent-forgetting model for list-structured P-A learning. In this model, the amount forgotten from a short-term state of learning between successive reinforcements was proposed to decrease as the trial number increased, since the intervening items became better and better known as the experiment proceeded. While they found the trial-dependent-forgetting model to fit the data

---

*In principle, the various sources of forgetting should be separable. For example, an experiment could be run in which items are compared which are tested at equal lags and have equal numbers of intervening new stimuli; the items would differ in that the interreinforcement lags of the intervening items would be low in one case and high in the other.

more closely than the alternatives, one cannot directly conclude that the finding applies to individual items; since a list design was used, the changes in forgetting could be the result of some sort of reorganization or integration of the entire list over trials.

Although Experiment I was not expressly designed to systematically vary the makeup of the intervening items at a given lag, a fair amount of chance variation occurred and it is possible to capitalize upon this fact. Every trial in the trial sequence was assigned a number "K" representing how well "known" was its stimulus-response pair as follows:

$$K = (\text{reinforcement number}) \times (20)/(\text{lag}+1). \qquad \text{Eq. II-1}$$

In this formula the reinforcement number and the lag refer to the stimulus tested on that trial. K is very highly correlated with the probability correct on each trial and therefore provides a reasonably valid measure. Next we compute for each item presented the average value of K during the preceding lag, and call this average $\overline{K}$. We can now compare the probability correct for each item with how well "known" were the items making up the preceding lag. Table II-4 presents the resultant data (on page 36) for items tested following their first and second reinforcement, at each of several lags. At each lag, all items are divided into two roughly equal groups, those with high $\overline{K}$ and those with low $\overline{K}$. Thus the items with lag 1 and reinforcement 1 are split into a high-group and a low-group, all items in the high-group having values of $\overline{K}$ greater than any items in the low-group. The mean probability correct is then computed for items in the high-group and for items in the low-group, and these means are listed in columns 2 and 3 of the table. Hence column two of the table gives the mean probability correct for items whose intervening items are relatively well-known.

There are a number of points to be made regarding Table II-4. First, there is a definite, highly significant effect in the expected direction: intervening items which are less well-known cause more forgetting.* Almost certainly the magnitude of the differences would have been even larger than those observed if variations in $\overline{K}$ had been larger; however, differences in $\overline{K}$ arose by chance rather than by design. Of particular interest is the result for lag 1. In this case there is only a single intervening item and $\overline{K}$ varies considerably from item to item; in fact, the mean probability correct for the intervening item was .31 for the low-group and .77 for the high-group. Nevertheless, only a difference of .05 was found in the measure tabled. If a rehearsal-type short-term process was causing the result, as in the Thompson study cited earlier, then this difference should have been far larger than was observed, and far larger than other differences in the table.** There is another feature of the data which makes this same point. The rehearsal model

---

*There is no question of significance. The results for reinforcements greater than 2 show essentially the same results as for those shown in the table. A sign test on the directions of the differences gives $p < .01$ and more rigorous tests would lower this probability considerably.

**The justification for this statement ultimately rests on a theoretical analysis in which the buffer model is applied to the data. It is beyond the scope of this report to go into the details of the analysis, but a buffer model was applied to the data of Experiment I. The best fit of the model was not adequate as a description of the data, and one of the major failings of the model was the extreme overprediction of the effects of known items at lag 1. Rather than the .05 difference at lag 1 which was presented in Table II-4, the buffer model predicted a difference of about .30.

explanation of the effect of known items holds that known items fail to cause decreases in performance because they do not enter rehearsal; if the intervening items do not enter rehearsal, then the target item will tend to stay in rehearsal in STS for a longer period of time, even until the moment of test. In this model, the first few items after the target item are crucial in determining the magnitude of the effect. In order to check this point, the analysis leading to the statistic in Table II-4 was repeated, except that $\overline{K}$ was calculated without including the K values of the first two intervening items. Nevertheless, the resultant pattern of results (excluding lag 1, of course) was virtually identical to that in Table II-4. A sign test on the direction of differences again gave a $p < .01$ as a level of significance. We therefore conclude that the $\overline{K}$ effect is not crucially dependent upon the K value of the first few intervening items. It seems reasonable, then, that the effect originates in the LTS retrieval process, rather than in a rehearsal mechanism. The explanation we propose, in terms of the theory of Chapter I, holds that the "age" of any code is dependent upon the number of new codes that are subsequently stored in LTS. Since the probability correct depends upon the "age" of a code, the effect found in Table II-4 follows directly.

Summary. There are several main results of Experiment I. First, the multiple-search nature of retrieval was established by a comparison of ranking and second-guessing effects on the same test trial. Second, performance was observed to tend toward chance as the lag increased; this and related findings demonstrated the inappropriateness of a model for this task which postulates a long-term memory absorbing state. Third, the forgetting of an item at a given lag, long or short, was observed to depend upon the degree to which the intervening items were known. Discussion of other results, and of the quantitative aspects of the data, will be reserved for Chapter III.

Experiment II

Experiment II was designed with the objective of providing a stringent test of the model used to predict the results of Experiment I. An integral feature of this model (to be discussed in detail in Chapter III) was the prediction of intrusion errors; i.e., incorrect retrievals from memory. In Experiment I responses were required on every trial, so that intrusions and pure guesses were not separable at the observable level. In Experiment II the response set size was increased and the subject was instructed to respond only when he felt he knew the answer. In this manner, intrusions may be observed directly. The ranking technique was not used - only a single first-guess was allowed - but second guesses were allowed following errors. A second objective of Experiment II was the collection of "interference" data which would allow for the natural expansion of the earlier model. Thus individual stimuli in the present experiment sometimes had their response assignment changed. Formally, a design was adopted which was the counterpart in a continuous paired-associate experiment of the standard proactive interference paradigm.

The design and procedure of Experiment II is in certain respects identical to that of Experiment I. Except where noted, the procedure was the same as in the previous experiment.

Design Justification. Each session involved an identical sequence of 400 trials; each trial consisting of a test phase followed by a study phase. The trial sequence, presented in Appendix 2, will be discussed shortly. As in Experiment I, the individual stimuli and responses were changed from one session to the next - only the sequence remained fixed. An individual stimulus could be presented on as many as 8 trials during the sequence, at varying lags. On some trials the response assignment of a stimulus was changed; on these trials the subject was notified following the test phase that the answer would be changing. The pair presented during the study phase would then contain the new response.

The item-types in the present experiment were constructed so as to provide a full test of proactive-interference phenomena with appropriate controls. Quite apart from considerations relating to the theory proposed in this paper, it is maintained that interference phenomena need reexamination in the context of continuous paradigms. Forgetting phenomena have been examined extensively for many years with the use of list-structured experiments: lists of paired-associates are successively learned, each list utilizing the same stimuli, but with response assignments shifted (i.e., the A-B, A-C design). The results of these experiments have been fairly successfully explained by some version of two-factor interference theory (Postman, 1961; Melton, 1963; Underwood, 1957; Keppel, 1968; etc.). The experimental effects are found to take place over whole lists, but it is often assumed that equivalent changes occur in individual stimulus-response assignments, the assumption based upon a seemingly natural inference. Thus, if, in an A-B, A-C design, it is found that increased training on the A-C list causes increased forgetting of the A-B list, it is then inferred that increased learning of a particular stimulus-response pair will result in increased forgetting of a previous pairing of that same stimulus with a different response. Recent research, however, has raised doubt about this inference (DaPolito, 1966; Greeno, 1967). Following A-B, A-C learning subjects were asked to give for each stimulus both responses previously paired with it; regardless of the presence of retroactive interference effects in the lists as a whole, it was found that the probability of a correct first-list response times the probability of a correct second-list response was equal to the combined probability of giving both responses correctly. This is a result to be expected if there were no individual item response interactions; i.e., if for a particular item the level of learning of the first list response does not affect the level of learning of the second list response, and vice versa. This implies that the usual inference from lists to items may not be valid, and theories of item interference should therefore be based on appropriate experiments which do not utilize a simple list structure.

Atkinson, Brelsford, and Shiffrin (1967) reported a continuous P-A experiment in which some indications of proactive interference were found for individual items. This finding was only incidental in that experiment, however, and could possibly have been caused by selection effects. Estes (1964) reported experiments in which proactive interference effects were sought for individual items buried in a list structure, but the results indicated no proactive effect. Peterson, Hillner, Saltzman, and Land (1963) reported a continuous task in which there were indications of retroactive interference. These experiments seem to delimit the current state of knowledge concerning individual item-interference: very little is currently established.

The present experiment was therefore designed to examine in depth the status of proactive item-interference. The item-types utilized for this purpose are listed in Table II-5. A stimulus is presented with its first response (R1) either 2 or 4 times for study. The response is then changed and 3 study trials are presented with the new response (R2), all at lag 10. The lags of the initial presentations are either (0-10) or (10-10) if there are two initial presentations, or (0-10-0-10) or (10-10-10-10) if there are four initial presentations. On the trial where the answer first changes, the test asks for the R1 response, the subject is then told the answer is changing, and the new pairing is presented. We denote these item-types by the initial sequence of lags. The column on the right margin of the table gives the number of instances of each item-type in the sequence of 400 trials.

A comparison of the first and second tests following the change of response, with the first and second tests before the change of response, should indicate any overall proactive effects. A comparison of the conditions in which the number of response 1 presentations varies (i.e., (10-10) vs. (10-10-10-10)) permits us to examine the probability of a correct R2 as a function of varying amounts of learning on R1. A comparison within the same number of initial presentations (i.e., (0-10) vs. (10-10)) should allow the same examination as above, but where the number of presentations is held constant (assuming that the 0 lags do not result in much learning). In this way it may be determined whether any proactive effect found is due to the amount learned about R1, or simply due to the number of presentations of R1.

The above item-types examine proactive interference only at lag 10. In order to study the effects of variations in lags, 16 other item-types were used. Each of these 16 item-types is given just three presentations; on the second presentation the response is changed. The lag between the first and second presentation is called lag 1; the lag between the second and third presentations is called lag 2. The item-types are listed in Table II-5a. Lag 1 takes on the values 0, 1, 4, 10; lag 2 takes on the values 1, 5, 10, 25. The entries in each cell of the 4 X 4 table are the number of occurrences of each item-type. These item-types will be denoted by their lag 1 and lag 2 separated by a comma: e.g. (4,25). Note that item-type (10-10) is different than item-type (10,10).

The subject is instructed to respond during each test with the response _most recently_ paired with the stimulus presented. He is told to "forget" any old pairings once the response has changed. The subject does not have to respond if he does not know the answer. If he does respond and is wrong, he is told so and given an opportunity to respond again.

Subjects. The subjects were 14 students from Stanford University who received $2.00 per hour for their services. Each subject participated in a minimum of 8 and a maximum of 11 experimental sessions plus one initial practice session. The sessions were conducted on weekday evenings and took approximately 55 minutes each. The subjects were procured without regard for sex through the student employment service. The apparatus was identical to that for Experiment I.

Stimuli and Responses. The stimuli were 1600 common English words either 3, 4, or 5 letters in length selected in random fashion from

TABLE II - 5

ITEM-TYPES FOR EXPERIMENT II

Item-type	Response 1	Response 2	No. of Types
0-10	P1 -Lag- 0 P2 -Lag- 10	P3 -Lag- 10 P4 -Lag- 10 P5	7
10-10	P1 -Lag- 10 P2 -Lag- 10	P3 -Lag- 10 P4 -Lag- 10 P5	7
0-10-0-10	P1 -Lag- 0 P2 -Lag- 10 P3 -Lag- 0 P4 -Lag- 10	P5 -Lag- 10 P6 -Lag- 10 P7	8
10-10-10-10	P1 -Lag- 10 P2 -Lag- 10 P3 -Lag- 10 P4 -Lag- 10	P5 -Lag- 10 P6 -Lag- 10 P7	7

In the above table P followed by a number represents the presentation number of a stimulus of that item-type.

TABLE II - 5a

ITEM-TYPES FOR EXPERIMENT II

First Lag \ Second Lag	1	5	10	25
0	3	3	4	4
1	3	3	3	3
4	3	3	3	4
10	3	3	4	3

In the above table the numbers in each cell are the numbers of instances of each item-type. Note that the first lag is previous to the changing of the response, and the second lag is subsequent to the changing of the response.

Thorndike (1921), with homonyms, personal pronouns, possessive adjectives, and the past tense of verbs eliminated. Ninety-five stimuli were randomly selected for use during each session, with the restriction that any stimulus used in a session could not be used in any succeeding session for that subject. Words were used as stimuli, rather than CCC's, in order to make the proactive interference comparisons meaningful. That is, the design does not use unique response pairings; hence the same response can be assigned to more than one stimulus. If two stimuli assigned the same response are not sufficiently different, it would be difficult to differentiate this case from the case where a single stimulus had a changed response assignment.

The responses were the 26 letters of the alphabet. At the start of each session all stimuli were assigned R1 and R2 responses randomly with the restriction that no word could be assigned its own initial letter as a response. Since no subject reported noticing this restriction, it may be assumed that the probability correct, if the subject decided to make a pure guess, would be 1/26.

<u>Instructions</u>. When a subject arrived for the first session he was given the following instructions to read:

> "This experiment will test your ability to remember responses to a series of common English words. The response will always be one of the letters of the alphabet. You must always try to remember the letter most recently paired with a particular word.
>
> The experiment will consist of a number of trials in succession and last about an hour (or less) each day. Each trial will begin when the word "test" will appear on the screen before you. Below the word "test" will appear an English word (which you may or may not have seen before on a previous trial.)

The task on this test trial is to give the response most recently paired with the word shown. If you have no idea what the answer is, then either type a "carriage return" (CR) or do not respond at all; if you have a guess, then type the letter you think is correct. Remember, the correct letter is the one most recently paired with a particular word.

If you type a letter and are wrong, the computer will tell you so and give you a second chance. Again, type a carriage return or do not respond if you have no idea as to the answer, and type the letter if you have a guess.

You must try to respond quickly, as there will be a time limit in which time you must give your response. If you exceed the time limit, the machine will go on to the study portion of the trial.

Following the "test" portion of the trial will be a pause. Then the word "study" will appear on the screen. Below the word "study" will appear the English word you were just tested on paired with the currently correct answer. This is always the correct response which you must try to remember. Feel free to use any coding mnemonics which help you to remember the response.

Sometimes the response presented for study will be different than the previously correct response associated with the given word. In this case, forget the previously correct response and learn the new response (the old one is now wrong). You will be warned just before the study trial if the response is being changed, so that you will never fail to notice that a change has occurred. This warning will be: "answer changes."

You will be given several seconds to study the current word-letter pair, and then, after a brief pause, the next trial will begin (i.e., a new test trial will occur). Each session will consist of a continuous sequence of these trials.

The experimenter will give you instructions regarding which booth to use, how to start each session, and what to sign each day."

The experimenter reviewed the instructions with the subject and then introduced him to the computer and its operation. The entire first session was used to familiarize the subject with the apparatus and instructions, and to give him practice at the task.

Procedure. As noted earlier, each session consisted of a sequence of 400 trials. Each trial involved a standard series of events. (1) The word Test appeared on the upper face of the CRT. Beneath the word Test appeared the member of the stimulus set indicated by the presentation schedule of Appendix 2. The subject then typed a letter if he felt he knew the response. If he was sure he did not know the response, then he could terminate the test trial by typing a carriage return. If an incorrect response was typed, then the words WRONG. TRY AGAIN appeared on the CRT below the previous response, which remained displayed. The subject could then respond, not respond, or type a carriage return, as for the first guess. If the subject had not typed a response within 3 sec. for the first-guess, or within 2.7 sec. for the second-guess, then the test phase was terminated. (2) The computer next determined whether the response to the current stimulus was to be changed; if so, the CRT was blanked momentarily, and then the following words appeared: ANSWER CHANGES. After 1/2 sec. the study phase began. If the response was not to be changed, then the CRT was simply left blank for 1/2 sec. until the study phase began. (3) The screen was blanked and then the word STUDY appeared at the top of the CRT. Beneath the word STUDY appeared the stimulus just tested along with the correct response to be remembered (changed or not as was appropriate). This display remained for 3.0 seconds. (4) The CRT was blanked for 1/2 sec. and then the next trial began. Using this procedure, the session of 400 trials took about 55 minutes.

