ABSTRACT

THE EFFECTS OF INTER-TRIAL SPACING IN PAIRED-ASSOCIATE LEARNING

by

Keith Frank Jones.

Paired-associate recall following two presentations of a pair to be remembered is heavily influenced by the spacing (in terms of intruding trials involving different pairs) between the two successive presentations. In particular, subsequent recall performance improves as the spacing between two successive presentations of a pair increases, at least up to some optimal interpresentation spacing interval. This effect is known as the spaced practice improvement (or SPI) effect, and is clearly of fundamental importance to our understanding of the relationship between repetition and learning. However, most of the recent research on the SPI effect has involved free-recall and Brown-Peterson paradigms, and there are grounds for suspecting that the SPI effect obtained with paired associates may have a different underlying rationale to the SPI effect observed in these other paradigms.

Although the extant data strongly suggest that pairs held in short-term memory at the time of their second presentation receive little or no benefit from that re-presentation, there has been no systematic work attempting to relate the effectiveness of a re-presentation with both interpresentation spacing and with the state of learning of a pair at the time of the re-presentation. This thesis was designed to investigate this relationship in an effort to derive constraints on an adequate theory of the SPI effect in paired-associate learning beyond those imposed by prior research.

To this end, three experiments were conducted, each of which employed a variation of the study-test, continuous paired- associate (CPA) paradigm. The basic condition common to all three experiments may be depicted as

$$P_1 \underset{i \text{ trials}}{\cdots\cdots\cdots} T_1 \quad P_2 \underset{8 \text{ trials}}{\cdots\cdots\cdots} T_2$$

where P_1 and P_2 are presentations of a pair to be remembered, T_1 and T_2

are tests of the pair, i represents the spacing between the two successive presentations in terms of intruding trials involving other pairs, and there were always 8 such trials between the second presentation P_2 and the final test of a pair, T_2. It will be noticed that T_1 always immediately preceded P_2, so that T_1 performance would give some insight as to the state of learning of a pair on entry to P_2.

In Experiment I there were ten conditions defined by i = 0, 1, 2 3, 4, 5, 6, 8, 12, or 16 trials. Common word stimuli were employed, paired with integer responses in the range 1-15. On each test trial, subjects were required to respond with the appropriate integer, guessing if necessary. Both study and test trials were paced at a 2-second rate.

In Experiment II, there were five basic conditions defined by i = 0,4, 8, 12 or 16 trials. The procedure followed that of Exp. I with two exceptions. In the first place, subjects were required on each test trial to make two responses; a stimulus recognition response ("old" or "new") followed by a recall response, again guessing where necessary. Secondly, because of the additional response required at test, both study and test trials were paced at a 3-second rate.

In Experiment III, there were five conditions defined by i = 0, 2, 4, 6, or 8 trials. Nonsense-syllable stimuli of low meaningfulness were paired with integer responses in the range 1-5. The procedure otherwise followed that of Exp. II with the important exception that, whereas study trials were paced at a 3-second rate as before, test trial duration was subject-determined (i.e. test trials were terminated only when the subject had completed his responding).

The principal findings of Exps. II and III may be summarized as follows. Although stimulus recognition appeared to be a necessary condition for correct recall, in that recall performance on any trial to which a recognition error had been made could be accounted for by a guessing hypothesis, there was no evidence that stimulus recognition otherwise influences the SPI effect on T_2 recall performance.

The results of Exp I strongly suggested that although T_2 recall performance following T_1 error improved sharply as interpresentation spacing increased from 0 to 1 or more trials, there was not subsequent systematic relationship with interpresentation spacing. On the other hand, T_2 recall performance following a correct recall on T_1 appeared to increase systematically with interpresentation spacing, and furthermore, this improvement appeared to be maintained over spacings in excess of those required to "wipe out" short-term retention effects at T_1.

The results of Exp I were subjected to a detailed analysis employing a Markovian learning model. Two major conclusions were drawn from the analysis. Firstly, the SPI effect on T_2 recall performance resulted entirely from an increased effectiveness with interpresentation spacing of the second presentation in reducing the subsequent decay rate of those items that were already moderately well encoded on entry to P_2. Secondly, this increase was maintained over interpresentation spacings in excess of those sufficient to remove short-term retention effects at P_2. These results appeared consistent with a differential encoding hypothesis based upon an encoding theory of paired-associate forgetting.

THE EFFECTS OF INTER-TRIAL

SPACING IN PAIRED-ASSOCIATE LEARNING

by

Keith Frank Jones.

A thesis submitted in fulfillment of

the requirements for the degree of

Doctor of Philosophy

in

The University of Stirling.

1976.

Ph. D. ~~awarded~~ conferred March 1977 (awarded December 1976)

PREFACE

It would be impossible to acknowledge the efforts of all the people who have helped me in the writing of this thesis. I would like to thank my supervisor, Dr. Michael F. Moore for his invaluable advice and encouragement. Special thanks also go to Angus Annan and Irene Lloyd, who helped design and assemble the apparatus employed in Exps I and II, and to Dr. Roger Henry, who gave me substantial assistance in programming the equipment employed in Exp III. I would also like to thank Klaus-Peter Stramm for helping me to run Exp II, and Dr. Ronald MacDonald and Dr. Alistair Watson for the many patient hours spent in discussion of various theoretical and practical problems which I encountered in the course of my research. Invaluable help and support of a different kind was provided by my wife, Adèle.

This research was supported and financed by an S.S.R.C. research studentship grant.

CONTENTS

LIST OF TABLES

LISTS OF FIGURES

CHAPTER ONE

INTRODUCTION

It has long been known that the repetition of a task to be learned usually has beneficial effects on performance, or in other words, practice facilitates learning. This is not to say that learning can be equated with practice; indeed, it has been shown that under some conditions, practice can impair learning, possibly because it induces fatigue (both physical and psychological) which retards performance and may also disrupt acquisition. It is probably true to say that no satisfactory definition of learning has as yet been propounded, but that in general psychologists are able to agree as to when learning is present in performance and when it is absent. Furthermore, it is generally accepted that the most straighforward method of studying learning is to set up some kind of practice schedule on the relevant task, and then to examine the resultant changes in performance with practice; if performance improves, learning has taken place.

During recent years, a great deal of research has been carried out in the field of human memory and verbal learning (for reasons which will become apparent a little later). Now, in any field of research, when a phenomenon is discovered that is not predicted by current theory, then it immediately becomes apparent that the theory is, if not totally incorrect, then at least inadequate. The opportunity then presents itself to advance our state of knowledge in the field in question. One such phenomenon that has emerged in the area of verbal learning and memory is known as the spaced practice improvement (or SPI) effect, which may be briefly summarised as follows. If a verbal item to be remembered is presented repeatedly for study, then subsequent recall of that item is better if periods filled with "interfering" activity (of the kind that would normally lead to a decline in recall if interpolated between study and test) intrude between successive presentations of the item, than when the successive presentations immediately follow each other with no such intruding activity.

Such a finding is clearly of immense importance to our understanding of the relationship between repetition and verbal learning. It seems reasonable to hope that a careful study of the phenomenon will be of great value, at best by suggesting a more comprehensive, unique theory of memory, and at the very least by reducing the range of plausible alternative theories. The issue will first be clarified, however, by an examination of the relationship between learning and memory, followed by a brief outline of current theories of memory, in order to provide a theoretical framework for later discussion. It is furthermore proposed to restrict research to an examination of the SPI effect in paired-associate memory, and to this end, a more detailed review of paired-associate memory will be undertaken.

1.1 Learning, Verbal Learning and Memory

Learning has long been an important area of study in psychology, due to a great extent to the recognition that a very large proportion of all behaviour is at least partially determined by the organism's experience. Because behaviour is so dependent on learning, the psychology of learning is a topic of fundamental theoretical importance. Most of the major learning theorists of the past, such as Tolman (1932), Guthrie (1935), Skinner (1938) and Hull (1943), have tended to concentrate on the motivational aspects of learning, following an approach which derives from Pavlov's (1927) classical work on conditioned reflexes. Research in this field characteristically involves the study of animal conditioning (both operant and classical), and is very much concerned with the role and nature of reinforcement.

Another major approach to the study of learning, which may be traced back to Ebbinghaus' (1885) book "On Memory", is concerned with the study of human learning. Work in this field has been primarily addressed to the investigation of the learning process per se. The difference between these two approaches may best be illustrated by outlining the general experimental procedures employed. In a typical conditioning study, an experimental (animal) subject is placed in a carefully controlled stimulus environment, and subjected to a training schedule which usually comprises a number of components, both discriminative and reinforcing in nature. Certain aspects of the training schedule are often contingent on the emission of a specific response by the subject. The training schedule is maintained until such time as the animal's behaviour is judged to have become stable, and a comparison is then made between aspects of the pre- and post- training behaviour of the subject. The emphasis is very clearly placed on investigating <u>what</u> is learned in relation to the applied training schedule, and theoretical approaches attempt to relate changes in behaviour to aspects of the training procedure via motivational hypotheses.

By contrast, in a typical human learning situation, the subject is set a learning task by the experimenter, whose interest in the subject's motivation is strictly limited; as long as the subject is sufficiently motivated to attempt to learn the task, he is satisfied. Furthermore, both subject and experimenter know in advance <u>what</u> the subject is trying to learn to do. The tasks employed in human learning studies may range from perceptual-motor tasks to verbal memory and problem solving. The subject's training usually takes the form of a sequence of practice trials, and aspects of the subject's performance are recorded on each trial. Emphasis is placed on an examination of the trial-to-trial changes in performance which occur; in this way, evidence concerning the nature of the learning process itself is assembled.

Thus, the animal conditioning and human approaches to learning may be distinguished both on operational grounds, and on the basis on theoretical emphasis. However, a more fundamental distinction between the two fields can be made; for example, Estes(1967) has pointed out that the manipulation of the delay and magnitude of a response-contingent reward in a human learning situation often has quite a different effect to a similar manipulation of response-contingent reinforcement parameters in animal conditioning studies. In general, it appears that the informational aspects of reward are critical in human learning, whilst reward magnitude has relatively little effect, in marked contrast to the effects of reinforcement in animal conditioning.

During the last fifteen years or so, by far the greatest proportion of work on human learning has been carried out in the fields of short-term verbal memory and verbal learning. Historically, the shift of emphasis towards verbal tasks was made on operational grounds, and was initiated by the most influential member of the human learning school, Estes (1960). During the late 1950's, psychologists began to recognise the shortcomings of the perceptual-motor learning paradigms traditionally employed in this field. Such tasks do not readily admit precise control over the various components involved in training, so that it is difficult to relate aspects of the subject's performance to a set of well-defined stimuli in the training situation, and furthermore, satisfactory performance measures are difficult to obtain, as such tasks often allow the subject to "trade-off" accuracy in his performance against speed. Verbal learning paradigms in contrast permit exact control of certain specific events, such as the presentation of an item to be memorised, and a means of relating performance to these events. In particular, one can say whether or not a response is appropriate, or correct, given a particular well-defined cue or stimulus. Thus the adoption of verbal learning paradigms allowed psychologists to exercise far greater control over the various task components involved in training, and furthermore admitted more satisfactory methods of measuring performance. In particular, the experimenter could "pace" the subject at any desired rate of presentation of verbal items to be learned, so that "trade-off" effects were largely eliminated.

A number of new verbal learning paradigms made their appearance at about the same time, which tended to break down the barriers between verbal and memory, and finally led to a re-direction of effort in both areas. Traditional human learning paradigms, dating back to Ebbinghaus' day, had usually involved subjects in undergoing repeated practice sessions on a list of verbal items to be remembered until recall performance had reached and maintained for some time a pre-determined criterion level, such as some fixed number of consecutive errorless trials. In other words, practice trials were administered until such time as it was considered, on fairly arbitrary grounds, that the subject had "learned" the list of items. Such a procedure naturally led to a tendency to measure learning performance in terms of trials-to-criterion. Such a measure is clearly unsatisfactory. In the first place, it is unduly sensitive to relatively small variations in error patterns (for example, a subject learning a paired-associate list who gradually reduces overall errors from trial to trial would be equated by this measure to a subject who makes only one error on each trial, but takes the same number of trials to reach an errorless criterion). Secondly, such an approach does not readily lend itself to a detailed analysis of trial-to-trial changes in performance.

This point is nicely illustrated by a study by Tulving (1964), who employed a traditional paradigm in which free recall performance on a 22-word list was examined as a function of the number of repetitions of the list. In order to ensure that all items were likely to be learned at the same rate, the list order was re-randomized on each trial. Tulving isolated four response categories on each repetition.
On, say, the n'th trial, these were:-

(Cn-1,Cn) - recall of an item which was also recalled on the previous trial

(Nn-1,Cn) - recall of an item which was not recalled on the previous trial.

(Cn-1,Nn) - non-recall of an item which was recalled on the previous trial.

(Nn-1,Nn) - Non-recall of an item which was not recalled on the previous trial.

The number of items falling into each category was measured on each trial. It was found that only the intertrial retention category (Cn-1,Cn) actually showed an increase as a function of the number of trials, n, whilst the category (Nn-1,Nn) decreased sharply. There was evidence of short-term intertrial retention, provided by the category (Nn-1,Cn) which declined slightly across trials, coupled with the category (Cn-1,Nn) which remained almost constant.

These results suggest that performance on any given trial consists of a fairly long-term component provided by (Cn-1,Cn) which depends strongly on trials, coupled with a more short-term effect (Nn-1,Cn) which depends only slightly on trials. These two processes can be readily identified with long and short-term memory components, which are known to depend heavily on the number of other words on the list which intrude between the presentation of a particular word and its subsequent recall. However, in this traditional paradigm, list position was randomized and much valuable information relevant to the underlying memory processes was lost. Experiments of this kind made it abundently clear that verbal learning and memory are strongly interrelated, and that any investigation of verbal learning must have regard to known memory phenomena. Indeed, if a definition of verbal learning were attempted then it would have to equate the state of learning of an item with the subject's ability to recall it from memory. Although few psychologists would attempt such a definition, there is general agreement nowadays that the distinction between verbal learning and memory is largely artificial. An operational distinction has long been made which defines a memory experiement as one in which items to be remembered are presented only once, and a verbal learning experiement as a study in which such items are presented repeatedly. Such a distinction may have appeared valid when repetition was equated with a trials-to-criterion procedure, but it can only be confusing in the modern context, when in a single experiment, some items may be presented only once, whereas others may be presented many times.

In the light of the difficulties of defining learning, and it should be borne in mind that repetition is neither a necessary nor a sufficient condition to produce learning, it would perhaps be better to dispense with the term altogether, and instead to talk about the effect of repetition on memory. Such an argument would gain weight were one to consider the results of recent studies involving a number of the new memory paradigms intimated earlier, such as the Brown-Peterson and continuous paired-associate procedures. The picture that is now emerging suggests that verbal memory phenomena are strongly task-related, in that performance depends on strategies employed by subjects in specific task situations which govern the organisation and representation of material in memory. The idea that verbal learning, let alone human learning in general, can be characterized as a collection of phenomena which can be easily described with reference to a set of fundamental psychological laws, is becoming increasingly untenable.

Thus, in summary, it can be said that the study of topics traditionally associated with a human learning approach can be more fruitfully conducted within a memory framework. The subject matter of the field described as verbal learning may be more accurately characterised in terms of the effect of repetition on verbal memory, so that consequently, a summary of current theory and research in the field of memory will be necessary in order to provide a firm basis for the interpretation of material related to the SPI effect, which will then be discussed at a later stage.

1.2 Current Theoretical Issues in Memory

Although there are many interesting issues in the study of memory, the following stand out as being the most relevant to the present thesis.

1.21. Interference Theory

The direction taken by a great deal of contemporary research has been strongly influenced by classical interference theory, which is perhaps the earliest hypothesis concerned with forgetting, and which inherits many of the concepts of associativity postulated by Ebbinghaus (1885). Interference theory regards the learning of an association between a stimulus and a response as the basic unit of memory. Forgetting is taken to be a consequence of an original association being followed by a subsequent conflicting association.

Suppose the association A-B is learned, and subsequently the association A-C. Then is it postulated that A-C learning produces a competing association and at the same time a weakening of the original A-B association, termed retroactive inhibition. With the passing of time, however, the original A-B association is hypothesised to recover some of its original strength, a phenomenon claimed to be analogous to spontaneous recovery in classical conditioning, and so it becomes increasingly able to compete with A-C, which consequently tends to be forgotten. This effect is termed proactive inhibition.

1.22 Number of Memory Stores.

One of the major theoretical controversies in the field of memory has concerned the question of whether there are one or two memory stores, corresponding to short-term memory (STM) and long-term memory (LTM). Such a distinction was first made by James (1890), who defined the terms primary and secondary memory introspectively : an event in primary memory has never left conciousness and is part of the psychological present, whilst an event in secondary memory has been absent from conciousness and belongs to the psychological past. James postulated that primary memory would extend over a fixed, but limited, period of time.

A similar dichotomy was proposed by Hebb (1949), who based his arguments on the discovery of the physiologist Lorente de Nó (1938) of neurological fibres arranged in close, possibly self-exciting circuits. Hebb postulated that in LTM, permanent structures or traces would be formed, which could only be disrupted by interference from other long-term traces, whilst in STM, traces would be short-lived, as a result of their dependence on reverberating, self-exciting neural circuits which would be readily subject to decay.

Broadbent (1958) inferred a similar mechanism from an information-processing approach to memory, and based his inferences on behavioural data. His conception of the memory system involved three components : a sensory memory store which was capable of holding a considerable amount of information for a very short period of time, a limited-capacity STM system, in which the memory trace was assumed to decay rapidly but could be maintained by rehearsal (the process of repeating to oneself the items to be remembered), and a long-term store, in which forgetting was attributable to interference. Work by Sperling (1960) and by Averbach and Coriell (1961) demonstrated the existance of a very short-term visual store or "iconic" memory (with a decay time of less than a second) in which a fairly literal trace of the stimulus is held, whilst Neisser (1967) argued that work by Treisman (1964) demonstrated the existance of a similar sensory store in the auditory, system, with a decay time of the order of 2 seconds. A similar store was postulated by Crowder and Morton (1969). Thus, there is considerable corroboration for Broadbent's idea of a very short-tem sensory memory store.

Interference theory made no distinction between long-and short-term memory, so that opponents of the two-store hypothesis can generally be associated with an interference - theory position. For example in an extremely influential paper, Melton (1963) attacked the dichotomous view of memory, arguing that there had been little functional distinction made between STM and LTM, and that furthermore, short-term forgetting could be explained in terms of the principles of interference theory. Defenders of the two-store hypothesis were quick to point out that the work on which Melton had based his arguments involved experiments which did not separate LTM and STM components in performance; the STM component was superimposed on long-term recall, which could account for the supposed similarities between the two memory stores, Furthermore, a variety of evidence that appears to be an embarrassment to an associative interference-unitary memory position has accumulated in recent years.

Many memory tasks appear to have two components which can be readily identified with STM and LTM. For example, in a free-recall task, the subject is presented with a list of words which he must subsequently recall in any order he wishes. The probability of recall of an item depends on its position in the list. In particular, the last few words presented are usually recalled particularly well; this phenomenon is known as the "recency effect". Glanzer and Cunitz (1966) have shown that when recall is delayed briefly, and the delay interval is filled with some task such as counting, in order to prevent rehearsal, then the recency effect disappears, whilst memory for ealier items is comparatively unaffected. Postman and Phillips (1965), observed similar results but interpreted them as evidence for an increase in proactive inhibition at the end of learning , in other words, for the spontaneous recovery of earlier items in the list. Glanzer and Cunitz have pointed out that if this were true, the recall of relatively early items in the list should improve in proportion to the decline in recall of later items. No such improvement was found, however.

When recall of the list is to be made in the order in which it was presented, so that the last items in the list are to be recalled after a fairly long period filled with the recall of earlier items, it is found that recency effects are much diminished, whilst the recall of items earlier in the list is unaffected (Raffel, 1936; Deese, 1957; Murdock, 1963b; Tulving and Arbuckle, 1963 and 1966). Although the recency effect is so sensitive to delay, it appears otherwise to be very stable; this contrasts with performance on the rest of the list, which is affected by a whole range of variables that leave performance on the last few items unchanged, such as presentation rate, word frequency, the subject's age, and many other factors (Raymond, 1969).

Studies of the recall performance of amnesic patients has also produced results which a unitary hypothesis would find difficult to explain. Milner (1967) reported the case of a patient with hippocampal lesions who suffered from an inability to remember any new information for very long; as soon as his attention was distracted, the new material was lost, although he could recall incidents that had happened before the brain damage had occurred, and appeared quite normal on tests involving previously acquired knowledge. This evidence suggests the existence of a short-term memory system in which items can only be retained if attention is concentrated on them, and a separate long-term memory. The patient had apparently lost the ability to form new long-term traces, although retrieval of traces already in long-term memory was possible. These observations have since been confirmed in other cases.

Shallice and Warrington (1970) have reported a patient who showed unimpaired retention of events in everyday life and normal learning ability who was, however, unable to report back sequences of more than two digits, and in a free recall test showed a recency effect of only one item. Two-store theory would claim a normal LTM but a defective STM in this patient. It is clear that a unitary memory system would have trouble in explaining such cases.

Two store theory also gains support from studies which suggest that there is some limit on the storage capacity of short-term memory, as suggested by Broadbent, both in terms of the span of immediate memory (Miller 1956), or the size of the recency effect (Craik, 1971). When the subject's information processing capacity was reduced by the requirement to sort cards during the presentation of a list of words, it was found that performance at long retention intervals deteriorated, whilst that at short lags was unchanged (Baddeley, Scott, Drynan and Smith, 1969). This result clearly suggests a two-component memory system, with different storage or acquisition properties.

Some functional differences between long and short-term memory have been pointed out by Baddeley. It was shown that STM was adversely affected by phonemic similarity within a list of words presented for free recall, but that it was relatively unaffected by semantic similarity (Baddeley, 1966a). When performance improvements due to short-term remembering were eliminated by testing memory at only relatively long retention intervals, the rate of learning a ten-word sequence was unaffected by phonemic similarity, but was reduced by semantic similarity (Baddeley, 1966b). Kintsch and Buschke (1969) studied the same question using a serial probe task (Waugh and Norman, 1965), in which the subject attempts to remember a list of items, and is tested by being given one of the items (the "probe") as a recall cue ; his task is to supply the item which followed the probe in the original list. It was found that phonemic similarity affected performance only on the last few items of the list, or those that would be recalled from short-term memory, whilst semantic similarity affected only long-term memory.

These results strongly imply a two-store memory system, consisting of a limited-capacity, rapidly decaying short-term store which contains predominantly phonemic characteristics, and a large capacity, long-term store with a slower decay rate, which operates predominantly on the semantic aspects of verbal material. Of course, the argument regarding the number of memory stores does not end here. For example, Wickelgren (1970) has presented an argument for a third store, called intermediate-term memory (ITM) with a rate of decay faster than that of STM and slower than that of LTM. Both Young (1971) and Pollatsek (1969) have postulated intermediate "fluctuation" memory states, from which recall is imperfect. Furthermore, two-store or multi-store theories can be attacked on the grounds that to be valid, they must adequately define what is meant by a store, and moreover they must define what is stored. The supposed functional separation of phonemic and semantic properties of short- and long-term memory has recently been called into question by a considerable body of experimental evidence which suggests that both phonemic and semantic memories are potentially available at both short and long retention intervals (Shulman 1971). Such results can only be an embarrassment to multi-store theories, which can only find credence if a clear functional distinction can be made between the various stores. Nevertheless, the evidence in favour of distinguishing between long-and short-term memories is also very convincing, and it is hard to see how a single-store, unitary theory of the memory system could explain simultaneously all the results listed above. However, before attempting to resolve this question, a more detailed examination of long and short-term forgetting will be undertaken.

1.23 Short-term forgetting

It was stated earlier that many memory tasks contain two components which can be readily identified with STM and LTM. In particular, performance is often virtually perfect if recall is tested immediately after presentation, and it thereafter declines fairly rapidly with increasing retention intervals. However, this decline is not maintained. Beyond a retention interval of a few seconds, performance deteriorates only very slowly as the retention interval increases. In this section, interest is confined to the rapid forgetting which occurs over the initial part of the retention curve.

Broadbent (1958) claimed that all short-term forgetting was due to a spontaneous decay of the STM trace with time coupled with the limited nature of the capacity of STM, whilst unitary theorists such as Melton (1963) held that STM was subject to the same laws of interference as LTM. Recently, however, the issues have become less clear. Rejection of a unitary hypothesis, for example, would not necessarily imply acceptance of a simple trace-decay theory of short-term forgetting.

The serial probe technique of Waugh and Norman (1965) allows the experimenter precise control over the length of the retention interval. In the original study, subjects were presented with 16 digits at the rate of 1 or 4 digits per second. One of these was then repeated (the probe) and the subject was required to respond with the digit which followed the probe. A simple time-decay theory would predict much better retention of the more rapidly presented list. It was found, however, that the number of items between presentation and test was the most important determinant of recall probability. Shallice (1967) pointed out that the rapidly presented digits did show less marked forgetting than the slower items, and that, furthermore, a higher degree of initial learning would be expected with the slower rate, which would in turn reduce the apparent rate of forgetting. This study, then is in broad agreement with the hypothesis of a limited capacity STM and that displacement of earlier items by later ones is the main cause of forgetting. Shallice's observations can be accounted for if it is assumed that displaced items are not immediately lost, but decay rapidly over time.

Other experiments on the effects of presentation rate on forgetting from STM have produced conflicting results. Aaronson (1967) has reviewed many of these studies, and has pointed out that slower presentation rates often imply higher initial retention rates, whilst in many studies, STM and LTM components of recall performance are difficult to separate. This difficulty arises because on a two-store hypothesis, some of the items that are recalled from STM at short retention intervals are also held in LTM (these are precisely those items which can be recalled at long retention intervals), and it is therefore impossible to determine which items are held only in STM, and those which are held in both stores. Nevertheless, a recent extensive study by Glanzer, Gianutsos and Dubin (1969) employing a free recall paradigm showed that the displacement hypothesis was the most likely factor involved in eliminating the recency effect, with a small effect of decay over time.

However, such arguments do not establish the existence of a limited capacity short-term store. For example, short-term memory effects could to some extent result from rehearsal processes. Rehearsal is the term used to describe the process of sub-vocalising material to be remembered, and will be dealt with more comprehensively in a later action. However, several investigators have proposed that rehearsal might serve to maintain material in immediate memory (e.g. Waugh and Norman , 1965), and such a hypothesis is supported by the results of Brown-Peterson studies. For example, Peterson and Peterson (1959) required subjects to retain a sequence of three consonant letters. In order to prevent rehearsal, they had their subjects count backwards in threes from a randomly determined starting point and in time with a metronome, immediately following the presentation of the consonant trigram. Following this retention interval, subjects were required to recall the trigram.

It is clear that the retention interval was filled by a difficult task that would certainly preclude rehearsal of the trigram, but would not semantically interfere with it in the classical associationist interference theory sense. However, it was found that substantial forgetting occurred, following the usual pattern. Performance declined rapidly over retention intervals of a few seconds, and thereafter more slowly. As will be seen in Chapter Three, these results have been replicated many times. It is also interesting to note that performance is typically nearly perfect with no retention interval, and such a condition corresponds exactly to an immediate memory span situation. Miller (1956) has reported that subjects can typically recall sequences of up to 5 or even 9 verbal items (depending on the material) without error immediately after presentation, but that beyond this critical length, errors are made. This critical length is known as immediate memory span, shorter lists are designated as sub-span lists, and larger areas as super-span.

The rapid decay of material from memory which occurs in tasks which include the presence of some interfering (or rehearsal-preventing) activity between presentation and recall suggests that the span of immediate memory is limited by rehearsal capacity. There is some evidence that sub-vocal rehearsal is a sequential process, similar to vocal rehearsal, but somewhat faster (Landauer, 1962), so that the limited capacity of the rehearsal process can only be explained in terms of a rapid decay of unrehearsed items, which are not held in long-term memory. Thus in attempting to remember a sub-span list, the subject is seen as sub-vocally rehearsing the entire list in sequence, jumping back to the beginning each time he reaches the end. If the list is too long, then the unrehearsed early items would have decayed, leading to retention loss, or alternatively the subject might opt to rehearse a sub-span portion of the list to the detriment of the remaining items. In any case, it is clear that too long a list will cause disruption of such a cyclic or "rote" rehearsal process, as will an interfering task of the kind employed in the Peterson and Peterson study. It should be stressed that this argument has been advanced to explain only the rapid initial decline in performance that occurs when subjects' ability to rehearse a sub-span list is curtailed. For example, immediate memory span can be increased by using more meaningful material, so that performance on such a task cannot be explained entirely in terms of rote rehearsal, whilst much of the slower, long-term decay in Brown-Peterson studies may well be due to semantic interference from previously-presented material, as will be seen in Chapter Three. These reservations, however, do not invalidate the argument presented above.

However, another problem is posed: in order to account for the limited capacity of short-term memory in terms of rehearsal, and for the retention of some amount of rapidly-decaying information without rehearsal in Brown-Peterson studies, it was necessary to postulate that "unmemorised" items which are not currently being rehearsed are not immediately lost, but are retained subject to rapid decay. If this rapid decay were time-dependent, then it would appear that a separate short-term memory store must be proposed, whereas if this decay were dependent on interference of some kind from other items, then there would be little need to propose a separate short-term store. Landauer (1962) has reported that sub-vocal rehearsal takes place at a rate of about 3 syllables per second, although it may in rare cases be as rapid as 6 syllables per second. With a typical memory span of around 7 monosyllabic items, this would suggest a decay time for unrehearsed items of about 2 seconds. This figure agrees very well with Neisser's (1967) proposal for an acoustic sensory store, or "echoic" memory. The articulatory nature of sub-vocal rehearsal would clearly suggest that rehearsal might give rise to an echoic memory trace, even it items were presented visually. Such a store could well account for the short-term decay over time found by Glanzer et al (1969) and implied by Shallice (1967).

Wickelgren (1973) has argued that rapid short-term decay of unrehearsed material might well result from phonemic interference which would presumably reach far greater proportions than semantic interference from a given number of interfering items, and would even occur when attending semantically unrelated material as in Brown-Peterson studies. In other words, it is argued that short-term forgetting results from interference in a similar way to long-term forgetting, and that therefore, there is no structural difference between short and long-term memory, and hence no reason for making such a distinction. Such an argument, however, would have difficulty in explaining the slower rate of short-term forgetting found with faster presentation rates in a probe task by Shallice (1967), which strongly implies some amount of spontaneous decay over time. However, if Wickelgren's argument is accepted in conjunction with the limited-capacity rehearsal hypothesis and Neisser's sensory store, most short-term memory phenomena can be explained.

Perhaps the most outstanding evidence for two-store theory lies in the study of amnesic patients. It will be remembered that Milner's (1967) patient appeared to have normal STM and defective LTM; the above arguments, however, explain STM in terms of a sensory store and active rehearsal processes, neither of which is really part of the memory system as such. If the subject was merely unable to form new memory traces, only a slight inpairment of STM performance would ensue (due to the loss of phonemic information) which might well have been too small to detect. Shallice and Warrington's (1970) patient, who apparently suffered from a defective STM may either have lost the ability to rapidly subvocally rehearse material, or he may have had poor access to decaying sensory information. Although this argument is not so satisfactory as the two-store hypothesis in accounting for this patient, most of the other phenomena ascribed to short-term memory can be accounted for by rehearsal and sensory storage. Wickelgren's argument regarding the rapid decay and phonemic information by interference can explain the sensitivity of short-term retention to phonemic similarity, but this phenomenon can also be explained in terms of rehearsal errors and sensory memory decay. Indeed, Shulman's (1971) conclusion that phonemic memories are available at long retention intervals suggests that some phonemic information can be retained far longer than a separate short-term memory structure would suggest.

To conclude, it appears that evidence from short-term forgetting studies by no means establishes the existence of a separate short-term memory store, but that the phenomena observed can be explained in terms of an active rehearsal process coupled with an "echoic" sensory store, neither of which can be described as being fundamentally part of the memory system, and by the rapid decay of phonemic information in memory.

1.24 Long-term Forgetting

In studying short-term forgetting, there is always a problem in interpreting results since it cannot be determined which items recalled at a short retention lag would subsequently have decayed rapidly, and which would have been recalled even after a relatively long retention interval. Such difficulties do not apply when examining long-term forgetting, since the rapid decay items can be "wiped out" by ensuring that recall is made after a sufficiently long retention interval.

Of course, classical interference theory was originally advanced to account for long-term forgetting phenomena, but more recently, a number of observations have come to light which cast doubt upon its validity. It would firstly appear that interference theory would have difficulty in predicting forgetting in a situation where the material to be remembered consists of items such as nonsense syllables, which a subject is unlikely to have encountered in other situations. Underwood and Postman (1960) proposed that such material might include improbable letter combinations which would conflict with the subjects previously acquired language habits. An extension of this hypothesis, however, should predict that high-frequency words should be more prone to proactive interference from previously acquired language habits than low frequency words. Attempts to demonstrate faster forgetting rates for low frequency letter combinations or high frequency words have nevertheless proved unsuccessful (Keppel, 1968).

Another major difficulty for classical interference theory is raised by what Martin (1971) has called "the independent retrieval phenomenon". If associations A-B and A-C learned, interference theory claims that forgetting occurs as a result of mutual interference between the A-B and A-C associations, and so, when recall of both is required, a negative correlation in the recall probabilities of the two associations would be predicted. However, the recall probabilities of two such conflicting associations have been found to be independent across a wide range of experimental conditions (Greeno, 1969; Martin, 1971).

Results of this nature are extremely embarrassing to classical interference theory, and it is unlikely that the traditional stimulus-response associationist position will survive. Nevertheless, no-one would claim that all long-term forgetting is due only to the decay of the memory trace over time. There are several feasible hypothesis of long-term forgetting : the simple "overwriting" of memory traces, response competition in situations where only one response may be given to a recall cue, inadequate initial storage of information leading to subsequent confusion and competition, or even failure to retrieve the appropriate information even when it is adequately stored. Furthermore, performance can also be affected by the subject's organisation of the material to be remembered.

Early work on the rate of organisation in memory followed the classical associative tradition ; Bousfield (1953) showed that words belonging to certain categories tended to be clustered together in the free recall of randomized lists, and Jenkins and Russell (1952) demonstrated that pairs of words that tend to be highly associated (such as table - chair, bread - butter) also show clustering in free recall. Results of this kind take advantage of pre-existing associations, and are therefore consistent with the traditional associationist position.

More recent experiments, however, suggest that subjects can and do actively organise material in memory. Tulving (1962) defined subjective organisation as a tendency to recall groups of words in the same order on successive learning trials in a free-recall situation. He found that subjective organisation was a significant phenomenon, and that it increased on successive learning trials and was correlated positively with the amount recalled. Whereas Tulving employed an information-theoretic measure of subjective organisation, Bousfield, Puff and Cowan (1964) merely counted the number of words which were recalled in the same order on successive trials and based an index on this total. Nevertheless, they obtained essentially identical results with those of Tulving.