At the start of each session, the computer randomly assigned each subject 95 stimuli he had not seen in previous sessions. Each stimulus was then randomly assigned two different letters as responses, with the restriction that the first letter of a stimulus could not be used as its response. The first 14 trials consisted of 10 filler items, items which appeared only seldom thereafter.

Altogether 147 subject-sessions of data were collected (not counting the practice sessions). Due to computer shutdown and other extraneous factors, only 122 of these sessions were entirely completed, the remainder being close to completion. The data collected consisted of the entire sequence of events within each session, including the latencies of the responses. At the conclusion of the experiment each subject filled out a written questionnaire.

Results of Experiment II

A large amount of data will be presented in the present section. As it is rather difficult to grasp without a theoretical basis, detailed discussion will be put off until the next chapter. An attempt will be made here to limit discussion to certain highlights. In the following the first response given by the subject is termed a "first-guess," and the second response when given by the subject is termed a "second-guess." Table II-6 presents the probability of a correct first-response for each subject, lumped over all trials and sessions. The results are listed in order of increasing probability correct. It is evident that there is a wide range in subject ability at this task. Despite this, the remaining data is presented in a form averaged over all subjects in order to gain precision of estimates. This should not overly distort the observed effects, since a subject by subject breakdown of the data seemed to show the same qualitative effects holding for individual subjects as for the group average.

TABLE II - 6

MEAN PROBABILITY CORRECT
FOR SUBJECTS OF EXPERIMENT II

Subject Number	7	6	2	14	3	13	11	8	9	12	5	1	4	10
Probability Correct First-guess	.29	.30	.34	.36	.41	.49	.51	.51	.51	.51	.53	.56	.68	.69

TABLE II - 7

MEAN PROBABILITY CORRECT
FOR SUCCESSIVE DAYS OF EXPERIMENT II

Day Number	1	2	3	4	5	6	7	8	9	10
Probability Correct First-guess	.52	.48	.44	.48	.45	.50	.47	.42	.49	.52

Table II-7 gives the probability of a correct first-guess on successive days of the experiment (practice day not included). There is no evidence for a trend over days. Apparently, as in Experiment I, proactive interference from session to session was not an important factor. The data to follow will be averaged over all sessions. In the following discussion an error will be taken to mean the absence of a correct response; the term intrusion will be reserved for overt errors.

First-Response Data. Figure II-7 presents, in the top panel, the probability of a correct first-guess for each of the item-types listed, at each of their presentations. Figure II-8 presents the same probability for the remaining item-types. Consider first the top panel of Figure II-7. The observed data is represented by open circles; ignore the predictions for the present. The vertical line in each graph delineates the point at which the R1 response is changed. Following the change of response all lags are 10. The successive lags previous to the change are presented in the item-type name at the top of each graph. There are slightly more than 1000 observations at each point shown. The most important features of these data relate to the question of proactive interference. In conditions (10,10), (10-10), and (10-10-10-10), the probability correct after one reinforcement is about .55. The first test after the response changes, however, has a probability correct of about .41. Hence an overall proactive effect is present. A comparison of all five conditions reveals that the proactive effect is not dependent upon the number of reinforcements prior to the change of response, nor upon the terminal probability correct just prior to the change. This is true despite a reasonable range in both variables:

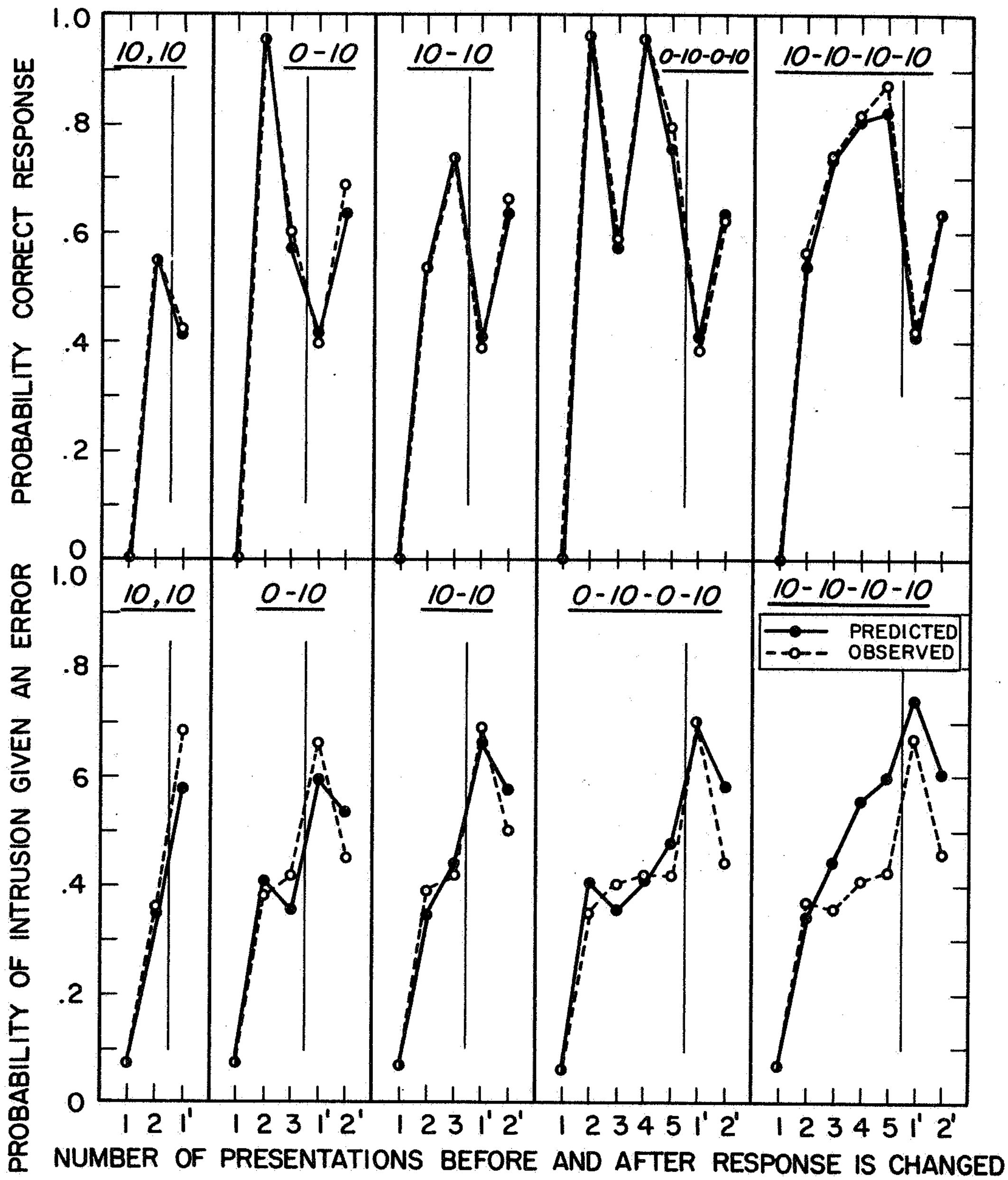

Figure II-7. Probability of First-Guess Correct Responses and First-Guess Intrusions, for the Major Item-Types.

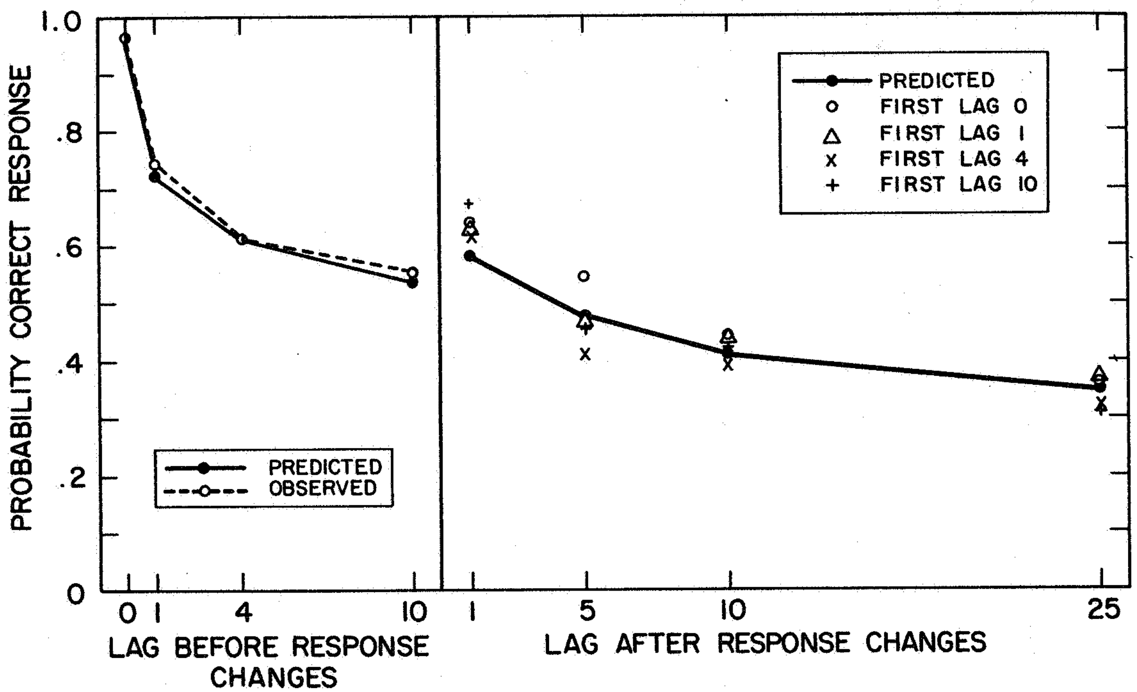

Figure II-8. Probability of Correct First-Guesses as a Function of Lag, for the Matrix Item-Types.

the number of initial reinforcements takes on the values 1, 2, and 4; the terminal probability correct takes on the values .55, .61, .74, .80, and .87; the probability correct after the change of response takes on the values .42, .40, .39, .39, .42. A similar result appears to hold for the second test following the change of response. This lack of dependence upon the degree to which the first response is learned raises some questions about the source of the overall proactive effect. In particular, one must consider the hypothesis that the subjects, having been informed that the response is changing, attempt to code the new pairing with a probability smaller than for an R1 reinforcement. This hypothesis, and a number of models which can account for the observations, will be dealt with in the following chapter.

Figure II-8 presents much the same pattern of results as those just discussed. This figure gives the probability of a correct first-guess for the test before and after the response is changed, where the lag previous to, and following, the change of response is varied. The left-hand panel presents the first-reinforcement lag curve for lags 0, 1, 4, and 10. The observations are the open circles. Following each of these lags the response is changed and a second lag of 1, 5, 10, or 25 ensues. The right-hand panel in the figure presents the results for the 16 resultant conditions, henceforth termed the "matrix" item-types. If variations in the first lag did not have a differential proactive effect, then the four observations at each lag in the second panel should not differ from each other, which seems to be the case. The data are somewhat more unstable than in the previous figure because each point in the right-hand panel is based on approximately 400 to

500 observations. Points in the left-hand panel are based on about 1800 observations.

Figure II-7 presents, in the bottom panel, the probability that a false intrusion response was given, conditionalized upon the fact that a correct response was not given (the unconditional probability of an intrusion was divided by 1.0 minus the probability correct). In the following we refer to a response given in error which had previously been associated with the tested stimulus as an _old_-intrusion. Other intrusions are called _new_-intrusions. In Figure II-7 both types are lumped. The observed points are represented by open circles. Several points should be noted concerning these graphs. The intrusion rate for newly presented items is above zero (about .07), but well below that observed on succeeding trials. If the subject searched his memory for an answer on _every_ new trial, it might be expected that an intrusion rate higher than those on succeeding trials would result. The relatively low rates observed would be expected if the subject was often recognizing quickly that the stimulus presented was new, and thereby ceasing further memory search. Note also that there is a considerable increase in intrusions following the change of response - in fact, the increase in number of intrusions is considerably larger than the decrease in probability correct at those points. Most of the increase in intrusions following change of response is of course in _old_-intrusions. Table II-8a gives the probability of an old-intrusion for the major item-types, conditional upon the fact that a correct response was not made. The numbers in parentheses are predictions which may be ignored for the moment. Before the change of response the probability of an old-intrusion

TABLE II - 8

FIRST-GUESS INTRUSIONS
(Predicted Values in Parentheses)

Table II - 8a: Probability of Old-Intrusion Given an Error

		Item Type			
		0-10	10-10	0-10-0-10	10-10-10-10
Number of presentation after change of response	1	.461 (.345)	.517 (.443)	.552 (.471)	.514 (.470)
	2	.171 (.216)	.225 (.153)	.171 (.238)	.236 (.238)

Table II - 8b: Probability of Intrusion Given an Error

First Test

First Lag		
	0	.38 (.41)
	1	.40 (.31)
	4	.40 (.34)
	10	.37 (.35)

Second Test

	Second Lag			
0	.84 (.78)	.60 (.73)	.64 (.68)	.60 (.65)
1	.69 (.65)	.74 (.62)	.66 (.59)	.61 (.57)
4	.63 (.62)	.71 (.61)	.68 (.58)	.64 (.56)
10	.59 (.60)	.62 (.59)	.69 (.58)	.65 (.55)

Table II - 8c: Probability of Old-Intrusion Given an Error

		Second Lag			
		1	5	10	25
First Lag	0	.65 (.66)	.34 (.53)	.39 (.47)	.33 (.39)
	1	.42 (.49)	.57 (.40)	.47 (.35)	.33 (.30)
	4	.40 (.46)	.45 (.37)	.47 (.33)	.44 (.29)
	10	.36 (.41)	.40 (.35)	.44 (.32)	.41 (.28)

is zero, so these trials are not tabled. Note that in the table the old-intrusion rate shows a tremendous decrease from the first to the second test of R2. This might be explained if the subject was learning on the first trial that the old-intrusion he had given was wrong - this intrusion would then be repressed on the next trial. The intrusion results for the item-types where the lag was varied are presented in Table II-8b and II-8c. Table II-8b gives the lumped results, and Table II-8c the old-intrusion results. Discussion of these tables are reserved until the next chapter.

For a number of reasons it might be felt that intrusion rates should increase as the duration of the session lengthened. This possibility may be examined by considering intrusions on items presented for the first time at different locations in the trial sequence. Figure II-9 presents these results. Intrusion rates are averaged for successive groups of eight new items during the trial sequence. The graph demonstrates that a fairly orderly increase in intrusion rates occurs, though not of large magnitude.

Second-Guess Data. Figure II-10 presents data for second-guesses following *new*-intrusions on the first guess. The top panel presents the probability of a correct second-guess for the major item-types. Table II-9a presents the same probabilities for the item-types on which the lag was varied. It may be observed that the second-guess curves follow the first-guess curves in general form: there is a rise before the change in response and then a sharp drop after the change. Furthermore, across conditions, variations in presentation schedules prior to the change do not seem to affect the second-guessing rate following the change; this fact conforms to the first-guess finding.

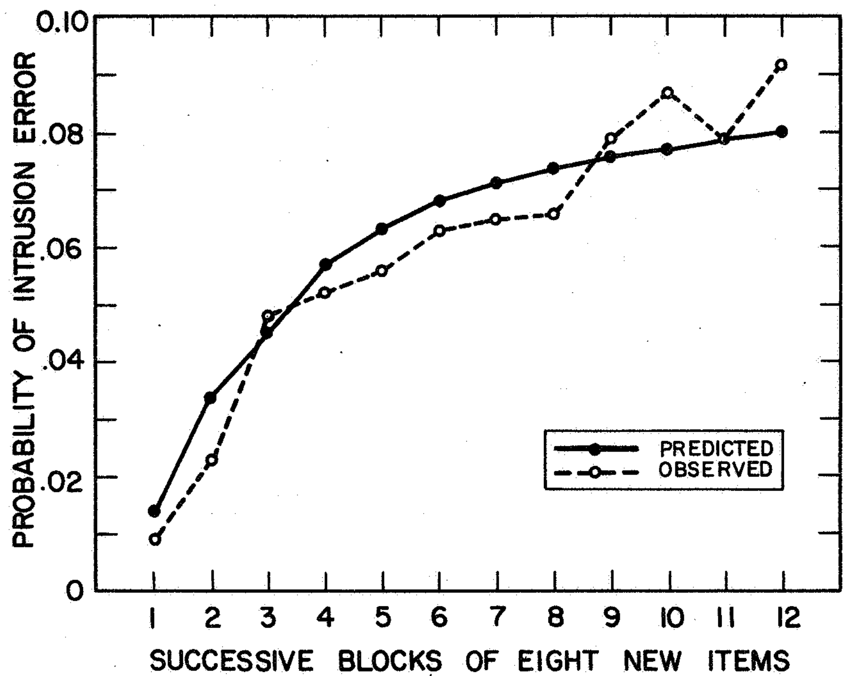

Figure II-9. Probability of Intrusions for New Items, as a Function of Duration of Session.

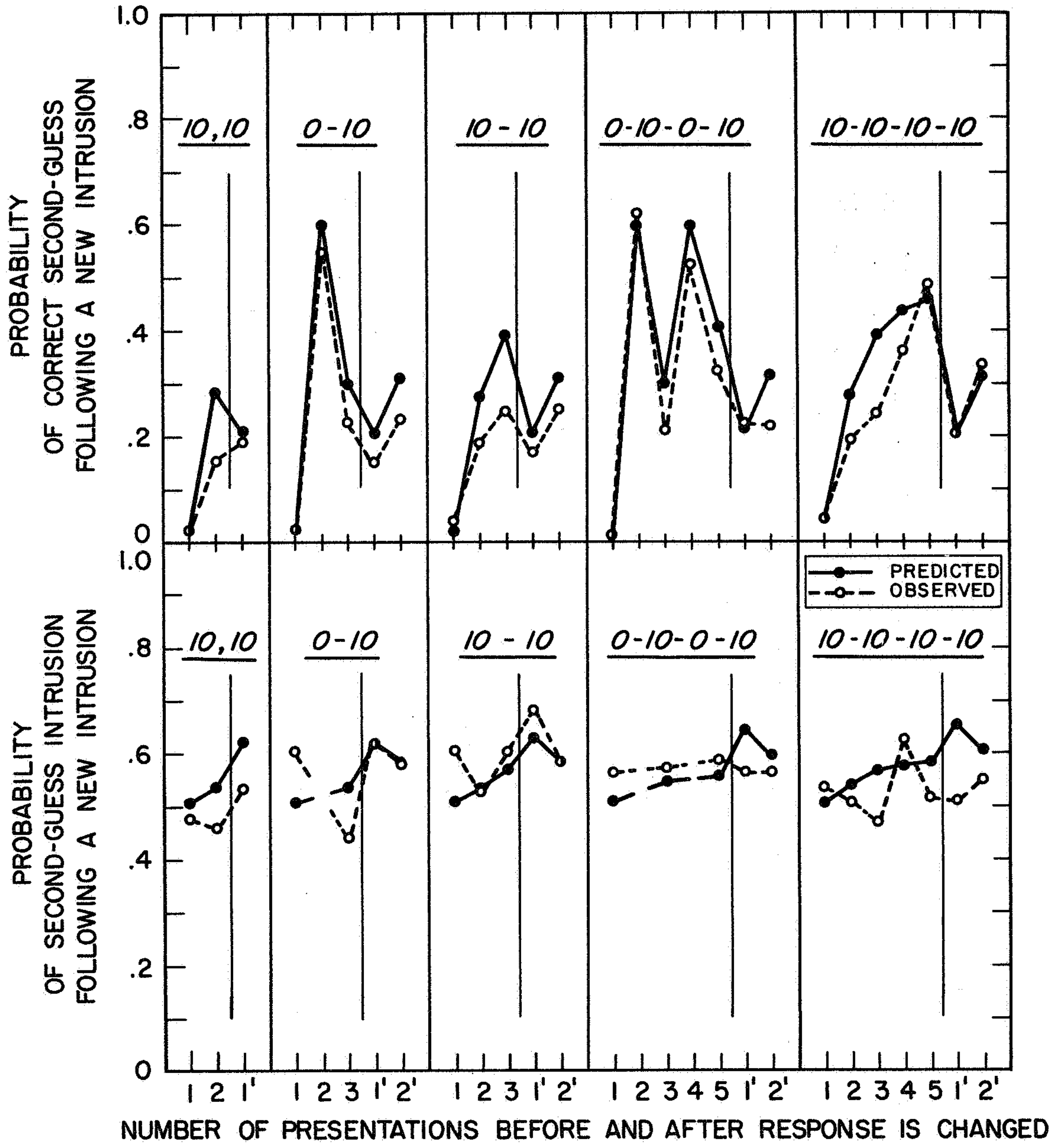

Figure II-10. Probability of Second-Guess Correct Responses and Second-Guess Intrusions, for the Major Item-Types.

TABLE II - 9
SECOND-GUESS INTRUSIONS
(Predicted Values in Parentheses)

Table II - 9a: Probability of Correct Second-Guess Following a New Intrusion

First Test

First Lag	
0	.65 (.60)
1	.25 (.32)
4	.15 (.33)
10	.20 (.28)

Second Test

First Lag \ Second Lag	1	5	10	25
0	.15 (.24)	.19 (.24)	.18 (.23)	.13 (.17)
1	.35 (.21)	.17 (.23)	.10 (.21)	.11 (.17)
4	.18 (.21)	.18 (.23)	.16 (.22)	.07 (.18)
10	.15 (.20)	.28 (.23)	.20 (.21)	.11 (.18

Table II - 9b: Probability of Second-Guess Intrusion Following a New Intrusion, Conditional Upon a Second-Guess Error. Top Matrix for Second-Guess New Intrusions. Lower Matrix for Second-Guess Old-Intrusions.

First Test

First Lag	
0	---
1	.45 (.46)
4	.45 (.51)
10	.44 (.53)

Second Test

First Lag \ Second Lag	1	5	10	25
0	.53 (.30)	.37 (.40)	.36 (.45)	.44 (.49)
1	.50 (.35)	.37 (.43)	.42 (.46)	.39 (.51)
4	.28 (.39)	.39 (.44)	.52 (.47)	.51 (.50)
10	.36 (.42)	.32 (.45)	.48 (.50)	.41 (.51)

First Lag \ Second Lag	1	5	10	25
0	.06 (.32)	.12 (.25)	.15 (.21)	.19 (.16)
1	.18 (.19)	.18 (.18)	.12 (.15)	.17 (.12)
4	.16 (.17)	.19 (.16)	.11 (.14)	.15 (.12)
10	.11 (.14)	.13 (.14)	.06 (.13)	.09 (.11)

Table II - 9c: Second-Guess Old Intrusions

	0-10	10-10	0-10-0-10	10-10-10-10
1	.12 (.13)	.17 (.17)	.23 (.18)	.22 (.22)
2	.07 (.06)	.10 (.07)	.10 (.07)	.07 (.08)

The lower panel in Figure II-10 presents the probability of any intrusion on the second-guess following a new-intrusion on the first-guess. The probability plotted is conditional upon a second-guess error. Table II-9b presents the same data for the item-types on which lag was varied. Table II-9c presents the second-guess old-intrusion rate for the item-types in Figure II-10. The first point to notice about the observations is the rather high rate of intrusions as compared with the rates observed on the first guess. Whereas the intrusion rates on the first-guess lie at about the .40 level, the second-guess intrusions are between probabilities of .5 and .6.* One possible interpretation of this finding would hold that the subject's decision criterion for output of responses found during memory search has been lowered on the second-guess. Particularly interesting is the intrusion rate for new items: Having made a wrong first-guess on a new item, subjects will then make a wrong second-guess with a probability of almost .60 (which can be compared with the first-guess new-intrusion rate of .07). An implication of this result is that once a decision has been made to search LTS on the first-guess, a search will always be made on the second-guess.

Table II-10 presents the data dealing with second-guesses following old-intrusions given on the first-guess. The results should be noted carefully because they are rather crucial to the model used in Chapter III. Table II-10a gives the probability correct following an old-intrusion. This probability is quite high -- higher even than that

*A part of this rise might have been due to subject selection, but a subject-by-subject breakdown showed 13 out of 14 subjects to have higher overall second-guess than first-guess intrusion rates.

following a new-intrusion. Table II-10b gives the probability of second-guess new-intrusions following first-guess old-intrusions. We shall merely note for the present that this new-intrusion rate is lower than the new-intrusion rate following first-guess <u>new</u>-intrusions.

<u>Latencies</u>. It is beyond the scope of this report to make a thorough analysis of the latency results. Tables II-11 through II-15 present the mean latencies for all item-types for the following conditions: a) correct first-guess responses, b) first-guess old-intrusions, c) first-guess new-intrusions, d) correct second-guesses following old-intrusions, and e) correct second-guesses following new-intrusions. We mention here the following results. (1) The latencies of a correct response decrease as the number of reinforcements increase; i.e., for the (10-10-10-10) condition the mean latencies are successively 1.52, 1.42, 1.36, 1.33. (2) The longer the lag, the longer the latency of a correct response. For initial lags of 0, 1, 4, and 10, the mean latencies of a correct response are 1.03, 1.37, 1.50, and 1.56. This result would have a natural interpretation if memory search were temporally ordered to some degree, but could also be handled if there were a significant amount of correct retrieval from a fast access short-term store at the shorter lags. (3) The latencies of a correct response following the change of response are slower than the corresponding latency for the first response. Nevertheless, these latencies after the change of response do not vary as a function of the type of sequence prior to the change. This result is in good accord with the response data; i.e., the change of response has an effect, but an effect independent of the history preceding it.

TABLE II - 10

SECOND-GUESSES FOLLOWING OLD-INTRUSIONS AS FIRST GUESSES

Table II - 10a: Probability Correct

Number of Presentations	1	2
0-10	.31	.54
10-10	.27	.50
0-10-0-10	.23	.51
10-10-10-10	.27	.39

First Lag \ Second Lag	1	5	10	25
0	.35	.28	.28	.29
1	.42	.33	.41	.24
4	.43	.34	.29	.22
10	.42	.30	.24	.29

Table II - 10b: Probability New Intrusions Conditional Upon a Second Guess Error

Number of Presentations	1	2
0-10	.36	.52
10-10	.36	.51
0-10-0-10	.30	.44
10-10-10-10	.34	.44

First Lag \ Second Lag	1	5	10	25
0	.31	.33	.38	.36
1	.29	.30	.25	.37
4	.26	.36	.27	.32
10	.35	.33	.35	.32

TABLE II - 11

MEAN LATENCIES FOR CORRECT FIRST-GUESSES

First Response Test

	P1	P2	P3	P4
0-10	1.04	1.51		
10-10	1.55	1.45		
0-10-10-10	1.04	1.53	1.14	1.42
10-10-10-10	1.52	1.42	1.36	1.33

Second Response Test

	P1	P2
0-10	1.66	1.54
10-10	1.63	1.57
0-10-0-10	1.63	1.54
10-10-10-10	1.67	1.59

P= number of previous presentations of the stimulus-response pair

---

First Response Test

First Lag	
0	1.03
1	1.37
4	1.50
10	1.56

Second Response Test

Second Lag

First Lag	1	5	10	25
0	1.46	1.56	1.52	1.61
1	1.42	1.57	1.56	1.73
4	1.48	1.72	1.64	1.67
10	1.37	1.60	1.64	1.63

TABLE II - 12

MEAN LATENCY OF FIRST-GUESS
OLD INTRUSIONS

First Response Test

	P1	P2
0-10	1.60	1.83
10-10	1.63	1.83
0-10-0-10	1.67	1.77
10-10-10-10	1.62	1.94

P= number of previous presentations of the stimulus-response pair

Second Response Test

Second Lag

First Lag	1	5	10	25
0	1.52	1.63	1.68	1.77
1	1.60	1.56	1.59	1.65
4	1.57	1.55	1.60	1.57
10	1.43	1.57	1.60	1.65

TABLE II - 13

MEAN LATENCIES OF FIRST-GUESS NEW-INTRUSIONS

First Response Test

	P1	P2	P3	P4
0-10	1.40	2.00		
10-10	2.03	2.06		
0-10-10-10	1.85	2.03	1.56	2.03
10-10-10-10	2.04	1.93	1.98	1.94

Second Response Test

	P1	P2
0-10	2.05	2.11
10-10	2.03	2.05
0-10-0-10	2.05	2.07
10-10-10-10	2.07	1.92

P= number of previous presentations of the stimulus-response pair

---

First Response Test

First Lag	
0	1.44
1	1.99
4	1.98
10	2.07

Second Response Test

Second Lag

First Lag	1	5	10	25
0	1.79	1.87	2.02	2.12
1	1.85	2.08	2.01	2.06
4	1.91	1.97	2.10	2.06
10	1.93	2.10	1.92	2.17

TABLE II - 14

MEAN LATENCY FOR CORRECT SECOND-GUESSES
FOLLOWING OLD-INTRUSIONS

First Response Test

	P1	P2
0-10	1.54	1.29
10-10	1.61	1.00
0-10-0-10	1.49	1.27
10-10-10-10	1.74	1.26

P= number of previous presentations of the stimulus-response pair

---

Second Response Test

Second Lag

		1	5	10	25
	0	1.50	1.73	1.66	1.46
First	1	1.52	1.83	1.52	1.57
Lag	4	1.59	1.52	1.55	1.50
	10	1.59	1.53	1.50	1.48

TABLE II - 15

MEAN LATENCIES OF CORRECT SECOND-GUESSES FOLLOWING NEW-INTRUSIONS

First Response Test

	P1	P2	P3	P4
0-10	0.70	1.36		
10-10	1.37	1.23		
0-10-10-10	1.08	1.28	0.81	1.53
10-10-10-10	1.44	1.35	1.16	1.21

Second Response Test

	P1	P2
0-10	1.33	1.55
10-10	1.40	1.20
0-10-0-10	1.35	1.12
10-10-10-10	1.33	1.35

P= number of previous presentations of the stimulus-response pair

---

First Response Test

First Lag	
0	0.63
1	1.34
4	1.37
10	1.33

Second Response Test

First Lag \ Second Lag	1	5	10	25
0	0.93	1.43	1.19	1.32
1	1.50	1.06	1.30	0.82
4	1.60	1.65	1.15	1.42
10	1.42	1.64	1.20	1.27

We next turn to the intrusion latencies. The mean latencies of intrusions, both old and new, are slower than the corresponding correct latencies in all cases; however, the latencies of new-intrusions are markedly longer than those of old-intrusions. This result, as will be seen in the next chapter, has important implications regarding the temporal ordering of the memory search. The latency of new-intrusions, as opposed to the correct latencies, does not vary as the number of reinforcements of R1 increases. The latency of a new-intrusion seems to be slower the longer the lag since the correct response, but the effect is essentially eliminated if lag = 0 is not considered. Finally, turning to the second-guess results, we will mention here only the following fact: after the change of response, the mean latency for a correct second-guess is shorter following new-intrusions than following old-intrusions. This would be surprising if the source of first-guess old-intrusions arose in confusion of the old and new responses. That is, if the old and new responses were confused and the subject chose one to output, then it might be expected that it would not take long to output the other after a wrong first-choice.