Further evidence that there is a causal relationship between subjective organisation and learning was produced by Tulving (1966). His first experiment involved the free-recall learning of a 22-word list. Half his subjects had previously read through the list 6 times, whilst control subjects had 6 readings of a completely unrelated list. There was no difference in the rate at which the two groups learned the critical list, which shows that rote repetition alone does not facilitate free-recall learning of well integrated items. In a second study, a list of 18 unrelated words was to be learned. Half the subjects had previously learned a 9-word list made up from items on the critical list, whilst control subjects had previously learned a totally unrelated 9-word list. Despite the initial advantage of the former group, it was found that after trial 7, the control group performed better. It would thus appear that the learning of irrelevant word clusters was actually deleterious to the learning of the critical list, where a different organisation would presumably be optimal. It is interesting to note that the experimental group would have produced more inter-word association, and should therefore have performed better according to associative theory.

These results, however, pose another problem. In assigning a currently-presented word to what is presumably an idiosyncratic, semantically determined word cluster, the subject must have access to the previously-presented words that form that cluster. In general, however, these words will have left consciousness in that they are unlikely to be currently undergoing rehearsal, so the question remains as to what form this access takes. For example, does the subject maintain some kind of functional semantic representation of word clusters in a conscious rehearsal loop and actively add some representation of the current word to the appropriate cluster description, or is a particular cluster retrieved as a result of some kind of recognition process triggered off by some property of the current word? When a word is added to a cluster, is the representation of the entire cluster in memory updated, or is the word merely given a representation in memory that is somehow similar to that of other words which are subsequently recalled as a group? Some attempt to answer such questions will be made in the next section.

It has recently become clear that a distinction must be made between learning and performance in memory. Performance only gives an indication of what can be retrieved at a particular time. In a study by Tulving and Pearlstone (1966) subjects were required to learn a list of 48 words, comprising 12 categories of 4 words each. Each category was presented as a group and preceded by its name. Cued subjects were given a list of the category names during recall, whilst uncued subjects were not. Although cued subjects recalled more words overall, it was found that uncued subjects recalled as many words per category; they merely missed some categories out. When uncued subjects were subsequently given the list of category names, further words were recalled, almost entirely from the omitted categories, and the overall performance of the uncued group became almost equal to that of the cued group. A similar phenomenon was reported by Tulving (1967), who conducted a free-recall experiment following which retention was tested on three successive trials. Although the number of words recalled on each of these trials remained roughly constant, only about one-half of the words recalled occurred on all three trials. It is clear that the subject's performance on each trial was somehow limited by his retrieval ability.

These results suggest that long-term forgetting is due to some extent to the difficulty of locating and retrieving information that has, in fact, been stored in memory. In other words, there is a lot more material "available" (actually stored) than is "accessible" (or able to be retrieved) at the time of recall. Furthermore, it has been shown that pre-existant or active subjective organisation of the material to be remembered influences the storage , and probably the retrieval, of that material. Several questions are posed by these results, For example, to what extent is forgetting due to the inaccessibility of information as opposed to its unavailability, and is unavailability caused by interference from other information in memory in a similar way to inaccessibility? How are these factors affected by organisation, and how exactly does organisation facilitate performance? Recent research has suggested that such questions may best be answered within the framework of encoding theory.

1.25 Encoding Theory

Psychologists have recently recognised that a distinction must be drawn between the nominal stimulus, that is the stimulus as the experimenter presents and defines it, and the functional stimulus, which is the form in which the stimulus is stored in memory. The act of transforming the physical or normal stimulus into a functional one is known as coding or encoding, and a functional stimulus is known as a code or an encoding.

Two types of encoding are distinguished (Baddeley and Patterson, 1971) and are known as reduction and elaboration coding. Reduction coding operates to reduce the amount of material the subject has to process. For example, it may take the form of selecting one from amongst many attributes of a presented stimulus item (e.g. the CVC nonsense syllable VJP might be encoded in terms of a phonemic representation of its initial letter V). A second form of reduction coding takes the form of "rewriting" several items into a single coding. A classic example of this is given by Miller (1956) who trained subjects to recode long sequences of binary digits (O's and I's) by splitting them into groups of three digits, each of which was then substituted by an octal digit (0-7). The subjects thus had only to remember a far shorter sequence of octal digits, which were decoded into binary triples during recall, and a far greater than normal immediate memory span for binary digits resulted. Richardson (1972) has produced convincing evidence that stimulus selection coding takes place in certain situations, whilst further support for the process of "rewriting" of items into a single code, or hierarchical coding has been produced by Johnson (1970, 1972). Both types of reduction coding are believed to be used when material is presented at a fast rate, and when the items to be recalled are not hard to discriminate from a large number of related items.

When items are difficult to discriminate, elaboration coding is seen to be useful, since this form of coding provides enough attributes of an item to be remembered to distinguish it from other related, but not-to-be-remembered items. For example, an item such as "apple" may be encoded in terms of the fact that it occurred in a list after the word "table", in terms of its sound attributes (e.g. it is disyllabic and starts with A) and in terms of meaning (e.g. it is an edible fruit). Items may not only be elaborated by being coded in terms of attributes that the subject extracts, other features may also be added to them. For example, a verbally presented item may be coded into a complex visual image (Paivio, 1969; Bower 1970).

Craik and Lockart (1972) have distinguished various depths of encoding. At the so-called surface level, the nominal stimulus is seen as entering some kind of sensory store, such as the iconic and echoic stores described in section 1.22, and it is thought that certain acoustic-articulatory-phonemic features of the stimulus may become part of the ensuing functional stimulus, although such features are probably prone to a great deal of interference and rapid decay. At a deeper level episodic attributes might be encoded (in other words attributes derived from the episode of presentation) such as whether the item was presented visually or auditorily, where is appeared in the list, how many times it appeared, and so on. Evidence that such information is often encoded has been reported by Hintzman (1970), Hintzman and Block (1971) and by Hintzman, Block and Inskeep (1972). At the deepest encoding levels, semantic attributes referring to the item's meaning resulting from general past experience and maturation are thought to be encoded. Craik and Lockart argue that the deeper forms of encoding are less prone to interference and therefore less subject to decay, but on the other hand more difficult to construct and decode (Elias and Perfetti, 1973; Wood 1972; Gardiner, 1974). The subject is seen as exercising some degree of control over the level of coding applied, and it is postulated that his choice of coding strategy will depend upon task variables.

Coding theory has been of considerable help in examining the question of accessibility and availability and evidence has recently come to light that suggests that in free-recall, organisation plays an important part. It will be remembered that in the Tulving and Pearlstone (1966) study (See 1.24) the superiority of subjects cued with category names was in terms only of the number of categories from which items were recalled, and not in terms of the number of items per category. This result suggests that although subjects had stored items by means of category codings, they could not provide themselves with these codings at recall. A study by Tulving and Psotka (1971) has shown that interference between two lists each of which are composed of a number of semantic categories is at the category, and not at the word level. These interference effects may be counteracted by producing category cues (Strand, 1971). Furthermore, Cohen (1966) has pointed out that if a category of items is recalled at all, then several of its member items are recalled; single items from a category are seldom remembered. It is possible, however, that if one item from a category is recalled, then the subject may be able to deduce the category and hence recall more of its items.

These results suggest that in the free-recall of categorized lists, subjects employ some kind of hierarchical encoding scheme, whereby individual items are encoded in terms of semantic categories. It furthermore appears that coding is less deep (and more prone to interference) at the superordinate or relational level than at the individual item level. However, such studies are rather artificial in nature, and it is not at all certain whether organisation in unrelated free-recall list learning is based upon semantic relationships, and indeed the method of detecting organisation in such situations by "clustering" would tend to favour the detection of episodic relationships. Herriot (1974) has argued that in the case of categorized lists, the relations between items are encoded very early on in the presentation trials, because the attributes by which items are related are very obvious. In the case of unrelated lists, however, it is argued that relations between items will only become apparent when the items have been coded by several attributes (increasing the likelihood that several items will share an attribute in common). Overlaps may well be multiple, in that different relations may become apparent for a single item. By this argument, organisation is seen as operating principally during retrieval, the recall of one word "throwing up" an attribute that is shared by another which hence acts as a recall cue. It is only with repeated practice that these overlaps become encoded as relations, and some form of active organisational encoding is initiated.

Forgetting in situations other than free-recall has often been explained in terms of the "encoding specificity" hypothesis (Thompson and Tulving, 1970) which maintains that the coding used for retrieval of an item has to be the same as that used for its storage. Their experiements, which vary the context in which the critical item appears during presentation and recall, show poorer recall when the item is probed in a different context to that in which it was presented. However, there are many ways that context may change from presentation to recall outside the context of the experimenter. For example, the subject may be daydreaming about different things at the two phases of the experiment. Nevertheless, the hypothesis is useful in providing an explanation for the kind of forgetting that appears in, say, Brown-Peterson studies, where the subject appears to experience difficulty in discriminating between the current to-be-remembered item and previous ones. It may well be that forgetting in this situation is largely due to the loss of episodic information that would presumably enable the subject to make such a discrimination.

1.26 Rehearsal.

An argument has already been advanced that proposes rehearsal as fulfilling the role of an active short-term memory store, operating principally on the acoustic representation of the nominal stimulus. This interpretation is rendered difficult, however, by results of Craik (1968) who showed that words could vary in length without having any effect on the recency effect, and Glanzer's (1972) finding that lists of proverbs show a recency effect. However, some evidence as to the active nature of short term memory is provided by studies which show that repetitions of the same item one after another do not have any effect on short-term capacity as measured by the recency effect (Glanzer and Meinzer, 1967; Waugh and Norman, 1968). This implies that repetitions are recognised for what they are, and are filtered out by an active selection process.

However, a considerable body of evidence now exists to suggest that rehearsal increases the likelihood that material will receive a deeper encoding. Howe (1967) presented two groups of subjects with a 9 consonent list. One group was specifically instructed to rehearse the items aloud in groups of three. When recall was tested after an interfering digit reading task, the recency performance of the control group was depressed, in agreement with the studies cited earlier (see section 1.22). However, for the rehearsal group, all items were slightly affected to the same extent, suggesting that they were all similarly encoded and thus were equally vulnerable to interference effects.

Bernbach (1967b) showed his subjects eight different colour cards which were placed face down in a row. A test card was then shown which the subject was required to match by selecting the appropriate face-down card. The performance of adults and young children was compared. Apart from the overall superiority of the adults, both performance curves showed a recency effect which spanned more items for adults, but adults also showed a primacy effect (superior recall of early items) which was completely absent for children. It was argued that these differences were due to the adults ability to rehearse the names of the colours, which were unknown to the children, and that the primacy effect was caused by a deeper encoding of the early items resulting from the greater amount of rehearsal they would receive in comparison with later items.

When the children were taught names for the colours and retested, it was found that although their performance was still inferior to that of adults, their performance curves now showed increased recency effects (indicating that they were rehearsing) and a pronounced primacy effect, indicating that rehearsal had facilitated deeper encoding of the early items in some way.

More direct evidence has been produced by Rundus (1971), who found that forcing subjects to rehearse aloud each item in a free recall list presented at a slow rate improved recall on the asymptotic, and not on the selection process.

However, a considerable body of evidence now exists to suggest that rehearsal increases the likelihood that material will receive a deeper encoding. Howe (1967) presented two groups of subjects with a 9 consonent list. One group was specifically instructed to rehearse the items aloud in groups of three. When recall was tested after an interfering digit reading task, the recency performance of the control group was depressed, in agreement with the studies cited earlier (see section 1.22). However, for the rehearsal group, all items were slightly affected to the same extent, suggesting that they were all similarly encoded and thus were equally vulnerable to interference effects.

Bernbach (1967b) showed his subjects eight different colour cards which were placed face down in a row. A test card was then shown which the subject was required to match by selecting the appropriate face-down card. The performance of adults and young children was compared. Apart from the overall superiority of the adults, both performance curves showed a recency effect which spanned more items for adults, but adults also showed a primacy effect (superior recall of early items) which was completely absent for children. It was argued that these differences were due to the adults ability to rehearse the names of the colours, which were unknown to the children, and that the primacy effect was caused by a deeper encoding of the early items resulting from the greater amount of rehearsal they would receive in comparison with later items.

When the children were taught names for the colours and retested, it was found that although their performance was still inferior to that of adults, their performance curves now showed increased recency effects (indicating that they were rehearsing) and a pronounced primacy effect, indicating that rehearsal had facilitated deeper encoding of the early items in some way.

More direct evidence has been produced by Rundus (1971), who found that forcing subjects to rehearse aloud each item in a free recall list presented at a slow rate improved recall on the asymptotic, and not on the recency part of the performance curve. Results of this nature suggest that at least in some situations, rehearsal may have the effect of holding items in attention whilst deeper encodings of those items are constructed. It would therefore appear to follow from this argument that any rehearsal process whose function is to transform the acoustic-articulatory nominal stimulus into a functional encoded form must be able to deal with functional items. In other words, the units of rehearsal may not be simply acoustic-articulatory syllables, but rather whatever functional units the subject is encoding the nominal stimulus into. It has been established that, for example, rehearsal of visually encoded material can occur (Hintzman and Summers, 1973; Shaffer and Shiffrin, 1972).

Thus, the fact that active rehearsal may operate upon functional stimuli provides an explanation for the results of Craik (1968) and Glanzer (1972) cited at the beginning of this section. The length of the nominal item in acoustic-articulatory terms may not be the prime determinant of rehearsal capacity, but rather the decay time of the functional representation of that item.

It may be necessary to distinguish between "passive" rehearsal, whose function is merely to recirculate and hold nominal stimuli, and "active" rehearsal whose function is to encode nominal items. Such a distinction may well depend on task variables and their effect on the subject's depth of coding strategy. Nevertheless, rehearsal and surface coding can still be advanced as an explanation of short-term forgetting effects, and it is still doubtful whether it is necessary to postulate a separate short-term storage structure. Nevertheless, Wickelgren (1973) has established convincingly that short-term forgetting does proceed at a faster rate than long-term forgetting, and for this reason it may be useful to retain the term STM as providing an operational, as apposed to a functional, description of the rapid-decay portion of retention curves.

1.27 Repetition and Practice

A number of hypotheses have been advanced to explain the improvement in performance that often occurs when items to be remembered are repeated or when the subject is given adequate time to rehearse them. In this sense, a rote rehearsal of the nominal stimulus could be regarded as forming a repeated presentation. Three main positions may be defined. The first of these claims that memory traces are either formed or they aren't ("all-or-none"), so that repetition would increase the probability that a trace was formed. A second approach postulates that memory traces can assume values on some kind of strength continuum, so that practice would have the effect of increasing the strength, and thus the resistance to interference, of the memory trace. Finally it could be postulated that a number of independently-decaying memory traces of a stimulus are formed ("multiple copy") and that this number increases with practice. There are any number of intermediate positions within this framework. For example, a multiple-copy-strength model could be proposed, or a two-store theory with, say, an all-or-none STM and multiple copy LTM.

A closely related issue is that of consolidation, which may be generally defined as the hypothesis that a memory trace has some chance to increase in strength or permanence each moment it is held in the memory system. A more specific version of this hypothesis may be traced to Hebb (1949), who postulated that if a short-term reverberating trace were not interfered with, but simply allowed to run its course, then it would consolidate into a more permanent long-term structural trace. This has given rise to the question of whether repetition and rehearsal of an item can lead to its consolidation into long-term memory, and to the subsequent controversy as to whether material can be rehearsed whilst the subject is ostensibly paying attention to the presentation of another item.

A different approach to the effect of a repetition is the differential encoding hypothesis, which claims that if the same nominal stimulus is presented to the subject on two or more different occasions, then that stimulus may be perceived and encoded in different ways on these occasions. These different encodings would presumably have an elaborative effect, in that they would produce more potential retrieval routes to the critical item, and might also operate to increase its discriminability. Both these processes would facilitate performance, Coding hypotheses differ from the mechanistic proposals listed above in that they ascribe a different function to rehearsal, in that with increased rehearsal time, there may still be some limit on the range and depth of attributes which will be encoded, whilst a repetition, especially in a new context, might well serve to increase this range.

Of course, it is beyond the scope of this brief review to resolve all the issues outlined above. Nevertheless, many of these issues will be encountered again in later chapters, and will be dealt with within the contexts in which they arise. The brief theoretical framework which has been established will serve as a sound basis for the interpretation and understanding of the experimental material to be presented later on in this thesis.

The chapter will now be concluded with a review of the paired-associate literature, which will fall into two parts, Firstly, after a brief outline of experimental techniques and paradigms, an examination of the factors affecting paired-associate forgetting will be made; in other words, of the field traditionally known as memory. Following this, the effects of various kinds of practice on paired-associate memory will be discussed, an area which may be roughly described as paired-associate learning.

1.3. Paired-Associate Memory

Paired-associate tasks have traditionally been regarded as the ideal method of investigating the formation of associative connections between pairs of items. However, in the light of the inadequacy of associative interference theory, less emphasis has been placed upon paired-associate memory in comparison with techniques such as the free recall, serial probe and Brown-Peterson paradigms. This is unfortunate for several reasons.

In the first place, it has been shown that performance in any memory task is influenced by task-specific, active organisational and encoding processes. Therefore there appears to be little justification for according greater importance to one experimental technique over another; each paradigm places its own unique demands upon the subject, and it is likely that he will react differently to different memory tasks. Concentration upon too narrow a range of memory tasks could very well lead to a confounding of subject strategy, with the result that active, organisational effects are mistaken for underlying memory mechanisms and structures.

However, there are more convincing arguments in favour of pursuing paired-associate studies. It has been shown that free-recall lists are actively encoded into related groups to some extent. However, it is extremely difficult to detect and identify such encodings - mere clustering in recall may well reflect shallow, episodic encodings as opposed to deep semantic grouping. Brown-Peterson studies typically employ stimulus triples, since even shorter sub-span stimuli produce very little forgetting. It is quite possible that such triples are associatively encoded in such a way that recall of one item of the triple may assist retreival of its other items. However, the effects of such encodings tend to be swamped by the difficulty experienced by the subject and producing adequate episodic traces that will enable him to discriminate between the current triple and previous ones. Serial probe techniques clearly require the subject to encode items in terms of episodic, serial relationships.

Thus, many memory tasks probably involve the subject in encoding relations between successive items, but the lack of specific controls makes it nearly impossible to detect these encoding aspects. In free-recall and Brown-Peterson studies, the subject's main task is recall, so that there is no guarantee that a relational encoding will be of any use. Studies with categorised free-recall lists suggest that entire related groupings may be omitted in recall if the relation itself cannot be recalled. There is no way of determining to what extent active relational encoding strategies will be employed in such situations, and to what extent relational encodings merely result from the encodings of overlapping attributes. However such relational encodings occur, it is fairly certain that they do occur, and they may therefore have a profound effect upon performance.

Serial probe techniques also present the subject with a formidable task. Not only must each list item be sufficiently well encoded as to permit its recall, but the subject is also faced with the task of producing a sequential relational encoding between successive item pairs. Forgetting in such a task could result from inadequate recognition of the probe item, inability to retrieve the appropriate response, or an inability to determine which response is appropriate.

Paired-associate techniques therefore provide a valuable tool for the investigation of relational encoding, which must occur to a lesser or greater extent in all these other situations. They compare favourably with the serial probe technique in that in a paired-associate task, the subject is only required to produce relational encodings between specific and well defined item pairs, and not between each successive item pair. Furthermore in a probe task, each list item is equally likely to be the required response, so all items must be encoded for recall, whilst in a paired-associate task, the subject knows exactly which item of each pair to encode for recall, and which one to encode for recognition. This certainly reduces his information-processing load, since it is generally accepted that recognition is far easier than recall (e.g. Kintsch 1970a).

Paired-associate techniques also permit the experimenter to independently vary stimulus and response material, and therefore alter the encoding requirements of "probe" and "target" items, whereas similar material must be employed throughout in other paradigms. Paired-associate methods therefore admit a very precise definition of the encoding requirements demanded of subjects, and could be reasonably expected to provide an admirable research tool for the investigation of the various theories of long-term forgetting such as trace overwriting and response competition which is, after all, the main object of memory research.

1.31. Paired-associate paradigms

The earliest paired-associate paradigms typically involve the learning of a repeated list of stimulus response pairs, usually to some pre-determined criterion level of performance. The subject is instructed that the first, or stimulus item of each pair will serve as a recall cue for the second, or response, item. In the study-test method, the list of pairs is presented for study, one pair at a time, following which the stimulus items alone are presented one at a time and the subject attempts to respond with the appropriate response item. The experiment consists of an alternating sequence of a block of presentation or study trials followed by a block of test trials, and usually the serial position of items within a block is randomised from each block to the next.

In the anticipation method each pair receives one anticipation trial (comprising a test trial followed by an immediate presentation of the same pair) on each presentation of the list. Thus, an anticipation trial on the pair GREEN-MAN would take form: GREEN -, GREEN-MAN. The subject attempts to respond with the appropriate word "MAN" when the stimulus word "GREEN" is presented alone, and is then immediately presented with the correct pairing. Both the study-test and anticipation learning paradigms suffer from the drawback that the list order is re-randomised from one trial to another, and although care is taken to ensure that short-term retention effects are "wiped out" by providing adequate numbers of trials on other items between the presentation and test of each item, the experimenter has no other effective control over the retention intervals between study and test.

The relatively new continuous paired-associate (CPA) paradigm has provided a more flexible tool for the investigation of paired-associate retention. In this paradigm, the number of trials on other pairs intruding between successive trials on each particular pair is determined in advance by the experimenter. This is achieved by "interleaving" pairs to form a list in such a way that the interval between successive trials on each particular pair consists of study and test trials of other pairs. When all the pairs of interst, or "critical pairs" have been assigned to list positions, any remaining vacant list positions have unanalysed "filler" pairs assigned to them. These fillers are constructed in such a way as to be indistinguishable from the critical pairs, so that they will receive as much of the subject's attention as the critical pairs.

The main features of the interleaving process may be depicted diagramatically as follows:-

1) Pair A receives two trials, A_1 and A_2, with four intruding trials between them:-

$A_1 \ldots . A_2 ..$

ii) Pair B receives two trials with five intruding trials between them:-

$A_1 B_1 \ldots A_2 . B_2$

iii) Pair C receives two trials with two intruding trials between them:-

$A_1 B_1 . C_1 . A_2 C_2 B_2$

iv) Finally, the two vacant list positions are occupied by presentations of the filler pairs Y and Z:-

$A_1 B_1 X C_1 Y A_2 C_2 B_2$

Each list position may be occupied by either a study trial or a test trial, or alternatively, anticipation trials may be employed throughout.

One drawback to the CPA paradigm is that in general, the experimenter has little control over the specific sequence of trials that fill the intertrial intervals of particular items. In the diagram above, for example, the first trial on each pair may constitute a study trial, and the second a test trial. Pair A is therefore tested at a retention interval of four trials each of which is a study trial, whereas pair C is tested at a retention interval of two trials, one of which is a study trial, and one of which is a test trial. There is no guarantee that an intruding test trial will have the same effect on retention as an intruding strudy trial, so that even if each retention interval is tested many times with different pairs in the hope that these different effects will "average out" between retention intervals, an element of uncontrolled variation is introduced into the experiment. Difficulties of this kind arise even if anticipation trials are employed throughout; although this method guarantees equal numbers of intruding tests and presentations at each retention interval, there is still the possibility that say, an intruding short-lag test which results in a correct response has a different effect on the retention of the critical item than an intruding long-lag test that results in an error.

This problem is avoided by the paired-associate (or PA) "probe" task in which a list of pairs is presented for study, one pair at a time at a fixed rate, following which the stimulus item of one of the pairs is presented as a recall cue. Thus only one pair from the list is tested. Since each pair in the list is equally likely to be the one tested, and since the retention interval comprises only intruding presentation trials, this technique does not suffer from heterogeneity in the effect of the retention interval. By varying both list length and the position of the critical (or subsequently probed) pair the experimenter can control both the retention interval in terms of the number of subsequent pairs, and furthermore the number of previously-presented pairs, so that the PA probe procedure is a useful method of examining the effects of proactive interference. However, this procedure is clearly far more expensive in terms of time and material than the CPA method.

Within the framework of these methods, there are a number of variables pertaining to the stimulus and response material which lend themselves to experimental control. Thus, for example, the response items may be totally unfamiliar to the subject (e.g. nonsense syllables), extremely familiar and from a finite, well-defined set (e.g. the integers 1,2,3,4 and 5), or familiar, but from a poorly-defined, potentially very large set (e.g. five-letter, common nouns). In the first instance, the subject would be expected to experience a great deal of difficulty in recalling responses since the task would involve learning all the responses "from scratch", whilst in the second case, very little response learning would be required, and in the third example, the subject's task would involve the learning of which items from a familiar set actually belong to the response pool. The difficulty of encoding the stimuli may be similarly manipulated.

This thesis is primarily concerned with the relational encoding of the stimulus-response pair, and it has been argued that this is the major object of paired-associate studies. Failure to produce the appropriate response to a stimulus may result in three ways; failure to correctly identify the stimulus, failure to associate the appropriate response with the stimulus, and failure to retrieve the response. The third possibility can largely be eliminated by employing well-defined "compatible" response sets, such as a range of integers, thereby removing response learning components from the task. There seems little justification for increasing the complexity of a task when there are aspects of performance on the most simple form of the task which are not fully understood. Unfortunately, material involving response learning cannot be excluded from this review, since there are many important studies of this kind which have not been replicated with the response learning component of the task eliminated.

1.32 Short-term retention of paired associates.

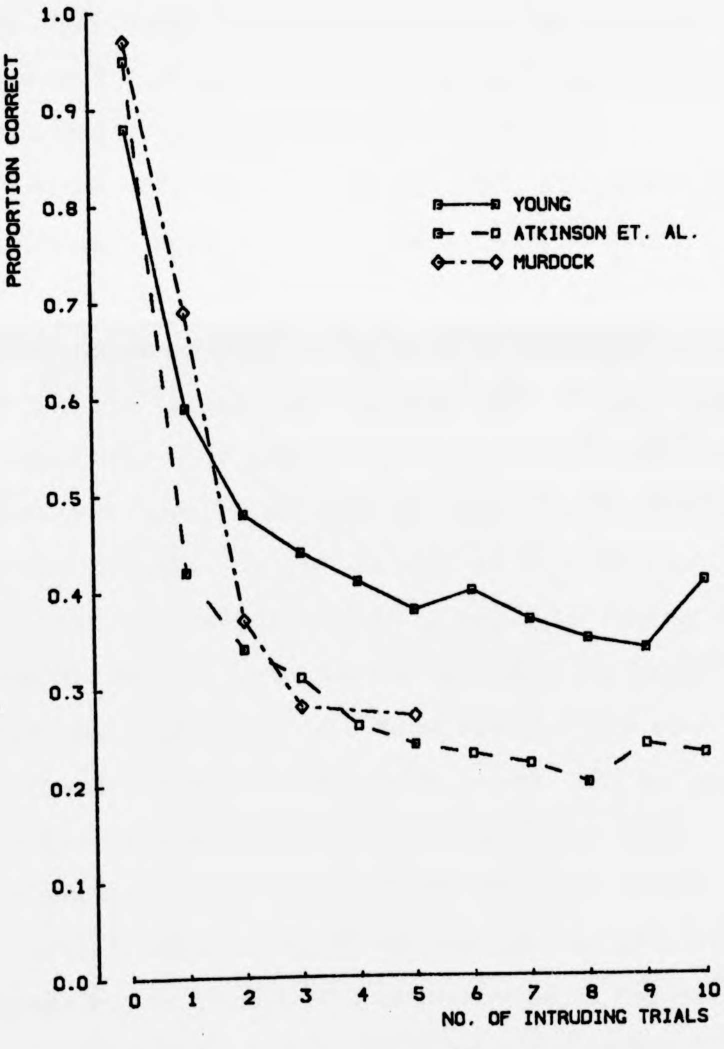
The most marked aspect of short-term retention of paired associates is that a substantial amount of forgetting occurs over quite short lags (in terms of the number of intruding trials) between the presentation of a pair and its subsequent test. Performance tends to decline rapidly up to a lag of about 3 intruding trials, and thereafter far more slowly. This type of relationship has been found to hold over a wide range of stimulus and response material, and across a variety of presentation rates, and is typical of both CPA and probe procedures.

The three retention curves displayed in Figure 1 are fairly typical, but despite their obvious similarities, they result from widely differing experimental procedures. The data collected by Young (1966) emerged from a complex CPA study in which stimuli were consonant trigrams (CCC'S) and responses were the digits 0-9. A randomized interleaving technique was employed to prepare a separate list for each subject. Dummy pairs were used to provide a "primacy buffer" (in other words, dummies began each list to eliminate primacy effects from performance on critical items), and such pairs were also employed to fill vacant list positions. Study trials were of 4 seconds' duration, whereas that of test trials was subject-determined (i.e. trials were terminated only when the subject had made his response). The data displayed in Fig. 1. is that for critical items which were presented once, and tested after a retention lag of 0 -10 intruding trials.

A CPA procedure was also employed by Atkinson, Brelsford and Shiffrin (1967), in two studies employing two-digit numbers as stimuli, and the 26 letters of the alphabet as responses. Anticipation trials were employed throughout; the test phase of such trials lasted for 3 seconds, followed by a 2-second blank interval, a 3-second study phase on a new item, and a further 3-second blank period preceded the onset of the next trial. It is clear that this procedure allowed the subject ample time to rehearse previously-presented pairs during the 11-second anticipation trial on the two current pairs. The data in Figure 1 emerged from the investigators' Exp.II, in which each critical pair was presented once, and subsequently tested after lags ranging from 0 to 20 intruding anticipation trials; only data for lags 0 to 10 are shown, as performance over longer lags declined only slightly, and followed the general trend of the portion of the curve displayed. Subjects were instructed that each pair would only receive one presentation and a subsequent test, and that therefore any pair just tested could safely be forgotten. Despite this, no difference in performance was found between three experimental conditions in which the number of different pairs that a subject would be required to remember at any point in time was 4,6 or 8 and the results displayed are averaged over these three conditions.

FIGURE 1

Some typical paired-associate retention curves.



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
6
7
8
9
10
NO. OF INTRUDING TRIALS
YOUNG
ATKINSON ET. AL.
MURDOCK

A paired-associate probe procedure was employed by Murdock (1963a, Exp.I) with stimuli and responses consisting of common English words presented at a 2-second rate. Critical items were preceded during presentation by 0,1,2 or 3 other pairs, and were tested at lags of 0,1,2,3 or 5 subsequently presented pairs. Subjects were allowed 15 seconds for recall. The retention data displayed in Figure 1 were obtained by averaging proportions correctly recalled across the number of prior pairs at each retention lag. Similar results were obtained in a further study (Murdock, 1963a Exp.II).

Peterson, Saltzman, Hillner and Land (1962, Exp. I) employed a CPA procedure in which stimuli were common 3 - and 4 - letter, monosyllabic English words, and responses were the digits 1-9. Each critical item was presented once for study and subsequently tested after 0,1,3 or 8 intervening trials on other items. An interleaving technique similar to that employed by Young (1966) was used in the preparation of lists, but otherwise this experiment differed from Young's in that the duration of both study and test trials was 2 seconds. The important difference here lies not so much in the slightly faster presentation rate, but in the fact that the subject was forced to respond in 2 seconds. Nevertheless, the obtained retention curve appeared not too different to those discussed earlier. Greeno (1967) has pointed out that similar exemplars could be extracted from two further studies by the same authors (1962 Expts II and III) and from a study by Peterson and Brewer (1963, ExpIII) which differed from the current experiment only in the number of alternative numerical responses; by correcting all data for guessing, Greeno was able to construct a retention curve over a wider range of study-test lags which is essentially identical to Young's data in Figure 1.

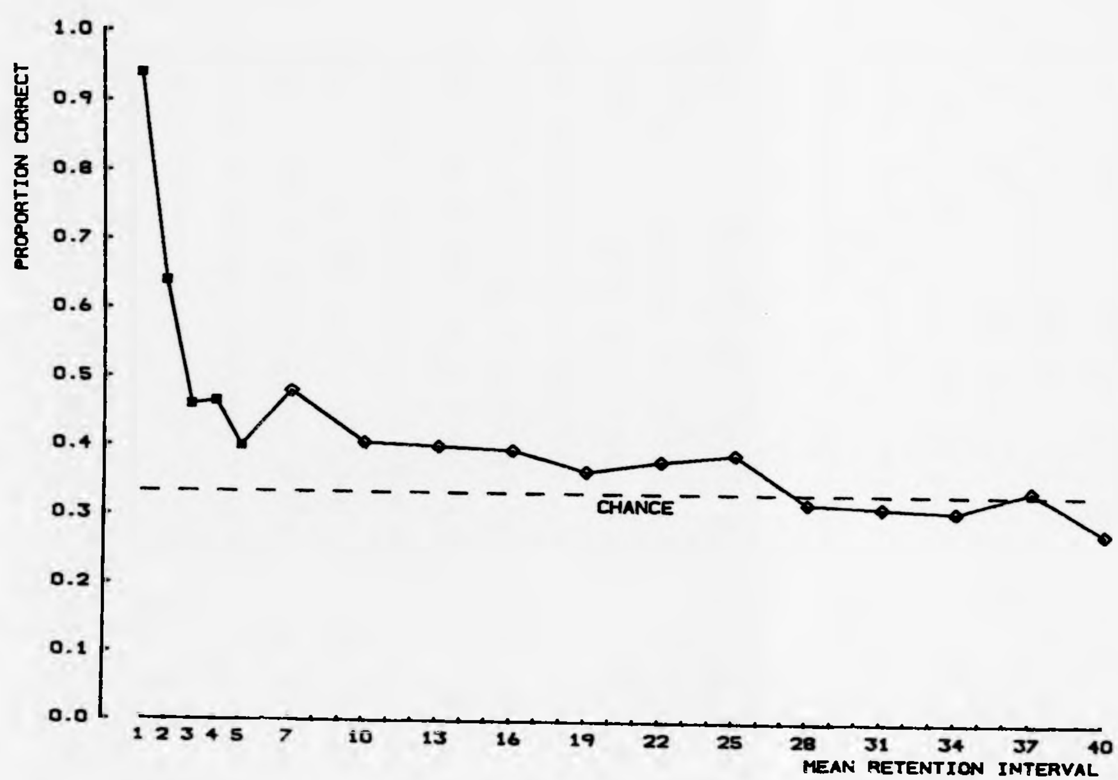
It would appear at this stage that the effect of the retention interval on paired-associates memory can be easily interpreted in terms of a dichotomous STM-LTM view of memory. The relatively rapid decline in performance from lags 0 to 3 could be identified with a rapidly decaying STM component, and the more gradual subsequent decline with LTM. A CPA study by Bjork (1966) has provided more direct evidence on the nature of short-term forgetting. Sequences of anticipation trials were constructed in such a way that all intertrial lags from 1 to 40 intruding trials were equally likely to occur. Although all the pairs in this study were presented at least 12 times, Bjork argued that if long-term forgetting were relatively slight, then retention due to STM alone could be measured as a function of lag if attention were restricted to those items on which a subsequent error was made, since any item on which an error is subsequently made is very unlikely to be currently held in LTM. The data are reproduced in Figure 2.