Conclusions

A rather large amount of diverse data has been collected in the two experiments. The variables examined include lag between study and test, number of reinforcements, second-guessing, rankings, negative transfer, intrusion rates for both first- and second-guessing, and latencies of response. A storage and retrieval model of long-term memory was described in Chapter I which, at least theoretically, had the capacity to deal with these variables simultaneously. In the next chapter it will be seen whether an explicit model based on the general theory can deal quantitatively with the data.

CHAPTER III

THEORETICAL ANALYSIS: A STORAGE AND RETRIEVAL MODEL

The derivation of a quantitative model from the theory presented in Chapter I involves a large number of individual decisions. The number of possible models that could be derived is extremely large, and this report cannot compare and contrast them all. Rather, an attempt will be made to construct the simplest possible model consistent with both the overall theory and the data. A few variations of the resultant model will also be discussed.

A model will first be presented for the data of Experiment I. This model will then be extended, but not altered, in an attempt to predict the data of Experiment II, data involving a number of additional variables.

Experiment I

The Short-Term System. The subject is assumed to pay some attention to each item presented for study, and thereby enter it into STS, at least momentarily. Therefore a test at lag 0 should result in nearly perfect performance (since the study phase and the test phase of the next trial are separated by only 3/4 sec.). We do not wish to involve ourselves in predicting just how good performance on such a zero-lag test should be (we would have to consider typing mistakes, and so forth) and therefore will treat the few zero-lag trials that occur as special cases. The first-guess and second-guess predictions for performance at zero-lag are simply set equal to the mean probability which was observed in all such instances, .97 and .50 respectively.

The present task was designed so that the short-term control processes utilized would tend to be single-trial coding mechanisms, rather than multi-trial rehearsal operations. That the design was successful in this regard is indicated both by subject reports and by the relative lack of an effect due to the type of intervening item at a lag of 1. Nonetheless, some items are undoubtedly maintained in STS beyond the trial of presentation -- this could occur if the subject takes more than one trial to encode certain items, or if some items previously encoded are given a small amount of additional rehearsal. It is therefore proposed that any item for which a storage attempt is not made decays rapidly from STS and is lost by the termination of the following trial. On the other hand, items which are coded decay from STS at a rate independent of the type of intervening items. Specifically, let $P(A)$ represent the probability that a storage attempt is made for a particular item; note that $P(A)$ includes the probability that the item is already in STS when presented on a trial. Let $P(R_i)$ represent the probability that the item will be present in STS at a lag of *i*. Then we have the following:

$$P(R_0) = .97 \qquad \text{Eq. III-1}$$

$$P(R_i) = P(A) \times (1-\alpha_1)^i$$

where α_1 is a parameter governing decay from STS. It might be asked whether there is a reason other than intuitive for including a decaying short-term process in the model. As it will be seen later, it is through the action of this process that a distributed learning effect is predicted by the model.

There is one important exception to the stated results concerning lack of organized rehearsal. The design of the experiment was such that a test of an item at lag 1 was almost always followed by a sequence of further tests of that item at lag 1. All subjects reported noting this fact, and a majority of them reported specifically rehearsing these items when they were noticed. As a result, performance on Type 1 and Type 2 items was abnormally high for presentation numbers 3, 4, 5, and 6. Rather than add to the model a specific rehearsal process to account for these observations, we will merely comment that it would be easy to do so.

Storage. When a currently unretrievable item is presented for study, an attempt may be made to store it. Let α be the probability of attempting to store such an item. The information stored will involve three components: stimulus, response, and associative information ($F(I_s)$, $F(I_r)$, and $F(I_a)$). As the present experiment is not designed to emphasize the differences between these information measures, we will characterize the amount of information transmitted to LTS by a single measure, $F(I)$, where the components of $F(I)$ include the three measures above. The exact form of $F(I)$ is not crucial to the model, but a reasonable spread in its distribution is necessary (a spread in the distribution is needed to predict both the first-guess lag curve and the rather low, and invariant, second-guessing performance over lags). For the purpose of simplifying calculations $F(I)$ will be approximated by a two-point distribution as follows. $F(I)$ is divided at its median; codes with strengths above the median will be called hi-codes and defined to have strength σ_H; codes with strengths below the median will be called lo-codes and defined to have strength σ_L ($\sigma_H > \sigma_L$). Thus an attempt to store information will result in a lo-code with probability .5 and will result in a hi-code with probability .5. The information stored will be placed in a location determined by stimulus characteristics, but because the present experiment uses a continuous task with homogenous items, the placement will not be ordered from the point of view of the model. Hence the model will treat placement as an essentially random process.

There are a number of decision rules which determine whether a storage attempt will be made for a particular item. Basically, a storage attempt will be made with probability α only when a correct response has not been retrieved from STS or LTS on the test phase of the trial. The only exception to this rule occurs at zero-lag. Term the state in which an item enters STS only momentarily, and is not coded, as the null-state. Items in the null-state at test, even though in STS, are treated as if a successful retrieval had not occurred. Thus an attempt may be made to store these items with probability α. These decision rules imply that a code which has just resulted in a successful retrieval will not be disturbed by further storage attempts, a reasonable strategy for the subject to adopt. On the other hand, the act of successful retrieval itself could reasonably be expected to make future retrieval easier. For this reason, lo-codes which have been successfully retrieved from LTS are treated thereafter as hi-codes (the alternative model, in which retrieved lo-codes are not altered, will be discussed later). One final informational change occurs in a code that has been successfully retrieved from LTS: the code is updated temporally to the present.

There are two processes which may occur when an item is given a reinforcement beyond the first. In one, a code which has not been retrieved from LTS will be left untouched, and a new and different code will be introduced during the study phase of the trial. In the other, the unretrieved code will be retrieved while a new storage attempt is made during the study phase, since the correct response is supplied at that time. If the code is retrieved during study, then it may be assumed that the ongoing storage attempt will consist of amending or changing the retrieved code; thus only a single code will result. Most likely, a mixture of these processes will take place during an experiment of the present type. However, because it greatly simplifies matters computationally, we shall assume that only the second hypothesis occurs; thus only a single code can exist for an item at any one time in LTS.*

The proportion of times a coding attempt is made, based on α, should be closely related to the decay rate from STS, α_1; that is, the more coding effort expended on intervening trials, the more likely is an item's loss from STS. For simplicity, we shall assume $\alpha_1 = \alpha$ in the remainder of this chapter.

Retrieval. At zero lag the subject is correct with probability .97 and second-guesses correctly with probability .50. The following

*The extended model, in which a mixture of the two possibilities occurs, will necessarily predict the data more closely than the restricted model actually used. However, the type of data collected in the present experiments is such that the extended model will not be better to an appreciable degree. As it will be seen, the restricted model fits quite well.

discussion does not deal with the zero-lag case. At test, a search is first made of STS; if the item is found, then it is reported correctly with probability 1.0. If the item is not found in STS, a search is made of LTS. We continue to use the terminology of Chapter I: if the stimulus currently being tested has a code stored in LTS, this code is termed the c-code; the other codes stored in LTS are termed i-codes.

For any stimulus tested, only a small subset of the codes stored in LTS will be examined during the search. This subset (termed the examination-subset) will be defined by the characteristics of the stimulus presented, characteristics that lead the subject to examine certain memory regions rather than others. Of course, once the search begins, the successive members of the examination subset will be determined to a large degree by associative factors. For the current experiment, however, the associative factors must be treated as essentially random, and the probability that a c-code will be in the examination subset depends only upon the "age" of the code, and the strength of the code.

Although the search through memory proceeds one code at a time, the clearest exposition results if we consider the search process in two stages. First we define a potential examination-subset, containing all those codes that will eventually be examined if the search continues long enough. In the second stage we define the order of search through the subset, and the probability of terminating the search and emitting a response at some point. Let $P(Z_i)$ be the probability that a c-code will be in the examination-subset, if the current test is at lag i. Then

$$P(Z_i) = \frac{\sigma}{\sigma + \beta\ (\text{age})} \qquad \text{Eq. III-2}$$

where σ is the strength of the c-code (either σ_H or σ_L), age is some function of i, and β is a parameter $(0 \leq \beta < \infty)$ governing the dependence of $P(Z_i)$ upon age. Since evidence was presented in the previous chapter that the probability correct depended upon the degree to which the intervening items were "known," the age of an item is defined to equal the mean number of new codes that were stored during the lag since the item's last presentation. The mean is taken over all possible realizations of the experiment; it is used rather than the actual number of new codes stored as an approximation to make the mathematics of the model tractable. The particular function presented in Eq. III-2 was utilized because it conforms to the criteria mentioned in Chapter I, and because of its simplicity. At large i, the value of $P(Z_i)$ decreases quite slowly as i increases, but at small i an appreciable decrease occurs.

If a c-code is examined during the search two processes can occur; first, a response may be recovered; second, the subject engages in a decision process to decide whether to emit any response recovered. In the following, the possibility that a response other than the one encoded will be recovered from the c-code will not be considered; this possibility will be taken up instead in the intrusion rate from i-codes. The probability of recovery and output should then be a straightforward function of the strength of the code: designate ρ_1 as the probability of recovery and output on the first-guess search, given a code was examined. Then,

$$\rho_1 = 1 - \exp(-\sigma) \qquad \text{Eq. III-3}$$

where exp is the exponential function ($\exp(k) = e^{-k}$) and σ is the strength of the code examined.

Next we turn to a consideration of intrusions, where an intrusion refers to the recovery and output of a response, as the result of the examination of an i-code during the search. The probability that an i-code will be in the examination-subset will depend in part upon the similarity of its stimulus to the stimulus being tested, but on the average this probability will be considerably smaller than for a c-code. Similarly, the probability that examination of an i-code results in the recovery and output of a response is considerably less than for a c-code. Each of these possibilities may be incorporated into the model by introducing the concept of effective-strength of an i-code, σ_I, where σ_I is less than either σ_H or σ_L. The degree to which σ_I is less than σ_H or σ_L should depend upon the similarity, or amount of generalization, between the stimuli used in the experiment. Note that it does not matter whether an i-code is a hi-code or a lo-code; its strength is σ_I in both cases. (While on the one hand a hi-code will be in the examination subset and lead to response recovery more often than a lo-code, on the other hand a hi-code is more likely to contain information which will inhibit intrusions during response-production.) Equations III-2 and III-3 can now be generalized to include i-codes: depending on the code being examined, σ in these equations will take on the value σ_I, σ_H, or

σ_L. Note that the age in Equation III-2 applies to the code under examination, and not necessarily to the item being tested.

The final component of the search process to be specified is the order of search through the examination-subset. To begin with, note that the experimental design utilized does not induce an order in the search (as might be the case if the stimuli were grouped in some obvious manner). In Chapter I it was suggested that an item would tend to be examined earlier in the search, the greater its strength and the lesser its age. We choose here to assume a strictly temporal search, independent of the strength of the codes. While this assumption cannot be entirely accurate, it should prove instructive to see how far it can be carried. Furthermore, it has the advantage of making the mathematics of the model tractable.

The memory search is assumed to be terminated when the first response is recovered and output; this seems reasonable if responding is required to be fairly rapid. As noted in Chapter I, this assumption leads to predictions that rankings and rerankings beyond the first choice will be at the chance level, which is close to the effect observed. If every code in the examination-subset is examined without a response being recovered and output, then the subject guesses randomly.

Following an error (an incorrect first-ranking) the subject engages in a second search of LTS. The second search is identical to the first, except that the decision criterion for output of recovered responses is lowered. This assumption is based on the results of Experiment II, where it was observed that the intrusion rates were considerably higher for second-guesses than for first-guesses. The change in decision criterion is assumed to apply to all codes, and is governed by a parameter γ as follows: let ρ_2 be the probability of recovery and output on the second-guess search, given that a code was examined. Then,

$$\rho_2 = 1 - \exp(-\gamma\sigma), \quad \gamma > 1. \qquad \text{Eq. III-4}$$

Equation III-4 is of course the counterpart of Equation III-3 for the first-guess search. The second-guess search is assumed to proceed independently of the first-guess search, but a c-code examined and rejected on the first-guess cannot give rise to a response on the second-guess.*

Review of the Model. The model utilizes six parameters:

α: governs the probability of a coding attempt, and decay from STS;

β: adjusts the degree to which an item's probability of being examined during the search depends upon age;

σ_H: the strength (amount of information stored) for a hi-code;

σ_L: the strength for a lo-code;

σ_I: the strength for an i-code (a code for an item other than the item currently being tested)--governs intrusions;

γ: adjusts the decision criterion for output of a recovered response during the second-guess search.

When an item is presented for test, a memory search commences. At zero-lag the probability correct is .97 and the probability of a correct second-guess is .50. Otherwise, if the item is currently present in

---

*In fact, this assumption makes almost no difference in the predictions for the data of Experiments I and II, compared with the complete independence assumption. It was used here because it seemed reasonable that the same c-code examined twice within a second or two would seldom give rise to differing results. The same does not apply to i-codes because σ_I is low enough that the change in decision criterion on the second-guess will make a significant difference.

STS, then a correct response is output. If the item is not in STS, then a search of LTS begins. The search takes place through a subset of the codes stored in memory, termed the examination-subset. The probability that a particular code will be in the examination-subset is given by Eq. III-2. The subject considers each code in the examination-subset in temporal order, the most recent first. The probability of recovering and outputting a response while considering a particular code is given by Eq. III-3. If all the codes in the subset are examined, but no response is emitted, then the subject guesses randomly. Whenever a response is recovered and emitted, the search is terminated and the subject ranks the remaining alternatives randomly. If the first-ranking proves to be incorrect, then a second search is initiated. This search is identical to the first, except that the decision criterion for output of a recovered response is lowered. In addition, a c-code examined and rejected during the first search cannot give rise to a response on the second search.

During the study phase of a trial the following events take place. If a successful retrieval had been made from LTS, then the code utilized is temporally updated to the present; in addition, a lo-code retrieved successfully becomes a hi-code. If a retrieval had been made from STS, then no new code is stored. Following any incorrect retrieval, or a pure guess, or a retrieval at zero-lag from the null-state, an attempt is made to store with probability α. If a storage attempt is made, then a hi-code will result with probability .5, and a lo-code will result with probability .5. Following a storage attempt, an item will leave STS with probability α on each succeeding trial.

In the following sections of the paper the model will be used to predict second-guessing data, among other phenomena. It should be noted that these data are conditional upon first-guess errors, and therefore are subject to considerable selection effects due to subject-item differences. The model predicts such selection effects since codes are assumed to be stored which have differing strengths. Thus selection due to subject-items should present no difficulties. This is not true, however, if items are selected on the basis of their performance on previous trials. Large subject differences are observed in both experiments; these differences will result in a considerable distortion of sequential phenomena which will not be predicted by the model. For this reason, this paper will not deal with sequential phenomena (such as two-tuples of errors on successive reinforcements, etc.).

Mathematical Analysis. The following discussion will be facilitated by a number of definitions. Let $c_{i,j}$ represent a correct response on the *i*th trial and the *j*th guess (i gives the trial number in the sequence of 439; j = 1 implies the rankings; j = 2 implies the rerankings). Let $e_{i,j}$ represent the corresponding error function. Let $\Omega_{i,k}$ represent the state of the memory system at trial i, for some realization of the experiment, k. The state of the system is described by three lists: the stimuli which are currently in STS, the stimuli which have lo-codes stored in LTS, and the stimuli which have hi-codes stored in LTS.

We shall deal in the following only with $P(c_{i,j})$, and not with the rankings and rerankings beyond the first choice -- the model predicts these to be at the chance level. We therefore have:

$$P(c_{i,j}) = \sum_{k=1}^{\infty} P(c_{i,j,k} \mid \Omega_{i,k})\, P(\Omega_{i,k}) \;, \qquad \text{Eq. III-5}$$

where summation is taken over all realizations of the experiment, denoted by k. For certain models this sum would be unwieldy to work with, but for the present model in which search is strictly temporally ordered and in which age is approximated by the mean number of intervening new codes, it is possible to bypass the summation and deal with the average state of the system at each trial, called $\bar{\Omega}_i$. $\bar{\Omega}_i$ may be iteratively calculated trial by trial, and $P(c_{i,j})$ is a relatively simple function of $\bar{\Omega}_{i-1}$. The details of the calculations, which are straightforward but require a cumbersome amount of notation, are reserved for Appendix 3. We note here only the following observation, which has not been stressed previously. When generating the predictions for the second-guess data, one must take into account the selection effect on the proportions of hi- and lo-codes introduced by the first-guess error. For example, many more errors occur if the item being tested has no code stored, or a lo-code stored, than if a hi-code is currently stored. As a result, the second-guess rates conditional on an error can be surprisingly stable over reinforcements and lags.

Using the computational methods described in Appendix 3, predictions can be generated from the model for any given set of parameter values. These predictions consist of the following vector for each of the 439 trials of the experiment: $[P(c_{i,1}); P(c_{i,2}); 1-P(c_{i,1})-P(c_{i,2})]$. Note that $P(c_{i,2})$ is not conditional upon a first-guess error; the numbers graphed in Figures II-5 and II-6 are conditional and equal $P(c_{i,2})/P(e_{i,1})$.

Given predictions for any given set of parameter values, we next define a goodness-of-fit measure. Corresponding to the predicted probabilities above, we define three observational quantities. $O_{i,1}$ is defined to be the observed number of correct first-guesses on the *i*th trial; $O_{i,2}$ is defined to be the observed number of correct second-guesses on the *i*th trial; $E_{i,2}$ is defined to be $N_i - O_{i,1} - O_{i,2}$, where N_i is the total frequency of all responses on the *i*th trial. The goodness-of-fit measure to be used is termed π^2 (Holland, 1967), and is calculated identically to χ^2 as follows:

$$\pi^2(\alpha,\beta,\sigma_H,\sigma_L,\sigma_I,\gamma) =$$

$$\sum_{i=1}^{439} \left\{ \frac{[N_i P(c_{i,1}) - O_{i,1}]^2}{N_i P(c_{i,1})} + \frac{[N_i P(c_{i,2}) - O_{i,2}]^2}{N_i P(c_{i,2})} + \frac{[N_i P(e_{i,2}) - E_{i,2}]^2}{N_i P(e_{i,2})} \right\} . \qquad \text{Eq. III-6}$$

N_i in the above equations decreases from 83 when i=1, to 58 when i=439. Although the π^2 distribution is not identical to that of χ^2 because certain independence assumptions are not satisfied in the above sum, a crude approximation to the levels of significance of π^2 can be made by use of the χ^2 tables. In using the tables, the degrees of freedom (d.f.) is equal to twice the number of trials, i, over which the π^2 is summed, minus the number of parameters being estimated (6 in the present case).

The next step is to estimate parameters by minimizing the π^2 function over all possible sets of parameter values. A grid search procedure was used to accomplish the minimization; i.e., a reasonably exhaustive search was made through the possible sets of parameter values, the computer generating predictions and computing π^2 for each set. The set of parameters giving rise to the lowest value of π^2 is assumed to generate the best fit of the model to the data. We will first state that the minimization carried out over all 439 trials resulted in predictions that consistently underestimated presentations 3 through 6 for item-types 1 and 2. As pointed out earlier, however, this was expected since the subjects reported rehearsal schemes for these trials. Therefore, in order not to bias the predictions for the remaining data, the 32 trials of the above type were deleted from the π^2 sum. Thus the π^2 function in what follows is summed over only 407 trials.

Predictions of the Model. The values of parameters which minimized the π^2 function for Experiment I were $\alpha = .68$, $\beta = .286$, $\sigma_H = 10.5$, $\sigma_L = 1.16$, $\sigma_I = .17$, $\gamma = 2.3$. The minimum π^2 value was 871.4, and the number of d.f. = (407)(2)-6 = 808. Since for large d.f. $\sqrt{2\chi^2} - \sqrt{(2)(d.f.)-1}$ is approximately normally distributed with a one-tailed test appropriate, a χ^2 value of 871.4 would be just above the .05 significance level. This is a strong indication that the model and the data were in close agreement on a trial-by-trial basis (if we ignore the abnormal points for item-types 1 and 2). The predictions of the model for the lag curves and the various item-types are shown in Figures II-2 through II-6 (pages 43 through 47) as the solid black points connected by unbroken lines. Except for the central portions of the Type 1 and Type 2 curves, the predictions are quite accurate. Even for the Type 1 and 2 curves the predictions are quite accurate for presentations 1 and 2, before rehearsal has begun, and for presentations 7 and 8, after rehearsal has ceased. Particularly noteworthy are the second-guess predictions, since only a single parameter, γ, has been utilized for adjustment of the second-guessing probability. It is instructive to note how the model predicts the maximum in the second-guess lag curve in Figure II-2 (page 43). At very small lags, all stored c-codes are likely to be retrieved correctly, so that most of the errors will occur when no c-code is stored in LTS; hence second-guesses will not be accurate. At longer lags, more and more intrusions occur before the c-code is reached in the first-guess search, hence more and more c-codes are available in LTS during second-guessing. At very long lags, even though many intrusions occur before the c-code is reached in the first-guess search, and therefore many c-codes are available during second-guessing, the lag is so long that the probability correct drops again. Note also that the distributed practice effect is predicted by the model. Such an effect arises from a short-term decaying store from which little learning takes place (Greeno, 1964). In the present model recovery from STS maintains lo-codes which would otherwise probably be transformed to hi-codes.

We may ask how the model performs under various restrictions and alterations. If $\gamma = 1.0$, which implies that the same bias applies during second-guessing as first-guessing, the predictions of the second-guessing probability are consistently above the observations, and the minimum π^2 almost doubles in value. Hence the altered output criterion implied by $\gamma = 2.3$ is necessary in the model. No restrictions among the three strength parameters, σ_H, σ_L, and σ_I can come close to fitting the data; that is, no two of the strength parameters may be set equal without losing accuracy of the model. An interesting alternative model results if we eliminate the assumption that successfully retrieved lo-codes become hi-codes. The minimum π^2 for the resultant model is 1020.4; the primary reason this model mispredicts is that very little learning is predicted to take place over the first few reinforcements of an item. Reference to Figure II-3 (page 44) shows a large rise in probability correct over the first few reinforcements. The transforming of retrieved lo-codes to hi-codes should not be misconstrued as antithetical to the finding from 3-state Markov models (Greeno, 1967a) that learning from the intermediate state is minimal. There is no simple correspondence between the three states of the Markov models, and the various states of the present model; rather they overlap each other. In any event, the present model does have a state from which little learning occurs: STS. To the extent that one is willing to equate this state and the intermediate Markov state, there is no conflict.

Finally, we may ask how the model predicts if "age" is based upon the number of intervening trials, rather than the number of intervening new codes. The minimum π^2 for this model is 920.0, perhaps not a dramatic increase, but one which confirms the empirical finding in Chapter II that "unknown" intervening items cause more forgetting.

The fit of the model to the data of Experiment I is quite good. The model is able to deal quantitatively, and simultaneously, with variations in number of reinforcements and in lag, with first-guesses and second-guesses, and with rankings and rerankings (in a sense). Nevertheless, the model as it stands has the power to deal with a considerably richer set of data. To be precise, an integral feature of the model is the prediction of intrusions, but intrusions were not observable in Experiment I. Experiment II, therefore, should provide a considerably more stringent test of the model. In addition, the model is extended to predict phenomena relating to the changing of response assignments for individual stimuli.

Experiment II

Before discussing Experiment II we wish to reiterate some important terminology. The term "intrusion" denotes the emission of an incorrect response. Two types of intrusions are possible: "new-intrusion" is used to denote the emission of a response which has never been paired with the stimulus being tested; "old-intrusion" is used to denote the emission of a response which is incorrect but has been paired at some earlier point in the session with the stimulus being tested. That is, an old-intrusion denotes the emission of the R1 response, if the R2 response is currently correct. The term "first-guess" denotes the subject's response during the initial portion of the test trial. If a first-guess intrusion is given, then the subject is given another chance to respond called the "second-guess." Thus, for example, the results of a hypothetical test trial might be described as a "second-guess old-intrusion following a first-guess new-intrusion." This terminology should be noted carefully, since it will be used throughout the remainder of this chapter.

There is one extension of the model that is not related to the change of response. As seen in Figure II-7 in the lower panel (page 65) there is a considerable rise in the intrusion rate following the first presentation of an item. The most likely interpretation of this finding is the one outlined in Chapter I. When the stimulus is presented for test, it is presumably scanned for salient characteristics. If a very salient characteristic is found, a search is then made in the memory location indicated by that characteristic, and if appropriate information is not found there, then the stimulus is identified as new and the search ceases. We therefore introduce a parameter δ to govern this process. Let δ be the probability that a normal search is made for a new item. Thus with probability $1 - \delta$ the stimulus is recognized as new and no search is made. We assume that no previously presented item is recognized as new (presumably old stimuli with high-salient characteristics always have enough information stored in the appropriate location that a recognition occurs and the search continues).

The model must now be extended to account for change-of-response phenomena. In order to make the following discussion clear, we define an o-code to be the code which encodes the R1 response for the item being tested, if the R2 response is currently correct. Thus the image encoding the previously correct response is called an o-code. It will be assumed that when a change of response occurs the o-code, if it is present in LTS, will not be updated temporally, it will simply remain in LTS and may be found during a later search. During a later search of LTS the probability that an o-code will be in the examination subset, and the probability that the R1 response will be recovered, will be the same as for a c-code at that same age. That is, since the stimuli are the same for the two codes, the same strengths apply in Equations III-2 and III-3: σ_H if a hi-code is stored, and σ_L if a lo-code is stored. However, the probability of output of the recovered response must depend upon whether information has been added to the o-code that it is "old" and hence wrong. We shall assume that whenever an R1 response has been retrieved, output, and is incorrect, that this information will be added to the o-code, so that the o-code cannot give rise to an old-intrusion on following trials. During the trial on which the answer is changed, however, the R1 response is correct when given. We therefore introduce a parameter κ defined as the probability that an o-code is tagged as wrong. The tagging is a result of the message ANSWER CHANGES which appears on the CRT, and a result of the changed pairing which is then presented for study. Note that κ applies only on the trial on which the answer changes, and applies only to o-codes which were correctly retrieved during the test phase of the trial.

The model as it now stands, due to the strictly temporal search characteristic, predicts no proactive effect. This is true because the c-code will always be encountered in the search before the o-code, if both are in the examination-subset. It was seen in Figure II-7 (page 63), however, that an overall proactive effect existed: the probability correct following the change of response was less than the probability correct following the first presentation of the R1 response. A parameter α_0 is therefore defined as the probability of attempting to encode the R2 response during the trial on which the change of response occurred, where $\alpha_0 < \alpha$. It is assumed that α_0 applies because the message ANSWER CHANGES appears on the screen. On trials where this message does not appear, α is assumed to apply in the usual way. Presumably the message sometimes induces the subject to pass by the new pairing, perhaps as a result of fear of confusion.

The extended model to be applied to Experiment II has three parameters not used in the model for Experiment I: δ, the probability of searching LTS when a new stimulus is tested; κ, the probability of tagging an o-code with the information that the response has been changed; and α_0, the probability of attempting to store on the trial when the response changes. Note that κ and α_0 apply only on the trial on which the response changes. When a search is made of LTS and no response is recovered and output, then the subject refrains from responding -- he does not guess.

Mathematical Analysis. For a given set of parameter values, the predictions of the model are generated in a manner quite similar to the method used for Experiment I. Appendix 4 presents the alterations in the iterative procedures used that enable us to predict the data for Experiment II. A natural next step would be the definition of an appropriate π^2 function, followed by a minimization routine. Unfortunately there is too much observed data for an attempt to minimize π^2 to succeed in a reasonable length of time, if all the data is considered simultaneously. Therefore, as a first step, we will fit the first-guess data only. The resultant parameter values, except for γ, will then be fixed. As a second step, the model will be applied to the second-guess data, but only γ will be estimated freely; the other parameters will retain the values giving the best fit to the first-guess data. The reason for estimating γ from the second-guess data is that γ is most sensitive to this data.

Let N_i be the total number of observations at the ith trial; let O_i be the observed number of correct first-guesses at the ith trial;

let Z_i be the observed number of intrusions (both old- and new-) at the ith trial. Let $P(c_i)$ be the predicted probability of a correct response at the ith trial; let $P(z_i)$ be the predicted intrusion probability at the ith trial (unconditional, and including both old- and new-intrusions). Then the following π^2 function is defined as a goodness-of-fit measure.

$$\pi^2(\alpha,\alpha_0,\beta,\sigma_H,\sigma_L,\sigma_I,\delta,\kappa,\gamma) =$$

$$\sum_{i=1}^{400}\left\{\frac{[N_iP(c_i) - O_i]^2}{N_iP(c_i)} + \frac{[N_iP(z_i) - Z_i]^2}{N_iP(z_i)} + \frac{[N_i(1-P(c_i)-P(z_i)) - (N_i-O_i-Z_i)]^2}{N_i(1-P(c_i) - P(z_i))}\right\}. \qquad \text{Eq. III-7}$$

The general comments made regarding Equation III-6 apply here also. N_i in the above π^2 function varies from 147 when i=1 to 122 when i=400. The number of degrees of freedom of π^2 in this instance is $(2 \times 400)-9 = 791$.

A grid search procedure was used to minimize π^2 over the possible sets of parameter values. When the parameters giving rise to the minimum value of π^2 were found, the second step of the estimation procedure was carried out. First a new π^2 function called π_1^2 was defined; π_1^2 was identical to π^2 except that all quantities were redefined to apply to the second-guess (thus N_i became the total number of intrusions, both new and old; etc.). All of the parameter values giving rise to the minimum value of π^2 were fixed except for the value of γ. Then $\pi_1^2(\gamma)$ was minimized. The minimum value of π_1^2 was 937.4 which occurred when $\gamma = 4.9$. This value of γ, along with the fixed values of the other parameters, was then used to recalculate π^2. The resultant value of π^2

was not appreciably higher than the minimum value based only on the first-guess data. As a result, we shall accept as "best" the predictions as generated by the parameter set with $\gamma = 4.9$. The values of the other parameters giving rise to the minimum π^2 are as follows: $\alpha = .94$, $\alpha_0 = .74$, $\beta = .25$, $\sigma_H = 45.1$, $\sigma_L = 1.25$, $\sigma_I = .117$, $\delta = .33$, and $\kappa = .30$. The minimum π^2 value was 872.6 (treated as a χ^2 this value would correspond to a level of significance between .05 and .01).