Although the greater proportion of forgetting appears to occur between lags 0 and 3, performance shows a subsequent gradual decline to guessing at around lag 20. This would imply a small amount of retention due to STM far beyond retention lags which a forgetting-by-displacement hypothesis would predict. Of course, there is the possibility that some of the analysed items were held on LTM and subsequently forgotten. Furthermore, it must be borne in mind that much of the data results from items that had been presented several times; there is some convincing evidence, mainly from Brown-Peterson experiments, that the repetition of material to be remembered markedly retards the rate of short-term forgetting. These studies will be discussed in Chapter Three.

Bjork's results can be more easily understood in the light of the results of a study-test CPA experiment by Peterson, Saltzman, Hillner and Land (1962, Exp.II). Stimuli were consonant-vowel-consonant trigrams (CVC's) of 99 - 100% Archer (1960) meaningfulness, and responses were the digits 1 to 15. Both study and test trials were of 2 seconds' duration. Critical pairs received one presentation, and were tested at a lag of i = 0,2 or 4 intruding trials, and were then subsequently tested again at a lag of j = 2 or 4 trials after the first test. A typical sequence may be represented diagramatically as follows:-

FIGURE 2.

Retention of paired associates as a function of lag, prior to the trial of the last error (Bjork, 1966)



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
CHANCE
1 2 3 4 5 7 10 13 16 19 22 25 28 31 34 37 40
MEAN RETENTION INTERVAL

$$\underset{\text{lag } i}{P \ldots\ldots T_1} \underset{\text{lag } j}{\ldots\ldots T_2}$$

The results of this study are tabulated below, where C_k means "correct on T_k" and W_k means "wrong on T_k".

		$P(C_2)$		$P(C_2/C_1)$		$P(C_2/W_1)$	
	$P(C_1)$	$j = 2$	4	$j = 2$	4	$j = 2$	4
$i = 0$	.99	.48	.30	.49	.30	x	x
2	.37	.34	.27	.6[illegible]	.46	.14	.29
4	.30	.24	.25	.71	.52	.06	.10

It is clear from this data that there is a certain amount of reminiscence from T_1 to T_2; in other words, some items that were wrong on T_1 were correct on T_2. In all but one case (when $i = 4$ and $j = 2$) the observed proportion $P(C_2/W_1)$ was significantly superior to guessing, on a 2 - tailed Z test. It appears in general that especially at short lag, certain items are difficult to recall, and that furthermore, this difficulty dissipates over a subsequent retention interval, at a faster rate than that at which forgetting occurs.

These results suggest that although some items are accessible, or stored in memory, they are not always available for recall (see 1.24). This phenomenon could be explained in terms of the encoding specificity hypothesis (Thompson and Tulving, 1970; see 1.25) whereby the code used for retrieval must be the same as that used in storage. More specifically it is postulated that certain episodic attributes included in the original coding are not always available at the time of testing, possibly because the test episode itself posesses different attributes to the encoding episode. Consequently, the subject may initiate his response search on the basis of functional aspects of the stimulus which may not correspond to those functional aspects used in producing the relational stimulus-response encoding at the time of presentation. On a subsequent test, however, episodic features may be available which correspond closely to those employed during original encoding, leading to correct recall, and thus the reminiscence effect described by Peterson et al above.

In Bjork's (1966) study, of course, many of those features that would aid recall at test might well occur during the presentations of other pairs, leading to inter-pair interference and forgetting. Nevertheless, and episodic encoding hypothesis could well account for the slow decline to chance of performance on items on which a subsequent error was made. In other words, it is claimed that a proportion of such items were encoded in a fashion that relied substantially upon episodic information, although it is clear that the greater proportion of them may well have progressed no further than a short-term "passive" acoustic rehearsal loop (see 1.26). The latter items would account for the rapid decline in performance from lags 1 to 3, and the former for the relatively slow subsequent decline to chance at around lag 20. Clearly, Bjork's data are more satisfactorily explained in terms of functional encoding theory than by a simple LTM - STM dichotomy.

1.33 Prior Activity

The so-called primacy effect in free-recall illustrates very nicely the effect of prior activity; items in intermediate list positions are less well recalled than items at the beginning of the list. In other words, prior activity tends to inhibit performance on later items. This phenomenon is termed proactive interference (P.I), although it must be borne in mind that this terminology does not imply acceptance of traditional interference theory; it is just a label for describing a widely observed phenomenon. Of course, the free-recall paradigm is not ideally suited for studying the effects of prior activity, since the retention lag is not under the experimenter's control.

The paired-associate probe paradigm has the advantage that the experimenter can control both the amount of prior activity and furthermore the retention interval. A number of studies by Murdock (1963a Exps. I and II; 1963c. Exp.I) employed a PA probe procedure with pairs consisting of common English words. Lists of varying lengths were

FIGURE 3

The effect of prior activity on the retention of paired associates (Murdock, 1963a)

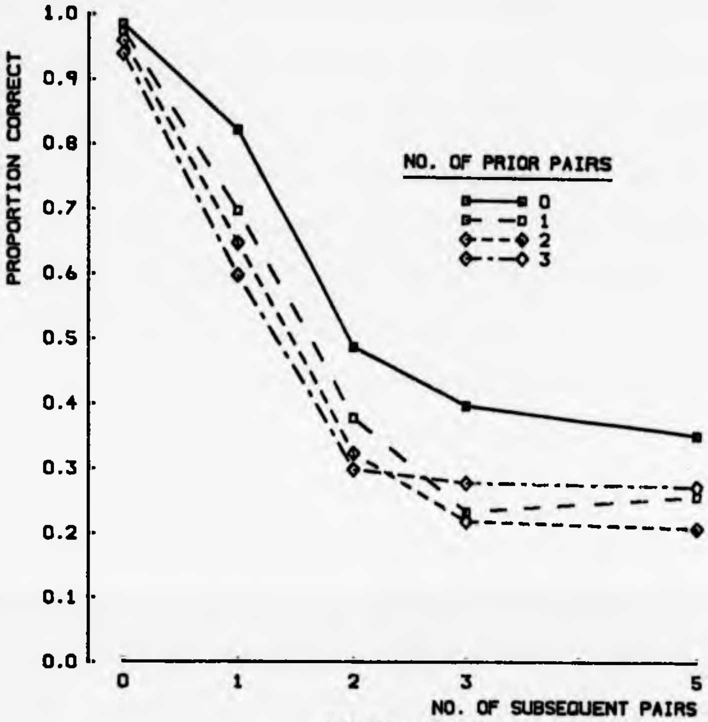


FIG.3A - EXP.I.

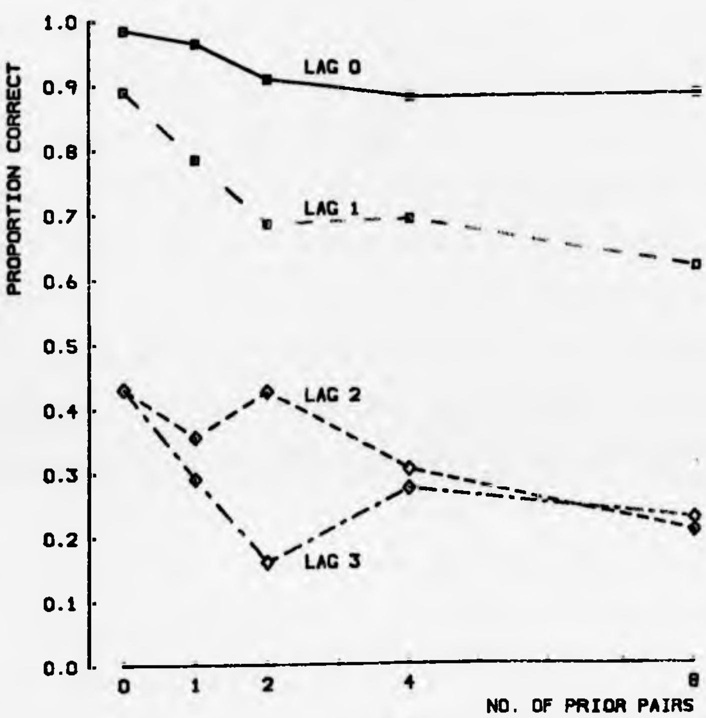


FIG.3B - EXP.II.

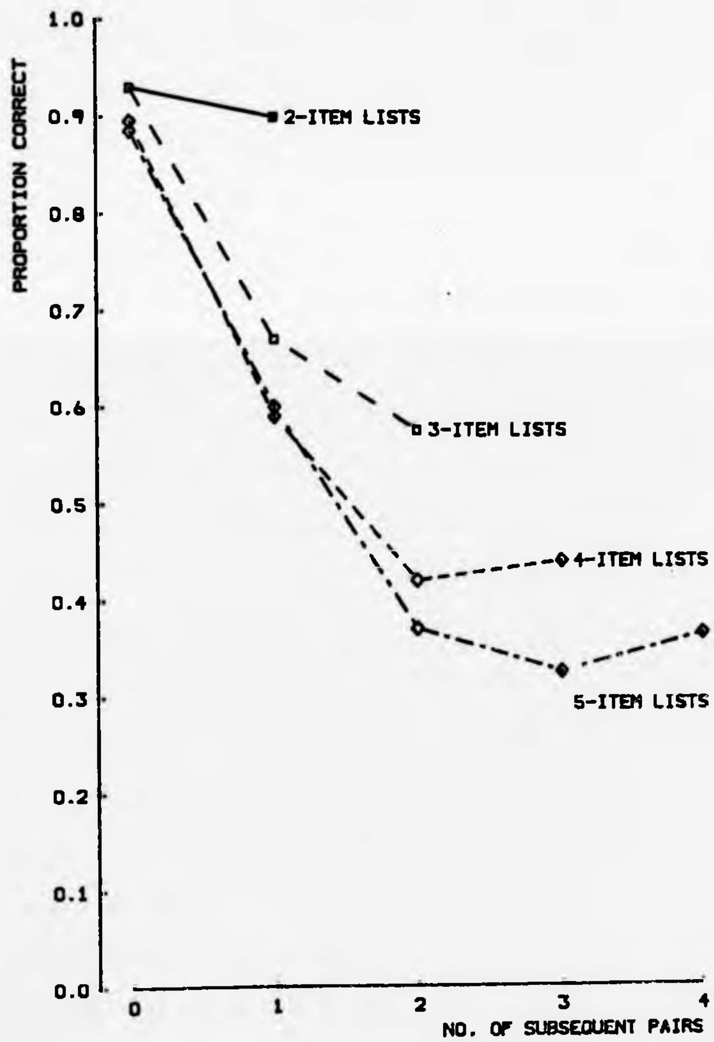
presented at a 2-second rate, with each subject being tested many times in each of a number of conditions defined by the number of pairs presented prior to and after the critical item. In this way the effects of varying amounts of prior activity upon retention at a given lag were examined. Although results in all three studies were very similar, there are several ways of presenting them; Figure 3 shows the data for Murdock's (1963a) Exps I and II as a function of retention lag and the number of prior pairs respectively. It is clear from the figure that prior activity appears to have a sharply deleterious effect up to and around two prior pairs, and thereafter, the effect of prior activity increases only very slightly.

One problem with studies of this kind lies in the fact that retention lag and prior activity are confounded with list length. On closer examination, the PA probe procedure is very similar to the serial probe paradigm of Waugh and Norman (1965); it can, indeed be thought of as a more simple version, since the subject knows in advance which items to learn (the responses), and which ones he merely needs to recognize (the stimuli). In turn, the serial probe procedure is a simpler version of the serial recall paradigm, and one would expect nearly perfect serial recall of lists of around 1 to 5 words (Miller, 1956). This probably results from the subject's ability to rehearse all the items in a very short (or subspan) list. Thus, one would expect a fairly sharp drop in performance as the paired-associate list length increases beyond two or three items, or about four to six individual words.

Retention data from Murdock's (1963c) Exp. I are presented in Figure 4 as a function of list length. It can be seen that, in general, for a given retention lag, performance declines with list length, or in other words with the number of prior pairs. This is especially true if items tested at a retention lag of zero are discounted on the grounds that performance on such items will be enhanced by short-term retention effects. There is indeed a very sharp drop in performance on items tested at lag 1 from a two - to a three - item list, but it is also clear from the figure that at a given retention lag, the largest

FIGURE 4

Retention of paired associates as a function of list length.
(Murdock 1963c, Exp.I)


PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
NO. OF SUBSEQUENT PAIRS
2-ITEM LISTS
3-ITEM LISTS
4-ITEM LISTS
5-ITEM LISTS

discrepancy in performance occurs between items with no prior pairs, and items with 1 prior pair.

These observations appear to suggest that the deleterious effect of PI is primarily due not so much to some kind of trace competition or interference from early items that depresses performance on later ones, as to an enhancement of performance on early items. It can be argued that if the subject were to adopt a primarily passive rote rehearsal strategy of the kind that would be optimal in, say, a serial memory span task, than on the majority of Murdock's lists (which are super-span) the earlier list items would by cycled through the rehearsal loop more often than later ones, due to the limited capacity of the loop (see 1.26). Indeed, the extra processing time available for early items due to the rehearsal loop not being full might well be used in "active" functional rehearsal, leading to deeper encoding, thereby leading to a marked primacy effect. However, there are problems with such an interpretation. An examination of Figures 3 and 4 will confirm that performance on supposedly sub-span lists of two pairs (i.e. four words) was not perfect, although the probe PA task should if anything be easier than a memory span task in which all four words must be recalled in their correct serial order, and on which subjects could be expected to perform perfectly (Miller 1956). Furthermore, the curves presented in the figures show a very small primacy effect compared with, say, typical free-recall curves.

These results can be explained if it is borne in mind that Murdock's subjects were each tested on a large number of lists of varying length the majority of which were beyond their immediate memory span. It is postulated that subjects would quickly give up an initial passive acoustic rehearsal strategy after encountering a super-span list, and instead adopt a strategy of, say, activity rehearsing each item as it occurred in order to produce a deeper encoding. Such a strategy would tend to improve overall performance on long lists, but might well yield less than perfect performance on short lists due to the greater difficulty of producing these deeper encodings, and would certainly reduce the primacy effect. Alternatively it could be argued that the subject's initial strategy would predominantly involve encoding at the episodic level; primacy in this case would result from the enhanced distinctiveness and resistance to interference of the episodic cues available for the first few items on the list. The strategy-change hypothesis would then claim that because of the rapid build-up of episodic interference in superspan lists, the subject would tend to opt for a semantically-based encoding strategy, which would likewise improve overall performance due to the added resistance of such encodings to interference, and result in a reduction of the primacy effect, and generally imperfect performance on very short lists due to encoding difficulties.

The strategy-change hypothesis is supported by a number of studies by Murdock involving 6-pair PA lists throughout. When subjects were tested repeatedly on such lists (1963a, Exps. III and IV; 1963b Exps. I, II and III; 1963c, Exps II and III) serial position curved displayed only a very small primacy effect, and were comparable to that displayed in Figure 4 for the 5-pair list. However, when naive subjects were tested on only one 6pair list, the resultant serial position curve displayed a far more marked primacy effect (Murdock, 1963a, Exp. IV). An enormous primacy effect was found by Tulving and Arbuckle (1963) using a 10-pair PA probe list with digit-nonsense syllable pairs (which would be difficult to encode semantically, and might therefore make subjects even more prone to employ a rote rehearsal strategy); again, subjects were tested on only one list.

Furthermore in a similar word-pair PA probe study in which PI between lists was examined (Murdock 1964), it was found that when subjects were tested on six 6-pair lists, there was a marked primacy effect on the first two lists so tested, but on subsequent lists this was found to disappear. Indeed, on the fifth and sixth lists, no items at all were correctly responded to in the first list position, and performance was found to improve monotomically with list position (i.e. as the number of subsequent pairs declined). No evidence of a decline in overall performance was found across lists; if anything, overall performance was found to improve with lists, although this improvement was maintained only up to list 4, and thereafter overall performance fell back to a level comparable with that on list 1.

There was little evidence to suggest that this subsequent decline in performance was in any way due to long-term interference or confusion with items from previous lists. The proportion of extra-list intrusion errors made on the last two lists was no higher than that on earlier lists, and there was no significant difference in the probabilities of an extralist intrusion given an error between the first four lists and the last two (Z = .396).

These results strongly suggest that there is little if any interaction of long-term memory traces between lists (incidently, another nail in the coffin of traditional interference theory), and so it can be deduced that the decline in performance on the last two lists was probably due to fatigue, or to a loss of motivation as the experiment proceeded. Furthermore, the relatively low level of extra-list intrusion errors (as compared with intrusions from items within the same list) was found by Murdock in all his studies involving the repeated testing of subjects on many lists. The fall-off in primacy coupled with an improvement in performance found across the first four lists strongly suggests a practice effect caused by a stategy change of the sort suggested - a move away from a shallow encoding strategy that would favour items in short lists or in early positions in longer ones, to a more balanced, deeper encoding strategy that would improve overall performance on superspan lists.

To return to the PI effects within lists, a re-examination of Figures 3 and 4 at this stage would suggest that with practiced subjects the effect of intra-list PI reaches an asymptote after about two prior pairs, and this deleterious effect occurs only when performance is measured at non-zero retention lags : in other words, it is a long-term rather than a short-term memory effect. There is some evidence that the deleterious effect is only apparent, in that it may result from a strategy that favours early list items being employed by subjects on their first two or three lists. However, it is unfortunate that Murdock's lists are in general so short as to limit the retention data for large numbers of prior pairs to critical items tested at relatively short lag. The fact that intralist intrusion errors were found to occur with a consistently higher frequency than extralist intrusions would suggest that some kind of long-term trace competition within lists was taking place, although it could equally well be argued that the subject could somehow distinguish response items from the current list from those of previous lists, and that intra-list intrusion errors were therefore a result of accurate guess-work. At this stage, there is no way of determining how much of the intralist PI effect is due to rehearsal on early lists, and how much is due to some kind of intralist long-term memory trace competition, confusion or interference.

A modified version of the PA probe paradigm was employed by Peterson and Peterson (1962) in an attempt to compare directly the deleterious effects of prior and subsequently presented pairs. Lists consisted of two pairs, designated pair A and pair B respectively, each of which comprised a 3-letter stimulus word, and a 4-letter response word. The following (visual) presentation scheme was employed : stimulus A was presented alone, followed by response A alone, then stimulus B alone and finally response B. Stimuli were presented in red, upper case type, whilst responses consisted of black, lower case letters, and all presentation trials were of 1 second's duration.

Immediately after the of set of response B, a number appeared on the screen; subjects were required to count backwards from this number until the onset of the probe stimulus, which occurred after

4,8 or 16 seconds of backward counting. A comparison condition in which only one pair was presented in a list completed the design. The observed proportions of correct responses were as follows:

	Retention interval (seconds)		
	4	8	16
Single item list	.94	.89	.84
Two item list, test A.	.67	.63	.60
Two item list, test B.	.57	.46	.43

These data clearly illustrate the deleterious effect of a prior or subsequently presented pair. Recall in both two-item list conditions was inferior to that of a single item at all retention intervals. Furthermore, the second pair (B) was recalled markedly less well than the first pair (A) at all retention intervals, suggesting strongly that prior activity is more deleterious than subsequent activity.

These results contrast sharply with those of Murdock involving two-item lists (e.g. Figure 4), although Murdock also employed word pairs presented at a 2-second rate. However, Murdock's probe occurred immediately after the presentation of the second pair, and it is probable that performance on the second pair would thus be augmented by STM effects to a greater degree than that on the first pair. In the present study, backward counting during the retention interval would effectively "wipe out" this STM component.

A more meaningful comparison can be made by examining the recall of the first few items, on, say, a 6-item list in Murdock's experiments treating the latter part of the list as a retention interval whose effect would be to "wipe out" STM effects on the recall of the critical item. Thus, in Murdock's (1963,a) Exp.I, the proportion of correct responses to items with no prior pairs and 5 subsequent pairs was .350, whilst that on items with 2 prior pairs and 3 subsequent pairs was only .219. Thus with a retention interval of 3 subsequently presented pairs, an additional 2 pairs presented prior to the critical item proved more deleterious than an additional 2 pairs presented subsequently.

Many similar examples may be isolated from Murdock's data, to which can be added the results of serial position experiments to be discussed in the next section, which suggest that long-term paired-associate memory is adversely affected to a greater extent by prior activity than by subsequent activity.

It should be pointed out, however, that the apparently larger difference in the two effects found by Peterson and Peterson suggest that the above comparison with Murdock's data does not provide a complete explanation of their results. Subjects in the Peterson and Peterson study were never presented with lists longer than two pairs (i.e. four words), and it could be argued that since all lists were subspan, it is probable that some kind of rehearsal strategy would be employed throughout. Thus, rehearsal of the first pair would interfere with attention to the second, to the detriment of the long-term encoding of the second pair.

Even if this explanation in terms of a rehearsal strategy is not accepted, a more convincing argument can be constructed from a more detailed examination of Peterson and Peterson's experiment, and in particular, of the unusual presentation procedure employed. Clearly, the subject's attention will be focussed on pair A, the first pair in the list, during the presentation of stimulus A, and that of response word A. However, when stimulus word B appears, the subject has not yet seen response word B - it is surely likely that the subject would prefer to continue attending to pair A (by rehearsing it, or otherwise "working" to encode it) rather than switching attention to the stimulus half of pair B, which alone wouldn't be of much use to him. Thus, the relatively large difference found between the deleterious effects of prior and subsequent activity in favour of the latter in this study can be explained in terms of the subject's devoting more attention to the first pair of the list than to the second pair, as a result of the presentation procedure employed.

A different approach to the examination of PI in paired-associate memory was followed by Bjork (1970 Exp.I) A probe procedure was employed, in which pairs consisting of nonsense-syllable stimuli in the range of 40-60 Archer (1960), and word responses drawn from Thorndike and Lorge (1944) with a G rating in the range 18-A were presented visually at a 3-second rate. One half of the pairs appeared as a green background, and the remainder on a yellow background. The lists employed consisted of 0,1,2 or 3 pairs shown on a first colour (colour A) followed by 1,2,3,4 or 5 shown on the second colour (colour B). Test slides consisted of a probe stimulus shown by itself on a white background.

The 48 subjects were instructed that any time a list contained a colour change (from green to yellow or vice-versa), they could forget the pairs shown in the first colour, and indeed, the probe stimulus came from the colour B part of the list. Each subject was tested on 60 lists (thus every serial position in the colour B part of the list was tested once for every value of the number of colour A pairs). One list in four contained no colour A items, so that since the list presentation order was randomized, the subjects could never know if a list would contain a colour change (or forget instruction), and would thus have to try to learn each pair as it was presented.

The results of this study were startling. At every level of retroactive interference (in terms of the number of subsequently presented colour B pairs), colour B items presented prior to the critical item resulted in a marked decrease in performance, whereas previously presented colour A items had no effect whatsoever. An analysis of intrusion errors (inappropriate responses) showed that overall, there were roughly 17 times as many colour B response intrusions as colour A response intrusions. Furthermore it was found that the number of colour A response intrusions remained roughly constant, or independent of the number of colour A pairs in the list, whereas colour B response intrusions increased with the number of colour B pairs in the list.

Bjork's (1970) ExpII followed the same procedure as his Exp. I.

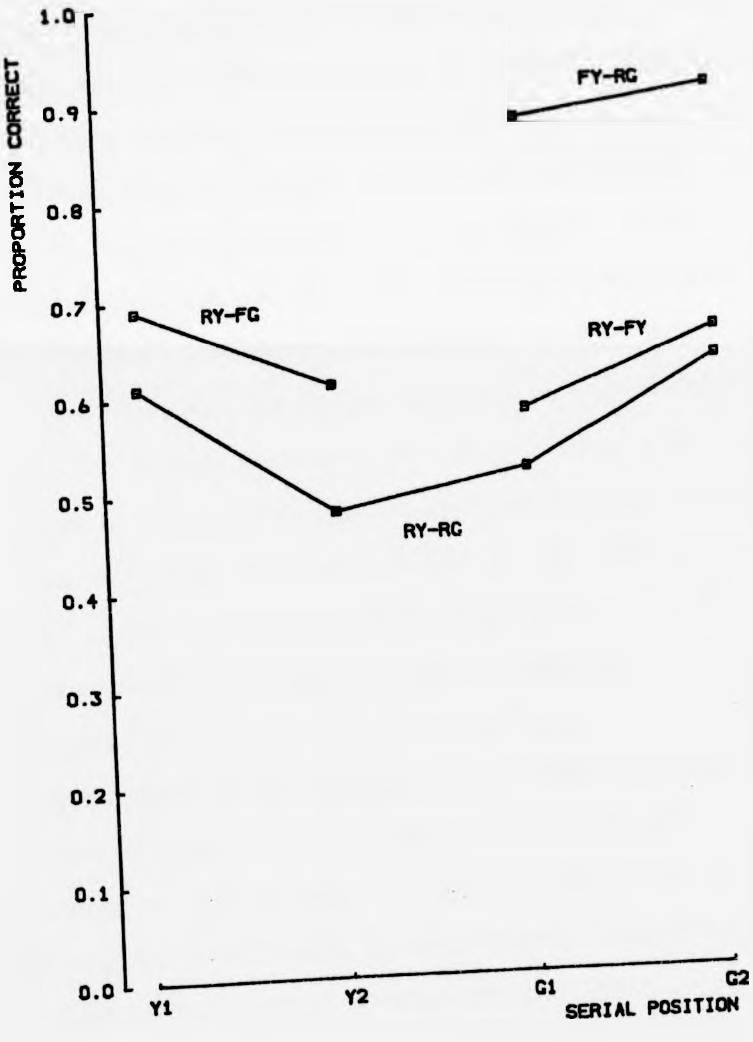
except that no "forget" instructions were given (in other words, colour changes had no significance whatsoever), and all items were equally likely to be tested. It was found that in this case, the deleterious effect of previously presented colour A pairs did not differ from that of a similar number of previously-presented colour B pairs. No difference was found in the recall of items from a two-colour list from that of similarly-positioned items in a one-colour list of the same length - thus colour in itself was not aiding performance. Taking the results of these two experiments together, Bjork concluded that the effect of a forget instruction was to effectively truncate the list, or in other words to "wipe out" the PI effect of the colour A portion of the list. Thus "both in terms of performance level and in terms of the nature of errors, a list of n colour A items followed by m colour B items in Experiment I is functionally a list of m items".

Bjork conducted a further experiment along similar lines (1970 Exp III) in an attempt to clarify a number of theoretical issues concerning intentional forgetting. A procedure similar to that of Exp I and II was employed, excpet that every list consisted of two pairs on a yellow background, a first instruction, two pairs on a green background, a second instruction, and a test trial consisting of a probe stimulus item (i.e. a CVC) from the list on a white background. The first instruction told subjects either to forget or remember the two yellow pairs. If the first instruction was to remember, the second instruction told subject's either to forget the yellow pairs, forget the green pairs, or remember the green pairs. If the first instruction was to forget the yellow pairs, the second instruction told subjects either to forget the green pairs or remember the green pairs.

For every combination of instructions, recall of items in each serial position of the to-be-remembered part of the list was tested four times per subject, although in the case of the instructions being

FIGURE 5

Correct response proportions as a function of list type (Bjork,1970, ExpIII)



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PROPORTION CORRECT
FY-RG
RY-FG
RY-FY
RY-RG
Y1
Y2
G1
G2
SERIAL POSITION

to forget yellow and then to forget green, the test slide said "no test". This condition was included to prevent subjects from predicting that a "forget yellow" instruction would be followed by a "remember green" instruction. The results of this study are presented in Figure 5.

Performance improves in every case from a "forget" instruction, relative to performance on the RY:RG list in which all pairs on the list are to be remembered. It is also very clear from the figure that the positioning of a forget instruction is important, in that performance on the green pairs is much better in the FY:RG condition than in the RY:FY condition. A similar pattern emerged from an analysis of intrusion errors; there was only one yellow response intrusion out of a total of 31 errors in the FY:RG condition, whereas 54 errors out of 110 were yellow response intrusions in the RY:FY condition.

Bjork has argued that the data suggest a two-process theory which asserts that subjects are able to take advantage of a forget instruction in two ways. Firstly, they organise the items to be remembered in a grouping that functionally separates them from the items they are to forget, and secondly, they devote all rehearsal, mnemonic and integrative activities following the forget instruction to the items they are to remember.

Bjork argued that both processes are necessary to explain the results of Exp III, since the grouping notion alone would imply that performance in the RY:FY condition should be equal to that in the FY:RG condition. On the other hand, in order to predict the results of Exps I and II, the rehearsal process alone would have to be coupled with the assumption that the "forgotten" items were lost. That this is not the case has been shown by Reitman, Malin, Bjork and Higman (1973) who essentially replicated Bjork's Exp I except for one innovation : subjects were tested very infrequently on one of the to-be-forgotten pairs in a list. When such a test occurred, the probe stimulus was marked with an asterisk. Subjects were, however, instructed to make every effort to intentionally forget when instructed to do so.

Out of the original 82 subjects, 50 were excluded when a postexperimental interview revealed that they had not followed the forget instruction to the letter. The data of the remaining 32 were essentially identical to those of Bjork's Exp. I. for to-be-remembered items, and it was found that the "forgotten" items were recalled well above chance level, but not as well as "remember" items. It was also found that although "forget" items interfered with each other, as did "remember" items, no interference effects were evident between the two groups of items.

If Bjork's arguments are accepted, then it follows that in a normal PA probe procedure, subjects must do some work on the early items in the list during the presentation of later ones. This is a highly controversial prediction, and before such an explanation is accepted, an alternative theory must be sought.

A re-examination of Bjork's (1970) Exp III (Figure 5) suggests that subjects have some ability to discriminate between the "remember" and "forget" items in the list even when the first instruction is to "remember yellow". The improvement in performance yielded in the RY:FG and RY:FY conditions over and above the RY:RG condition can be explained if it is remembered that no cue as to the colour of the probe stimulus was given during testing. In the first two conditions listed above, subjects were able to determine, from the instruction sequence, where to direct their search; in the RY:FG condition, to the earlier or yellow portion of the list, and in the RY:FY condition, to thelatter portion of the list. No such information was available in the RY:RG condition. If anything at all is surprising in these results, then it is surly the fact that the additional benefit derived from the additonal recall cue was so small.

A more detailed examination of the errors made in this experiment however, suggests that the beneficial effects of cuing are not as straightforward in their effect as has been suggested. Table 1 shows the number of observations in each condition out of a total of 2?8 which fall into five response categories : correct, yellow intrusion, green intrusion, extra list intrusion, and omission. The figures in parentheses following the error frequencies are the conditional probabilities observed, given that an error is made, of the particular type of error.

TABLE 1.

Error Analysis for Bjork's (1970) Exp. III

Condition	Yellow Intrusion	Green Intrusion	Other Intrusion	Omission	Correct
Uncued, test G (RY:RG-test G)	59(.48)	37(.30)	18(.15)	9(.07)	165
Cued, test G (RY:FY-test G)	54(.49)	25(.23)	18(.16)	13(.12)	178
Uncued, test Y (RY:RG-test Y)	32(.24)	44(.34)	32(.24)	24(.18)	156
Cued, test Y (RY:FG-test Y)	36(.36)	16(.16)	24(.24)	24(.24)	188
Forget Y test G (FY:RG-test G)	1(.03)	20(.65)	10(.32)	0(.00)	257

Consideration will first be given to the four conditions in which the first instruction is to remember yellow. Clearly, the subject's behaviour during the presentation of the four pairs will not differ between these conditions, and the second instruction can only affect the subject's retrieval, or search strategy. When yellow pairs are tested, the data follow very much the pattern that one would expect. The number of correct responses increases, and the number of green intrusions decreases, when the second instruction gives the subject information about where to direct his search; in other words, when a comparison is made of the uncued, test yellow condition (RY:RG-test Y) with the cued test yellow condition (RY:FG-test Y). These results suggest that (i) yellow stimuli are not well recognised as such when presented as a probe (so that the information implicit in the FG second instruction is of use) and (ii) the subject has some information about response colour, which enables him to retrieve the correct response more often, and to avoid clearly inappropriate green intrusions when he knows that a yellow is under test.

However, the picture is not so clear when the two test green conditions are examined. Indeed, there is no significant difference in the number of correct responses in the two conditions (RY:RG - test G and RY:FY - test G)or in the patterns of responses across the five categories ($Z = 1.05$, $\chi^2_4 = 3.76$, both not significant). There are two possible explanations for the lack of a facilitatory cuing effect. Either colour information is of no use to the subject when green items are tested, or colour information is already available when green stimuli are presented for test (i.e. they are recognised as green stimuli) so that cuing provides no additional information. Since it has already been argued that response colour information is available to the subject, because yellow cuing enables him to avoid green response intrusions, the latter of these two explanations is accepted as the most likely. However, if the subject recognises green stimuli as such, and has information regarding the colour of a retrieved response, why are there almost exactly twice as many yellow intrusions (113) as green intrusions(62)when the data for green recall folowing an initial RY instruction are pooled?

Given that the correct response has not beenretrieved, there are two inappropriate yellow responses to one inappropriate green one. Thus, if no colour information were available about retrieved response words, one would expect twice as many yellow responses as green ones. However, it has already been argued that there is some response colour information available to the subject. There is, however, an alternative explanation of the data. Suppose material is encoded in such a way that there is very little colour information available on yellow stimuli and responses and a great deal of information available on green stimuli and responses.

If the colour information on green responses is available only <u>after</u> the response has been recalled (and does not materially aid recall of the response) then the following predictions can be made : (i) yellow cuing will have a beneficial effect, because colour information is not always available on yellow stimuli, and such information will lead the subject to avoid making a green response, which upon retrieval will be recognised as inappropriate. (ii) Green cuing will not be very helpful, because green stimuli are nearly always recognised as such anyway. (iii) When responding to a green stimulus, the subject is roughly twice as likely to retrieve an inappropriate yellow response as an inappropriate green one, but will not surpress such a yellow response since he lacks positive information that it is inappropriate.

Attention is now turned to the fifth condition in Bjork's (1970) Exp. III, that is, the condition in which the first instruction is to forget yellow. It is reasonable to suppose that the subject will employ a different strategy during the presentation of the green pairs to that employed in the previous four conditions, and that furthermore, this strategy will be based on the expectation that the second instruction will be to remember green, and that a green stimulus will constitute the probe. If the second instruction is, in fact, to forget green, the slide "no test" appears instead of a probe stimulus, and the subject will have lost nothing by adopting such an approach.

The enormous improvement in the level of recall of G_2 (the second green pair) when tested in this condition as opposed to the two conditions where the first instruction is to remember yellow can be easily explained. All PA probe experiements typically yield an enormous recency effect; for example, Murdock's data in Figure 4 show that the last item of a list is recalled with a probability in the region of .9 whilst the figure for the fourth item of a four-item list with no forget instruction in Bjork's (1970) Exp.I is .94. Why, then, is G_2 recalled so badly in the RY:RG and RY:FY conditions? The answer clearly lies in the fact that in these two conditions the subject has to attend and process the second instruction, which comes after G_2 and before the probe onset. This processing will prevent rehearsal of G_2, and will effectively "wipe out" the STM component responsible for the recency effect. When, however, the first instruction is to forget yellow, the subject can effectively ignore the second instruction, and continue rehearsing G_2, since the second instruction conveys no useful information. (It is interesting to speculate as to whether the subjects could so easily have ignored an auditory, as opposed to a visual, signal). If a probe stimulus occurs, the subject can respond, whilst if a "no test" appears on the screen, he will not have lost anything by rehearsing. Another glance at Figure 4 will show that the first item of Murdock's 2-item lists was also very well recalled, as was the first item of a 2-item list with no forget instruction in Bjork's Exp I (recall probability of 0.81). It is probable, therefore that even when subject's are not principally employing a "memory span" rehearsal strategy, they continue to use such a strategy on the first one or two items of any list, and discontinue it when the list becomes super-span. It is likely that such a strategy is employed following a first FY instruction, on the two green pairs. Thus the overall improvement in performance on green items following an initial FY instruction can be explained in terms of the subject employing a rehearsal strategy which is effective because there is no further processing required of the subject (which would disrupt the strategy) between presentation of the two pairs and the onset of the probe stimulus. Clearly, then, the results of Bjork's Exp. III do not unequivocably support his discrimination-concentration of effort theory.

However, it will be remembered that Bjork's theory is strongly supported by the results of his (1970) Exp.1, and by the results of Reitman, Malin, Bjork and Higman (1973). In the latter study lists of nonsense-syllable-word pairs were presented at a three-second rate. Pairs were presented on a green or a yellow background, each list comprising 0,1,2,3 or 4 colour A pairs (green or yellow) followed by 1,2,3,4 or 5 pairs on the other colour background (colour B). Subjects were instructed to forget all the items prior to a colour change. After a short practice session, subjects were informed that occasionally, tests would occur of colour A pairs (i.e. "forget" items) and that such tests would be signalled by an asterisk next to the probe stimulus, but that they should continue to forget colour A items as before. Out of 82 subjects tested, 50 were rejected on the basis of a post-experimental interview as not adequately following the instructions to forget.

In so far as tests on colour B pairs, the 32 critical subjects produced data almost identical to that of Bjork's (1970) Exp.1. In particular, recall of to-be-remembered items (colour B) was unaffected by the number of previously-presented "forget" (ColourA) items, whilst colour A intrusions were independent of the number of colour A pairs, and were outnumbered by colour B intrusions in the ratio of 16:1. When pre-signal pairs (i.e. "forget" or colour A items) were tested, and it must be borne in mind that such tests were cued with an asterisk, it was found that the probability of recall declined with the number of forget pairs (in other words, forget pairs were mutually interfering) and that recall of forget pairs whilst being significantly above chance level, was decidedly inferior to that of comparable pairs (i.e. with the same number of preceding and following items) from lists in which no forget signal was given. Since the subject would not know during presentation whether he would be subsequently instructed to forget first colour items, this result rules out the possibility that subjects were not devoting much effort to first colour items in the expectation of a "forget" signal. It thus appears that "forget" items were, to some extent, actually forgotten. However, it was also found that the level of recall of "forget" items decreased as the number of subsequently-presented "remember" items increased, and furthermore, the ratio of appropriate to inappropriate colour response intrusions to "forget" items was only about 3:1 (in comparison with a similar ratio of 16:1 found when "remember" items were tested). Thus, "remember" items were interfering retroactively with "forget" items.

The fact that "forget" items are recalled less well than comparable items from a single colour list even though they are cued on test (so that the subject knows where to direct his search) lends strong support to Bjork's hypothses, which would predict that "forget" items would receive less processing than comparable remember items. However, the high level of "remember" response intrusions to "forget" stimuli would suggest that "remember" items were interfering in a more complex way than a simple "displacement from processing" hypothesis would predict. However, these data do admit an alternative explanation.

Encoding theory would claim that in a PA task, the response is encoded in relation to the functional, or encoded form of the stimulus (e.g. Martin 1972). Thus, in a normal PA probe list learning study it is probable that in encoding a particular pair, the subject becomes aware that certain aspects of his encodin scheme are shared by the encodings of earlier pairs, by some sort of recognition process, and that he will therefore be motivated to elaborate his current encoding in order to discriminate it from earlier ones. An elaboration of the stimulus encoding would fulfil this requirement admirably. To the extent that the subject has some access to previous encodings during the encoding of a current pair, it can be said that he is devoting attention to earlier pairs during the presentation of a current pair. This does not, however, imply that he is actively operating on these earlier pairs, as Bjork's hypothesis suggests, but rather that he is taking them into account whilst encoding the current pair.

Proactive interference would be a measure of the extent to which the subject was forced to elaborate the encoding of a currently-presented pair in order to discriminate it from earlier, similarly encoded pairs. The additional processing time involved in operating on the functional stimulus would thus reduce the processing time available for, and thus the efficiency of, the encoding of the response and the relational encoding of the functional stimulus - functional response association. Retroactive interference would result from a failure to adequately discriminate the encoding of a current pair from that of an earlier one. The most recent encoding might to some extent "overwrite" the earlier one;

a more satisfactory argument would suggest that since the episodic aspect available at test are more likely to correspond to those used in the encoding of a more recent item than an earlier one, then search at test is far more likely to lead to the retrieval of the more recent encoding. In this way, subsequently presented material would retroactively interfere with performance on earlier items.

When a subject is instructed to forget a list of items just presented, it is clear that the "forget" items will mutually interfere in the normal way. It is postulated that encodings of subsequent pairs, however, are then made without reference to the previous encodings of "forget" items, but with reference to previously encoded "remember" items, at least to the extent to which sufficient episodic information is available to discriminate the two groups. Colour cues as employed in the studies above would certainly provide a distinctive episodic cue. As a result, "forget" items will not proactively interfere with "remember" items and furthermore, "remember" items will interfere retroactively to a greater extent with "forget" items than with comparable "remember" items.

An attempt can now be made to explain the error analysis of the Reitman et al. (1973) study. It is fairly certain that their subjects were unable to recognise first colour "forget" stimuli adequately without being presented with such information at test, since performance on uncued (unasterisked) tests of "forget" stimuli was found to be at a negligible level; such a phenomenon could be well explained in terms of the encoding specificity hypothesis. This result suggests that stimuli were encoded in such a way that cuing as to whether the stimulus were a "remember" or a "forget" stimulus would be necessary for recognition, and would not be available subsequent to recognition. The relatively low rate of inappropriate colour, "forget" responses to "remember" stimuli can be explained in terms of two processes. In the first place, by a similar argument employed to explain the results of Bjork's (1970) Exp.III, it is postulated that response encodings often contained a colour component. Such a component would be more likely to be available upon retrieving a "remember" response than a "forget" response, either because more emphasis was placed upon encoding such a component for "remember" responses, or because "remember" response encodings would retoactively interfere with similar "forget" response encoding. Thus, were the subject to employ a directed guessing strategy, a far greater number of positively-identified "remember" responses would result than unidentified "forget" responses. Furthermore, the encoding hypothesis would also predict that a mis-recognised "remember" stimulus is almost certain to be mis-recognised as another "remember" stimulus, since two such stimuli are more likely to share episodic aspects that are present at test than say, a "remember" and a "forget" stimulus, resulting in a high level of appropriate-colour response intrusions.

On the other hand, when applying a directed response guessing strategy to a cued "forget" stimulus, the subject would, as in Bjork's Exp III, surpress responses known to be of inappropriate colour (i.e. "remember" responses). However, such information would be less likely to be available on every "remember" response in the Reitman et.al. study than in Bjork's Exp. III, since initial encoding by colour would virtually be essential in Bjork's task. Furthermore, such an encoding would be easier to maintain in that task, since only two second-colour responses were present in each list, as compared with anything up to 5 in the Reitman et.al. study. In addition, were a "forget" stimulus mis-recognised, although it is highly likely that it would be mis-recognised as another "forget" stimulus, there is still a good chance that it would be mis-recognised as a "remember" stimulus whose encoding would be more likely to share a number of other (i.e. non-colour) episodic attributes with the test episode. These arguments would account for a higher level of inappropriate response intrusions to cued "forget" stimuli than to "remember" stimuli.

It should be stressed that the directing guessing strategy was postulated to explain the error analysis of Bjork's Exp.III in those conditions in which the subject was trying to remember items of both colours during the presentation of the second colour items. The encoding hypothesis would predict such behaviour, since the subject would certainly attempt to encode second-colour items by colour in order to distinguish them from first-colour items (especially as colour cues were often available at test). Such a strategy is not necessarily predicted when encoding second-colour items after an instruction to forget first-colour items, and it is thought that stimulus mis-recognition provides a more satisfactory explanation of the Reitman et. al. error analysis.

The encoding hypothesis advanced to explain directed forgetting can be expanded to produce a general theory of paired-associate forgetting. When the attributes employed in encoding a currently-presented pair are recognised as being similar to those employed in encoding a previously-presented, to-be-remembered pair, then an attempt is made to elaborate the current stimulus encoding in such a way as to discriminate it from the earlier stimulus encoding. To the extent that this detracts from the processing time available for completing the encoding of the current pair, earlier to-be-remembered pairs will interfere proactively with performance on later ones. Nevertheless, the effect of this PI on performance is thought to be slight, and the major effect of a forget cue (either explicitly within a list or implicitly between two successive lists) may well be to afford a little extra processing time to early items in the currently to-be-remembered list. Encodings of currently-presented pairs are seen to interfere retroactively with previous similar encodings since they are more likely to share episodic attributes with the test sequence, and are therefore more likely to be retrieved than the earlier encoding. This effect will not be so pronounced when the current stimulus has been differentially encoded to the earlier one (as when the earlier pair is still to-be-remembered), as when the current stimulus is not differentially encoded in this way (when the earlier encoding is of a to-be-forgotten item). One problem with such an interpretation results from the finding by Reitman et. al. that uncued "forget" stimuli were not recognised as such at test. How, then, could subjects discriminate between which earlier encodings to take into account when encoding current stimuli, and which to ignore?

Two possible explanations are advanced. In the first place, it is possible that encoding similarities do not become apparant until encoding has progressed beyond the stimulus encoding phase, and that therefore more information from the earlier encoding is available than stimulus information alone upon which to base such a discrimination. Secondly, and more convincingly, it could be argued that earlier encodings became available only if they share episodic aspects with the current presentation episode. "Forget" items would clearly not be nearly so likely to share such aspects with currently-presented "remember" items as previously presented "remember" items. Thus, the subject does not so much choose to ignore the encodings of previously presented "forget" items having recognised them as such, as he fails to have access to such encodings in the first place.

Such a hypothesis would predict that PI would not extend beyond some limit determined by the rate at which episodic aspects change from trial-to-trial, and is supported at least by Murdock's failure to find such effects beyond about two prior items, although Bjork (1970 Exp.1) found within list PI effects extending over four prior items. This prediction would be difficult to test experimentally in any case, since one would need to be certain that episodic information were being attended and encoded in the first place. With the long lists required to find a PI limit, there is a very real chance that subjects would opt for a deeper (semantic?) encoding strategy, and a far greater extension of PI effects would be expected in this case.

It should also be pointed out that an episodic access hypothesis would predict an increased effect of PI with increasing retention interval, since in this case, in addition to increasing the encoding requirements of later items, similarly encoded early items would be equally <u>unlikely</u> to share episodic aspects with the test sequence as more recently encoded items, and hence equally likely to be retrieved at test. The increase of PI with retention interval is a well established phenomenon (e.g. Koppenaal, 1963), and indeed results of this nature led interference theorists to postulate the "spontaneous recovery" of initial associations (see 1.21). The above theory, then, gives a satisfactory account of paired-associate forgetting phenomena.

1.34. Interpolated Recall.

So far, attention has been restricted to PA probe studies in which the retention interval between the presentation of an item and its subsequent test is filled with presentations of other pairs. The question now arises as to the effect of recalling other items during the retention interval. Two studies by Murdock (1963,b, Exps I and II) employed a PA probe procedure with lists consisting of 6 common-word pairs presented at a two-second rate, which were followed by 3(Exp I) or 6(ExpII) probes on different list pairs. In Exp II, it was not possible to include every possible testing order, so a counterbalancing procedure was employed to ensure that the items tested prior to the test of any given item were equally likely to come from any list position, so that, on average, intruding tests should be of equal difficulty irrespective of the current critical item. All tests were subject-paced; in other words, the onset of the next probe stimulus was delayed until the subject had responded to the current probe.

Murdock's findings were straightforward. Both studies suggested that interpolated recall had a detrimental effect, and that this effect was most pronounced for later serial positions. Thus, the proportions of correct responses for the first list item after 0,1 or 2 interposed recalls were .266, .315 and .237 respectively, whereas the corresponding figures for item 6 were .864, .409 and .285. The results of Exp. II suggested that from 3 to 5 interpolated recalls had the same effect as 2, so that additional interpolated recalls beyond 2 had little or no effect.

A similar study by Tulving and Arbuckle (1963) employed 10-item lists with the digits 0-9 as stimuli, and nonsense-syllable responses. They found that interpolated recall had no effect on the recall of the first four items, and an increasingly detrimental effect on items 5 to 10. Furthermore, the effect of interpolated recall was found to asymptote after about four such tests; in other words performance remained roughly constant whether there were 5 or 10 interpolated recalls. These data agree very well with Murdock's results.

A study by Tulving and Arbuckle (1963) employed a PA probe procedure with common-word-digit pairs presented visually at a 2-second rate. Items in serial positions 1 to 5 were tested following two succeeding presentations with no interpolated tests, and two interpolated tests with no successive presentations, allowing a comparison of the detrimental effects of a two-test retention interval with that of two presentations. It was found that a two-presentation retention interval had a more pronounced deleterious effect on recall. The authors argued that interpolated tests were primarily effective in preventing rehearsal of items not being tested and in "wiping out" active short-term memory of the most recent items (including, possibly, sequential information), whereas interpolated presentations would, in addition, affect the more deeply encoded components of previously-presented items, resulting in poorer performance.

This hypothesis is supported by Murdock's (1963b) Exp III, in which test trials were of fixed duration, either 2 or 8 seconds. It was found that one or two interpolated 8-second tests proved more deleterious to the recall of a critical item than the same number of 2-second tests. Murdock also found (1963b, Exp. I) that in 6-item lists, an incorrect response to an interpolated test of item 5 was more deleterious to the recall of item 6 than a correct response to such a test of item 5. Not only would a correct recall imply a shorter search time on the intruded test trial, but might even permit the retrieval of response 6 via a sequential encoding involving pair 5. This type of process would presumably not interfere with the sequential encoding of item 6, so that this process could also underly the result.

1.4. The Effects of Practice on Paired-Associate Memory

In the context of this thesis, the term practice will be used rather loosely to refer to any experimental procedure that potentially allows the subject to allocate additional processing time to the material to be remembered. Such procedures usually take the form of either allowing the subject additional rehearsal time by increasing the duration of presentation trials, or of increasing the overall exposure time of selected pairs by presenting them repeatedly.

1.41 Rehearsal

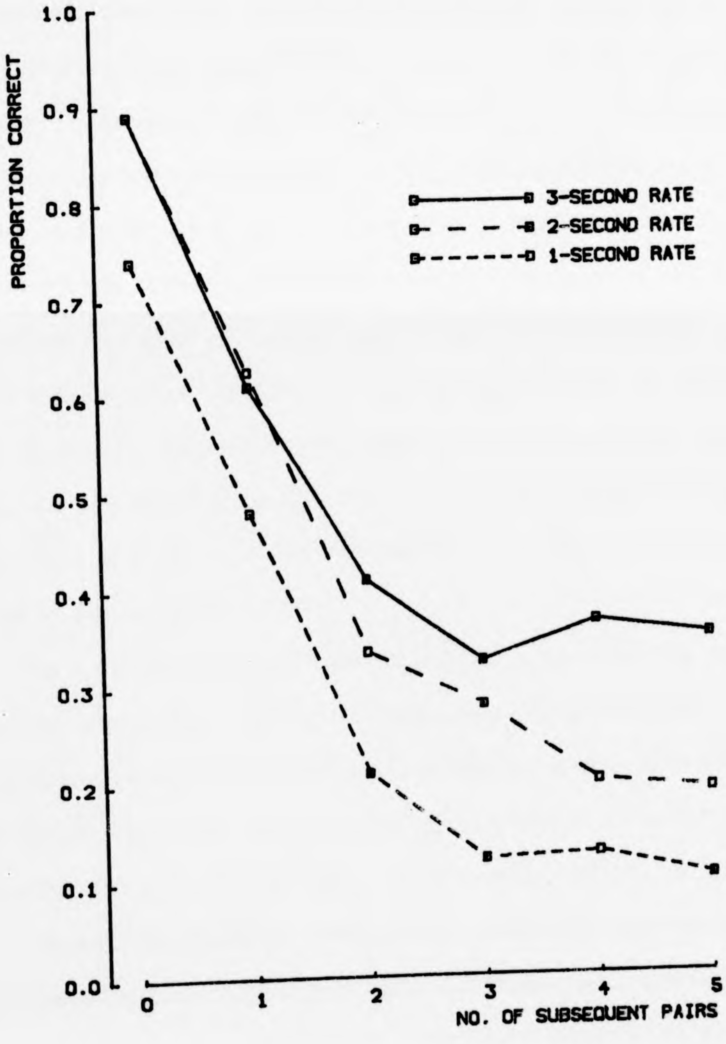
It is certainly a well established result that giving subjects more time to study paired-associates improves performance. For example, Keller, Thompson, Tweedy and Atkinson (1967) employed a list-learning paradigm in which stimuli were 2-digit numbers, and responses were the letters A,B and C. The list was repeated 15 times, and pairs were presented for $\frac{1}{2}$,1,2 or 4 seconds. Three pairs received each of these presentation durations throughout (making a total of 12 pairs receiving a fixed presentation duration) whilst a further 12 pairs had the duration of their presentation trial randomly assigned with each repetition of the list. It was hoped that this procedure would prevent subjects from adopting special strategies in this task, such as using some of the presentation time of, say, well-learned 4-second items to rehearse previously-presented short presentation items, and furthermore this method allowed data from "all-same" presentation duration items to be averaged with that from "random duration" items, thereby minimising such effects in the overall results.

The total proportion of erros made over the 15 trials was recorded for each presentation duration, and it was found that this error rate decreased with increasing exposure time, falling from about .66 for a $\frac{1}{2}$-second rate, to .58 with a 1-second rate, and thereafter to about .53 for 2- and 4- second rates. It therefore appeared that there was some limit beyond which an increase in exposure time would not produce a significant performance improvement. Several processes could account for such a limit. In the first place, there is no guarantee that subjects were actively employing all the exposure time available on each pair to actively process that pair; for example, subjects may have been reluctant to maintain attention on any particular pair for more than about 2 seconds. It is also possible that the number of attributes available for encoding would somehow be limited by the episodic aspect of the presentation trial; beyond a certain point, all the available attributes would be adequately encoded, and extra processing time would therefore be of little benefit.

Of course, this study only measured retention from trial-to-trial but no control was available over the retention interval (see 1.31). In a study by Murdock (1963 c, Exp.II) the presentation rate of a list of 6 word pairs was varied, taking values of 1 second per pair, 2 seconds per pair, or 3 seconds per pair. A probe technique was employed to examine retention of items at each serial position. The data from this study are depicted in Figure 6. It is clear that improvement in performance with exposure rate is most marked at long retention intervals, and although the number of prior pairs is confounded with the number of subsequent pairs in this study, the lack of a primacy effect in this data suggests that PI effects are negligible (again, subjects were tested repeatedly on many lists). Differences in short-term retention appear only between the 1-second and slower rates, and may result from difficulties in attending every item at such a fast rate; in other words, at very fast rates, a proportion of items don't even get into an active rehearsal loop, possibly because the subject is still actively encoding the previous item when the next one is presented. Such an interpretation would account for the differences in performance found by Keller et. al. between the ½,1 - second and slower rates, although their failure to find significant differences between 2- and 4- second rates conflicts sharply with the Murdock study.

FIGURE 6

Retention of paired associates from a 6-item list as a function of presentation rate (Murdock 1963c, Exp. II)



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
5
NO. OF SUBSEQUENT PAIRS
3-SECOND RATE
2-SECOND RATE
1-SECOND RATE

It should also be pointed out that in Murdock's study, presentation rate is confounded with the length in time of the retention interval, which would operate to the detriment of performance at slower presentation rates in two ways. Firstly, it would reduce the likelihood that episodic temporal cues encoded at presentation would be present at test, and furthermore, at slower rates, more attributes per item might be encoded, increasing the chance that subsequently-presented pairs would share encoding aspects with earlier ones, leading to interference. These processes would serve to decrease the differences in long-term retention performance in Murdock's study, which makes the results of Keller et.al. doubly surprising. However, it should be borne in mind that Keller et. al. measured performance across 15 repetitions of the list, and there is a very real chance that, as more items become very well encoded after a few repetitions, subjects would utilise the presentation periods of such items to continue the active encoding of the previous item, thereby reducing overall differences in performance between those items that at least enter active rehearsal (i.e. those presented at a 2- or 4- second rate).

The most important point to emerge here is that there is no guarantee that the experimenter - controlled presentation time necessarily corresponds to the subject - controlled processing time allocated to a particular item. This appears to be especially true in situations where pairs receive many presentations. In such situations, it is postulated that subjects may choose to disregard the presentations of items that they know they know particularly well, and utilise such a presentation interval to more adequately encode a previously-presented pair.

1.42 Unreinforced test-trials

In a typical study-test paired-associate learning procedure, the subject is repeatedly presented with cycles comprising a study list followed by a test list, the order in which items appear in the list being re-randomized each time (see 1.31). A number of studies by Izawa have made use of a more elaborate version of this paradigm.

Izawa constructed elaborate schedules for her items, involving reinforcements, or study trials, (R), tests (T), neutral trials (N) and blank trials (B). On a blank trial, nothing would appear on the screen for a period equal to that of a study or test trial, whilst on a neutral trial, the subject would be required to fill such a period with an activity such as colour naming to prevent rehearsal of previously-presented material.

Izawa also distinguished between mixed and unmixed list designs. In a mixed list, items receiving different schedules would be tested together in the same lists, as follows: suppose for example, that a number of items receiving an RT schedule are to be mixed with a similar number receiving an RTT schedule. Each cycle of the experiement involves one trial on each item as follows:-

	Cycle								
	1	2	3	4	5	6	7	8	...
RT items	R	T	R	T	R	T	R	T	...
RTT items	R	T	T	R	T	T	R	T	...

Thus, the experiment would begin with a list of study trials on all the items. Following this, a list of test trials on all the items would be presented. The third list, however, would involve study trials on the RT items mixed together with test trials on the RTT items, whilst the fourth list would comprise test trials on the RT items and study trials on the RTT items, and so on. In an unmixed list design, however, different groups of subjects would be tested on either lists made up entirely of RT items, or on lists made up entirely of RTT items, so that on each cycle, the list presented would consist entirely of study trials, or entirely of test trials. Under such a scheme, then, the normal study - test procedure would constitute an unmixed - list RT design. It should also be pointed out that Izawa employed consonant - vowel - consonant (CVC) nonsense syllables as stimuli in all her experiments.

In her 1966 study, Izawa compared the standard study - test RT condition with a double - test RTT condition in a mixed list design, employing 2 - digit number responses. It was found that in the RTT condition, there were no essential differences in the overall proportion of correct responses made between the two successive tests following a presentation, which suggested that test trials had no effect whatsoever, in that they led neither to forgetting, nor to learning. However, an important effect of unreinforced test trials was found : performance following the second or subsequent presentation of an item in the RTT condition was significantly superior to that of an RT item which had received the same number of presentations. Similar results have also been found with both number and noun responses in both mixed lists (Izawa, 1967) and unmixed lists (Izawa 1970).

In another mixed list design (Izawa, 1968) performance was measured following the eight presentation of items that had received either a schedule of eight consecutive study trials (R), eight study trials on an RT schedule, or eight study trials on an RTTTT schedule. It was found that the four-test condition led to the best performance, followed by the single-test condition. In an attempt to establish whether there was some upper limit beyond which additional unreinforced test trials would have no further beneficial effect, Izawa (1969) compared performance following 1,2,3 and 4 presentations on items tested under four different schedules in a mixed list design. The four schedules were as follows:-

1) RT 2) $RT_1,\ldots,T_5$ 3) $RT_1,\ldots,T_{11}$ and 4) $RT_1,\ldots,T_{15}$

Both number and noun responses were investigated. It was found in each case that although the single test condition (1) resulted in the poorest performance, the best performance resulted from the five-test condition (2) whilst performance in the eleven and fifteen-test conditions (3) and (4) was if anything slightly inferior to that in condition (2). This result suggested at least an upper limit of between 5 and 11 unreinforced test trials beyond which no further improvement in performance would take place.

In her 1967 study, Izawa also compared RTB and RBT schedules, and RTN and RNT schedules, for both noun and number responses, in mixed lists also including the RT and RTT schedules discussed above. Little difference was found between performance in these lists containing blanks and those containing neutral trials. However, this does not necessarily imply that subjects were not using blank trials to rehearse previously presented material; in a mixed list design, almost any item (under any schedule) could receive the benefit of this extra processing time, although there was no significant difference in overall levels of performance in lists containing blanks and those containing neutral trials. Blanks and neutral trials had the same effect on those items whose schedules included them, in that performance in condition RTB was superior to that in RBT, and a similar comparison held for conditions RTN and RNT.

Izawa argued that these results suggested at least two functions of unreinforced test trials. In the first place, they clearly acted in some way to "potentiate" learning on the next study trial, and secondly, they had the effect of preventing forgetting, which accounted for the superiority of RT - over R-T conditions (where - represents a B or an N trial). It was found, furthermore, that the superiority of the RTB to RBT conditions was also maintained in an unmixed list design (Izawa 1970) (when subjects would almost certainly make use of the long blank period to rehearse and practice items), for number responses, and for noun responses up to about four reinforcements.

Izawa (1971) argued that the role of an unreinforced test trial operated principally in increasing the number of encoded stimulus attributes that would be relationally linked to the response. To account for the phenomena that the more T's per replication, the better the learning, she postulated that during the active search made on test trials, most of the stimulus attributes that would not normally be available (i.e. were that trial a presentation) are sampled, and may consequently still be available for association with the response during the next presentation trial. On the other hand, those stimulus attributes that were available at the start of the test trial but were not of use in retrieving the appropriate response would be "discarded" by the subject in some way, and would be less likely to be sampled and used in an attempt to retrieve the response on a subsequent test trial. This hypothesis would account for the lack of forgetting across unreinforced test trials. A mathematical model based on this theory proved quite successful (Izawa,1971) but as will be shown in Chapter Four, there is considerable evidence that stimulus encoding during on-going paired-associate learning is reductive in nature, and not elaborative as Izawa suggests.

1.43 Repeated Presentations.

A number of CPA studies suggest that repeated presentation of a pair leads to superior performance across a variety of retention intervals. For example, Peterson, Saltzman, Hillner and Land (1962 Exp.I) employed a study-test CPA technique in which trials involving common word stimuli and numerical responses in the range 1 - 9 were visually presented at a 2-second rate. Some pairs received a single presentation and were tested at a retention interval of 0.1,3 or 8 intruding trials on other pairs, whilst in another condition, items received two presentations, one after another, and were tested at a retention interval of 3 intruding trials. The retention data for the single-presentation items has been described elsewhere (see 1.32), but it was found that retention after 3 intruding trials for single items yielded a proportion correct of 0.40, whilst the corresponding figure for the double items was 0.55.

In a further study, Peterson et.al. (Exp. III) employed a similar design, in which stimuli were 99-100% Archer (1960) nonsense syllables, and responses were numbers in the range of 1 - 10. Four conditions were employed in a mixed-list study-test CPA design : (1) double presentation. Items in this condition received two successive presentation trials, one immediately after the other. (2) presentation-test. Items in this condition were presented once, and then immediately tested. (3) Rehearsal. Items in this condition were presented, and on the following trial, a pair of ditto marks (") appeared on the screen. (4) Single presentation. Items in this condition received just one presentation. All items were then tested after a retention interval of three intruding trials on other pairs had followed the schedules described above.

The proportions of correct responses obtained in the four conditions were as follows: (1) .42, (2) .44, (3) .40, (4) .35. As would be expected, the double presentation, presentation test and rehearsal conditions all produced superior performance to the single-presentation condition.

Brelsford, Shiffrin and Atkinson (1968) employed a modified version of the CPA anticipation procedure, in which stimuli were two-digit numbers, and responses were letters of the alphabet. Their visually-presented anticipation trials lasted for a total of 11 seconds each, and comprised a 3-second test phase, followed by a 2-second blank period, a 3-second study phase, and finally a 3-second blank period before the onset of the next trial. Other aspects of their procedure are described more fully elsewhere (see next section: 1.44). The proportions of correct responses observed for various retention intervals since the most recent presentation (measured in terms of the number of intruding anticipation trials on other items) are depicted in Figure 7. It should be pointed out that in general, the successive presentations of a particular pair would not generally have occurred on successive anticipation trials, so that these data are averaged over the various intervals separating the successive presentations of each particular item. Nevertheless, the curves obtained are remarkably similar to that of Atkinson, Brelsford and Shiffrin (1967, Exp. II) depicted in Figure 1. The data furthermore suggest presentations are of benefit at all non-zero retention intervals.

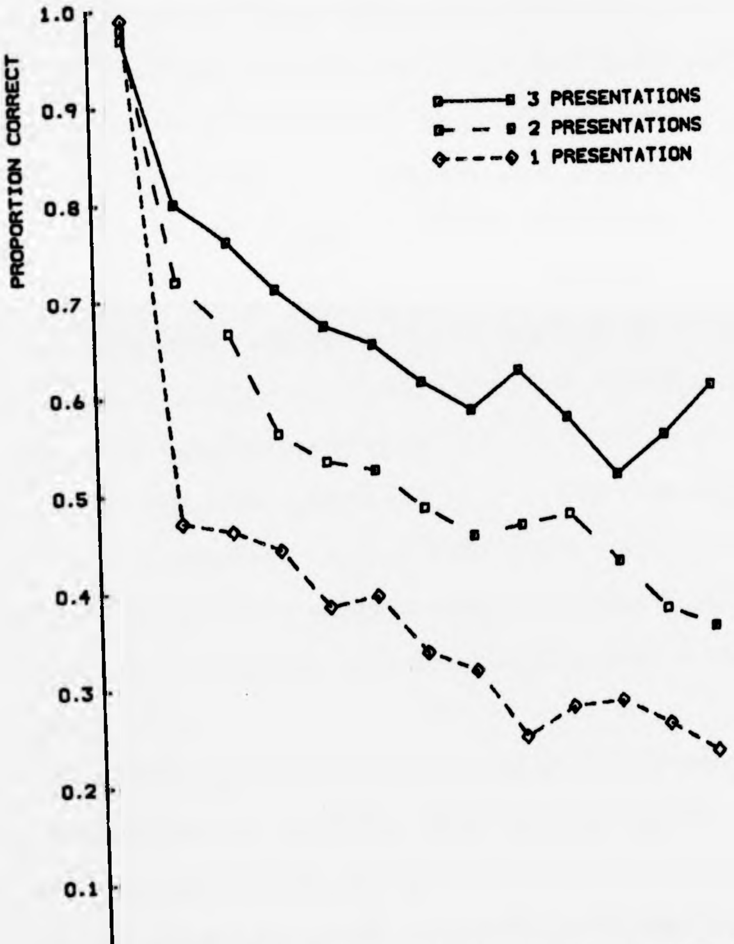
A number of list-learning studies, however, have not produced

FIGURE 7

Retention of paired associates as a function of the number of presentations (Brelsford, Shiffrin and Atkinson, 1968)

FIGURE 7

Retention of paired associates as a function of the number of presentations (Brelsford, Shiffrin and Atkinson, 1968)



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3 PRESENTATIONS
2 PRESENTATIONS
1 PRESENTATION
0 1 2 3 4 5 6 7 8 9 10 11 12
RETENTION INTERVAL

such convincing results. Peterson and Brewer (1963, Exp.II) employed a study test paradigm in which common-word stimuli were paired with numerical responses in the range 1 - 12. Three conditions were compared in a mixed list design. Pairs in the single presentation condition were presented once only in each study cycle. In the double presentation condition, pairs received two presentations, the second following immediately after the first, on each study cycle. Pairs in the interference condition received a presentation with an incorrect response, immediately followed by the presentation of the correct pairing. Performance (in terms of the proportion of correct responses) was measured on each of 15 study-test cycles.

It was found that the double presentation condition produced only slightly superior performance to the single presentation condition at all stages of learning, and indeed, performance following the fifteenth study trial was almost equal in the two conditions. Furthermore, although the interference condition was enormously disadvantaged in comparison with the other two over the first six cycles, performance in this condition thereafter rapidly overhauled that in the other conditions, and was only slightly inferior after the final cycle.

Calfee (1968) suggested that the surprisingly slight differences in performance found in such studies might well be an artifact of the experimental design, in that, if items in the double - presentation condition are learned faster, then on later blocks, subjects might well devote the presentation time of such items to processing items that are not so well learned. Such a hypothesis would account for the small differences found between the conditions in this study after seven or more study cycles. Furthermore, Calfee (1968, Exp. III) convincingly demonstrated an inverse relationship between the rate at which items were learned and the number of unlearned items in the list.

Greeno (1964) employed an anticipation paradigm with short word stimuli and numerical responses in the range 1 - 5. He arranged his items into blocks of 30 anticipation trials. Half his items received only one presentation per block, whilst the remainder received two anticipation trials, separated by zero or one intervening trial on another item. Thus, single- and double- presentation conditions were compared in a mixed list. Greeno found hardly any difference between performance on single- and double- items in terms of inter-block retention (i.e. when performance was measured on the first anticipation trial on an item in each block), although performance on the test phase of the second anticipation trial of double items in a particular block was nearly perfect. Greeno argued that within-block retention of double items resulted from short-term memory components, whilst inter-block retention implied long-term performance (although the inter-block retention intervals were not directly controlled, they were sufficiently long to "wipe out" short-term retention effects). Although Calfee's objections would apply to this study, Greeno's data showed no signs of performance on single items overhauling that on double items as the number of blocks increased; on the contrary, differences in performance were non-existent from the first block of trials onwards.

Similar studies were carried out by Greeno and White and by Greeno and Rumelhart (reported in Greeno, 1970a) but with an important difference to the study reported above: in both these studies, an unmixed list design was employed. In other words, separate groups of subjects were tested either on blocks of trials containing only single-presentation items, or on blocks containing only double-presentation items. Greeno (1970a) aggregated the results of these two studies, and compared them with his 1964 experiment. It was found that inter-block performance on double items in unmixed lists was markedly superior to that on unmixed single items. Furthermore, this difference became more pronounced as the number of blocks (in other words, the number of successive single or double presentations) increased.