Predictions of the Model. The predictions of the model for the first-guess data are presented in Figures II-7, II-8, and Table II-8 (pages 65, 66, 69). The predictions, overall, are quite accurate; intrusion rates and correct guesses are predicted accurately both before and after the response changes, as a function of the number of reinforcements, and as a function of lag. The model predicts the overall proactive effect (due to the parameter α_0), and also the lack of a proactive effect as a function of the sequential history before the change of response (due to the strictly temporal search). There are several discrepancies that should be examined, however. First, note that the probability correct is considerably underpredicted after four reinforcements in the (10-10-10-10) condition (the discrepancy is .05 which is equivalent to a z-score of about 4.2). The model in general will underpredict after a large number of reinforcements for the following reason. Because the search is strictly temporally ordered, there is always a minimum average number of intrusions which occur before the c-code is ever examined, no matter how well the c-code is stored. Thus there is a ceiling for the probability correct at a given lag, as long as new items are continually introduced. In Experiment I some items were given up to 7 reinforcements, but the lags in these cases were large, and the probability correct never got near enough to the arbitrary ceiling for discrepancies to occur. In the present experiment, there are only four consecutive reinforcements before the response changes; as a result only a single discrepant point occurs. Thus it is not safe to conclude without further experimentation with greater numbers of reinforcements that the model definitely fails in predicting such a ceiling effect. (However, we will shortly examine evidence of a rather different character which will definitely show that the strictly ordered search hypothesis is in error.) A second discrepancy of the predictions occurs in the intrusion rates following the change of response, especially old-intrusion rates. Even though a proactive effect is not predicted for the probability correct, old-intrusions are predicted to rise as the amount of learning concerning R1 increases. The data, however, show a quite stable old-intrusion rate over conditions.

The above points notwithstanding, the predictions for the first-guess data are quite accurate. There is another statistic which bears this out. The model predicts that the new-intrusion rate will increase during the session, since more and more items are available to give rise to new-intrusions. This is easiest to check for new items. The observations and predictions are given in Figure II-9 (page 71). The overall level of the predictions in the Figure is governed by the parameter δ, and its accuracy is not surprising; however, the form of the predicted increase is quite close to that observed. The meaningfulness of this statistic is difficult to determine. The overall reduction in intrusion rates (reflected by δ) is assumed to occur because new items are recognized as such; it might seem logical that this recognition process would be a function of the duration of the session. It is possible to argue, however, that recognition via extremely salient stimulus characteristics is not appreciably affected by the number of stimuli input. This question should prove susceptible to further experimental research; for the present, it is not unreasonable to accept the second hypothesis above, an hypothesis in accord with the model.

Before turning to the second-guess results it would be instructive to consider the values attained by several of the parameters. It has been suggested earlier that the value of σ_I should be reflective of the amount of inter-stimulus generalization in the experiment. Since Experiment I utilized highly confusable consonant trigrams, and Experiment II utilized words, the value of σ_I should be smaller in the second experiment. The values attained were in the expected direction (.18 and .117 respectively). At first glance, the value attained by σ_H, 45.1, seems far too high; for example, this value would lead to predictions that the probability correct at a lag of near 300 would be as high as .30 (depending upon the condition). Fortunately this prediction can be roughly checked in the data since there were a few instances of very long lags. For example, stimulus number 10 (in the trial sequence of Appendix 2) was given successive reinforcements on trials 13, 39, and 389. The predicted probability correct for trials 39 and 389 was 44.6 and 28.5 respectively. The observed values on these trials were 42.1 and 42.4 respectively. Thus, the observed values were even higher than those predicted. Similarly, stimulus number 47 was given its final two reinforcements on trials 77 and 380. The predicted values for these trials were 35.4 and 26.3; the observed values were 35.3 and 42.3.

These results indicate that the high value of σ_H estimated in the present case was quite appropriate.

The second-guess predictions are presented in Table II-9 and in Figure II-10 (pages 73, 72). Figure II-10 gives the probability correct in the top panel and the overall intrusion rate in the lower panel, both following first-guess new-intrusions. In addition, the predictions in the lower panel are conditional upon a second-guess error. In both panels the fit is fairly accurate. The high intrusion rates predicted occur because $\gamma = 4.9$, considerably lowering the decision criterion for output of second-guess responses. A very high intrusion rate is predicted even for new items, items not previously presented. The model predicts this effect because the rates shown are conditional upon a first-guess error; an error implies that during the first-guess the subject did not recognize that the item was new, and made a decision to search LTS. Under these circumstances, a second-guess search will also be made, and since the stimulus being tested is new, this search will quite often result in intrusions (there is no c-code in LTS to lower the intrusion probability). Table II-9c gives the breakdown of the predictions in the lower panel of the figure, i.e., it gives the second-guess old-intrusion probability for the major item-types, following new-intrusions on the first-guess (the combined old- and new-intrusion rates were given in the figure). The predictions for these cases seem quite accurate, lending support to the hypothesis that o-codes and c-codes are quite similar, even with respect to their probability of being given following an extraneous intrusion.

Tables II-9a and II-9b give the second-guess predictions for the matrix item-types, following a new-intrusion on the first-guess. The first comment to be made is that the predictions in these tables are consistently high; this results from a failing of the model to be discussed shortly (under-predictions following old-intrusions on the first guess); if the second-guess data following first-guess old-intrusions were not part of the π^2 minimization, then these data would be fit more closely. Qualitatively, the effects predicted are observed with several minor exceptions. For example, in Table II-9a, a maximum probability correct is predicted at a second lag of 5: this prediction is observed if one ignores the observation at (1,1). In fairness to the model it should be pointed out there are very few observations in the (0,1) and (1,1) conditions. Similarly, in Table II-9b, the predicted increase in second-guess new-intrusions as a function of the second lag is observed if one eliminates the (0,1) and (1,1) points. More serious are the deviant predictions for second-guess old-intrusions after the second lag. The old-intrusion rate is predicted to rise as the second lag increases; this is observed for first lags of 1, 4, and 10, but just the opposite is seen for a first lag of 0. This misprediction could be rectified by assuming that the zero-lag is a special case that results in a very high probability of coding the old-response as being wrong. In the previous model, this coding only occurs after a non-null-state retrieval.

As a whole the predictions discussed so far are quite accurate. We turn now to a prediction which conclusively demonstrates that the assumption of a strictly temporally ordered LTS search is not adequate. These predictions are the counterpart to the observations presented in Table II-10 (page 76). The predictions were not given there because they are so extremely discrepant from the observations. The observed probability of a second-guess correct response following a first-guess _old_-intrusion is quite high -- about .30. Without giving the predictions cell by cell, we can state that the predicted probability correct varies between .02 and .05, depending upon the condition. The model predicts such low probabilities following first-guess old-intrusions because a c-code will always be examined before an o-code, if both are in the examination subset. This occurs because the LTS search is strictly temporal, and the c-code is always more recent than the o-code. If an old-intrusion first-guess is given, then it is certain that the c-code is either not present or has been bypassed in the search. A c-code present in LTS is not bypassed often, but when it is, it is almost always a lo-code; thus the probability of recovering it correctly during second-guessing is very low. The predicted second-guess intrusion probabilities following first-guess old-intrusions are also fairly deviant. Because the probability correct is predicted to be quite low, the intrusion predictions are quite high, about .45.

These failures of the predictions of the model make it clear that the assumption of a strictly temporal LTS search must be altered. The precise manner of alteration, which will still allow prediction of the previous observations, is not trivial and will be discussed later.

The failure of the temporal search assumption would make it presumptious to extend the present model to the latency results. Nevertheless, there are a number of theoretical remarks that may be made concerning the observed latencies. A simple model which can be used as a base for speculations holds that items retrieved from STS have a relatively short mean latency; items retrieved from LTS have a latency proportional to the number of codes examined before the response is output. The observed increase of correct response latency with lag can be explained either by considerations of recovery from STS (which decreases with lag) or by a partially temporal LTS search. The decrease in correct response latency with the number of reinforcements cannot be explained by a strictly temporal search; however, a search that examines codes in an order partially dependent upon the code's strength can predict this effect nicely. As the number of reinforcements increase, more and more of the c-codes stored will be hi-codes; hi-codes will tend to be examined earlier in the search than lo-codes because of their greater strength and hence will result in lower latencies. Previous studies have reported latency decreases with increases in reinforcements (i.e., Rumelhart, 1967), but responding in these studies was required on every trial. The results could therefore be explained as the result of averaging guesses and retrievals. Rumelhart also found that the latency of correct responses decreased after an item's terminal error, a result not explicable by guessing considerations. The effect is predicted quite easily by the present model, however. The same assumption regarding order of search can help explain why correct response latencies after the change of response are higher than before the change: The o-code will be examined occasionally before the c-code; even when the o-code response is inhibited, the latency of giving the c-code response will be lengthened by the prior consideration of the o-code. At first glance, it might appear that an occasional prior consideration of an o-code will not significantly alter the latency predictions, but this is not so. The predicted mean number of i-codes in the potential examination-subset is only 5.0 for the present model, even on the last trials of the session. The mean number actually examined prior to a correct response is considerably less than this figure, perhaps less than 1.0. In these circumstances, only a small proportion of o-codes additionally examined prior to emission of the correct response will greatly affect the predicted latency of such a correct response.

That intrusion latencies would be larger than correct response latencies would not be unexpected even in the strictly temporal search model. The model in which the search order depends upon the strength of the codes, however, does not only explain this result, but also why the latencies of old-intrusions are markedly smaller than those of new-intrusions (since the strength of i-codes is much less than that of o-codes, the o-codes will be examined earlier in the search). The fact that latencies of old-intrusions are greater than those of correct responses, even though in most cases there is a higher proportion of high strength codes for o-codes than c-codes, indicates that there is at least some temporal component to the search.

In the absence of a specific model, we will not discuss the latency results further. The major import of these results is that the order of the LTS search through the examination-subset must be only partially temporally ordered, and partially dependent upon the strength of the codes in the subset. This is the same conclusion arrived at through a consideration of the probability of a correct second-guess following an old-intrusion on the first guess. We might turn then to a discussion of the necessary features of such a model.*

Extensions of the Model. The most reasonable extension of the model lets the order of search through the examination-subset depend upon both the strength and temporal position of the codes. However, as soon as the strictly temporal search is altered, a proactive effect will be predicted which depends upon the amount of learning of the R1 response. That is, in the extended model the proportion of times the o-code is encountered prior to the c-code will be greater the more often the o-code is stored, and will be greater the larger the strength of the o-code. Similarly, the number of old-intrusions should be markedly affected by the level of learning of the R1 response, but neither of these predictions is observed. Apparently what is needed in the model is a mechanism by which well-known o-codes are marked as being wrong (old), but in which the number and strength of the unmarked o-codes remain very nearly constant over a wide range of reinforcement histories. The formulation of such a process would undoubtedly entail the use of several new parameters, but several parameters of the current model could very probably be eliminated, namely α_0 and κ. The precise formulation of an appropriate model to deal with the change-of-response data is beyond the scope of the present report; it must await further research to verify the results found, and to extend the range of variables studied. The major change

*The entire question of order of search can probably be settled unconditionally by engaging in further research in which each stimulus has a unique response assignment. Then all intrusions could be precisely placed temporally.

of response result, that proactive item interference does not depend upon the degree of learning of the R1 response, is certainly surprising in the light of the list structure results, and from the point of view of two-factor interference theory. This alone is sufficient reason for engaging in further research dealing with individual item-interference.

Concluding Discussion

We may summarize the major results of Experiment I as follows. First, it was found that the second-guessing probability could be considerably above chance even when responses ranked after the first choice were correct at the chance level. This result was interpreted as implying that the subjects used a retrieval strategy which output the first acceptable response recovered in the memory search. If this strategy is adopted, then the subject will give the recovered response as his first-ranking and guess for the remaining three rankings. Thus only the first-ranking will be above chance. Second-guessing, on the other hand, is based upon the result of an additional search of memory and may therefore be above chance. Second, it was found that performance in a continuous task decreased toward the chance level as the study-test interval became very large; in addition, when the lag between reinforcements was large, learning curves did not asymptote at a probability correct of 1.0, but rather seemed to stabilize at some intermediate value related to the size of the lag between reinforcements. These results demonstrated that any model which assumes a long-term absorbing state is not an appropriate representation of the memory process for tasks of the present type. In order to predict the above results, it was proposed that codes of varying strength are stored in LTS, and that the probability of retrieval at test is dependent upon the age and strength of the stored codes. This model was able to predict the learning, forgetting, and second-guessing data quite accurately. Third, it was found that the amount of forgetting at a given lag was dependent upon how well-known were the intervening items. The model predicted this result because the "age" of an item was made dependent upon the number of new codes that were stored during the intervening period.

The primary empirical results of Experiment II were concerned with proactive interference. It was found for both the probability of a correct response and the probability of an intrusion that an overall proactive effect was present. The magnitude of the effect, however, was not dependent upon the reinforcement and lag history prior to the change of response. The model predicted this proactive effect for probability correct because it assumed a strictly temporally ordered memory search. However, it was found that the probability of correctly second-guessing following an old-intrusion was about .30, markedly higher than the predictions of about .05. This latter finding demonstrated that the memory search could not be strictly temporally ordered; it was argued that search order is dependent upon the strength of codes as well as their age. This hypothesis was given further support by the analysis of response latencies. First, the latency of a correct response decreased with the number of reinforcements; second, the latency of a correct response was greater following the change of response than prior to the change. These latency results would be expected if codes of greater strength tended to be examined earlier in the memory search. Although this extension of the model seems quite natural, it results in the prediction that proactive effects will depend upon the reinforcement and lag history prior to the change of response. Since this prediction was not confirmed, further extensions of the model were suggested which would handle the observations.

Because an important feature of the storage and retrieval model was the prediction of intrusions, Experiment II was designed to examine intrusion probabilities over a wide range of conditions. In general, the model predicted the intrusion probabilities quite accurately. Two findings are especially noteworthy. First, the intrusion probabilities during second-guessing were found to be considerably higher than those during first-guessing; this result was taken to imply that the criterion for output of recovered responses was considerably lowered during second-guessing. Second, the intrusion probability when a new stimulus was presented for test was very much lower than that observed for previously presented items. This result reflects a recognition process in which certain new stimuli are recognized as being new; when presented stimuli with very salient characteristics do not trigger a recognition response in the expected location, it is assumed that a decision is made to cease further memory search. However, if a decision is made to search LTS, then a second-guess following an error should result in a very high intrusion probability, and this was also observed.

Taken as a whole, the predictions of the model were quite accurate. The model proved capable of dealing quantitatively and simultaneously with a wide variety of data, including lag, number of reinforcements, second-guessing performance, intrusion rates on first- and second-guessing, and change of response phenomena. The primary way in which this model differed from its predecessors was its emphasis upon an ordered search through a small subset of the codes stored in LTS. The value of such a process was confirmed by the analysis of the data; in fact, the analysis gives considerable support to the theory outlined in the first chapter of this report.