Greeno (1970a) examined two possible explanations of these results, The first of these was a time-sharing hypothesis (Greeno, 1967) which stated that on the second of two closely-spaced presentations, the subject may be unwilling to use the second presentation to process the current item, on the grounds that he had just seen it and processed it anyway, and may therefore use the time available to process other items currently held in short-term memory. In a mixed list design, the items which would benefit from this extra processing would be equally likely to be single- or double- presentation items, so that overall, both types would receive an average equal amounts of processing (about $1\frac{1}{2}$ trials' worth per block). In unmixed lists, however, double items would receive twice as much processing as single items per block, since the available processing time on the second presentation of a double item would be used to process another item from the list, which in this case would be another double item.

There are two objections to such a hypothesis. Greeno pointed out that if processing time were shared in this way, then the most likely items to benefit from the extra processing would be those that were presented just prior to the second presentation of a double item. Potts (1969) found no evidence of improved performance on such items, although as Melton (1970) has pointed out, it is possible that the items that would benefit from extra processing would be those that shared the same response as the current item (on the hypothesis that subjects were grouping items by response). Potts employed numerical responses, and in general, many pairs in his study shared each response number, so this objection is certainly pertinent. Greeno also pointed out that a closely spaced double presentation was of little value in Brown-Peterson studies in which only one item has to be learned at any one time; in this situation, the time-sharing hypothesis would certainly not apply, since there would be no other item to benefit from extra processing. Such an objection appears somewhat superficial, however, since different organisational strategies might well apply in the two paradigms.

There are two, more fundamental, objections to the time-sharing hypothesis in this context. In the first place, such a hypothesis would predict an improvement in performance on double items in unmixed lists over that on comparable items in mixed lists, as on average, the former items would receive 2 trial's processing per block, and the latter $1\frac{1}{2}$. Although Greeno made no attempt to statistically compare performance on double items in his 1964 study with that on double items in the unmixed list study, an examination of the data suggests that if anything, performance on double items was poorer in the unmixed list condition.

Admittedly, performance on single items in the unmixed list studies was poorer than that on single items in the mixed list condition, despite the fact that, if anything, inter-block forgetting should have been more pronounced in the mixed list study, when there were 30 trials per block as compared with only 15 in the unmixed list studies. However, a similar argument applies to performance on double items; there were 30 trials per block in the mixed condition (comprising 10 trials on single items and 20 trials on 10 double items, giving a total of 20 different pairs) as compared with 30 trials comprising 15 double items in the unmixed condition. Again, if anything, there should have been a more pronounced inter-block forgetting effect on double items in the mixed list condition.

Thus, if a between-study comparison of single item performance is accepted as valid evidence supporting the time-sharing hypothesis, then by the same token, the results of the mixed - unmixed list comparison of double item performance provide an equally valid basis for rejecting the hypothesis. Greeno (1970a) has pointed out that such inter-study comparisons may not have been reliable due to a number of small procedural differences between the various experiments. Therefore, there is little evidence to support the time-sharing hypothesis whichever way these inter-study comparisons are regarded.

The second objection to the time-sharing hypothesis has already been hinted at. If subjects applied a time-sharing strategy during the presentation of items that they regard as adequately processed, then unless a very strong form of Greeno's hypothesis applies (i.e. that subjects always time-shared on every second presentation of a double item) then there should have been an initial advantage in favour of double items in Greeno's (1964) study which would have disappeared as the number of blocks increased. In other words, Calfee's argument that subjects would also time-share on items that had been presented many times (and would therefore be regarded as adequately processed) should also be expected to apply. However, it has already been reported that Greeno found no evidence of such an effect.

Greeno (1970a) suggested an alternative hypothesis to account for his data. If the between-study comparisons are regarded as being unreliable, than the weight of evidence suggests that the time-sharing hypothesis is incorrect. It was postulated that subjects may occasionally need to "take a rest" from processing, perhaps in order to maintain the level of efficiency of processing over a substained period of time. The subject would clearly take such a rest when he considered that the item currently being presented had probably received adequate processing in the past. Greeno suggested that, in particular, items which received closely-spaced double presentations would in general still be held in short-term memory at the onset of the second presentation, and that, consequently, the subject would recognise them as items that had just recently been processed and would therefore not require processing again on the current trial. This would clearly account for Greeno's (1964) mixed list data.

This processing-attenuation theory can also be extended to account for the results of Peterson and Brewer and Calfee mentioned earlier. As the number of trial blocks increases, so the double items become better learned. Thus, a double item presented in a relatively late block may well be regarded as very well learned by the subject, and in need of no further processing. Hence in a relatively late block, double items would receive hardly any additional processing, whilst single and interference items would still be receiving a great deal of processing, and would therefore tend to "overhaul" double items as the number of blocks increased. Calfee's results could be explained by the argument that, as the number of unlearned items in the list decreased, so the number of learned items would increase, and the subject would be increasingly able to avoid resting from processing on trials involving unlearned items.

Of course, it follows from Greeno's version of the processing-attenuation hypothesis that multiple presentations would have a far larger beneficial effect on performance were they spaced out to the extent that on each presentation, short-term retention from the previous presentation would be "wiped out". There is a good deal of evidence that the spacing of repetitions does have a beneficial effect, as will be seen in the next section.

1.44 The Spacing of Repetitions

In an attempt to find a spacing effect such as Greeno predicted, Calfee (1968, Exp.I) employed a mixed - list anticipation procedure in which CVC - number pairs received one or two anticipation trials per block. The two presentations of double items were separated by 0,1,2 or 3 intruding trials on other items. This meant that inter-block and within-block spacings of double items would be confounded (since a shorter within-block spacing would imply a longer inter-block spacing), but Calfee argued that since 27 trials appeared in each block, the between-block differences were small relative to the within-block differences in spacing.

Calfee measured both inter-block retention, and the mean number of presentations required for an item to reach a criterion of three consecutive correct responses. Inter-block retention was found to be almost equal in the 1,2 and 3 spacing conditions, but performance in these conditions was superior to performance in the zero spacing condition, which was in turn slightly superior to performance in the single presentation condition. Both the time-sharing and processing-attenuation hypotheses could account for these results, and the fact that all non-zero spacings of double items produced comparable results supports Greeno's hypothesis that items still in short-term memory receive little or no processing when presented.

However, the trials-to-criterion data produced a surprise. Although, in the case of double items, this data followed an almost identical pattern to the inter-block retention data, it was found that single items reached criterion in fewer trials than double items. However, when these scores were converted into blocks-to-criterion, (by halving the scores of double items which received two trials per block), it was clear that the single items required more blocks to reach criterion. Of course, Calfee hypothesised that single items received more processing on later blocks because by then, all the double items had been learned, and such a hypothesis is clearly consistent with his results. However, these results could be explained both by a time-sharing and by a processing-attenuation theory, and on balance, the list-learning data does not appear to discriminate between the two theories.

In a further study, Calfee (1968) Exp.II, attempted to obtain a clearer picture of the effects of the spacing of repetitions by balancing out the rates of learning of the various types of item. He divided his material into two groups of six blocks each; spaced items occurred once in each block, whilst massed items occurred either twice in each second block or three times in each third block. Thus, over a group of six blocks, each item of any type appeared six times. Retention was tested after the sixth block and again after the twelfth. No consistent differences in performance were found between any of the item types. Spaced items and both kinds of massed item produced very similar performances on each retention test. Furthermore, no consistent differences were found in the rates of learning of the various types of item, as measured by a trials-to-criterion score.

Although Calfee failed to find a consistent effect of interpresentation interval, a number of studies exist which establish such an effect beyond all doubts. In a mixed list study, Izawa (1968) found that retention after eight repetitions was far better on an RNNNN schedule than on a simple R schedule. Although Calfee's objections regarding differential learning rates apply to this study, the differences in performance found in this study were relatively large. Furthermore, since 12-item lists were employed in the mixed design, it should be pointed out that in both conditions, interpresentation intervals would have been relatively large (averaging 11 trials on other items in the R condition, and around 55 in the RNNNN condition).

Apart from the differential learning rate effects present in such list learning studies, there is the added problem that such studies do not permit the experimenter to control retention intervals, and the fact that performance at any stage in such a study is averaged across a distribution of retention intervals may well account for the inconsistencies observed in their results. Fortunately, most of the really convincing evidence concerning the effects of the spacing of repetitions involves studies employing variations of the CPA procedure. In general, such procedures do permit precise control over retention interval and furthermore, the subject's learning load (in terms of the number of items currently to be remembered) usually remains fairly constant across each experimental session.

Bjork (1966) employed a CPA anticipation procedure in which there were 21 items with CVC stimuli, and the digits 3,5 and 7 as responses. Presentation sequences were constructed in such a way that all interpresentation intervals from 1 to 40 were equally likely to occur, and a different sequence was employed for each item. All subjects received the same sequences but with different items. The shortest sequence involved 12 repetitions, and the longest 29.

As has been described elsewhere, Bjork found that prior to the trial of the last error, performance immediately after a presentation was almost perfect, but it then rapidly fell away towards the guessing level of 1/3 (see 1.32 and Figure 2). However, in addition Bjork found more rapid learning and better performance at long retention intervals with those sequences which involved predominantly well-spaced, as opposed to closely-spaced presentations. Although Bjork's study confounded the number of learned and unlearned items with the number of presentations, this confounding was not so complete as in, say, a rigid list-learning situation, and in any case, he discovered improvement with interpresentation spacing relatively early in his lists. In fact, Bjork's procedure is fairly untypical of CPA experiments in general, when the number of items to be remembered and their overall state of learning remains roughly constant throughout the experimental session.

A number of CPA studies have employed a procedure whereby the retention interval is held constant, and items receive two presentations with various interpresentation intervals. For example, Peterson and Brewer (1963 Exp III) employed a study - test CPA procedure which made use of an interleaving process similar to that described earlier (see 1.31). Their stimuli were common, monosyllabic words of 3- or 4-letters, and responses were numbers in the range 1 - 9. Each item was allocated to one of four schedules, which may be represented as follows:-

$$P_1 - i - X - 30 - T$$

where P_1 represents the first presentation trial involving that item, T represents a final test trial involving that item, and X represents one of the four sequences described below. The first presentation (P_1) and the sequence X were separated by i trials, either presentations or tests, involving other items, whilst the sequence X and the final test T were separated by a similar interval of 30 trials.

The four conditions were defined as follows:-

I : information — X represents a test trial followed by an immediate presentation of the critical item i.e. X = TP

N : neutral — X represents two consecutive test trials on the critical item i.e. X = TT

D : double — X represents two consecutive presentations of the critical item i.e. X = PP

S : single — X represents two intruding trials on other items

Thus, S item formed a comparison condition, against which performance in the other three conditions could be measured. All trials were presented visually at a 2-second rate, and the interval i between P_1 and X took values of 1 or 4 trials on other items. The results of this study are presented in tables 2 and 3.

TABLE 2

Proportions correct at second presentation
(Peterson and Brewer, 1963 Exp.III)

	Interval i.	
Condition	1 trial	4 trials
Information (I)	.82	.24
Neutral (N)		
First test	.77	.27
Second test	.79	.30

TABLE 3

Proportions correct at final test
(Peterson and Brewer, 1963, Exp III)

		Interval i.	
Condition	Second occurrence (X)	1 trial	4 trials
Information (I)	correct	.34	.70
	wrong	.16	.22
Neutral (N)	Both correct	.31	.70
	Both wrong	.02	.06
Double (D)	Two presentations	.31	.36
Single (S)	None	.19	.21

In the neutral condition (X = TT) it is clear from table 2 that there was a slight reminiscence effect from one test to the next, similar to that observed by Peterson, Saltzman, Hillner and Land (1962, Exp. II - See 1.32). Performance at final test in this condition following two correct responses at X (=TT) was similar to that in the information condition following a correct response at X(=TP), so that the second test appeared in this case to fulfil the same role as a re-presentation.

Overall proportions correct at final test in the information condition (irrespective of whether a correct or a wrong response was given just prior to the second presentation) were .31 and .34 when the interval i was 1 and 4 trials respectively. These figures compare with those obtained for retention at final test in the double condition (See Table 3), so that, on the whole, it appears that the two massed re-presentations in the double condition were of no more value than the single re-presentation in the information condition. Both sets of data, however, show a slight improvement in final test performance with an increase in the interval between the first and subsequent exposures, or in other words, a spaced practice improvement (SPI).

A glance at Table 2 should convince the reader that rapid short-term forgetting occurred, following the initial presentation, between retention intervals of 1 and 4 trials. Thus, in the information condition, a considerable proportion of the items that were correctly recalled just prior to the second presentation at lag 1 would have been held in short-term memory. On the other hand, the majority of the items correct at lag 4 just prior to the second presentation would probably have been recalled from memory proper (or long-term memory). Thus, the startling differences in performance with interpresentation lag at final test following an initial correct response in the information conditions (see Table 3) are certainly consistent with the hypothesis that items held in short-term memory when re-presented receive no further processing and so subsequently decay. However, once again, the data does not tell us whether the subject employs the additional processing time made available by ignoring such items on re-presentation, or not.

It is also of interest to note that performance at final test in the information condition following a wrong response just prior to re-presentation showed a far smaller SPI than that following an initial correct response. Furthermore, performance following an initial error was roughly comparable to that on singly -presented items tested at a slightly longer retention lag. It is postulated that the small SPI effect observed in this case may have resulted from items that were available but not accessible at first test (and such items would similarly account for the small reminiscence effect described above in the neutral condition) . However, the negligible size of the reminiscence effect suggests that the majority of the items that were wrong at first test were simply not available, and the data imply that upon re-presentation, such items produced equivalent performance to brand-new single items. In other words, the majority of the pairs that were recalled wrongly just prior to re-presentation were essentially equivalent to brand-new items receiving their first presentation.

Unfortunately, the small range of interpresentation intervals investigated in this study doe not give a very clear picture of the extent of the SPI effect in paired-associate memory. However, Peterson, Wampler, Kirkpatrick and Saltzman (1963, Exp.1) conducted a similar CPA study in which common word-number pairs received two presentations separated by 0,1,2,4,8 or 16 intruding trials on other items, and were tested at a retention interval following the second presentation of 8 such trials. The results of this study are depicted in Figure 8, along with data from a similar study by Young (1966) described below.

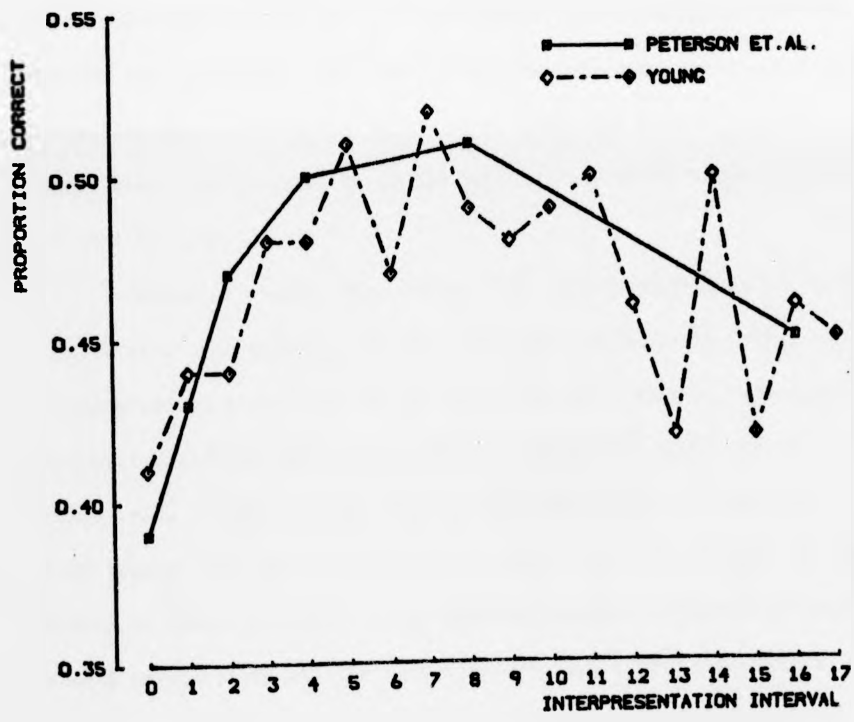
Young (1966) employed an interleaved study-test CPA procedure, in which stimuli were consonant trigrams (CCC's) and responses were numbers in the range 0-9. All trials (both study and test) were presented visually at a 4-second rate. Two types of schedule were employed. In the first of these (PPT) an item received two presentations (i.e. study trials) separated by an interpresentation interval of from 0 to 17 intruding trials (both study and test) on other items; retention was investigated on a test trial which occurred after a retention interval of 10 intruding trials on other items following the second presentation. The second type of schedule (PTPT) differed from that above only in that an additional unreinforced test trial was intruded between the two successive presentations of an item, at various positions in the interpresentation interval. The proportions of items correctly recalled at final test in the PPT condition as a function of interpresentation interval are depicted in Figure 8, along with the results of the comparable items examined by Peterson, Wampler, Kirkpatrick and Saltzman. The figure suggests that in both studies, retention performance following a fixed retention interval after the second presentation showed a rapid improvement with an increase in interpresentation spacing from zero to about eight trials. With increased spacing, however, there is a suggestion that performance declines again (although neither investigator actually found a statistically significant decline). However, these results have been shown to reliably establish an SPI effect.

Young's retention data following a single presentation have been discussed elsewhere (see 1.32 and Figure 1) although it is interesting to note that in this case, short-term retention effects appear to have dissipated after about 2 trials, whereas the SPI effect appears to continue well beyond such an interval. In other words, it would seem that although Young's data are consistent with the hypothesis that items in short-term memory do not receive additional processing when re-presented, performance may continue to improve beyond the range of inter-presentation intervals that an explanation solely in terms of such a short-term memory hypothesis would predict.

The Peterson and Brewer (1963 Exp. III) data examined above would suggest that an examination of performance following a correct response to a first test immediately prior to the second presentation would provide a far more sensitive way of looking at the SPI effect, and such data is available from a number of PTPT items (which had their first test just prior to the second presentation) in Young' study. Unfortunately

FIGURE 8.

Retention of paired associates as a function of the spacing of two study trials.



PETERSON ET.AL.
YOUNG
PROPORTION CORRECT
0.55
0.50
0.45
0.40
0.35
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
INTERPRESENTATION INTERVAL

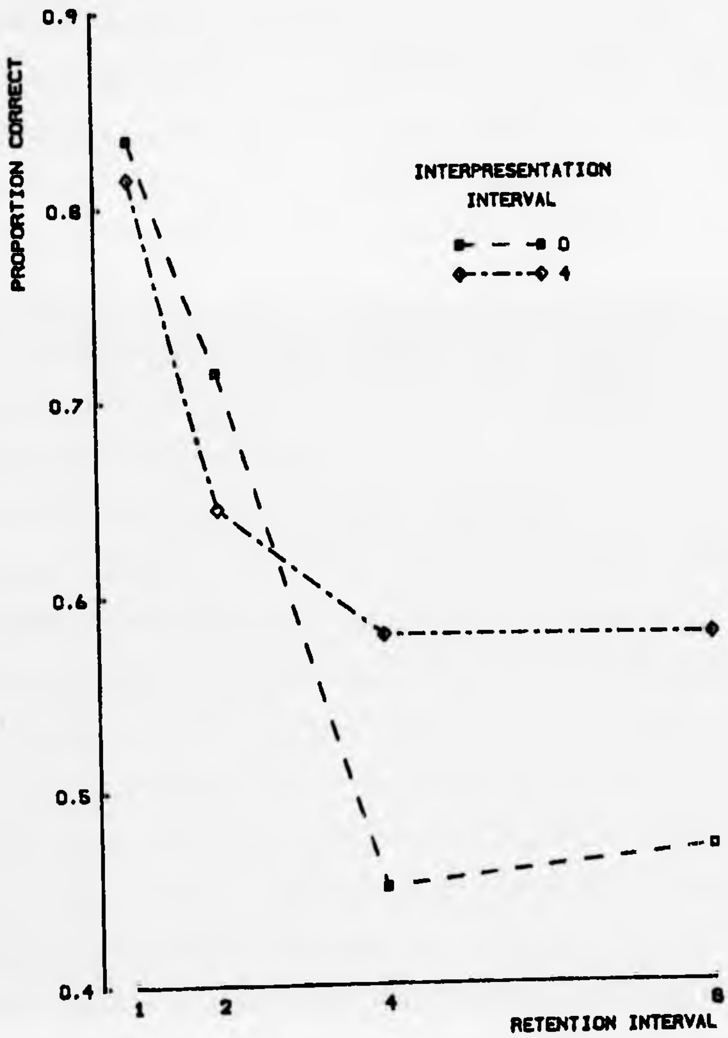
Young tested insufficient replications of such items to generate data stable enough to give a reliable picture of the extent of the SPI effect.

Indeed Young's examination of performance in all his PTPT conditions was handicapped by an inability to find reliable trends in performance at final test conditional on correct or wrong responding on the first test, for this reason. However, taking marginal performance at the final test, it was found that in general, PTPT items resulted in superior performance to comparable PPT items, and that furthermore, performance at final test on PTPT items was generally better if the first test occurred in the middle of the interpresentation interval rather than at one end or another. The result that an intruded unreinforced test trial improved performance following a subsequent presentation is consistent with the results of the various studies by Izawa discussed in section 1.42.

A number of studies have varied both interpresentation and retention intervals. For example, Peterson, Hillner and Saltzman (1962) employed an interleaved study-test CPA procedure in which stimuli were common, monosyllabic 3- and 4- letter words, and responses were numbers in the range 1-10. Items received two presentation trials followed by a test trial. The interpresentation interval consisted of zero or four intruding trials on other items, and the retention interval between the second presentation and the test comprised 1,2,4 or 8 such trials. All trials were presented visually at a 2-second rate. The results of this study are presented in Figure 9. These data clearly demonstrate an interaction between interpresentation and retention intervals: short interpresentation intervals are better when the retention interval is relatively short, and long interpresentation intervals produce superior performance at long retention intervals. A similar result has been established for word-word pairs presented at a 2-second rate, although when such pairs were presented at a 4-second rate, both short and long interpresentation intervals led to roughly equal performance at short retention intervals (Peterson, Wampler, Kirkpatrick and

FIGURE 9

Paired-associate retention as a function of retention interval and of the spacing between two successive presentations. (Peterson, Hillner and Saltzman 1962)



PROPORTION CORRECT
0.9
0.8
0.7
0.6
0.5
0.4
INTERPRESENTATION
INTERVAL
0
4
1
2
4
8
RETENTION INTERVAL

Saltzman 1963, Exp III). Rumelhart (1967) found an analagous interaction between spacing and retention intervals in a CPA task in which each individual pair was given six anticipation trials separated by various sequences of inter-trial intervals.

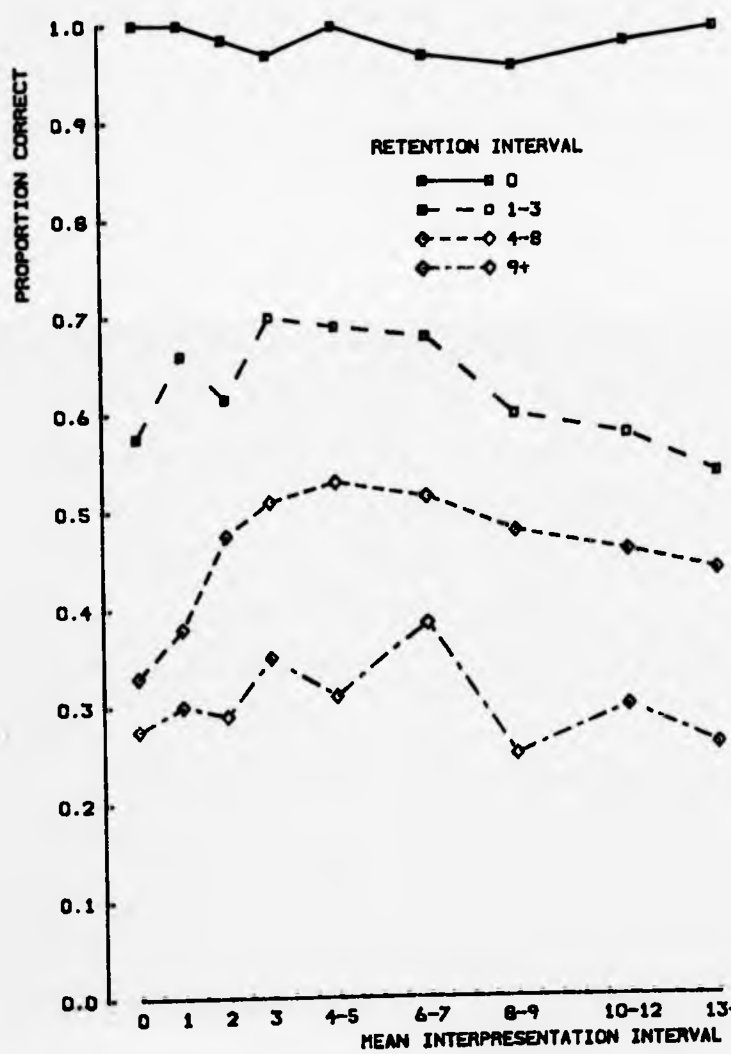
Brelsford, Shiffrin and Atkinson (1968) employed a modified CPA anticipation procedure in which stimuli were eight randomly-selected two-digit numbers, and responses were letters of the alphabet. Their lists were constructed in the following way. Each of the eight stimuli was randomly paired with a response, and each pair so formed was then allocated a 1-,2-,3-, or 4- reinforcement schedule with probabilities of 0.3, 0.2, 0.4 and 0.1 respectively. When a pair had received its final presentation, it was subsequently tested in the normal way, but the study phase of the anticipation trial in which the final test occurred was used to present a new pair, consisting of the stimulus just tested paired with a new response (i.e. one that it had not been paired with earlier in the session). The new pair so formed was assigned a reinforcement schedule in accordance with the probability distribution described above. Thus, each of the eight stimuli occurred many times during the session, paired with a number of different responses.

The stimulus that would be involved on each anticipation trial was determined at random, so that the interval (in terms of intruding anticipation trials involving other stimuli) between successive trials on the same stimulus was geometric, with a parameter of $\frac{1}{8}$. Each (visually-presented) anticipation trial lasted for 11 seconds, and comprised a 3-second test phase, followed by a 2-second blank period, a 3-second study phase and finally a further blank period of 3 seconds. Retention data from this study averaged across the various interpresentation intervals has already been described elsewhere (see 1.43 and Figure 7).

Retention data on pairs which had received their first two presentations are presented in Figure 10, as a function of retention interval and of interpresentation interval. Various intervals have

FIGURE 10

Effect of retention interval and the spacing between two presentations on paired-associate memory (Brelsford, Atkinson and Shiffrin , 1968)


PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
RETENTION INTERVAL
0
1-3
4-8
9+
0
1
2
3
4-5
6-7
8-9
10-12
13+
MEAN INTERPRESENTATION INTERVAL

been averaged together in producing this data, due to the smaller numbers of observations at long lag. It is quite clear, however, that these data do no exhibit the kind of interpresentation - retention interval interaction described by Peterson, Hillner and Saltzman (1962) discussed above. This may be due either to the fact that data from very short non-zero retention intervals are averaged in with data from longer ones, or to some difference in the subjects' short-term rehearsal strategy. It should be pointed out that that rate of presentation was remarkably slow in the present study, and that furthermore, the results obtained are similar in form to those obtained by Peterson, Sampler, Kirkpatrick and Saltzman (1963, Exp. III) when using a slower rate than that employed by Peterson, Hillner and Saltzman. Thus, presentation rate appears to affect the interaction between interpresentation and retention interval. Otherwise, the data for non-zero retention lags resemble very closely those of other investigations discussed above. In other words, for a given non zero retention interval, performance improves with interpresentation spacing. Again, the data suggest that there is some limit to this improvement; beyond an interpresentation spacing of 4 to 5 anticipation trials, performance appears to decline slightly.

A comparison of the curves in this figure with the short-term retention curve (following a single presentation) obtained in the same study (See Figure 7) again implies that items still held in short-term memory when they are re-presented do not receive much benefit from the second presentation. However, the data do no appear to discriminate between the hypothesis that the time made available during the second presentation of such items it used to process other items, and the hypothesis that this free time is used to "take a rest" from processing. Furthermore, the single-presentation retention data suggest that very few items will still be held in short-term memory when re-presented at any interpresentation interval greater than zero. Consequently, the continued improvement in performance observed with increased interpresentation spacing appears to be maintained well beyond the interpresentation spacing that an explanation solely in terms of the short-term retention hypothesis would predict.

A study by Landauer (1969) suggests that the improvement in performance achieved with interpresentation spacing may well be maintained over extremely long retention intervals. Landauer employed 4 lists, each consisting of 12 pairs : one list comprised nonsense syllable-integer pairs, a second employed body - part stimuli and colour-name responses, a third adjective stimuli and consonant letter responses, whilst the final list comprised common first names paired with the months January to December. Within each list, some pairs were presented only once, and others received two presentations separated by intervals of 0,1,3 or 5 trials on other pairs.

The lists were presented to subjects in booklet form, one presentation to a page. Subjects read the booklet at a paced rate, in that the pages were turned at a 2 -second rate in time with a metronome. A test was administed by having subjects go through a booklet of stimuli at a paced 5-second rate, circling the desired response in a list of all twelve possible responses. One test of retention was administered after an interfering free-recall task, at an average retention interval of 3 minutes, and a further test was administered after 3 days. Landauer found no essential differences in the patterns of performance on the various item lists; the proportions of correct responses observed, aggregated over the four lists are presented in Table 4.

TABLE 4.

Proportions of paired associates correctly recalled (Landaver, 1969).

Retention Interval	Single Items	Double items - interpresentation interval 0	1	3	5
3 minutes	.51	.54	.58	.67	.66
3 days	.37	.39	.42	.46	.47

Double presentations were clearly superior to single presentations at both retention intervals. Furthermore, at both retention intervals, performance improved with interpresentation spacing up to an asymptote at about 3 intruding trials ; no additional improvement was observed at a spacing interval of 5 trials.

Finally, mention must be made of a CPA study by Bjork and Abramowitz (1968) in which some paired-associate items received sequences of four anticipation trials. The first and third trials were separated by an interval of 21 anticipation trials, 20 of which were intruding trials on other items, and one of which was the second anticipation trial on the current item. Retention was measured on the fourth trial of each item sequence; this followed the third trial after a retention interval of 2, 8 or 20 trials on other items. The investigators were principally interested in the effects on retention on the fourth trial of the positioning of the second presentation of an item between its first and third presentation.

It was found that at all retention lags, performance was optimal at final test when the second presentation fell half way between the first and third presentations. This result is clearly analagous to Young's (1966) finding with an unreinforced test trial similarly intruding between two presentations. In addition, it was found that at all retention intervals, performance at final test remained roughly the same if the intervals between the first and second, and between the second and third presentations were interchanged. In other words, when performance was measured at a fixed retention interval following three presentations, the first and third of which were always 21 trials apart, a second presentation intruded between the first and third had the same effect whether it was nearer the first presentation, or equally near to the third presentation. The spacings between the first and second, and between the second and third presentations were thus found to be commutative in their effects on subsequent performance. Unfortunately, it is not known whether such an effect holds for all intervals between the first and third presentation.

1.5. Summary

Following the development of a tentative theory of paired-associate forgetting (in 1.33) based on a careful consideration of proactive and retroactive interference effects, a number of phenomena concerned with the effects of practice on paired-associate memory have been described. The most important effects described may be summarized as follows:

1) Allowing subjects additional time in which they can potentially process items to be remembered generally results in improved performance. This suggests that subjects process paired-associate items in real time.

2) Unreinforced retention tests (i.e. where no immediate feedback is given to the subject as to the correctness or otherwise of his response), appear both to retard the rate of forgetting, and to render subsequent presentations more valuable.

3) Repeated presentations of a pair appear to have little value if they occur one after the other in immediate succession.

4) However, if such repeated presentations are spaced out with intruding trials on other items between them, subsequent performance is improved.

5) If performance is measured after very short retention intervals, then massed presentations appear to be no worse then spaced presentations. Indeed, if rapid presentation rates are employed, massed presentations may produce superior performance at short retention intervals.

6) There appears to be some limit to the improvement of performance with spacing; as the interval between two presentations increases, performance appears to improve to a point, and thereafter no further improvement, and perhaps a decline, is observed.

7) It appears that the majority of items which are currently held in short term memory when re-presented receive little benefit from that additional presentation. It is not clear, however, whether such items receive processing on that presentation or not. If it is assumed that such items are not processed, it is possible that the subject may use their

presentation trial either to process other items, or to take a rest from processing altogether.

8) Similar observations may be made regarding items which have received many presentations, and may be regarded by the subject as adequately learned on their next presentation.

9) There is some evidence that performance continues to improve with interpresentations spacings far in excess of those which would be sufficient to "wipe out" short-term retention of items at their second presentation.

10) When three presentations are given, it appears that performance is optimal when they are equally spaced. Unequal spacings appear to be commutative in their effects on subsequent performance.

The next step is clearly to extend and refine the proposed theory to take account of these practice effects, which are of obvious and fundamental importance to our understanding of paired-associate memorizing and learning, and then to devise experimental tests of the theory. However, these practice effects, and in particular the spaced practice effect, are not unique to paired-associate tasks. Psychologists have known about the beneficial effects of spaced practice in a variety of human learning situations for nearly a century. Furthermore, spaced practice improvements have been found more recently in a variety of memory tasks.

Thus, all these results must be carefully examined and taken into account before attempting to derive an adequate theory of paired-associate memory. In Chapter Two, a brief history of the spaced practice effect will be presented, whilst Chapter Three examines contempory results concerning repetition and practice in a variety of different memory situations. The task of postulating a theory, or a range of theories, which take account of the results outlined above will be returned to in Chapter Four.

CHAPTER TWO

THE SPACING OF PRACTICE IN HUMAN LEARNING

- A BRIEF HISTORY

As was pointed out at the end of the previous chapter, psychologists have long realised that performance in a variety of learning tasks is often improved if an interval is allowed between successive repetitions and practice trials. Of course, many psychologists today would claim that a human learning approach is far too broad to establish really meaningful results; however, there is certainly a very real danger of approaching the SPI effect in paired-associate memory from too specific a standpoint. This brief review will attempt to establish the major findings in the area of the spacing of practice of traditional human learning theorists.

2.1 Jost's Law

Perhaps the earliest discovery that performance could be improved by spacing practice trials was made by Ebbinghaus (1885), who found that he could memorize lists of nonsense syllables and stanzas of Byron's "Don Juan" in fewer readings if three days were allowed to elapse between successive practice sessions rather than one day.

Jost (1897) found that subjects could master paired-associate lists in fewer repetitions if the lists had been part-learned on the previous day, as opposed to lists that had been part-learned a few minutes prior to the learning session, despite the fact that only 5% of the items were correctly recalled at the start of the session in the former condition as opposed to 40% in the latter. Jost summarized these and other similar findings in the following hypothesis, often known as Jost's Law: "If two associations are now of equal strength but of different ages, then further study will have greater value for the older one". Youtz (1941) has produced a notable review of many studies which lend experimental support to Jost's Law.

When viewed in the light of the increased amount of forgetting that would take place with the increase of inter-trial spacing, results of this nature were very surprising indeed. In the light of their fundamental importance to the understanding of learning, it was inevitable that psychologists would make an intensive effort to gain a better understanding of the role of interpresentation spacing in the learning process. Two basic experimental procedures were commonly employed to this end. In the first of these, single practice sessions were separated by various lengths of time, whilst the second procedure consisted of holding the time interval between successive practice sessions constant and systematically varying the number of practice trials per session.

2.2. The nature of the learning task

Generally speaking, it was found that a wide variety, and indeed the great majority of learning tasks yielded results which favoured the distribution of practice. Thus, for example, Calvin (1939) found that, for the acquisition of conditioned responses, 3 trials per minute were superior to 9 or 18 per minute. A similar results was obtained by Humphreys (1940) ; two blocks of 48 trials with an interspersed rest period yielded better results than a single block of 96 trials without a rest period. As has been indicated above, studies of verbal learning generally favoured the distribution of practice. To the studies of Ebbinghaus and Jost can be added those of Bumstead (1943) and of Hovland (1938, 1939 and 1949).

The majority of studies involving perceptual-motor learning tasks also favoured the distribution of practice, with the following reservation : when equal numbers of learning trials were employed in each practice session, results often indicated an optimal length of rest pause between successive sessions, beyond which performance either stabilised or actually declined. For the pursuit rotor task, using 1 - minute trials, Dore and Hilgard (1937) found that 11 minutes between trials was better than 3 minutes, which was in turn better than 1 minute. Lorge (1930), using a mirror drawing task, found that both

1-minute and 1-day intertrial rest periods were better than completely massed trials, but found no difference in the two conditions. Kientzle (1946) obtained similar results with an inverted alphabet task; performance improved as the intertrial rest period increased up to about 1 minute, beyond which duration no further improvement in performance was obtained.

In a pursuit rotor task with an oscillating target and 5-minute practice periods, Travis (1937) found that 20-minute rests gave the most rapid trial-to-trial improvement in performance, 5-minute rests next, and rests of 2 days the least rapid improvement. It would seem likely that had a more comprehensive range of rest periods been employed in the motor-learning experiments above, a similar pattern of performance improving with rest periods up to an optimal spacing, and then declining, would have emerged. A curious result that also belongs in this section was observed by Warden (1923). In a maze experiment with rats, he found that the maze was learned in fewer trials if a 12-hour intertrial rest period was employed, than in conditions with a 6- or 24- hour intertrial rest period.

An analagous result was found using the alternative experimental design; when rest pauses were kept uniform there was generally an optimal length of practice period (or number of trials per practice session) peculiar to each task. Thus, for example, Pyle (192[illegible]) found that in a substitution task, a 30-minute practice period was optimal, being superior to either a shorter (15-minute) or longer (40-or 60-minute) session. Snoddy (1945) found that, in learning to trace a star pattern reflected in a mirror, subjects improved more rapidly over trials if 1 trial per day was given, as opposed to a group of 10 trials every 2 days.

2.3 Factors favouring massed practice

It was found that in some situations, massed practice yielded a faster rate of improvement per trial than did distributed practice. Bell (1942) has shown this to be the case when a period of time is required to "get set" or "warm up" to a task. If the amount of material to be learned is very small, then massed practice may be superior to distributed practice. This was demonstrated by Lyon (1917) who needed less time to memorise a list of 12 digits in continuous readings than in one reading per day. With longer lists, the advantage shifted to daily reading, and became a very great when the lists were very long (100-200 digits). He hypothesised that the important factor here was probably the effect of short-term memory, or perhaps covert rehearsal. Very little short-term forgetting would occur during massed readings of a short list, and it is also possible that rehearsal capacity would have improved due to a warm-up effect in a few massed readings.

Somewhat surprisingly, a similar result was found to hold for rats learning mazes of different lengths. It was found that short mazes were learned in fewer massed than distributed trials, whilst larger mazes were learned in fewer distributed trials (1 per day) than massed trials (Pechstein, 1921 and S.A.Cook, 1928).

A "spider maze", with six alleys, five of which were blind, at each choice point was employed by T.W. Cook (1944). His human subjects learned it much more quickly in massed trials than with one trial per day. A similar result was obtained using a "mental" maze, which offered six choices at each choice point (subjects had to discover by trial-and-error which of the numbers 1-6 was correct at each choice point.) These results were explained in terms of the serious consequences of forgetting in such a task.

A further factor which may have influenced Cook's results was demonstrated by Erickson (1942), who used a puzzle box that allowed for a great variability of attack. This was learned more quickly with massed practice; Erickson produced evidence to show that distributed practice in such a task tended to produce a fixation of response , whereas massed practice resulted in a greater variability of response.

required to "get set" or "warm up" to a task. If the amount of material to be learned is very small, then massed practice may be superior to distributed practice. This was demonstrated by Lyon (1917) who needed less time to memorise a list of 12 digits in continuous readings than in one reading per day. With longer lists, the advantage shifted to daily reading, and became a very great when the lists were very long (100-200 digits). He hypothesised that the important factor here was probably the effect of short-term memory, or perhaps covert rehearsal. Very little short-term forgetting would occur during massed readings of a short list, and it is also possible that rehearsal capacity would have improved due to a warm-up effect in a few massed readings.

Somewhat surprisingly, a similar result was found to hold for rats learning mazes of different lengths. It was found that short mazes were learned in fewer massed than distributed trials, whilst larger mazes were learned in fewer distributed trials (1 per day) than massed trials (Pechstein, 1921 and S.A.Cook, 1928).

A "spider maze", with six alleys, five of which were blind, at each choice point was employed by T.W. Cook (1944). His human subjects learned it much more quickly in massed trials than with one trial per day. A similar result was obtained using a "mental" maze, which offered six choices at each choice point (subjects had to discover by trial-and-error which of the numbers 1-6 was correct at each choice point.) These results were explained in terms of the serious consequences of forgetting in such a task.

A further factor which may have influenced Cook's results was demonstrated by Erickson (1942), who used a puzzle box that allowed for a great variability of attack. This was learned more quickly with massed practice; Erickson produced evidence to show that distributed practice in such a task tended to produce a fixation of response, whereas massed practice resulted in a greater variability of response.

Similar results were obtained by Garrett (1940).

Of course, there is always a possibility that studies reporting the superiority of massed practice are invalid because the spacing in the distributed practice condition was too long, when, in fact, a short spacing between trials would have been advantageous, as has been shown for some motor-learning tasks.

2.4 Factors favouring distributed practice

Undoubtedly, many of the results that favour distributed practice, especially those of motor-learning techniques, are attributable to fatigue and work decrement, rather than to learning factors. Hull (1943) introduced the concept of "reactive inhibition" in an attempt to bridge the gap between work decrement and learning principles. He postulated that every effortful response, whether reinforced or not, produced a tendency to avoid a repetition of that response, with the amount of inhibition becoming greater the more effortful the response. Hull believed that this reactive inhibition would dissipate as a simple decay function of the time allowed for rest.

The concept of reactive inhibition has important implications; differences in performance may not necessarily mirror differences in the amount learned; in ability, knowledge, or what Hull called "habit strength" and designated by the symbol $S^{H}R$. Whether one agrees with Hull's theories or not, this point is still valid, and it is important to discriminate between performance (which as observed in a function of $S^{H}R$ and inhibition I, in Hull's terminology) and learning. This point is beautifully illustrated by a study conducted by Kimble and Shatel (1952).

A pursuit rotor task was employed, in which two groups of subjects received trials of 15 seconds duration, over a 10-day period, 15 trials being given per day. The "massed" group had only a 5-10 second rest between trials, whilst the "spaced" group received an intertrial rest period of 65-70 seconds. During each day's work, the performance of the massed group was found to lag far behind that of the spaced group. It was discovered, however, that on the first and second trials of each day's work, the massed group, although still inferior in performance, had nevertheless significantly "caught up" with the spaced group in comparison with the difference in performance levels at the end of the previous day's work. This result would suggest that some kind of fatigue or inhibition was accumulating during each day's work, especially for the massed group, and dissipating with the passage of time to the start of the next day's session, and that furthermore, the difference in actual learning between the two groups was not as great as straightforward performance measures would suggest. Nevertheless, some difference was still apparent after a 24-hour rest, which should probably have been ample to dissipate any fatigue effects in performance, and probably most of Hull's reactive inhibition, which demonstrates that these effects were certainly impeding learning for the massed group. Similar results were obtained for the pursuit rotor task by Adams (1952), for the inverted alphabet task by Kientzle (1949), and for a substitution task by Epstein (1949).

In many verbal learning situations, the superiority of distributed practice may be due to extra practice in the form of covert rehearsal during the rest pause. This can be prevented by filling the interval with controlled activity of some kind, provided this does not interact with the task itself. In many such situations (such as the experiments of Hovland reported below) distributed practice is still beneficial, which shows that conscious rehearsal is not likely to provide a sufficient explanation of the superiority of distributed practice.

Hovland (1938) believed that distributed practice was favoured by the dissipation during rest pauses of interferences built up during the practice sessions. Nonsense syllables were presented at two different rates, either 4 seconds or 2 seconds, in an attempt to isolate this factor. Massed and distributed practice were compared under the two conditions, and it was found that although overall performance was superior at the slower rate, the advantages of distributed practice were greatly reduced. Hovland argued that at the slower rate, greater processing time was available to subjects, and interference was thus less likely to occur, which accounted for the overall improvement in performance. However, as less interference was present, less advantage was gained with distributed practice, as the decay of interference was a much less significant factor.

In another study, Hovland (1939) compared the effects of massed and distributed practice on a serial learning task, and on a paired-associate learning task. The anticipation procedure was employed in both cases; the serial list appeared in a fixed order, of course, whilst the paired-associate list was re-randomised on each trial. It was found that the advantage of distributed practice was more marked with the serial material when performance was measured in terms of the number of trials taken to reach an errorless criterion run. Again, Hovland applied an interference hypothesis, claiming that interference would be more resistant to decay with the fixed order serial-learning list. He also found (Hovland 1949) that in paired-associate learning the advantages of distributed practice were more pronounced with a faster rate of presentation, and again accounted for this result in terms of within-list interference. At faster presentation rates, within-list interferences would have less time to decay during presentation trials, and so consequently the advantages of distributed practice would be more pronounced.

Motivational hypotheses may also be advanced to explain the superiority of distributed practice. Prolonged practice might well result in reduced motivation (due to boredom, etc) with attendant reduction in performance. Unfortunately, hypotheses of this kind appear to result in identical predictions to Hull's inhibition theory, and it is therefore practically impossible to discriminate experimentally between the two approaches. However, all three factors of reduced motivation, increased inhibition and increased interference should tend to operate in the same direction when the quantity of the material to be learned is increased. Thus, in a verbal learning situation, one would expect the advantage due to distributed practice to increase with list length. This has been shown to be the case by Lyon (1914) and by Hovland (1940).

2.5. Summary

At first sight, there appear to be substantial similarities between the effects of the spacing of practice trials in traditional learning studies and the effects of the spacing of presentations on paired-associate memory. For example, in both cases there appears to be an optimal spacing period beyond which performance not only shows no further improvement, but actually declines. In both cases, this result could be accounted for in terms of the increase in inter-trial forgetting with very long inter-trial spacings.

However, there are some important differences. In conditioning and perceptual-motor studies, practice trials are separated by rest periods: intervals of time during which the subject's activity is not controlled by the experimenter. It is quite possible that human subjects in perceptual-motor studies may have been able to use these rest periods to practice some aspects of the task, perhaps by employing some kind of active attentional procedure analagous to rehearsal (although it is certain that animal subjects in conditioning studies were unable to fruitfully use these rest periods in such a way!) Nevertheless, it is true in general that performance in perceptual-motor and conditioning tasks declines across a time interval between practice and test, so that in both learning and memory studies, the spacing interval is filled with some kind of activity that usually results in forgetting. However, the superiority of spaced practice in these learning studies was explained in terms of dissipation of muscular fatigue and reactive inhibition across the spacing interval, and the hypothesis that, with spacing, a higher level of motivation might be maintained than in massed practiced trials. Clearly, none of these explanations is valid in the case of the spacing of presentations in paired-associate memory, since both massed and spaced items are equally likely to occur in such a situation in any part of the experimental session, and the factors above would only affect performance across the session.

In those situations where massed practice was superior, it was postulated that short-term forgetting between repetition might be responsible, in that such an effect would counteract the benefits of spacing. However, in paired-associate memory studies, performance improves with interpresentation spacings which guarantee short-term forgetting from one trial to the next. Indeed, the beneficial effect of interpresentation spacing in paired-associate memory may well be enhanced by short-term forgetting between successive presentations! It was also suggested that massed presentations often lead to a greater "variability of attack" in that the subject would be more likely to vary his learning strategy in such a situation. The weight of evidence from paired-associate studies suggests, however, that with massed presentations, the subject rather ceases his "attack" on the current item.

Hovland certainly controlled his subject's behaviour during the intertrial period to the extent of requiring them to perform some task that would preclude conscious rehearsal. However, his studies were very different in nature to those described in Chapter One. In Hovland's case, the item to be learned was an entire list rather than a single paired-associate, and the interval between successive presentations of this item was filled with activity on a task that did not conflict with his learning material. In contrast, in the paired-associate memory experiments described in Chapter One, the interpresentation intervals were filled with items to be learned that certainly did conflict with the critical item. Consequently, Hovland's explanation of the superiority of spaced practice in terms of the dissipation of interferences across the spacing intervals certainly does not apply to the effects of within-list spacing of successive presentations of items. If anything, more "interferences" should be built up with spacing in the latter case.

Hovland's studies are far more similar to the Brown-Peterson procedure, in that retention and interpresentation intervals consist of rehearsal-precluding activity that does not conflict with the material to be remembered. However, in Brown-Peterson studies, the items to be remembered usually consist of subspan word lists (such as noun trigrams), not great complex paired-associate or serial recall lists of the kind Hovland employed. The fact that performance in Brown-Peterson studies also improves with spacing (as will be seen in Chapter Four) casts doubt on Hovland's interference hypothesis, since intra-list interference effects in a 3-noun list must be negligible. It should also be pointed out that, in addition to the untenability of classical interference theory for reasons described elsewhere (see 1.24), Hovlands interference hypothesis suffers from a logical inconsistency, in that, if interferences are associative, and memory traces are associative, then why should interfering associations be forgotten more rapidly over an interpresentation interval than those appropriate to correct responding?

Thus, in conclusion, it appears that on the whole, classical learning studies do not contribute much to our understanding of the superiority of spaced presentations of paired-associates within a list of other pairs to be memorized. The early verbal learning studies of Hovland differ so much in procedure as to be irrelevant, and indeed his tasks were so complex that it is unlikely that a satisfactory explanation of his results will ever emerge. It is still possible, however, that the superiority of spaced presentations of paired-associate items within a list reflects some underlying basic property of verbal memory common to performance in other tasks. Consequently, the effects of the spacing of presentations in a variety of memory situations will be examined in the next chapter.

CHAPTER THREE

THE SPACING OF PRESENTATIONS IN MEMORY TASKS OTHER THAN PAIRED ASSOCIATES

Performance improves with the spacing of presentations of an item to be remembered in a number of memory tasks. Recent research has tended to concentrate on three basic areas besides paired-associate memory: namely free recall, recognition memory, and Brown-Peterson studies. In this Chapter, the effects of interpresentation spacing in these situations will be examined.

3.1 Free Recall

The experimental paradigm that has received most attention in recent research is almost certainly the free-recall procedure. Before describing results in detail, an outline is given of the basic experimental procedures employed.

3.11 Experimental procedures

In a normal free-recall procedure, a list of verbal items is presented, one at a time, to the subject, who then, upon completion of the list, attempts to write down as many of the items just presented as he can, in any order he wishes. A fundamental drawback to this procedure is the lack of experimenter control over the retention interval; for example, subjects may, if they wish (and frequently do) report items from the end of the list before earlier ones. This means that list position is frequently confounded with the retention interval, in addition to the numbers of previously- and subsequently-presented items.

The procedure employed for studying the effects of interpresentation spacing was first developed by Waugh (1963). Within a single presentation of the list, a number of items are repeated. These repetitions may occur in successive list positions, known as a massed practice (MP) schedule, or in list positions separated by presentations of other items, giving a distributed practice (DP) schedule. Distributed or spaced items may be tested under a variety of conditions defined by interpresentation interval (or "lag"); in other words, by the number of other items that are presented between successive presentations of the particular item. These intruding items are usually presentations of other critical items, and the lists are normally constructed employing an interleaving procedure similar to that described for CPA designs (See 1.31).

Of course, free-recall curves (which plot the proportion of items correctly recalled against list position) typically show marked primacy and recency effects. Thus, items from early list positions or from relatively late ones are recalled better than items in the middle portion of the list. The recency effect has been discussed elsewhere (1.22) whilst the primacy effect is often explained either in terms of the additional "active" rehearsal that early items receive in comparison with later ones, or in terms of the greater distinctiveness of episodic cues available at presentation (See 1.33 for a similar explanation of this effect in paired-associate probe lists). In order to avoid contamination of results by these effects, studies on spacing usually employ a number of dummy (unanalysed) items in the early and late list positions. The critical (analyzed) items are thus all presented in the middle portion of the list, where performance is relatively unaffected by list position.

Thus, studies involving the spacing of presentations in free-recall employ an analagous procedure to those concerned with such effects in paired-associate memory. In both cases, the presentations of items in a list are separated by presentations of other items to be remembered. Hence at first sight, one might well expect spacing phenomena in both instances to be very similar, and to reflect identical underlying memory processes.

3.12 Some negative results.

Although it generally appears that the relative improvement in recall performance with distributed practice (DP) in free recall is greater than that observed in other memory tasks, in some of the earliest studies on the spacing of presentations in free-recall lists no benefit at all was derived from DP. In two studies by Waugh (1963, 1967), lists of monosyllabic English words were presented auditorily at a rate of 1 word per second. Test items in the list each received two presentations; for massed items, these occurred in successive list positions whilst for distributed items the two presentations were separated by the presentation of one or more other words. No difference was found in the frequency of recall of massed and distributed items. Furthermore, in the 1963 study, the interpresentation lag for distributed items was systematically varied; no difference in the frequency of recall was found between items of different lags.

In an attempt to explain these nagative findings, Waugh (1970) conducted a further series of studies. In the first of these (Waugh, 1970, Exp. I.) lists of monosyllabic English words were presented auditorily at a 1-second rate. Test items received either two massed or two distributed presentations, with the interpresentation lag being systematica-lly varied. The proportion of words recalled in all conditions was about the same, which confirmed the earlier finding. When identical lists were presented auditorily at a slower rate (one word per 4 seconds) it was found that, whilst some improvement in recall was found for all items, recall for massed items was only slightly superior to that of items receiving only a single presentation, and furthermore, items receiving two distributed presentations yielded markedly superior recall to massed items (although not twice as good as that for massed items). Once again, no lag effect was found; all distributed items were recalled equally well.

In a second study (Waugh 1970 Exp.II) lists of common words were presented auditorily at a rate of 1 word per second. In the MP condition, lists consisted only of items receiving from 1 to 8 massed presentations; no distributed items occurred in these lists. In the DP condition items received from 1 to 8 distributed presentations. Lag was not controlled, and no massed items occurred in this condition. This paradigm will be referred to in future as the "unmixed lists" paradigm. Although there was no significant difference in the mean number of words recalled overall from MP and DP lists, there were differences in the relationship of recall to presentation frequency for the two conditions. MP facilitated recall relative to DP at presentation frequencies of 1, 2 and 3; gave equal recall for a frequency of 4; and poorer recall at frequencies of 6 and 8.

In both MP and DP conditions, it was found that the relationship between presentation frequency and recall was well described by a linear regression; the slope of the regression line was greater for DP, and in fact it was found that recall for 8 presentations was roughly 8 times as good as recall for a single presentation. In other words, the data for the unmixed DP list was well fitted by a regression line through the origin.

In a further study. Waugh (1970, Exp III) employed auditorily-presented lists of common words, each word being read at a 1-second rate. Each presentation was followed by a blank period of 1 - 8 seconds, designed to allow the subjects to rehearse covertly. Each item received only a single presentation. It was found that all items were recalled about equally well, irrespective of the length of the rehearsal period. Furthermore, the mean number of words recalled from a list of given duration of this type did not differ significantly from the mean number of words recalled from a list of the same duration in the previous study. In other words, the mean number of words recalled from a list of given duration remained constant regardless of whether the words received 1 - 8 massed presentations of 1 second each, 1 - 8 distributed presentations of 1 second each, or a single presentation of 1 second followed by a blank period of 1 - 8 seconds.

Waugh claimed that these results were in accord with the total-time law (TTL) which states that the amount learned from a list of items is a direct function of study time, regardless of how that time is distributed amongst items on the list. Cooper and Pantle (1967) have documented an impressive array of evidence in support of this hypothesis. Waugh suggested that all the preceding results could also be explained in terms of the TTL if it is supposed that subjects are unwilling to hold any item in attention (i.e. rehearse or otherwise process it) for more than a given length of time, and that when the duration of a presentation, or block of presentations, exceeds this critical period, then the excess time is devoted to the processing or rehearsal of previously-presented items. This hypothesis is identical to that proposed by Greeno (1967) to explain spacing effects in a paired-associate task (see 1.43).

This shared rehearsal hypothesis of Waugh has no difficulty in explaining any of her results. The negative findings for 1-second presentation rates with mixed MP and DP lists follows if one assumes that no free rehearsal time was available due to the rapid rate. For slower presentation rates with mixed lists, DP items would probably receive more processing time during actual presentation, and the same share of any other "spare" processing time as MP items. In the fast-rate unmixed lists condition, the enhanced recall of low-frequency MP items is taken to be a consequence of the extra processing time available during the presentations of high-frequency MP items, which would not be available to low-frequency items in unmixed DP lists.

Thus, Waugh's shared rehearsal hypothesis distinguishes between experimenter-controlled presentation time and subject-controlled processing time. An implicit assumption of the hypothesis appears to be that the probability of recalling an item is related in a very direct way, in fact varies as, processing time. The observed relationship between recall and processing time in rapidly presented unmixed DP lists lends support to this assumption. Unfortunately, the majority of findings concerning the distribution of presentations in free recall lists do not lend support to Waugh's interpretation of the DP phenomenon.

3.13 Differential rehearsal hypotheses

Underwood (1969) has reported the following series of studies whose results conflict very markedly with those of Waugh. In the first of these, the shared rehearsal hypothesis was tested directly (Underwood, 1969, Exps. I and II). Lists of common, unrelated monosyllabic nouns were presented auditorily at a rate of 1 word every 5 seconds. Two groups of subjects were tested; the control group received normal free recall instructions (Exp.I) whilst the experimental group were specifically instructed not to rehearse previously presented words, and to concentrate solely on the word being presented (Exp.II). A mixed lists design was employed, with both massed and distributed items receiving 2, 3 or 4 presentations. It was found that the proportion of distributed items recalled was considerably superior to that of massed items at each presentation level, for both groups of subjects, and that, furthermore, the performance of control and experimental groups was identical in every way.

In a further study, Underwood (1969, Exp.III) employed lists of common, unrelated, monosyllabic nouns which were presented auditorily at a rate of 1 word every 5 seconds, in an unmixed lists paradigm. Test items received 1 - 4 presentations. Different subjects were employed in each condition. It was found that the mean number of words recalled overall from DP lists was greater than that for MP lists. Furthermore, words that had only received a single presentation were recalled equally well in the two conditions, in sharp contrast to Waugh's (1970, Exp.II) findings. Underwood also found that on the whole, performance on both MP and DP items in this study differed very little from performance on comparable items in the previous (mixed list) study. The only major difference appeared to be a superiority in the recall of high-frequency DP items in mixed, as opposed to unmixed lists.

In addition, Underwood found that the overall probability of recall of massed items in both mixed and unmixed lists was identical; thus recall of massed items was not depressed by their occurrence in a mixed list. On the whole, these results to not support the shared rehearsal hypothesis. It should also be pointed out that in contrast with the results of Waugh's (1970,Exp.II) unmixed list study, the relationship between recall and presentation frequency in Underwood's unmixed list experiment was non-linear. Peformance showed an initial rapid improvement with frequency, but thereafter, the rate of improvement declined substantially. Indeed, hardly any difference was found in either MP or DP recall after 3 or 4 presentations.

As part of a more complex study, Underwood (1969,Exp.IV) attempted to find a systematic interpresentation lag effect. Lists of common, two-syllable unrelated nouns were presented auditorily at a rate of 1 per 4 seconds. Test items received 2, 3 or 4 presentations. The mean interpresentation lag was either 2, 8, 14 or 20 intervening presentations of other items. Significant effects of frequency, lag, and the frequency x lag interaction were found, but no orderly and systematic statement of the lag effects could be formulated.

A further series of experiments was reported by Underwood (1970). In the first of these Underwood, (1970,Ex ps. I and II) lists of short sentences were presented auditorily at a rate of 1 sentence every 5 seconds. Mixed lists were employed (Exp.I) as were unmixed lists of both MP and DP items (Exp.II). Test sentences received between 2 and 6 presentations. Performance was measured in terms of the number of sentences correctly recalled. It was found that in both the mixed and unmixed conditions, recall for DP items was superior to that for MP items at all presentation frequencies, and there was no significant difference in the relationships of recall to presentation frequency between mixed and unmixed DP conditions. Furthermore, singly-presented items were recalled equally well in all three conditions (i.e. mixed, unmixed DP and unmixed MP). It had been thought that sentences presented at a slow rate would be handled in a similar way to Waugh's words presented at a rapid rate, as there would be little spare processing time in either condition. These results are therefore evidence against this hypothesis, since DP was superior to MP in the mixed list, and in unmixed lists, there was no superiority of low-frequency MP item recall.

Under the assumption that it would take longer to establish a stable code for nonsense syllables than for words, Underwood (1970, Exp.IV) argued that an MP schedule might be less deleterious for syllables presented at a rapid rate than for syllables presented at a slow rate. An additional presentation at a rapid rate might allow the subject to establish a stable code that might not be established on the first presentation. Mixed lists of nonsense syllables of medium association value were presented visually at one of two rates; either 1 item per 2 seconds, or one item per 5 seconds. Syllables received from 1 to 4 massed or distributed presentations. It was found that recall for DP items was again superior to that of MP items presented the same number of times at the same rate. However, this difference in performance was not affected by the rate of presentation.

The purpose of a final study (Underwood 1970, Exp.V) was to examine whether rehearsal of an item, or the processing of such an item to give a long-term code, could be curtailed by having the subject perform a task after each presentation of an item in a free-recall list. Mixed lists of two-syllable nouns were presented visually; each presentation lasted for 4 seconds, and was followed by a 1-second interval. In the control condition, a blank appeared on the screen during this interval, whilst in the experimental conditions, two single digit numbers were displayed, separated by a plus sign, e.g. 3+5. In the "read" condition, subjects were simply required to read aloud the two numbers, whilst in the "add" condition the sum of the two numbers was required. The control and read conditions yielded almost identical results, with the usual MP-DP differences across presentation frequencies, whilst a similar, but much attenuated pattern was observed in the "add" condition.

Although both Waugh and Underwood failed to find a systematic effect of interpresentation lag, their results strongly disagree on all other counts. Underwood found that the recall of singly presented or low-frequency massed items did not depend on context; the same level of recall was found in both mixed lists and unmixed MP lists. The relationship between recall and presentation frequency was nowhere found to be linear. Specific instructions not to rehearse previously presented words did not affect the recall of either massed or distributed items in a mixed schedule. Finally, the mean number of words recalled from unmixed DP lists was always greater than the mean number recalled from unmixed MP lists of similar material, and of the same duration, in direct contradiction of the TTL.

One very obvious difference in procedure between the studies of Waugh and Underwood is to be found in the presentation rate. Waugh used a 1-second auditory presentation scheme throughout, except in a mixed schedule study which was presented at a 4 second rate, wherein items received only two presentations, and her results here did not conflict with those of Underwood. Underwood suggested that for slow presentation rates the differential rehearsal hypothesis might obtain in a modified form; subjects might be unwilling to hold any one item in attention for more than a given period of time, as before, but might not use any extra time thus gained to their fullest advantage (i.e. for the rehearsal of previously presented items). Performance on both MP and DP items would thus be independent of the type of schedule in which they appear, whilst DP would still produce superior performance, However, it is still unclear as to why the TTL should only obtain for very rapid auditory presentation rates. It should also be pointed out that Underwood's differential rehearsal hypothesis corresponds almost exactly to Greeno's (1970a) "resting" hypothesis (see 1.43).

Underwood (1970) confessed his dissatisfaction with this modified hypothesis; although it explains the broad pattern of his data, there are still one or two results which are somewhat embarrassing to the theory.

Firstly, it fails to predict the enhanced recall observed for high-frequency DP items in a mixed schedule, as compared to an unmixed schedule, and secondly it has very little to say about the effects of a disrupting task following each presentation trial. If subjects do not make the use of free time that has been suggested, then all three conditions (control, read and add) should yield very similar results. Performance in the add condition, however, was much poorer than performance in the other two conditions, for both MP and DP items. Two hypotheses are possible. Either the add condition affects retention, or else the task of performing mental arithmetic actually displaces items from STM, where they would otherwise be available for encoding into organised word groupings with later items.

Straightforward differential rehearsal hypotheses are unsatisfactory in another respect. There is very little reason to suppose that, say, Underwood's hypothesis should not apply equally well to other verbal learning paradigms, such as continuous paired-associate or Brown-Peterson type tasks. Melton (1970) has pointed out that the improvement in performance due to DP in these paradigms is usually much smaller than that obtained in free-recall procedures. If differential rehearsal only explains such small-scale improvements then there must be some other process present in the free-recall situation which accounts for the bulk of the improvement in performance due to DP.

Before leaving the problem of the conflicting results of Waugh and Underwood, it should be pointed out that Melton (1970) has suggested that Waugh's support of the TTL might be, to some extent, artifactual. Her lists were relatively short, and item sequences were highly predictable. Furthermore, each of her subjects was tested on many lists. It is therefore possible that they adopted a differential rehearsal strategy in the fast presentation situation as the best way of dealing with the material. This might also account for her failure to find a systematic lag effect.

Underwood's failure to find any systematic and orderly effect of interpresentation lag may well have been a consequence of his lack of control of the exact lag; only mean interpresentation lags were systematically varied. A paired-associate study by Bjork and Abramowitz (1968;See 1.44) has suggested that equally-spaced presentations are optimal in terms of recall performance. It is therefore feasible that the incomprehensible results of Underwood's study were a result of uncontrolled devations from equality of interpresentation spacings, which would result in large systematic biases in measuring the effects of mean lag.

Fortunately, there are some studies which show a very systematic and reliable effect of interpresentation lag in free recall situations. After a brief preliminary report (Melton, Reicher and Shulman, 1966) showing an increasingly beneficial effect of lags of 0,8, 20 and 40 for twice-presented items, Melton and Shulman (1967) reported the data shown in Figure 11. After a short practice task, each subject was given a recall test on each of three lists of four-letter nouns. In the middle portion of each list were 8 words that occurred once, and 4 words that occurred twice at each of the following interpresentation lags; 0, 2, 4, 8, 20 and 40. Different groups of subjects learned these lists by visual presentation at each of three different rates: one word per 1.3, 2.3, or 4.3 seconds. It is clear from the figure that the main effects of presentation rate and lag were significant. There was no significant rate x lag interaction. Therefore, in so far as visual presentation is concerned, rate of presentation is not a critical variable in determining the DP effect. It also appears from the figure that interpresentation interval has an increasingly beneficial effect well beyond those values that would be sufficient to "wipe out" short-term retention effects.

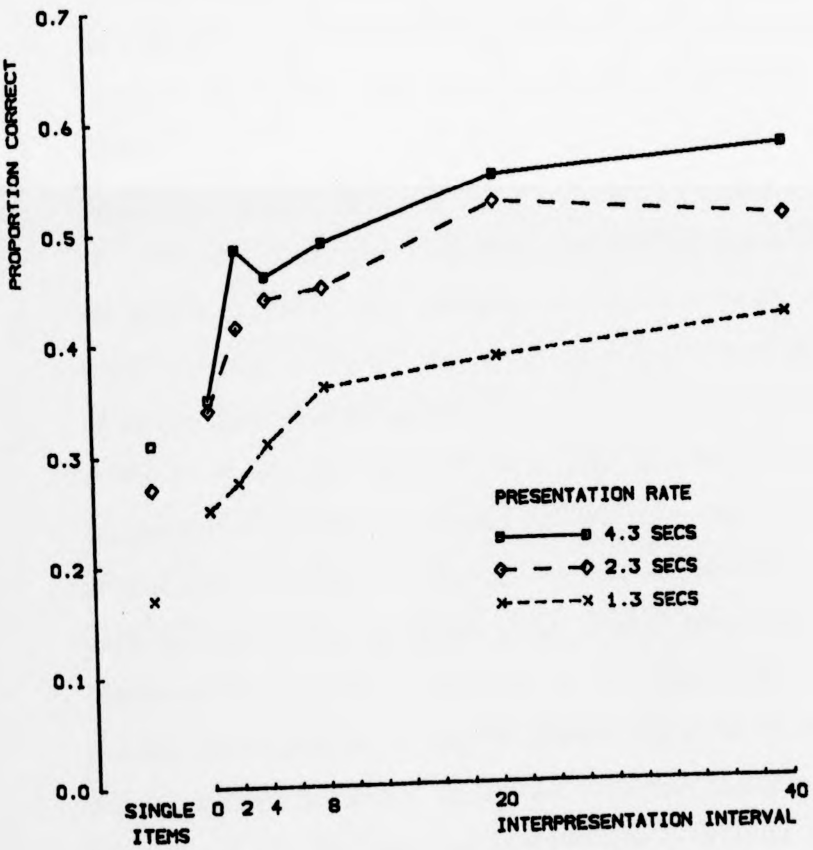
A very similar study by Melton and R.A.S. Adams was reported by Melton (1970). A simple factorial design was used, one factor being word class ("mixed" words as employed by Waugh, and high-frequency

FIGURE 11.

Free recall of words as a function of the interval between two successive presentations and of presentation rate (Melton and Shulman, 1967).

FIGURE 11.

Free recall of words as a function of the interval between two successive presentations and of presentation rate (Melton and Shulman, 1967).