APPENDIX 1

Column a = trial number
Column b = stimulus number
Column c = number of reinforcements of current stimulus

a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	1	0	45	24	0	90	30	0	135	36	1	180	58	0
2	2	0	46	23	1	91	26	3	136	36	5	181	41	0
3	3	0	47	19	2	92	21	5	137	20	4	182	37	6
4	4	0	48	21	1	93	56	0	138	37	1	183	39	5
5	3	1	49	13	3	94	27	2	139	11	5	184	42	0
6	5	0	50	16	4	95	56	1	140	32	1	185	19	7
7	6	0	51	25	0	96	24	1	141	12	7	186	20	6
8	7	0	52	25	1	97	30	1	142	22	1	187	38	4
9	2	1	53	51	0	98	31	0	143	38	0	188	13	7
10	8	0	54	19	3	99	31	3	144	27	6	189	32	3
11	9	0	55	52	0	100	57	0	145	37	2	190	58	1
12	10	0	56	26	0	101	57	1	146	35	2	191	57	2
13	5	1	57	16	5	102	26	4	147	23	1	192	41	1
14	11	0	58	20	1	103	29	6	148	24	2	193	21	6
15	12	0	59	21	2	104	30	2	149	39	0	194	56	2
16	13	0	60	13	4	105	27	3	150	40	0	195	40	2
17	12	1	61	19	4	106	60	0	151	25	2	196	31	2
18	14	0	62	53	0	107	61	0	152	37	3	197	38	5
19	12	2	63	53	1	108	62	0	153	51	1	198	24	3
20	15	0	64	11	2	109	59	0	154	38	1	199	39	6
21	12	3	65	52	1	110	20	3	155	63	0	200	43	0
22	16	0	66	54	0	111	30	3	156	39	1	201	29	7
23	12	4	67	14	1	112	26	5	157	35	3	202	43	1
24	17	0	68	19	5	113	32	0	158	16	6	203	41	2
25	12	5	69	26	1	114	11	4	159	37	4	204	43	2
26	17	1	70	21	3	115	33	0	160	53	2	205	44	0
27	13	1	71	13	5	116	27	4	161	20	5	206	43	3
28	17	2	72	27	0	117	14	2	162	32	2	207	42	1
29	16	1	73	28	0	118	30	4	163	39	2	208	43	4
30	17	3	74	28	1	119	59	1	164	11	6	209	20	7
31	18	0	75	28	2	120	15	1	165	38	2	210	43	5
32	17	4	76	29	0	121	34	0	166	37	5	211	26	6
33	19	0	77	55	0	122	55	1	167	14	3	212	44	1
34	17	5	78	29	1	123	34	1	168	35	4	213	32	4
35	20	0	79	26	2	124	35	0	169	28	2	214	41	3
36	16	2	80	29	2	125	30	5	170	39	3	215	33	1
37	21	0	81	21	4	126	36	0	171	52	2	216	40	3
38	13	2	82	29	3	127	27	5	172	40	1	217	45	0
39	11	1	83	27	1	128	36	1	173	54	1	218	14	4
40	19	1	84	29	4	129	18	1	174	28	3	219	44	2
41	12	6	85	19	6	130	36	2	175	64	0	220	15	2
42	22	0	86	29	5	131	37	0	176	38	3	221	46	0
43	16	3	87	20	2	132	36	3	177	39	4	222	34	2
44	23	0	88	13	6	133	17	6	178	55	2	223	47	0
			89	11	3	134	36	4	179	35	5	224	30	6

APPENDIX 1 (CONT.)

a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
225	41	3	270	48	5	315	33	2	360	63	1	405	64	1
226	44	3	271	40	5	316	75	0	361	86	0	406	69	2
227	43	6	272	45	5	317	14	6	362	87	0	407	80	0
228	45	1	273	28	4	318	70	1	363	53	4	408	71	6
229	55	3	274	49	1	319	49	6	364	75	2	409	62	1
230	47	1	275	47	6	320	15	3	365	52	4	410	74	6
231	18	2	276	46	5	321	40	7	366	74	2	411	81	0
232	46	1	277	52	3	322	34	3	367	14	7	412	72	6
233	44	4	278	54	2	323	59	3	368	54	3	413	68	2
234	17	7	279	35	6	324	76	0	369	71	3	414	1	2
235	42	2	280	53	3	325	30	7	370	72	3	415	33	3
236	41	5	281	49	2	326	1	1	371	61	1	416	75	4
237	47	2	282	55	4	327	43	7	372	76	2	417	2	2
238	36	6	283	37	7	328	4	1	373	74	3	418	70	2
239	42	2	284	42	4	329	55	5	374	88	0	419	49	7
240	44	5	285	66	0	330	50	6	375	47	7	420	15	4
241	32	5	286	67	0	331	18	3	376	7	1	421	59	4
242	22	2	287	48	6	332	42	6	377	8	1	422	34	4
243	46	2	288	49	3	333	71	0	378	55	6	423	76	4
244	47	3	289	45	6	334	60	1	379	35	7	424	3	2
245	27	7	290	32	7	335	72	0	380	71	4	425	77	0
246	40	4	291	58	2	336	36	7	381	67	1	426	4	2
247	28	3	292	21	7	337	73	0	382	72	4	427	77	1
248	24	4	293	46	6	338	6	1	383	74	4	428	55	7
249	59	2	294	56	3	339	73	1	384	66	1	429	77	2
250	45	3	295	49	4	340	44	6	385	89	0	430	50	7
251	47	4	296	40	6	341	75	1	386	9	1	431	77	3
252	25	3	297	38	6	342	22	3	387	48	7	432	18	4
253	41	6	298	24	5	343	76	1	388	75	3	433	77	4
254	46	3	299	31	3	344	30	5	389	45	7	434	5	2
255	51	2	300	39	7	345	71	3	390	58	3	435	77	5
256	65	0	301	57	3	346	72	1	391	71	5	436	6	2
257	16	7	302	49	5	347	23	4	392	74	5	437	82	0
258	47	5	303	68	0	348	24	6	393	46	7	438	83	0
259	42	3	304	50	0	349	30	6	394	72	5	439	73	2
260	48	0	305	69	0	350	28	7	395	76	3	440	44	7
261	45	4	306	50	1	351	25	4	396	56	4			
262	48	1	307	42	5	352	74	0	397	38	7			
263	32	6	308	50	2	353	41	7	398	24	7			
264	48	2	309	70	0	354	51	3	399	31	4			
265	46	4	310	50	3	355	69	1	400	90	0			
266	48	3	311	26	7	356	71	2	401	57	4			
267	49	0	312	50	4	357	42	7	402	10	1			
268	48	4	313	68	1	358	72	2	403	78	0			
269	14	5	314	50	5	359	74	1	404	79	0			

APPENDIX 2

Column a = trial number Column b = stimulus number
Column c = 0 for study of first response, 1 for second
Column d = number of reinforcements of latest response

a	b	c	d
1	1	0	0
2	2	0	0
3	3	0	0
4	4	0	0
5	2	0	1
6	5	0	0
7	6	0	0
8	4	0	1
9	7	0	0
10	8	0	0
11	9	0	0
12	1	0	1
13	10	0	0
14	3	0	1
15	12	0	0
16	12	0	1
17	5	0	1
18	45	0	0
19	13	0	0
20	45	1	0
21	8	0	1
22	14	0	0
23	14	0	1
24	43	0	0
25	3	0	2
26	15	0	0
27	12	0	2
28	40	0	0
29	40	1	0
30	13	0	1
31	40	2	0
32	16	0	0
33	6	0	1
34	14	0	2
35	14	0	3
36	43	1	0
37	15	0	1
38	12	0	3
39	10	0	1
40	41	0	0
41	13	1	0
42	41	1	0
43	16	0	1
44	41	1	1
45	45	1	1
46	14	1	0
47	43	1	1
48	15	0	2
49	12	0	4
50	47	0	0
51	47	1	0
52	13	1	1
53	42	0	0
54	16	1	0
55	17	0	0
56	17	0	1
57	14	1	1
58	42	1	0
59	15	0	3
60	44	0	0
61	44	1	0
62	41	1	2
63	13	1	2
64	42	1	1
65	16	1	1
66	46	0	0
67	17	1	0
68	14	1	2
69	49	0	0
70	15	1	0
71	46	1	0
72	44	1	1
73	18	0	0
74	49	1	0
75	13	1	3
76	16	1	2
77	47	1	1
78	17	1	1
79	51	0	0
80	48	0	0
81	15	1	1
82	50	0	0
83	52	0	0
84	18	0	1
85	19	0	0
86	10	0	1
87	50	1	0
88	20	0	0
89	17	1	2
90	51	1	0
91	48	1	0
92	15	1	2
93	48	1	1
94	52	1	0
95	18	0	2
96	46	1	1
97	19	0	2
98	19	0	3
99	20	0	1
100	49	1	1
101	21	0	0
102	21	0	1
103	22	0	0
104	53	0	0
105	53	1	0
106	18	0	3
107	58	0	0
108	15	1	3
109	19	1	0
110	20	0	1
111	53	1	1
112	50	1	1
113	21	1	0
114	22	0	1
115	51	1	1
116	57	0	0
117	18	1	0
118	58	1	0
119	52	1	1
120	19	1	1
121	20	0	3
122	59	0	0
123	59	1	0
124	21	1	1
125	22	0	2
126	60	0	0
127	57	1	0
128	18	1	1
129	23	0	0
130	23	0	1
131	19	1	2
132	20	0	4
133	57	1	1
134	59	1	1
135	21	1	2
136	22	0	3
137	60	1	0
138	54	0	0
139	18	1	2
140	54	1	0
141	23	0	2
142	23	0	2
143	60	1	1
144	58	1	1
145	24	0	0
146	54	1	1
147	22	1	0
148	55	0	0
149	56	0	0
150	55	1	0
151	25	0	0
152	25	0	1
153	23	1	0
154	56	1	0
155	24	0	1
156	56	1	1
157	62	0	0
158	22	1	1
159	61	0	0
160	61	1	0
161	55	1	1
162	61	1	1
163	25	1	0
164	23	1	1
165	56	1	2
166	54	1	2
167	24	1	0
168	62	1	0
169	22	1	2
170	62	1	1
171	26	0	0
172	26	0	1
173	61	1	2
174	25	1	1
175	23	1	2
176	27	0	0
177	64	0	0
178	24	1	1
179	28	0	0
180	28	0	1
181	22	1	3
182	64	1	0
183	26	0	2
184	26	0	3
185	25	1	2
186	63	0	0
187	27	0	1
188	63	1	0
189	24	1	2
190	63	1	1
191	28	1	0
192	65	0	0
193	64	1	1
194	65	1	0
195	26	0	4
196	67	0	0
197	67	1	0
198	27	0	2
199	66	0	0
200	65	1	1
201	66	1	0
202	28	1	1
203	76	0	0
204	29	0	0
205	64	1	2
206	26	0	5
207	70	0	0
208	76	1	0
209	27	0	3
210	30	0	0
211	30	0	1
212	66	1	1
213	29	1	2
214	29	0	1
215	24	1	3
216	69	0	0
217	26	0	6
218	70	1	0
219	76	1	1
220	27	1	0
221	69	1	0
222	30	0	2
223	30	0	3
224	67	1	1

APPENDIX 2 (CONT.)

a	b	c	d
225	29	1	0
226	31	0	0
227	69	1	1
228	70	1	1
229	32	0	0
230	32	0	1
231	27	1	1
232	68	0	0
233	68	1	0
234	30	1	0
235	72	0	0
236	29	1	1
237	31	0	1
238	71	0	0
239	68	1	1
240	72	1	0
241	32	1	0
242	27	1	2
243	71	1	0
244	30	1	1
245	71	1	1
246	72	1	1
247	29	1	2
248	31	0	2
249	77	0	0
250	77	1	0
251	33	0	0
252	32	1	1
253	79	0	0
254	34	0	0
255	34	0	1
256	30	1	2
257	23	1	3
258	73	0	0
259	31	0	3
260	73	1	0
261	27	1	3
262	33	0	1
263	32	1	0
264	79	1	0
265	34	0	2
266	79	1	1
267	34	0	3
268	30	1	3
269	37	0	0
270	31	1	0
271	75	0	0
272	75	1	0
273	33	1	0
274	75	1	1
275	78	0	0
276	77	1	1
277	78	1	0
278	34	1	0
279	78	1	1
280	35	0	1
281	31	1	1
282	80	0	0
283	80	1	0
284	33	1	1
285	73	1	1
286	36	0	0
287	36	0	1
288	70	1	2
289	80	1	1
290	34	1	1
291	35	0	2
292	31	1	2
293	81	0	0
294	33	1	2
295	81	1	0
296	73	1	2
297	82	0	0
298	36	1	0
299	82	1	0
300	34	1	2
301	81	1	1
302	35	0	3
303	37	0	0
304	88	0	0
305	38	0	0
306	38	0	1
307	86	0	0
308	86	1	0
309	36	1	1
310	82	1	1
311	34	1	3
312	81	1	2
313	35	0	4
314	37	0	1
315	88	1	0
316	87	0	0
317	38	0	2
318	38	0	3
319	86	1	1
320	36	1	2
321	87	1	0
322	39	0	0
323	87	1	1
324	35	0	5
325	37	1	0
326	88	1	1
327	83	0	0
328	89	1	0
329	38	1	0
330	95	0	0
331	95	1	0
332	93	0	0
333	39	0	1
334	34	1	4
335	35	0	6
336	37	1	1
337	93	1	0
338	83	0	1
339	89	1	0
340	38	1	1
341	11	0	0
342	11	0	1
343	95	1	1
344	39	0	2
345	89	1	1
346	36	1	3
347	37	1	2
348	82	1	2
349	83	0	2
350	94	0	0
351	38	1	2
352	81	1	3
353	11	0	2
354	11	0	3
355	39	0	3
356	90	0	0
357	90	1	0
358	85	0	0
359	85	0	1
360	83	0	3
361	94	1	0
362	93	1	1
363	84	0	0
364	89	1	2
365	11	1	2
366	39	1	0
367	92	0	0
368	74	0	0
369	92	1	0
370	85	1	0
371	83	1	0
372	94	1	1
373	74	1	0
374	84	0	1
375	71	1	2
376	11	1	1
377	39	1	1
378	91	0	0
379	91	0	1
380	47	1	2
381	85	1	1
382	83	1	1
383	90	0	1
384	74	1	1
385	84	1	0
386	89	1	3
387	11	1	2
388	39	1	2
389	10	1	0
390	91	0	2
391	91	0	3
392	85	1	2
393	83	1	2
394	71	1	3
395	92	1	1
396	84	1	1
397	94	1	2
398	47	1	2
399	10	1	1
400	10	1	2

APPENDIX 3

ITERATIVE PROCEDURES FOR CALCULATING PREDICTIONS FOR EXPERIMENT I

Let $b_{n,j}$ be the probability that the item being tested is in STS, at lag j.
Let $c_{n,k}$ be the probability correct on trial n, guess k.
Let $e_{n,k}$ be $1.0 - c_{n,k}$.

Let $\bar{\Omega}_n$ be the average state of memory at trial n. $\bar{\Omega}_n$ is equivalent to the status of the following five vectors, each of length n:

1) $code_i$ is the probability that a new code was stored on trial i.
2) buf_i is the probability that the item presented on trial i entered STS (but not the null-state).
3) hic_i is the probability that a hi-code for the item presented on trial i is temporally placed in memory at trial i.
4) loc_i is the probability that a lo-code for the item presented on trial i is temporally placed in memory at trial i.
5) q_i is a dummy variable; equals zero only if the stimulus tested on trial i is later tested on a trial previous to n, else equals one.

We now show how to derive $\bar{\Omega}_n$ as a function of $\bar{\Omega}_{n-1}$. Assume we have $\bar{\Omega}_{n-1}$.

We need the following definitions.
$CR1_n$ is the probability of a correct response recovery given a first-guess LTS search, on trial n.
$CR2_n$ is the same for a second-guess search.
$IN1_n$ is the probability of an incorrect response recovery given a first-guess LTS search, on trial n.
$IN2_n$ is the same for a second-guess search.
$CE1_n = 1-CR1_n - IN1_n$.
$CE2_n = 1-CR2_n - IN2_n$.
SC_j is the probability of a correct recovery in an LTS search given that the search has proceeded as far as the jth trial. (Note: the search proceeds backwards, from trial n to trial 1.)
SI_j is the same for incorrect recoveries from LTS during the search.
Let j* be the trial number of the c-code.
Let fpi be the probability of an incorrect intrusion between trials n and j*.