PROPORTION CORRECT
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
SINGLE ITEMS
0 2 4 8 20 40
INTERPRESENTATION INTERVAL
PRESENTATION RATE
4.3 SECS
2.3 SECS
1.3 SECS

nouns) and the second factor, modality of presentation (auditory or visual). Each subject had one list of each of the four types. The rate of presentation was always one word per 2.3 seconds. Significantly better recall was found with homogenous nouns, and with visual presentation. Significant lag effects, similar to those reported by Melton and Shulman were also found. There was also a significant lag x modality interaction, the slope of the lag functions being steeper for visual presentation. In particular, the slope of the lag function for auditorily presented mixed words was very small, perhaps another factor contributing to Waugh's failure to find a lag effect.

3.14 Differential encoding

Melton (1970) pointed out that a differential rehearsal hypothesis contains no provision for a systematic lag effect which extends far beyond the range of short-term memory, and suggested that some kind of differential encoding hypothesis might be a feasible explanation of the DP effect. The work of Tulving (1962 and 1966) has strongly suggested that free-recall learning involves subjective organisation of word clusters within a list, and that these subjective cluster units may serve as cueing systems at the time of recall (see 1.24). Therefore, the inclusion of a word in two different subjective clusters would increase the cues or "access routes" to retrieval. Melton proposed that as the lag between successive presentations of a particular item increases, the word contexts in which is occurs would become less and less correlated, and so the total number of different cues to its retrieval would increase.

This theory clearly accounts for the lag data presented above, and for the general observations concerning the superiority of DP to MP. Furthermore, the result obtained by Underwood (1969, ExpIII) that recall for high presentation frequency distributed items was enhanced if they occurred in a mixed, as opposed to an unmixed, schedule, can now be explained; for a given interpresentation interval the contextual cues surrounding a given word on its various presentation trials are less likely to be correlated if the word appears in a mixed list, where new context is supplied by new massed items, as opposed to an unmixed list, where only distributed items occur, and might appear several times in proximity to the given word. This difference would clearly become far more exaggerated with high presentation frequencies, as more "new" massed items would be available for contextual clustering in a mixed schedule.

Melton's differential encoding hypothesis is supported indirectly by the frequency judgement studies of Underwood (1969). In his Exp IV, two-syllable nouns were presented auditorily at the rate of 1 word every 4 seconds; items received 2, 3 or 4 presentations, with a mean interpresentation lag 2, 8, 14 or 20. Subjects were instructed to memorise the frequency of occurrence of each word. Although lag was not found to affect performance, mean judged frequencies were fairly accurate for twice-presented items, but fell off to about 3.3 for items which had received 4 presentations.

In a second study (Underwood, 1969, Exp. V) massed and distributed items were read at a 4-second rate in mixed lists. Two groups of subjects had both been instructed that a memory test of some kind would follow the list presentation, although word order would not be tested. The first group was then given a recall test, whilst the second group received a frequency judgement test. Test items were presented 2, 3 or 4 times. The usual MP-DP differences were found for the recall group. Frequency judgements for DP items were much as in the previous study, being about 2 for twice-presented items, rising to about 3.25 for items that had received 4 presentations. Frequency judgements for MP items were much poorer, being about 1.5 for items that had occurred twice, rising to about 2.0 for items that had been presented 4 times.

Although there was no control over what subjects were attempting to memorise in this study, the recall data suggests that normal list-learning behaviour was being employed. The frequency judgement data could be interpreted as the result of some kind of "context counting" behaviour during the test, as opposed to frequency encoding during the actual list reading. If such an interpretation is accepted, then this study clearly supports the multiple encoding hypothesis.

In a similar study, Hintzman (1969a) showed that mean judged frequency, although lower than true frequency, was an increasing function of lag. Madigan (1969) presented two groups of subjects with lists of nouns; presentation was visual, and at a rate of 1 word per 2.5 seconds. One group was instructed to recall the words, whilst the other was instructed to recall the lists, and also to give an estimate of how many times each word had occurred. Test items received two presentations, with an interpresentation lag of 0, 2, 4, 8, 10 or 32 items. The proportion of words recalled for the item recall group was an increasing function of lag; the same was true of recall for the frequency recall group. However, when the performance for the latter group was broken down into proportions of words which had been judged to have occurred once, and words which had been judged to have occurred twice, it was found that recall of the former category did not show a lag effect, whilst recall of the latter showed a marked effect of lag. This study illustrates very clearly the relationship between recall and judged frequency of occurrence in a free-recall situation, and substantiates the hypothesis that frequency judgements are made by some kind of "context counting" strategy.

In a second study, Madigan presented subjects with word pairs; presentation was visual, and at a rate of one pair per 4 seconds. Subjects were instructed that only the second word of each pair was to be remembered; the first word used was a "cue" word, included to facilitate recall. Each test item received two presentations, at a lag of 0, 4, 8 or 16 intervening pairs. One subject group performed in the same cue condition, wherein each to-be-remembered word appeared with the same cue word on each presentation, whilst a second subject group performed in the different cue condition where two different cue words were paired with each word to be remembered. When each list had been presented, subjects were given a 4 - minute free-recall test on to-be-remembered words, followed by a 4 - minute cued-recall test, during which they had a complete list of cue words before them.

For noncued recall, the same-cue group yielded results which showed a large effect of lag on the proportion of words recalled; the lag 16 condition showed a 90% improvement of recall over the lag 0 condition. Similar results were found for the different -cue group, although the benefit of longer interpresentation lags was less marked; the lag 16 condition produced about a 40% improvement in recall over the lag 0 condition.

For cued recall, the performance of the same-cue group improved through lags 0-8, and then declined at lag 16, whilst the performance of the different-cue group exhibited no systematic lag effect. Cued recall of words that had not been recalled in the uncued recall period was then examined; the performance of the same-cue group exhibited the same relationship with lag, that is, an improvement through lags 0-8, with a decline from lag 8-16. The performance of the different-cue group was found to decline slightly with increasing lag.

It is interesting to note that for both non-cued and cued recall, the overall performance of the same-cue and different-cue groups did not differ significantly; cueing seemed to affect only the lag function, with different cues facilitating recall for short lags and impairing recall for longer lags. It was also found that cued recall was superior to uncued recall in all conditions.

Partial support for the differential encoding hypothesis is afforded by the non-cued recall results. Differential cueing clearly reduced the slope of the lag function, as the hypothesis would predict, but did not produce an overall improvement in performance. This was possibly due to the fact that in the same-cue condition, cue words were presented twice, and were therefore more likely to be remembered than the once-presented different cues, and would also show a lag effect. Thus any improvement in recall of to-be-remembered words due to differential cueing would be counterbalanced by an improvement (due to superior cue recognition) in recall for the same-cue condition.

Although this argument also holds for the cued recall performance, two aspects of the data differ most strikingly from that for noncued recall. Firstly, differential cueing completely removed all effects of lag and secondly, recall in the same-cue condition did not improve beyond lag 8, and in fact declined from lag 8 to lag 16. This result is in complete agreement with results from paired-associate studies (which this condition strongly resembles) described in section 1.44.

Gartman and Johnson (1972) also studied the effects of context words on recall. Lists of common words were presented visually at a 2-second rate. Test items were homographs (words having two or more meanings) which were preceded on each of two presentations by two contextual words. Two conditions were employed: in the "same" condition, the context words were from the same context on each presentation, whilst in the "different" condition, two words from a different context occurred prior to the second presentation. For example:

Context	1st Presentation	2nd Presentation
SAME	metre inch foot	mile yard foot
DIFFERENT	metre inch foot	arm leg foot

Control items receiving two presentations were also included; no contextual factors were applied to these items. Interpresentation lags of 2-18 were employed. The results of this study are shown in Table 5. As can be seen, there is a large lag effect in the control condition, in contrast to the experimental conditions in which no significant lag effect was found.

TABLE 5

Recall of homographs and control words as a function of the lag between two presentations and context (Gartman and Johnson , 1972)

		HOMOGRAPHS			CONTROL WORDS		
CONTEXT	LAG:-	2	8-10	16-18	2	8-10	16-18
Same		.13	.20	.19	.13	.27	.51
Different		.61	.61	.60			

It does appear that there was a slight effect of lag in the same-context condition, but the authors have advanced no information on this point. Nevertheless, recall of homographs in the different-context condition was far superior to that of homographs in the same-context condition, and also superior to that of long-lag control items. Recall of control items and same-context homographs re-presented at lag 2 did not differ significantly from that of singly-presented items.

The results of this study do conflict markedly with those of Madigan (1969) discussed above, in that the overall level of recall in the different-context condition was far superior to that in the same-context condition, whereas in Madigan's study, no differences in the overall levels of recall were found between the same- and different-cue conditions. However, the two studies did differ procedurally, in that Madigan's subjects knew that they would not have to recall the cue words, whereas in the current study, subjects would presumably have attempted to learn the context words. The different encoding strategies which might have resulted from these different task requirements may account for this conflict in results.

The Gartman and Johnson study suffers from another difficulty, in that recall of homographs in the different context condition was more than double that of singly-presented items. The authors were unable to explain this result. It is suggested that subjects may have employed a categorisation strategy in this experiment, especially as cue words were not specifically pointed out as such, and were therefore indistinguishable from the items of interest. In such a case, homographs occurring in the "different" condition would be easily identifiable as homographs, and it is likely that subjects would therefore work differentially on such obviously "special" words. Furthermore, different-context homographs might be encoded as belonging to <u>three</u> categories; the two presented by the experimenter and the third, "homograph", discovered by the subject.

Thus the results of this study must be viewed with caution. However, it is difficult to see how the differential rehearsal hypothesis could be expected to account for these results, or those of Madigan. With the exception of Waugh's results, which have not to date been duplicated by any other observers, the results of free-recall experiments are well described by the differential encoding hypothesis. This hypothesis can explain the relationship between interpresentation lag and probability of recall, and judged frequency of occurrence, and is unique in offering an explanation of the results of cueing experiments. Furthermore, differential encoding can account for the greater magnitude of the lag effect in free-recall situations as compared with other paradigms which rely less on intra-list associations.

However, as a general explanation of the distributed practice effect it has shortcomings. One element of doubt is raised by Madigan's results on cued recall; performance appears not to improve with lag beyond a certain point. This is not accounted for by the hypothesis. The hypothesis would seem to imply that coding is purely determined by context, otherwise two massed presentations would allow a second chance to adequately code items inadequately coded on the first presentation. Furthermore, given that a code will support recall (and therefore recognition) of an item on its second presentation, what has the hypothesis to say about coding on the second presentation? For example, is the original encoding elaborated or is it ignored in so far as constructing a second code is concerned?.

Clearly such questions are of little importance when it is known that contextual coding is an important factor. However, they attain more relevance in the study of paired-associate learning and the examination of spaced repetitions in experiments of the Brown-Peterson type. In the former area, where recall is cued, results have suggested some limit beyond which lag is no longer beneficial, but detrimental to performance, whilst in the Brown-Peterson paradigm, only one subspan item must be remembered at any one time, so that there are no other to-be-remembered items to provide contextual cues.

3.2. Continuous Recognition Studies.

Kintsch (196) has shown that the probability of recognising an item correctly increases with the spacing of repetitions of that item in a continuous recognition task. Four digit numbers were presented visually at a rate of 1 item every 2 seconds. Each item was presented 6 times. On each trial, subjects were required to say whether the item had been presented before (by responding "old") or not (by responding "new"). Performance was measured by the proportion of items correctly identified as old, P("old"/old).

Four treatments defined by four spacing patterns were employed. If a sequence is represented as follows:-

$$P_1\ I_1\ P_2\ I_2\ P_3\ I_3\ P_4\ I_4\ P_5\ I_5\ P_6\ I_6$$

where $P_1 \ldots P_6$ are the six presentations of a particular item, and

$I_1 \ldots I_6$ are the interpresentation intervals, being trials involving other items, then the four treatments can be defined as follows:-

Short (S)	$I_1 = I_2 = I_3 = I_4 = I_5 = 1$ trial
Long (L)	$I_1 = I_2 = I_3 = I_4 = I_5 = 10$ trials
Long-short (L-S)	$I_1 = I_2 = I_3 = 10,\ I_4 = I_5 = 1$
Short-long (S-L)	$I_1 = I_2 = I_3 = 1,\ I_4 = I_5 = 10$

Kintsch's results are shown in Figure 12.

Performance on S items was clearly superior to that of L items at every repetition level. A comparison of S-L and L treatments at repetitions 5 and 6 clearly shows that three wide spacings were superior to three massed presentations when retention was measured at a long interval. A comparison of S and L-S treatments shows little difference in performance following three massed or three spaced presentations when retention is measured at a short interval, although once again spaced repetitions were slightly superior. Similar results were found with consonant-vowel-consonant trigram (CVC) material.

Kintsch compared a number of learning models having different acquisition and retention axioms, and found that his data were well explained by the LS-2 model of Atkinson and Crothers (1964). This model assumed three memory states: a long-term state, from which no forgetting occurs, a transitory short-term state, which leads to a correct response, but from which forgetting can take place, and a naive state, in which a correct response can be guessed - the probability of such a guess is taken to be the overall false recognition rate, P("dd"/new).

The model also assumes that the long-term state is equally likely to be entered from both the short-term and naive states. Learning and short-term forgetting parameters were estimated indepenndently for long- and short-delay schedules. It was found that short-delay schedules suffered less from short-term forgetting, as one would expect, but yielded lower values of the learning parameter. Performance shifts observed in L-S and S-L schedules were then pre-

$I_1 \ldots I_6$ are the interpresentation intervals, being trials involving other items, then the four treatments can be defined as follows:-

Short (S) $I_1 = I_2 = I_3 = I_4 = I_5 = 1$ trial

Long (L) $I_1 = I_2 = I_3 = I_4 = I_5 = 10$ trials

Long-short (L-S) $I_1 = I_2 = I_3 = 10$, $I_4 = I_5 = 1$

Short-long (S-L) $I_1 = I_2 = I_3 = 1$, $I_4 = I_5 = 10$

Kintsch's results are shown in Figure 12.

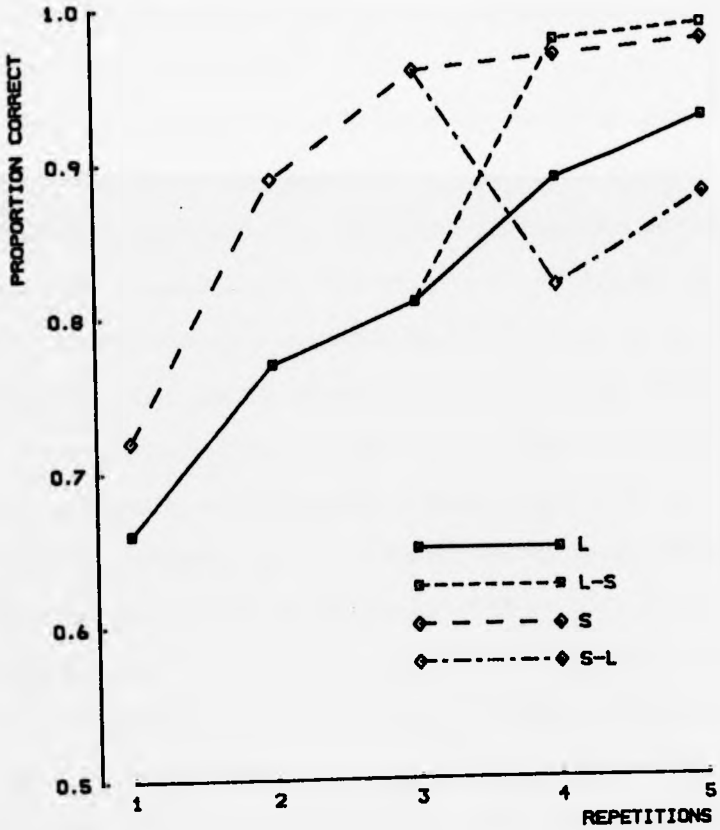
Performance on S items was clearly superior to that of L items at every repetition level. A comparison of S-L and L treatments at repetitions 5 and 6 clearly shows that three wide spacings were superior to three massed presentations when retention was measured at a long interval. A comparison of S and L-S treatments shows little difference in performance following three massed or three spaced presentations when retention is measured at a short interval, although once again spaced repetitions were slightly superior. Similar results were found with consonant-vowel-consonant trigram (CVC) material.

Kintsch compared a number of learning models having different acquisition and retention axioms, and found that his data were well explained by the LS-2 model of Atkinson and Crothers (1964). This model assumed three memory states: a long-term state, from which no forgetting occurs, a transitory short-term state, which leads to a correct response, but from which forgetting can take place, and a naive state, in which a correct response can be guessed - the probability of such a guess is taken to be the overall false recognition rate, P("dd"/new).

The model also assumes that the long-term state is equally likely to be entered from both the short-term and naive states. Learning and short-term forgetting parameters were estimated independently for long- and short-delay schedules. It was found that short-delay schedules suffered less from short-term forgetting, as one would expect, but yielded lower values of the learning parameter. Performance shifts observed in L-S and S-L schedules were then pre-

FIGURE 12.

Proportion of items correctly recognised as a function of short and long retention and interpresentation intervals (Kintsch 1966)



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
1
2
3
4
5
REPETITIONS
L
L-S
S
S-L

dicted moderately well. Nevertheless, Kintsch was unwilling to give any theoretical explanation as to why less learning should occur on a short delay schedule.

The underlying assumptions of the LS-2 model as applied to recognition memory are partially justified by a study conducted by Shepard and Teghtsoonian (1961). Their subjects were given a deck of about 200 cards, on each of which was a 3-digit number, and were instructed to go through the deck at their own rate, noting on a record sheet whether each item had been seen before (i.e. it was "old") or not (it was "new").

Each number appeared twice in the deck; the lag (in intervening items) between successive appearances of a particular item was systematically varied, and the proportion of correct recognition ("old"/old) responses plotted as a function of lag. It was found that performance declined rapidly up to a lag of about 10 items, and then much more slowly. A short-term decay component would account for the initial rapid decline in performance, although there did appear to be a significant decline in the long-term part of the curve. However, as a first approximation, the assumption of an absorbing long-term state appears justified.

Olson (1969) has demonstrated the transitory nature of short-term recognition memory. His subjects were presented with long lists of consonant trigrams (CCC's); each trial was of 3 seconds duration; an item was presented on a screen for 1 second, following which the screen remained blank for two seconds, during which period the subject was required to make a recognition response ("old" or "new") to the item he had just seen. Items were all presented eight times, at a variety of interpresentation lags between 0 and 70 intervening items.

Olson chose a criterion level of four successive correct recognition ("old"/old) responses, and examined the proportion of precriterion and postcriterion correct responses as a function of lag since the last presentation. Precriterion performance was found to decline with lag to an asymptote around the overall false alarm rate, P("old"/new) whereas postcriterion performance declined only slightly with lag. This result is clearly analagous with Bjork's (1966) results in a paired-associate study (see 1.32 and Figure 2).

Olson's study clearly justifies the assumption made by Kintsch that items forgotten from the short-term state re-enter the naive state; it further appears that for CCC's at least, the rate of forgetting from the long-term state is of negligible proportions.

Kintsch's (1966) results are therefore consistent with a differential rehearsal hypothesis which states that items still held in short-term memory when re-presented receive little benefit from the re-presentation. Such a hypothesis could well account for the slower rate of learning observed on a short-delay schedule, whilst the results of Shepard and Teghtsoonian suggest that short-term retention effects in recognition memory would be "wiped out" after about 10 intervening presentations of other items, so that items on Kintsch's long-delay schedules would not suffer at re-presentation.

It should also be pointed out that Melton's differential encoding hypothesis could also account for Kintsch's findings, especially if it is accepted that some kind of transitory short-term memory state can boost performance at short retention intervals, as suggested by the above studies. Short-delay schedules would then lead to less forgetting from this short-term state, but would also lead to less long-term learning (and therefore poorer performance at long retention intervals) than long-delay schedules as a result of there being fewer access cues coded on short-delay schedules.

Hintzman (1969b) presented subjects with lists of common words. Each item was presented three times; the interval between the first two presentations (the P_1-P_2 interval) was either 1,2,4,8 or 16 trials on other items, whilst the P_2-P_3 interval was always 16 such trials.

Response latencies were measured, and it was found that correct recognition ("old"/old) latencies were shorter than error latencies on all three presentations. An analysis of correct response times as a function of P_1-P_2 lag revealed that P_2 latency was an increasing function of P_1-P_2 lag, as one would expect; furthermore, the relationship of P_3 latency with P_1-P_2 lag clearly showed a spaced-practice improvement effect. In other words, correct response latency at P_3 declined as the P_1-P_2 interval increased.

A multiple encoding hypothesis could account for these results if it is assumed that access to one of two recognition codes is more rapid than access to a single code. The chance of possessing two such alternative encodings would clearly improve with P_1-P_2 spacing. If memory search is conceived as being made through a large set of codes at a constant rate, then it is feasible that the mean search time would be less if there were two target codes rather than one. The result that error latencies were higher than corect response latencies lends support to such a search hypothesis, since an unsuccessful search may not be terminated until a large number of inappropriate encodings had been drawn. The search process envisaged here would be initiated on the basis of some functional aspect of the current stimulus presentation. Encodings sharing these aspects would be drawn from memory, and then matched more carefully with the current item. Furthermore, recent work reviewed by Mandler (1972) suggests that contextual and semantic cues at test materially aids recognition performance, and results of this kind can only add credence to a differential encoding hypothesis that depends heavily on contextual attributes.

Unfortunately, a differential processing hypothesis can also account for the results of interpresentation spacing studies in recognition memory. No results exist which compare performance over a sufficient range of interpresentation intervals to determine whether performance continues to improve with spacing beyond the p int at which short-term retention ef ects would be "wiped out". Although Hintzman (1969b) employed a wide enough range of interpresentation intervals, his latency data did not even show a short-term retention ef ect at short lag.

3.3. Brown-Peterson Studies

The Brown-Peterson technique is one of the relatively new ex erimental met ods which appeared at about the same time that human learning theorists were making their move towards a verbal learning framework, and deserting their old, classical perceptual-motor tasks. It was first developed by Brown (1954) as a tool for the study of short-term retention, and later popul rised by Peterson and Peterson (1959). A single subspan item (i.e. one that can be held in STM and perfectly recalled if no distracting task intrudes between presentation and test), is presented for study, and after an interval filled with interfering activity of some kind (such as backward counting) the subject is asked to rec ll the item.

The procedure is frequently modified to admit various kinds of presentation-test sequences; the important feature is the nature of the interfering activity. This is generally chosen to preclude covert rehearsal of the test it m, but no to conflict with the test item (in the sense of not being material to memorise which might be confused with, or compete with, the item). This techni ue thus admits a very exact definition of the material that the subject is trying to learn at a given time - one specific subspan item.

3.31 Repeated presentations and rehearsal

Hellyer (1962) presented consonant trigrams (CCC's) 1, 2, 4 or 8 times in massed trials, and tested retention following 3, 9, 18 or 27 seconds of an interfering task, which in this case was a digit naming task. The proportion of complete trigrams correctly recalled was plotted as a function of retention interval, at each presentation

FIGURE 13

Proportions of consonant trigrams correctly recalled as a function of

a) the number of successive 1-second presentations of the trigram (Hellyer, 1962)
b) the amount of overt rehearsal of trigrams (Peterson and Peterson 1959)

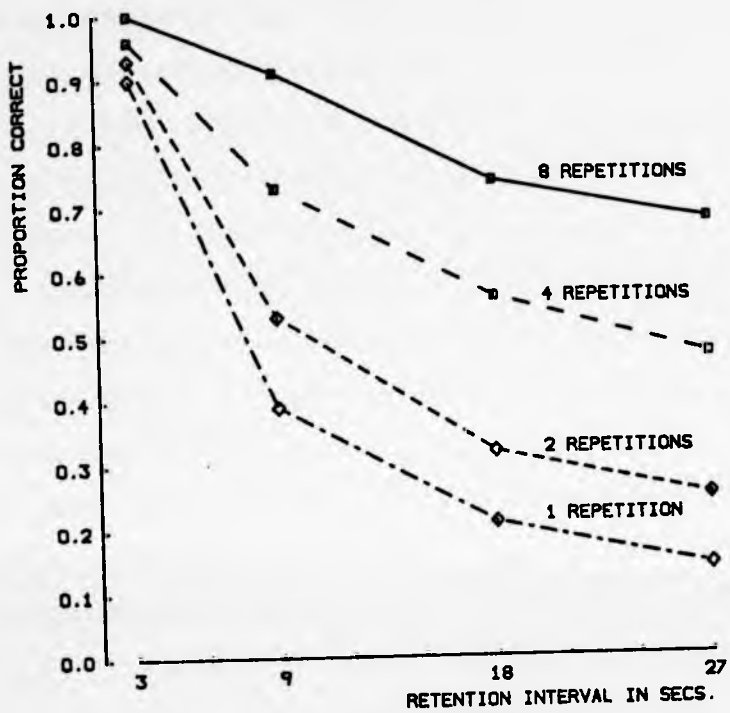


FIG.13A - HELLYER

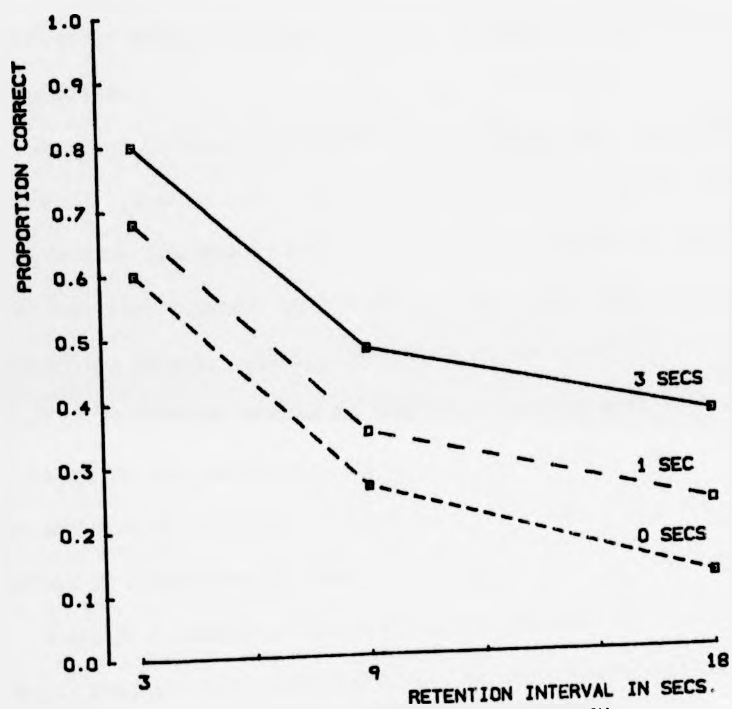


FIG.13B - PETERSON AND PETERSON

level. His results are shown in Figure 13a. Performance following a single presentation declined rapidly from 3 to 9 seconds, and then more slowly. The retention curve clearly demonstrates both short- and long-term decay components. As the number of presentations increases, it can be seen that both long and short-term retention improves, although performance after 3 seconds improves only slightly.

A study by Peterson and Peterson (1959) included two conditions: in the overt condition, subjects rehearsed the stimuli (3 consonant trigrams) aloud in time to a metronome, at a rate of 1 repetition per second, either 0, 1 or 3 times, whilst in thecovert condition subjects were given time to rehearse the stimuli to themselves for varying periods; no specific instructions were given to rehearse. A retention interval filled by counting backwards by threes for 3, 9 or 18 seconds followed in both conditions, followed by a test of retention. For the overt group, increasing the rehearsal time improved performance at all levels of interference time, but there was no effect of rehearsal time for the covert group. However, since there was no guarantee that the covert group were rehearsing, the negative results for this group are inconclusive.

Data for the overt group are shown in Figure 13b. The retention curves all appear to have the same shape, and again display an initial rapid decline followed by a slower decline. Furthermore, increasing overt rehearsal improves performance in both short- and long-term parts of the curves. Although Hellyer's curves flattened out as the number of repetitions increased, and the latter did not, this effect may be due to the increased difficulty of the material and retention task employed by Peterson and Peterson, rather than a difference in function of repetition and overt rehearsal.

A modified version of the paradigm was employed by Peterson (1963). Subjects were presented with a consonant-vowel-consonant trigram (CVC), and were then required to count backwards for 1, 3, 6 or 11 seconds. After this interference interval, the CVC was again presented. A second interference interval of 6 seconds followed this second presentation, and the subject was then cued for recall. This sequence may be presented as follows:

$$P_1 I_1 P_2 I_2 T$$

where P_1 and P_2 represent the first and second presentations of the CVC respectively; I_1 is an interference period of 1, 3, 6 or 11 secs., I_2 is an interference period of 6 secs., and T is the final test of recall.

Peterson's results are shown in the table below. Performance on the final test trial clearly improved with the length of the inter-presentation interval I_1. It is clear that once again, performance improves with the spacing of practice. It is difficult to see how the multiple encoding hypothesis could possibly apply in this case, as items do not appear in the context of other items to be remembered, and therefore the interpretation of 'access route' coding cannot apply.

	I_1 (seconds)			
	1	3	6	11
Proportion of trigrams correctly recalled	.66	.67	.74	.77

In order to gain some idea as to the underlying causes of the spaced practice effect in the Brown-Peterson paradigm, a more detailed examination of results is necessary.

3.32 Proactive interference

Observers have generally found that on the first trial of a BP task, forgetting is very rarely more than 10%, even after retention intervals of 20 seconds (Loess and Waugh, 1967). Performance then declines rapidly to a steady state with trials; it would appear therefore, that items studied on previous trials somehow interfere with the encoding or retention of items currently to be remembered. This effect will be labelled 'proactive interference' (PI); it must be noted, however, that this label does not in any way assume any particular theoretical viewpoint, such as classical interference theory, it is merely a title for the phenomenon. PI from preceding trials usually reaches asymptote in 3-6 trials, although this value is influenced by the inter-trial rest period; after a long rest, performance recovers to a very high level.

Similarity of the preceding items to the item currently to be remembered has been shown by Loess (1967) to have an important effect on PI; with the recall of word trigrams from different taxonomic classes (e.g. trees, birds, etc.,) a change of taxonomic class on a particular trial produced a dramatic improvement in performance on that trial (called by Loess 'release from PI'). For example, a control group presented with trigrams from the same class throughout recalled about 30% on trial 13; another group, who had received trigrams from one class only on the first 12 trials, and were then presented with a trigram from a different class on trial 13, achieved a recall score of 90%. Clearly, the effect was dramatic.

A variable that obviously depends on PI is the proportion of intrusions (recall errors that are previously-presented items, or parts thereof, as opposed to wild guesses and omissions). However, it would be rash to regard intrusions as a measure of PI, for two reasons. Firstly, there could be 'covert intrusions' that is, items that somehow interfere with the current item but are known by the subject to be incorrect, and are therefore not produced as a response. Secondly, some intrusions may not have interfered at all with the current item; they might just be reasonable guesses on the part of the subject.

Keppel and Underwood (1962) found that the proportion of intra-experimental intrusions (IEI's) increased with retention interval although at a slower rate than total errors. Loess (1968) found that 80-90% of IEI's came from the same category as items to be remembered, and of these 60-80% came from the most recent item of that category. However, the proportion of IEI'S from the most recent item of the same category decreased if there were other items between it and the to-be-remembered item. Pollatsek (1969) found that with items of the same kind (common four-letter noun trigrams with no category controls), 60-70% of IEI's came from the previous item, 10-20% from the item 2 back, and most of the remainder from items 3-4 back.

3.33 Some theories of the SPI effect in BP studies

In the light of results on the PI effect, three possible explanations of the spaced practice effect in the Brown-Peterson paradigm may be advanced.

1. Reduction of interference from previous items. The greater the spacing between two presentations, then the less should be any PI effects during the second presentation.

2. Consolidation of the memory trace. It is possible that the stimulus trace is in some sense getting more potent during the interference period.

3. Differential coding of items. This mechanism could have at least three different rationales (Pollatsek, 1969): (a). If the subject thinks he knows the item well on P_2, then he doesn't bother to work so hard; (b) If the subject doesn't know which items he knows well and which he doesn't, as after a short P_1-P_2 interval, then he will not know where to direct his effort; (c) If the subject already has a code which is "bad", then he is less likely to be able to think of a new one because of the interference from the old one.

Pollatsek (1969) employed five conditions in an attempt to discriminate between these hypotheses, as follows:-

I.	Simple (S)	P R I T
II.	Double Presentation (DP)	P_1 R_1 I_1 P_2 R_2 I_2 T
III.	Double Test (DT)	P_1 R_1 I_1 T_1 I_2 T_2
IV.	Forget (F)	As DP, with a new item for memory at P_2
V.	Control (C)	I_1P_1 R I_2 T

where P_j are presentations, I_j are interference intervals, T_j are tests of retention, and R_j are blank periods designed to allow covert rehearsal. In the DP condition, the two rehearsal periods R_1 and R_2 took values of 0, 3, or 6 seconds; the interpresentation interval I1 was 7 or 22 seconds, and the retention interval I_2 was 10 or 22 seconds (of paced forward counting in each case).

It was found that the proportion of items (word trigrams) retained increased with rehearsal time, and both R_1 and R_2 produced facilitation in this respect. This is in agreement with the assumed role of rehearsal deduced from the results of Howe (1967) and Bernbach (1967b) discussed earlier. It was also found that retention was poorer with the longer retention interval (I_2 = 22 secs.); the usual spaced practice effect was found however; performance was better in all conditions with I_1 = 22secs., as opposed to I_1 = 7 secs. Pollatsek's data, collapsed over both values of I_2, and all three values of R_2, are shown in Table 6.

TABLE 6

Proportion of complete word trigrams correctly recalled as a function of interpresentation interval and of rehearsal time (Pollatsek, 1969)

	R_1 = 0 secs	R_1 = 3 secs	R_1 = 6 secs
I_1 = 6 secs	.70	.80	.82
I_1 = 21 secs	.80	.88	.92

In the simple condition it was found that increasing overt rehearsal time generated a family of retention curves similar to those of Hellyer (Fig 13a), both in terms of complete trigrams recalled, and words recalled. This result is at variance with the negative findings for covert rehearsal of Peterson and Peterson (1959). In the double test condition, it was found that performance on T_2 was almost as good as performance on T_1, even with I_2 = 22 secs.; overt responding appeared to improve the coding of items.

In the forget condition, effects associated with the first (to-be-forgotten) item were quite small. The effects of R_2 and I_2 were comparable to those found in tasks S and DP. Performance was found to improve with I_1, as one would expect, but only very slightly. It was found, however, that increasing R_1, that is, the rehearsal period on the to-be-forgotten item, actually improved performance on the to-be-remembered item. The results for the control condition were similar to those for the simple condition, but retention was better, clearly illustrating a release from PI effect.

The data were analysed within a conceptual framework that distinguished between four types of storage: working memory (WM), a short-term acoustic store (STS), items in a long-term store (LTS) that were either uniquely coded (UC) and could be easily retrieved, and items in LTS that were only generally coded (GC) and suffered from response interference.

Within this framework, it was concluded that rehearsal allowed for more unique coding in LTS, since rehearsal of items to be remembered had a large beneficial effect, and rehearsal of to-be-forgotten items reduced their interference on to-be-remembered items. The retention curves obtained in task S also suggested that rehearsal was fortifying the STS trace.