Let $P(Z_k)$ be the probability that a code of type k is in the examination subset, where k= H,L, or I, depending upon the code type.
Let $P(P_k)$ be the probability that an examined code of type k gives rise to the response encoded, where k= H,L, or I, depending upon the code.

The status of a search of memory is defined by (SC_j, SI_j). This vector may be calculated recursively. If $j-1 \neq j^*$ then

$$SI_{j-1} = SI_j + q_j(1-SC_j-SI_j)(hic_j + loc_j)P(Z_I)P(P_I)(3/4).$$

$$SC_{j-1} = SC_j + q_j(1-SC_j-SI_j)(hic_j + loc_j)P(Z_I)P(P_I)(3/4).$$

But if $j-1 = j^*$ then,

$$SI_{j^*} = SI_j$$

$$SC_{j^*} = SC_j + (1-SC_j-SI_j)(hic_j P(Z_H)P(P_H) + loc_j P(Z_L)P(P_L).$$

In the above recursions, the age of an item at trial j is required (in $P(Z_k)$). The age is calculated as follows:

$$age_j = \sum_{i=j}^{i=n} code_i.$$

As the result of the recursion, we have (SC_1, SI_1). Then $CR1_n = SC_1$; $IN1_n = SI_1$.

We now have,

$$c_{n,1} = b_{n,j} + (1-b_{n,j})(CR1_n + IN1_n/4 + CE1_n/4), \text{ where } b_{n,j} = (buf_{n-j+1})\alpha^{n-j}.$$

Before the second-guess search predictions may be calculated, adjustment must be made for the selection effect due to the first-guess error. Hence, we must temporarily alter the proportions of hi- and lo-c-codes stored.

$$HIC_{j^*} = \left\{(1-b_{n,n-j^*+1})(hic_{j^*})(fpi + [1-(4/3)(fpi)][1-P(Z_H)][3/4]\,)\right\} / e_{n,1}.$$

$$LOC_{j^*} = \left\{(1-b_{n,n-j^*+1})(loc_{j^*})(fpi + [1-(4/3)(fpi)][1-P(Z_L)][3/4]\,)\right\} / e_{n,1}.$$

The second-guess recursion now proceeds identically to the first-guess recursion, except that the quantities above are substituted for hic_{j^*}, loc_{j^*}. The result is $CR2_n$, $IN2_n$, and $CE2_n$. Then we have,

$$c_{n,2} = (1-c_{n,1})(CR2 + IN2_n/3 + CE2_n/3).$$

This concludes the predictions on the nth trial; to calculate $\bar{\Omega}_n$, however, we must complete the nth trial of the five vectors making up the state of memory.

Let $Y = (1-b_{n,n-j^*+1})$; Let $W = CR1_n + IN1_n/4 + e_{n,1}(CR2_n + IN2_n/3)$.

Then,

$$code_n = Y(1-W)\alpha.$$

$$hic_n = Y(W+[1-W][\alpha/2]).$$

$$loc_n = Y(1-W)\alpha/2.$$

$$q_{j*} = 0.$$

$$buf_n = 1 - Y + Y(1-W)\alpha.$$

The above five equations transform $\overline{\Omega}_{n-1}$ into $\overline{\Omega}_n$. The iterative process then continues until the 439 trials are predicted. The boundary conditions on the above process, and special cases such as zero-lag, are not given here: they are straightforward, and their presentation merely increases the terminology needed.

APPENDIX 4

ITERATIVE PROCEDURES FOR CALCULATING PREDICTIONS FOR EXPERIMENT II

The iterations used for Experiment II are very close in character to those for Experiment I, and little purpose is served by repeating them here in full detail. Instead, we present only the equations which normalize the proportions of hi- and lo-codes for selection effects prior to second-guessing.

Before the answer changes, all intrusions are new, hence, there are just two conditions: HIC_{j*} represents adjusted hi-codes; LOC_{j*} represents adjusted lo-codes.

$$HIC_{j*} = \left\{Y(hic_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_H)][IN1_n(25/26)-fpi])\right\} /e_{n,1}.$$

$$LOC_{j*} = \left\{Y(loc_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_L)][IN1_n(25/26)-fpi])\right\} /e_{n,1}.$$

After the answer changes we must consider two possibilities: the intrusion could have been old- or new-. We denote the adjusted probabilities with primes (') if there was a new-intrusion; we denote the adjusted probabilities with quotes (") if there was an old-intrusion. Then,

$$'HIC_{j*} = \left\{Y(hic_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_H)][(1-f1)+f1(1-c2)(1-f2)]\right\}/n_{n,1}.$$

$$'LOC_{j*} = \left\{Y(loc_{j*})(fpi + [1-(26/25)(fpi)][1-P(Z_L)][(1-f1)+f1(1-c2)(1-f2)]\right\}/n_{n,1}.$$

$$''HIC_{j*} = \left\{Y(hic_{j*})(1-[26/25]fpi)(1-P[Z_H])(f1)(c2)\right\} /e_{n,1}.$$

$$''LOC_{j*} = \left\{Y(loc_{j*})(1-[26/25]fpi)(1-P[Z_L])(f1)(c2)\right\} /e_{n,1}.$$

The above equations use several definitions not used in Appendix 3.

Set $Y = (1 - b_{n,n-j*+1})$.

Let $n_{n,1}$ represent the probability of a new intrusion on the first-guess on trial n.

Let $e_{n,1}$ represent the probability of an old intrusion on the first-guess on trial n.

Let c2 represent the probability of giving the R1 response after examining the o-code.

Let 1-f1 be the probability of emitting a new intrusion as a result of examining a i-code temporally between the c-code and the o-code.

Let 1-f2 be the probability of emitting a new intrusion as a result of examining a i-code temporally older than the o-code.

APPENDIX 4 (CONT.)

Then the above equations give the correction for selection effects. The remaining calculations are straightforward, similar to those given in Appendix 3, and are therefore not presented.

REFERENCES

Atkinson, R. C., Brelsford, J. W., and Shiffrin, R. M. Multi-process models for memory with applications to a continuous presentation task. _Journal of Mathematical Psychology_, 1967, _4_, 277-300.

Atkinson, R. C., and Estes, W. K. Stimulus sampling theory. In R. D. Luce, E. Galanter, and R. R. Bush (Eds.), _Handbook of mathematical psychology_, Vol. 2. New York: Wiley, 1963, 121-268.

Atkinson, R. C., and Shiffrin, R. M. Mathematical models for memory and learning. Tech. Rep. No. 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965. (To be published in D. P. Kimble (Ed.), _The proceedings of the third conference on learning, remembering and forgetting_. New York: New York Academy of Science).

Atkinson, R. C., and Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), _The psychology of learning and motivation: Advances in research and theory_, Vol. 2. New York: Academic Press, 1968, in press.

Binford, J. R., and Gettys, C. Nonstationarity in paired-associate learning as indicated by a second guess procedure. _Journal of Mathematical Psychology_, 1965, _2_, 190-195.

Bjork, A. Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals. Tech. Rep. No. 106, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1966.

Bower, G. H. Notes on a descriptive theory of memory. In D. P. Kimble (Ed.), The organization of recall. New York: New York Academy of Science, 1967, 112-185.

Calfee, R. C., and Atkinson, R. C. Paired-associate models and the effects of list length. Journal of Mathematical Psychology, 1965, 2, 254-265.

DaPolito, F. J. Proactive effects with independent retrieval of competing responses. Unpublished doctoral dissertation, Indiana University, 1966.

Estes, W. K. Statistical theory of distributional phenomena in learning. Psychological Review, 1955, 62, 369-377. (a)

Estes, W. K. Statistical theory of spontaneous recovery and regression. Psychological Review, 1955, 62, 145-154. (b)

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), Psychology: a study of a science, Vol. 2. New York: McGraw-Hill, 1959, 380-491.

Estes, W. K. All-or-none processes in learning and retention. American Psychologist, 1964, 19, 16-25.

Greeno, J. G. Paired-associate learning with massed and distributed repetitions of items. Journal of Experimental Psychology, 1964, 67, 286-295.

Greeno, J. G. Paired-associate learning with short-term retention: mathematical analysis and data regarding identification of parameters. Journal of Mathematical Psychology, 1967, 4, 430-472.

Greeno, J. G. Some attempts to locate associative information. Prepublication, 1967.

Holland, P. W. A variation on the minimum chi-square test. *Journal of Mathematical Psychology*, 1967, *4*, 377-413.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon and D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968, 172-213.

Melton, A. W. Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 1-21.

Peterson, L. R., Sãltzman, D., Hillner, K., and Land, V. Recency and frequency in paired-associate learning. *Journal of Experimental Psychology*, 1962, *63*, 396-403.

Peterson, L. R., Hillner, K., and Saltzman, D. Supplementary report: time between pairings and short-term retention. *Journal of Experimental Psychology*, 1962, *64*, 550-551.

Postman, L. The present status of interference theory. In C. N. Cofer (Ed.), *Verbal Learning and Verbal Behavior.* New York: McGraw-Hill, 1961, 152-179.

Rumelhart, D. E. The effects of interpresentation intervals in a continuous paired-associate task. Tech. Rep. No. 116, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967.

Shiffrin, R. M., and Atkinson, R. C. Some speculations on storage and retrieval processes in long-term memory. Tech. Rep. No. 127, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1968.

Thompson, W. J. Recall of paired-associate items as a function of interpolated pairs of different types. Psychonomic Science, 1967, 9, 629-630.

Thorndike, E. L. The teacher's word book. New York: Teachers College, Columbia University, 1921.

Underwood, B. J. Interference and forgetting. Psychology Review, 1957, 64, 49-60.

(Continued from inside front cover)

91 P. Suppes. Information processing and choice behavior. January 31, 1966.

92 G. Groen and R. C. Atkinson. Models for optimizing the learning process. February 11, 1966. (Psychol. Bulletin, 1966, 66, 309-320)

93 R. C. Atkinson and D. Hansen. Computer-assisted instruction in initial reading: Stanford project. March 17, 1966. (Reading Research Quarterly, 1966, 2, 5-25)

94 P. Suppes. Probabilistic inference and the concept of total evidence. March 23, 1966. (In J. Hintikka and P. Suppes (Eds.), Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966. Pp. 49-65.

95 P. Suppes. The axiomatic method in high-school mathematics. April 12, 1966. (The Role of Axiomatics and Problem Solving in Mathematics. The Conference Board of the Mathematical Sciences, Washington, D. C. Ginn and Co., 1966. Pp. 69-76.

96 R. C. Atkinson, J. W. Brelsford, and R. M. Shiffrin. Multi-process models for memory with applications to a continuous presentation task. April 13, 1966. (J. math. Psychol., 1967, 4, 277-300).

97 P. Suppes and E. Crothers. Some remarks on stimulus-response theories of language learning. June 12, 1966.

98 R. Bjork. All-or-none subprocesses in the learning of complex sequences. (J. math. Psychol., 1968, 1, 182-195).

99 E. Gammon. The statistical determination of linguistic units. July 1, 1966.

100 P. Suppes, L. Hyman, and M. Jerman. Linear structural models for response and latency performance in arithmetic. (In J. P. Hill (ed.), Minnesota Symposia on Child Psychology. Minneapolis, Minn.: 1967. Pp. 160-200).

101 J. L. Young. Effects of intervals between reinforcements and test trials in paired-associate learning. August 1, 1966.

102 H. A. Wilson. An investigation of linguistic unit size in memory processes. August 3, 1966.

103 J. T. Townsend. Choice behavior in a cued-recognition task. August 8, 1966.

104 W. H. Batchelder. A mathematical analysis of multi-level verbal learning. August 9, 1966.

105 H. A. Taylor. The observing response in a cued psychophysical task. August 10, 1966.

106 R. A. Bjork. Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals. August 11, 1966.

107 R. C. Atkinson and R. M. Shiffrin. Some Two-process models for memory. September 30, 1966.

108 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the third year. January 30, 1967.

109 P. Suppes and I. Rosenthal-Hill. Concept formation by kindergarten children in a card-sorting task. February 27, 1967.

110 R. C. Atkinson and R. M. Shiffrin. Human memory: a proposed system and its control processes. March 21, 1967.

111 Theodore S. Rodgers. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. June 1, 1967.

112 Jack M. Knutson. Spelling drills using a computer-assisted instructional system. June 30, 1967.

113 R. C. Atkinson. Instruction in initial reading under computer control: the Stanford Project. July 14, 1967.

114 J. W. Brelsford, Jr. and R. C. Atkinson. Recall of paired-associates as a function of overt and covert rehearsal procedures. July 21, 1967.

115 J. H. Stelzer. Some results concerning subjective probability structures with semiorders. August 1, 1967

116 D. E. Rumelhart. The effects of interpresentation intervals on performance in a continuous paired-associate task. August 11, 1967.

117 E. J. Fishman, L. Keller, and R. E. Atkinson. Massed vs. distributed practice in computerized spelling drills. August 18, 1967.

118 G. J. Groen. An investigation of some counting algorithms for simple addition problems. August 21, 1967.

119 H. A. Wilson and R. C. Atkinson. Computer-based instruction in initial reading: a progress report on the Stanford Project. August 25, 1967.

120 F. S. Roberts and P. Suppes. Some problems in the geometry of visual perception. August 31, 1967. (Synthese, 1967, 17, 173-201)

121 D. Jamison. Bayesian decisions under total and partial ignorance. D. Jamison and J. Kozielecki. Subjective probabilities under total uncertainty. September 4, 1967.

122 R. C. Atkinson. Computerized instruction and the learning process. September 15, 1967.

123 W. K. Estes. Outline of a theory of punishment. October 1, 1967.

124 T. S. Rodgers. Measuring vocabulary difficulty: An analysis of item variables in learning Russian-English and Japanese-English vocabulary parts. December 18, 1967.

125 W. K. Estes. Reinforcement in human learning. December 20, 1967.

126 G. L. Wolford, D. L. Wessel, W. K. Estes. Further evidence concerning scanning and sampling assumptions of visual detection models. January 31, 1968.

127 R. C. Atkinson and R. M. Shiffrin. Some speculations on storage and retrieval processes in long-term memory. February 2, 1968.

128 John Holmgren. Visual detection with imperfect recognition. March 29, 1968.

129 Lucille B. Mlodnosky. The Frostig and the Bender Gestalt as predictors of reading achievement. April 12, 1968.

130 P. Suppes. Some theoretical models for mathematics learning. April 15, 1968. (Journal of Research and Development in Education, 1967, 1, 5-22)

131 G. M. Olson. Learning and retention in a continuous recognition task. May 15, 1968.

132 Ruth Norene Hartley. An investigation of list types and cues to facilitate initial reading vocabulary acquisition. May 29, 1968.

133 P. Suppes. Stimulus-response theory of finite automata. June 19, 1968.

134 N. Moler and P. Suppes. Quantifier-free axioms for constructive plane geometry. June 20, 1968. (In J. C. H. Gerretsen and F. Oort (Eds.), Compositio Mathematica. Vol. 20. Groningen, The Netherlands: Wolters-Noordhoff, 1968. Pp. 143-152.)

135 W. K. Estes and D. P. Horst. Latency as a function of number or response alternatives in paired-associate learning. July 1, 1968.

136 M. Schlag-Rey and P. Suppes. High-order dimensions in concept identification. July 2, 1968. (Psychom. Sci., 1968, 11, 141-142)

137 R. M. Shiffrin. Search and retrieval processes in long-term memory. August 15, 1968.