It was decided that the beneficial effects of spacing in task DP could not be explained by some kind of release from PI on the second presentation; a comparison with condition C was employed, and gave a significant underestimate of the spaced practice improvement. Furthermore, a consolidation explanation likewise underpredicted the size of the effect quite drastically. It was concluded that the major cause of spaced practice improvement in task DP was that the longer spacing interval allowed for more unique (and therefore more effective) encoding of the stimulus at the second presentation. However, on the basis of his data, Pollatsek was unable to discriminate between the three possible rationales which might underly such a mechanism.

Bjork and Allen (1970) also produced striking evidence against consolidation. The paradigm may be represented as follows:-

$$P_1 \; I_1 \; P_2 \; I_2 \; T$$

where P_1 and P_2 represent two presentations of a noun trigram; I_1 denotes the interpresentation task (digit shadowing), either 3 secs. (easy), 12 secs. (easy), 3 secs (difficult), or 12 secs. (difficult), and I_2 denotes the retention task of 8 or 12 seconds' moderate difficulty.

TABLE 7

Proportion of complete noun trigrams correctly recalled (Bjork and Allen 1970).

		CONTROL	I_2=8secs	I_2=20 secs
I_1=3 secs	Easy	.75	.87	.79
	Difficult	.62	.89	.79
I_1=12 secs	Easy	.66	.92	.82
	Difficult	.51	.94	.89

A control condition $P_1\ I_1$ T was included, as a check on the effects of the various I_1 conditions. Results in terms of complete trigrams are presented in Table 7. It is clear that the results for the control condition support the labelling of the interpresentation task as easy or difficult. The control results also show the level of recall just prior to the second presentation in the experimental conditions, and it is clear that performance following P_2 is generally better the lower the retention level on entering P_2. This result is clearly quite embarrassing to the consolidation position.

However, Pollatsek (1969) suggested that an underlying long-term trace might be strengthened by consolidation, whereas performance at short retention intervals might be 'boosted' by STS effects which are not subject to consolidation in the same way. However, one would expect more consolidation with an easy interference task in this case; it is clear that the 3 secs and 12 secs easy tasks do not produce the widely different retention levels in the experimental conditions that one would expect if the 12 second task allowed for 4 times the amount of consolidation as the 3 second task. Furthermore, difficult interpresentation tasks lead to better retention in double presentation conditions than comparable easy ones, especially in the 12 second case. This result poses extreme difficulties for any kind of consolidation hypothesis.

It would therefore appear that Pollatsek's differential encoding theory is the only one that is consistent with all the

Brown-Peterson results so far available. It is interesting to note that he rejected a simple 'release from PI' hypothesis, but did not include release from PI as a possible facilitative factor in finding a unique code on the second presentation, which appears an equally feasible underlying rationale for a differential encoding hypothesis to those that he actually did advance. Such a hypothesis might also explain what Pollatsek meant by a "bad" code.

3.4 Summary

The results of free-recall studies very strongly suggest that the benefit derived from spaced presentation in such a study springs mainly from additional contextual cues which become available for encoding with increased spacing. The Gartman and Johnson (1972) study, whilst superficially convincing, should be viewed with caution, since in this situation, experimental items were placed in semantically related contexts, and such contextual cues would not normally be available in most of the studies on spacing. Similarly, Madigan's (1969) study involving not-to-be-learned cue words was untypical of normal spacing studies in free-recall. There is a danger that in both these studies, the procedure employed would suggest a contextual encoding strategy that the subject wouldn't normally employ. However, both studies do establish beyond doubt that contextual information can be employed by subjects in the encoding of free-recall lists. Furthermore, the fact that same-context homogrpahs in Gartman and Johnson's experiment showed far less improvement with interpresentation spacing than controls presented in the same lists can only be interpreted as evidence that contextual information is normally employed in free-recall. In addition , their different-context homographs showed no spacing effects, and such effects were much attenuated in non-cued recall of Mandler's different-cue words, Thus, contextual encoding was accounting for a great deal of the benefit derived from the spacing of presentations.

When taken in conjunction with frequency judgement data, these results provide much more convincing support for the differential encoding hypothesis. This is not to say that other factors may not be operating to enhance performance with increased spacing. As Melton (1970) has pointed out, contextual cuing in free recall may only account for the greater benefit derived from spacing in this situation as compared with other memory paradigms.

The data from recognition memory studies are far more ambiguous and may be explained either in terms of a hypothesis of differential processing of items currently held in STM, or by a multiple encoding hypothesis. It is relatively easy to see how the encoding hypothesis would account for performance in free-recall; recall of a particular word would act as a cue for the recall of others. However, in a recognition memory study, a word presented at a particular test would generally be in a different context to those in which it appeared on previous trials, so that it is not at all obvious how contextual cues from earlier trials would become available to the subject to aid performance. The search hypothesis proposed in section 3.2 is at least a feasible explanation of how differential encoding resulting from spacing might operate. Neverthe less, on the whole, the data cannot be said to unequivocably support a differential encoding interpretation.

Brown-Peterson studies strongly indicate a differential encoding hypothesis, and indeed, the results of Bjork and Allen (1970) certainly render a consolidation approach untenable. However, once again, Pollatsek's (1969) conclusions do not discriminate between a number of rationales that could underly such a hypothesis, one of which was essentially a differential processing hypothesis. Furthermore, Pollatsek did not examine a sufficient range of interpresentation intervals to establish whether the effectiveness of a second presentation was dependent on an item'sretention in short-term memory. However, it is clear that the contextual cuing hypothesis, which is such a feasible explanation of spaced presentation effects in free-recall, cannot possibly apply to Brown-Peterson experiments. In conclusion, then, this chapter has been of great value in spelling out a number of alternative hypothesis that could account for the spaced practice effect in verbal memory, and furthermore, it has emphasised the point that the effect may well have different underlying rationales in different experimental situations.

CHAPTER FOUR

SOME HYPOTHESES CONCERNING THE SPACING OF PRESENTATIONS IN PAIRED-ASSOCIATE MEMORY

Results concerning the spacing of presentations in a number of other memory tasks, and in particular free recall, suggest that encoding processes may play a prominent role in determining the beneficial effects of distributed practice. Therefore, before going on to specify specific hypothesis concerning this effect in paired-associate memory, a brief examination of current theory and results in the field of encoding and organisational processes in paired-associate memorising will be undertaken.

4.1. Stimulus Encoding

A prominent line of reasearch in paired-associate memory has concerned the development and examination of associative interference theory by the use of negative transfer experiments based upon the classical paired-associate learning paradigms. For example, an influential study by Barnes and Underwood (1959) employed an A-B, A-C transfer paradigm in which a list of eight CVC - adjective pairs were learned to a criterion of one correct anticipation trial. Following the learning of the original list (A-B), 1, 5, 10 or 20 anticipation trials were administered on a new list, consisting of the original CVC stimuli re-paired with new adjective respons s (A-C). Subjects were then presented with the eight stimuli and instructed to recall both the first list and second list responses. It was found that as the amount of training on the second (A-C) list increased, so the proportion of correctly-recalled second-list responses increased, and there was a corresponding decrease in the recall of first list respons s. Furthermore, the level of recall of second list responses was generally lower than that of the r sponses of a similar list that had not been preceded by a competin A-B list. Results of this kind have since been replicated many times (e.g. Koppenaal, 1963; Postman, Stark and Fraser 1968). Such results were explained in terms of associative interference theory (see 1.21). Thus, the original A-B associations were seen as proactively inhibiting the formation of the new A-C associations, whilst these were in turn assumed to overwrite, or interfere retroactively, with the original A-B associations. The negative correlation between the recall of first and second list responses across the amount of re-training on the second list served to reinforce this view. However, these results presented only part of the picture. In particular, at each level of A-C training, it is probable that the same level of first list response recall would have been found regardless of whether the second-list response was correctly recalled or not - the so-called "independent retrieval phenomenon" (Martin, 1971; Greeno, 1969, See 1.24). As has already been stated results of this nature pose great difficulties for associative interference theory.

Although association theory has been found lacking in a number of respects, it does posses the attractive property of regarding the memory trace "as a constructive process relating to a cognitive act" (Neisser 1967); in other words, as a functional entity related to information already stored in memory - the mind is not seen as a blank slate. However, encoding theories also stress the functional nature of the memory trace, or code. Perhaps the major objection to association theory lies in the specific way in which it defines this functionality as deriving from direct word associations in memory. Furthermore, there is a growing body of evidence that the implicit assumptions underlying the acceptance of paired-associate learning as a straightforward paradigm for the learning of associations are false. In the first place, there is an implicit assumption that the nominal repetition of a stimulus or stimulus-response pair gives rise to an identical functional repetition; this is especially crucial in the interpretation of the results of transfer experiments. Secondly, it is assumed that learning takes place within pairs, and that therefore it should be unaffected by inter-pair organisation, and indeed should not be amenable to any form of subjective organisational process. There is considerable evidence to the contrary.

4.11 Stimulus meaningfulness and encoding variability

One of the most important recent developments in our understanding of paired-associate learning is the realisation that reptition of a particular stimulus, or stimulus response pair, does not necessarily lead to a repetition of the same encoded version of the stimulus or pair. Martin (1968b) has called this phenomenon "stimulus encoding variability", and has argued that many paired-associate learning phenomena can be explained in terms of a hypothesis that different stimuli have different numbers of possible perceptual/encoding responses that can be made to them.

Clearly in order to compare learning performance on stimuli with high and low encoding variability, and therefore test the predictions of a stimulus encoding variability hypothesis, it is necessary to have some ad - hoc method of classifying stimuli with regard to their degree of encoding variability. Martin has argued that consonant-vowel-consonant (CVC) and consonant trigram (CCC) stimuli constitute a ready-made set of stimuli classified in such a way, and that their encoding variability can be deduced from their meaningfulness (M).

Traditionally, M is associated with such variables as whether or not the stimulus elicits an association (Glaze, 1928, Witmer, 1935), how many associations it elicits (Nobel, 1952) how much like a word it is (Archer 1960) and how pronounceable it is (Underwood and Schulz 1960). Thus, high - M verbal units are seen to be better integrated serially (i.e. they are more word-like, as opposed to a random collection of 3 letters) and to elicit a greater number of associations with actual words. There is little controversy about the effects of meaningfulness on verbal learning; it has a facilitative effect. That is to say, high-M items are in general easier to learn than low-M stimuli in a single list learning situation (e.g. McGeoch, 1930; Cieutat, Stockwell and Noble, 1958)

The traditional view of how stimulus - M affects paired-associate learning is well represented by associative probability theory (Mandler, 1967, pp32-33; Underwood and Schulz, 1960, pp45-49) which claims that stimuli giving rise to a greater number of associations (high - M stimuli) are more viable in learning situations than are stimuli giving rise to fewer associations, because such stimuli are more likely to form an association with a response through the mediation of one of these existing associations. However, Martin places an almost completely opposite interpretation on the role of stimulus M.

In an examination of stimulus cue selection, Underwood (1963) demonstrated that subjects tend to associate overt responses with fewer than all aspects of the nominal stimuli than the learning task would appear to permit. Shepard (1963), in commenting on Underwood's paper, added the observation that stimulus M may be viewed in terms of the extent to which the stimuli may be analysed into components. Low -M stimuli were considered to be more fractionable and less well integrated into single word-like units than high-M stimuli. Martin (1968b) extended these observations into the hypothesis that there exists some kind of variability in the functional stimulus (that is, the perceived, encoded version of the nominal stimulus) which is inversely related to stimulus M.

Martin's hypothesis may be explained more fully with reference to the following schematic representation (somewhat idealised!) where S_H and S_L are, respectively, high-M and low-M nominal stimuli, and r-s corresponds to some central event composed of a perceptual response (r) plus the consequent functional encoding (s) of the nominal stimulus S.

$$S_H \longrightarrow r\text{-}s$$

$$S_L \begin{array}{l} \nearrow r_1\text{-}s_1 \\ \rightarrow r_2\text{-}s_2 \\ \searrow r_3\text{-}s_3 \end{array}$$

Thus, whereas S_H (e.g. the trigram MOP) is seen to have only one possible functional encoding, S_L (e.g. the trigram XCL) has three possible functional encodings, only one of which will be assumed to occur at any given time. Martin has called the event $S \rightarrow r - s$ (the elicitation of the perceptual response by the nominal stimulus together with the consequent encoding of that stimulus) the E - phase, or encoding phase, of the paired-associate process. The formation of associations between the momentary encoded version s of stimulus S and the overt response, and the elicitation of previously-formed associations by s are referred to as the A-phase, or association phase.

Of course, the above representation of the hypothesis is highly idealised, and it is not possible to determine exactly which perceptual/encoding responses may really be made to the trigram XCL, nor their relative probabilities of occurrence. However, in the case of alphabetical configurations, the work of Postman and Greenbloom (1967) and others (Underwood 1963) suggests that the initial letter of the stimulus is a high probability perceptual encoding. A rather more detailed (but still incomplete) distribution in terms of individual letter members could in theory be constructed from the results of a study by Yum (1931). Nevertheless, this aspect of the E-phase must remain rather indefinite; it is relatively unimportant, however, if the relative encoding variability of particular items can be deduced.

4.12 Repetition and Stimulus encoding

It has also been proved possible to gain some idea of the effect of repetition on encoding variability. In the case where repetition is of the stimulus-familiarization form, stimuli are rehearsed prior to, and independently of, paired-associate learning. Shulz and Martin (1964) have shown that 30 stimulus-familiarization trials have the same (facilitative) effect on subsequent paired-associate learning over a range of levels of stimulus M as do 30 familiarization trials on trigrams which do not appear subsequently in the paired-associate task.

In a study test paired-associate task, Martin (1966) forced one group of subjects to rehearse aloud the complete trigram stimulus during a short interval between the appearance of the stimulus and that of the response on study trials. A second group were required to count backwards. It was found that the deleterious effect of backward counting did not differ over four levels of stimuls M, and that there was no interaction between stimulus M and intervening activity. Therefore, increasing stimulus availability (in the sense of stimulus recallability) did not affect the acquisition of associations differentially over M. These results together suggest that although independent familiarisation with the nominal stimulus may assist in gaining experience with trigram stimuli, and perhaps a number of their possible perceptual encodings, thus increasing the overall probability that a perceptual encoding response (and hence an association with the overt response) will be made, it does not appear to differentially affect the relative availability of one alternative encoding to another for the purpose of incorporation into a response-producing association code.

In contrast, when repetitions of the stimuli occur in the context of on-going paired-associate learning, several experimental results indicate that the relative availability of the various functional encodings of a particular stimulus does alter, firstly towards degeneracy about the preferred functional version of the stimulus (a sort of "focussing in" onto the preferred functional version) and then with overlearning towards the inclusion of additional, alternative or more elaborate encodings. For example, a series of studies by James and Greeno (1967) employed compound stimuli composed of a word and a CVC, paired with digit responses. After pretraining with the compound stimuli-digit pairs to various criterion levels, subjects were given a number of trials on a new paired-associate task. The pairs in this second task were generated by breaking down the original pairs into a word-digit pair and a CVC - digit pair, each stimulus retaining the response digit that had previously been assigned to the corresponding compound stimulus. The proportions of word and CVC stimuli to which no errors were committed on the second task were compared. This proportion increased with pre-training for word stimuli, but remained constant at a very low level for CVC stimuli, except when a very large degree of pretraining had been administered, when it also showed an increase. It was also found that first-task learning could not be attributed to the word components of stimuli alone. These results suggest that, as stated above, with on-going paired-associate learning, subjects tended to "focus in" on specific aspects (functional versions) of the compound stimuli (to some extent on the word components), and it was not until a large amount of over-training had been administed that non-preferred functional versions (the CVC components) appeared to acquire cue function. These results were replicated several times by James and Greeno, incorporating additional controls.

It has also been shown that with learning, and certainly with overlearning, associations may form among the components of any given stimulus (James and Greeno 1967; Postman and Greenbloom, 1967), and so it is possible that in the above study, improved second task performance on the CVC components with overtraining may in part have been due to mediation through the word components. This possibility nevertheless still requires the relaxation of the selective process to allow encoding of the CVC components, and may be regarded as complementary to the encoding variability hypothesis.

4.13 Stimulus Recognition.

Given that the nominal stimulus S can be variably encoded, and given that a particular functional encoding is elicited on a given trial, then it follows that if the learner encodes the stimulus S differently on the next trial, he will fail to recognise that stimulus as the same stimulus that occurred on the previous trial. Melton and Martin (Martin 1967b) utilised the Shepard-Teghtsoonian (1961) continuous recognition memory paradigm (see 3.2) to study the effects of a between-groups manipulations of trigram M. High, medium and low- M lists were made up, each of 160 CCC or CVC trigrams, each of which was presented twice at lags of 1, 3, 6, 15 or 30 intervening items. Intralist formal similarity did not vary over M. It was found that recognition performance declined as lag increased, in agreement with the original Shepard and Teghtsoonian study, and that, furthermore, recognition increased as a function of list M at all lags, regardless of whether performance was expressed in raw frequencies, in a form corrected for false recognition, or in terms of the information-processing measure d'. These results suggest that the probability of making the same functional encoding response to a stimulus decreases as a function of the lag between presentation and test, but remains greater for high-M stimuli, as would be expected on the hypothesis that such stimuli posses fewer encoding alternatives.

More significantly, it was found that the false recognition rate for lists consisting of low-M stimuli markedly exceeded that of high-M lists. This result suggests that the functional encodings made to high-M stimuli contained more information about the nominal stimulus than those made to low-M stimuli, and supports the view expressed earlier that low-M stimuli are more fractionable and less well-integrated than high-M stimuli. Thus, for example, the low-M trigram XAL might be encoded as 'X - vowel - L', and would therefore give rise to a false recognition of the trigram XOL, whereas the high-M trigram MOP would almost certainly be encoded as the word "mop", and would be very unlikely to lead to the false recognition of the trigram MAP. Thus, stimulus M would appear to affect the E-phase in two ways; the perceptual fragmentation of low-M stimuli leads to poorer recognition due to encoding variability, and to an increased tendency towards false recognition due to incomplete coding.

The effects of E-phase variability on paired-associate learning may now be examined. The recognition hypothesis stated earlier may be extended to A-phase effects by making the observation that if a particular encoded version s of stimulus S is associated with the overt response

R on trial m, and if a different encoded version s' of that same nominal stimulus S is elicited on trial m + 1, then S will not be recognised, and furthermore, R will not materialise. This has been verified and replicated by Martin (1967a, 1967c) using a study-test paired-associate paradigm. Eight trigram-digit pairs were presented orally for silent study at a 2-second rate; on each trial, the eight study- trials stimuli were randomly intermixed with 16 new trigrams and presented at a 6-second rate. Subjects were required to make two responses to each test stimulus, firstly, a subject had to press one of the two buttons to indicate whether he recognised the trigram as a study-trials stimulus, and secondly, he had to respond with the first digit that came to mind.

In both studies (1967a, 1967c) it was found that although recognition memory for study-trial stimuli increased with trials, the probability of a correct response given a failure to recognise the stimulus remained at chance level. This was found to be so irrespective of the number of times the subject had responded correctly on previous test trials (1967c).

In a further report (Martin 1968a) it was shown that the false recognition rate of filler trigrams increased significantly if the filler trigram in question had its first letter in common with a study-trials trigram. It was furthermore shown that given false recognition of such a filler trigram, the probability of emitting the response which was paired with the study-trials trigram with the same initial letter increased with trials as did correct responding to study-trials trigrams, although it was slightly lower at all repetition levels. These results strongly suggest that the initial letter of the nominal stimulus was a high probability functional encoding, although the fact that intrusion responding to falsely-recognised fillers was slightly below the level of correct responding to study-trials trigrams with the same initial letter suggests that at least some false recognitions were not generated by initial-letter encodings.

The implication of these results is clearly that the overt response can only be elicited by an encoded, functional version of the nominal stimulus to which it has previously been associated. It is also worth noting that the recognition of study-trials stimuli in the above studies began in the 60% -70% range on test trial 1, and thereafter increased with trials to an asymptote. This suggests that the subjects' perceptual responses to stimuli, and consequent functional encodings, were more variable in early trials, and that thereafter this variability decreased with practice during on-going paired-associate learning.

Analagous results to those above were obtained by Bernbach (1967a) employing visual presentation in a continuous paired-associate task. Stimuli were trigrams made up from a set of nine consonants, and were counterbalanced for Witmer (1935) meaningfulness, whilst responses were the digits 1, 2 and 3. Each experimental item received 4 anticipation trials, with an interpresentation lag of 2, 5 or 10 trials. On each trial, subjects were required to make a stimulus recognition response (old/new) followed by a digital response. It was found that both the number of correct recognitions, and the number of correct responses increased with repetitions, and that both measures of performance were higher for short lag items at all presentation levels, as would be expected. Nevertheless, it was found that the response rate was no better than guessing if an item was called new, irrespective of the number of occurrences of the item. Furthermore, the probability of a correct response was found to be an increasing function of the number of consecutive correct recognition responses, independent of the number of presentations. This suggests that repetition is beneficial to the formation of associations only if the stimulus is perceived to be the same on each repetition; that is, the same functional version of the stimulus occurs on each repetition.

So far, no mention has been made of stimulus M effects in paired-associate learning. In the study on stimulus recognition in paired-associate learning reported above (Martin 1967c) stimulus M was varied within lists. It was found that high-M stimuli were better recognised especially at high presentation frequency, and that the probability of a correct response given a correct stimulus recognition was higher at all presentation levels for high -M stimuli. These results taken together yield the expected facilitation is paired-associate learning for high-M stimulus material. However, the latter result appears to deny that the effects of stimulus M can be isolated in the E-phase, but Martin has pointed out that with low-M stimuli, there are, owing to the existance of alternative encoding possibilities, more different processing routes via which recognition might ensue than there are processing routes via which correct responding might ensue. In other words, not every functional encoding that would give recognition has necessarily become associated with the response, whereas with high-M stimuli, there is a greater chance that the same functional encoding occurs on every repetition, and so repetitions will have more value in forming an association with the overt response, and there is a better chance that the encoded version sampled on a test trial is associated with the response.

It was also found in this study that the false recognition rate was not affected by stimulus M. This would appear to conflict with the observations of Melton and Martin (Martin 1967b) reported above. However, in the Melton and Martin Study, stimulus M was varied between lists, whereas in the Martin (1967c) study, stimulus M was varied within lists, and so it is possible that as many high-M as low-M stimuli could be falsely recognised as preceding low-M study trials stimuli by reason of the incomplete perceptual encoding of the low-M stimuli. This result would not follow if stimulus M were constant within lists.

Another apparent inconsistency is implied by these results, in that if stimulus recognition is a necessary antecedent to association activation and is greatly affected by M according to the proposed encoding variability hypotheses, then pre-familiarization with stimuli should affect stimulus recognition, and therefore paired-associate learning, differentially with respect to M, which as we have seen is not the case. However, Martin has argued that in the familiarization situation, the subject is attempting to learn the stimuli in such a way as to be able to recall them; thus the mode of operation of the E-phase is seen as not involving selective encoding of particular aspects of the stimuli, but rather a more uniform encoding of those aspects, to the end that stimuli become serially integrated and hence more verbally reproducible. This could also have been the case in the continuous recognition study of Melton and Martin, even though stimulus M effects did appear in performance. Therefore, although independent stimulus learning does impose ample experience with many (perhaps all) of the possible encodings of any given stimulus, and stimulus M may differentially affect this independent stimulus learning, the operation of the E-phase during independent stimulus learning is not seen as settling upon a consistent encoding. Thus, although in a subsequent paired-associate task stimulus recognition may be better for higher-M stimuli, as a result of pre-familiarization, the association activating power of any one encoding is initially weak owing to its infrequency of co-occurrence with the response term. Therefore, although pre-familiarisation should produce an overall facilitation of paired-associate learning, due to improved stimulus recognition, it does nothing in the way of selective encoding, and hence should not affect subsequent paired-associate learning differentially with respect to M. This argument is consistent with the results of the familiarisation studies reported earlier.

Despite its obvious success in accounting for the results reported so far, the encoding variability hypothesis is still a fragile structure, as most of these results could be accounted for by the more traditional view of stimulus M, that of associative probability theory. Although the work of Underwood (1963) and Shepard (1963) implies some validity for Martin's interpretation, the most convincing evidence that the encoding variability hypothesis is correct comes from studies of the effects of stimulus M in paired associate transfer experiments.

4.14 Paired-associate transfer studies

In a study by Martin (1968b) two transfer paradigms were employed; the A-B, A-Br task, and the A-B, C-B task. In the former task stimuli from an initial paired-associate learning task (A-B) appear in a second task, re-paired with different responses from the original response set. The scheme below is an example of this task:-

A - B	A - Br
XOL - 1	XOL - 3
ZAG - 2	ZAG - 5
BEX - 3	BEX - 2
HIN - 4	HIN - 1
KUJ - 5	KUJ - 4

The A-B, C-B task, employed as a control, involves an initial paired-associate task (A-B), followed by a second task in which totally new stimuli are paired with the original response set (C-B) for example:-

A - B	C - B
XOL - 1	TUZ - 1
ZAG - 2	FOQ - 2
BEX - 3	GEX - 3
HIN - 4	LAT - 4
KUJ - 5	QIH - 5

The encoding variability hypothesis predicts that although first-list learning (A-B) should reflect the usual effects of stimulus M (that is, facilitation for high-M stimuli), there should be markedly more negative transfer in the A-Br task for high-M than for low-M stimuli relative to the A-B, C-B control, as the low-M stimuli would be more amenable to recoding in the second task, and subjects would therefore not have to modify or overwrite any association code previously formed to link the first list encoded version of the stimulus with the first list response.

Martin's list were made up of 6 CVC - digit paires, stimuli being either high-M (100% Archer 1960) or low-M (21% Archer) Four conditions were employed: A-B, A-Br high-M; A-B, A-Br low - M; A-B, C-B high-M and A-B, C-B low-M. Twenty subjects were assigned to each of the four conditions. In each condition, the first (6 item) list A-B was learned to a criterion of two successive perfect trials, 2(6/6). The second list, A-Br or C-B was then learned to the same criterion, following which the original A-B list was re-learned to 2(6/6) and finally, the stimuli just seen, the A's were free-recalled.

As expected task 1 (A-B) learning was more rapid in terms of trials to criterion for high-M stimuli; this result shows the usual facilitative effect of meaningfulness. Percent transfer scores from task 1 to task 2 were calculated for individual subjects using the formula 100 $(T_1 - T_2)/(T_1+T_2)$ (where T_1 and T_2 are trials to criterion on tasks 1 and 2 respectively) and from these, mean transfer scores were calculated for the four conditions and compared statistically. In the A-B, C-B conditions, similar levels of positive transfer were found for high- and low-M stimuli, and so the facilitative effect of M was maintained in C-B learning.

In the case of the A-B, A-Br conditions, it was found that a small amount of positive transfer was present from task 1 to task 2 for low-M stimuli, and that negative transfer occurred for high-M stimuli; these transfer scores differed significantly. When transfer scores for the corresponding control conditions (C-B) were deducted, it was found that the relative amount of negative transfer in high-M conditions was even more pronounced than in low-M conditions.

It was also found that the variability among individual subject transfer scores from A-B to A-Br was significantly higher for low-M stimuli than for high-M stimuli, which suggests that the process underlying the smaller amount of negative transfer for low-M stimuli was nontheless more variable than that responsible forthe greater amount of negative transfer for high-M stimuli. Again, this conclusion accords well with the hypothesis that with low-M stimuli there is greater coding, and hence recoding, variability. By comparison, no difference between the variability of individual subject transfer scores was found between high- and low-M conditions in the A-B, C-B paradigm.

In task 3, the relearning of the original A-B lists,it was found that relearning in A-B, C-B conditions was more rapid for high-M stimuli, but that there was no significant difference in the amount of transfer from task 1 to task 3 between high- and low-M stimulus conditions; in both cases it was large and positive. In the case of the A-B, A-Br conditions, transfer from task 1 to task 3 was positive for both high- and low-M lists, but was significantly lower for high-M stimuli; in other words, it was harder to get back to the original A-B pairings after task 2 A-Br learning when the stimuli were high-M. Again, this suggests that low-M stimuli could be recoded, whereas associations made to high-M stimuli had to be re-learned.

The above transfer results are precisely in accordance with the predictions of the hypothesis that low-M stimuli can be encoded in a greater variety of ways than can high-M stimuli, and hence are able to provide a greater number of alternative recoding routes in a negative transfer situation. Notwithstanding the above utility of low-M stimuli in negative-transfer situations, far fewer low-M stimuli were successfully free-recalled in both the A-B, A-Br and the A-B, C-B paradigms. This serves to emphasise the observation made earlier; namely that stimuli to be recalled are encoded differently to those whose function is to signal an overt response, and that furthermore the relative unavailability of low-M stimuli in a free-recall situation can be explained in terms of the hypothesis that they are processed in a more fractionated, less well integrated fashion than are high-M stimuli.

It is worth noting that in the above study, the number of perseveration errors in the A-Br task (that is, wrong responses that would have been correct had the original A-B pairings still been in force) was significantly higher for low-M stimulus conditions. This result would appear to conflict with the encoding variability hypothesis and Martin was unable to explain it. This is because he made the implicit assumption that if two functional versions of the same nominal stimulus are associated with different responses, then the subject should be able to determine which of these functional encodings is relevant to the present task. It is quite possible that on the early trials of an A-Br transfer task, if the first task encoded version of the stimulus occurs, leading to the elicitation of the original(inappropriate) response, then the subject may be able to "tag" that association code as no longer relevant, as part of the association re-learning process. If the first task stimulus version is then elicited on a later trial, it is likely that in the absence of a new association with the relevant response, the subject will guess from the remaining response alternatives rather than knowingly make an inappropriate perseveration. This could account for the lower rate of perseveration to high-M stimuli, when the probability of the first-task stimulus version is very high (if not unity) in the transfer task. In the same way, with low-M stimuli, the subject may adopt a strategy of tagging the first list functional version of the stimulus as inappropriate. However due to encoding variability, he would have relatively fewer opportunities to do this to a particular functional encoding. Alternatively, he may prefer to search around for an alternative functional encoding rather than waste valuable processing time in tagging inappropriate first list stimulus encodings in this way. Whichever of these hypothesis is accepted,the net result would still be a greater degree of perseveration to low-M stimuli than to high-M stimuli during an A-Br transfer task.

Transfer studies are important in that they justify Martin's original hypothesis that low-M stimuli are more fractionable, and less well integrated than high-M stimuli; in other words, that they give rise to a greater degree of encoding variability. It was remarked earlier that associative probability theory can account for the results of single list learning situations with respect to stimulus M. However, if the argument that high-M stimuli are more viable in learning situations because they give rise to a greater number of associations that may mediate with the response is applied to the transfer study reported above, then it is clear that the ensuing predictions with respect to the effects of stimulus M will be totally opposed to the observed results. This strongly suggests that Martin's interpretation of stimulus M is correct.

The result that high-M stimuli lead to a higher degree of negative transfer in an A-B, A-Br paradigm when compared with a control A-B, C-B paradigm than do low-M stimuli has been replicated by Martin and Carey (1971); a similar, but non-significant effect was also found by Weaver, McCann and Wehr (1970). On the other hand, Postman and Stark (1971) found that high-M stimuli lead to less negative transfer in the above situation. Martin (1972) has argued that in some cases, learners may prefer to form new associations to old functional stimuli rather than part with established functional encodings, and that this preference may be to a large extent determined by task conditions.

Credence for this argument is provided by the studies of Merryman and Merryman (1971) and Schneider and Houston (1969). Both these studies made use of an A-B, AX-Br paradigm, wherein an additional redundant component was added to each stimulus during the learning of the transfer list. Merryman and Merryman found that their subjects opted to make use of the new cue, and in doing so were able to reduce interference between the two tasks, whereas Schneider and Houston found that their subjects effectively ignored the new cue. It is possible that in the latter task, subjects found it easier to modify their associations to the first task stimuli than to form new ones to the additional second task cues.

As a general theory of paired-associate learning phenomena, the stimulus encoding variability hypothesis still has a long way to go. There is at present an almost universal ignorance regarding the determinants of perceptual encoding. As intimated above, one relevant factor insofar as stimulus recoding is concerned may be the relative difficulties (imposed by task variables) of stimulus recoding and of the formation of new, conflicting associations. However, even in a single list situation, there are still many questions to be answered. For example, to what extent is perceptual encoding determined by the subject's reaction to task variables, as opposed to straightforward contextual effects? If the subject "focuses in" on a preferred encoded version of the stimulus during on-going paired-associate learning, to what extent is this determined by an active subject strategy of attempting to find an unambiguous set of functional encodings corresponding to the nominal stimuli, as opposed to same kind of passive, reinforcement process? Although Greeno (1970b) has argued (with data) that the response item of a paired-associate pair may be a factor in the determination of the functional encoding of the stimulus, these questions have still to be answered.

Although Martin's formulation of the encoding variability hypothesis depends heavily on the use of the meaningfulness of trigram stimulus material in order to form some kind of a-priori ranking of stimuli according to their encoding variability, so that experimental verification of the hypothesis is possible, it must be taken seriously as a general effect underlying paired-associate learning. The hypothesis could be applied to almost any form of unintegrated alphabetical or digital configuration, but clearly runs into trouble when attempting to deal with actual words. Obviously homographs possess alternative perceptual encodings, but words normally appear to be well-integrated units of 100% meaningfulness. One possible suggestion is that words may be encoded according to acoustic, episodic or semantic properties, or that perhaps some form of encoding variability may derive from the possibility of the perceptual encoding of a word being made at various levels within some kind of hierarchical semantic structure, such as that proposed by Collins and Quillian (1972). Whether these rationales are accepted or not, it would be dangerous to consider the encoding variability hypothesis as a theory solely concerned with alphabetical trigrams, and to reject it out of hand when dealing the paired-associate tasks involving word stimuli.

4.2 Inter-pair organisation

In the majority of paired-associate studies, there appears to be an implicit assumption that subjects attempt to memorize each pair individually, and that in particular, between-pair organisational processes do not exist. Battig (1966) has argued that a number of results involving paired-associate list learning provide evidence of active subject grouping during on-going learning on the basis of the "state of learning" of pairs.

For example, in a study by Brown and Battig (1962) the serial positions of items were randomly varied from one repetition of the list to the next in the usual way. However, when a subject had made his first correct response to an item, the position of that item in the list was thereafter held constant. This condition produced superior performance to the normal varied-order procedure. Reversal of this procedure, so that each pair was presented in the same serial position only until responded to correctly and was subsequently varied in serial position, also produced facilitation as compared with a normal varied-order condition (Battig, Brown and Nelson, 1963). In the latter study, the facilitation was slightly greater than that produced by a constant serial order on all trials for all items. These results suggest that it was not so much serial order *per se* that produced facilitation, but the fact that serial order could be used as a cue to the state of learning of an item.

In a study by Schild and Battig (1966), bidirectional conditions were employed, under which the stimulus and response items of each pair were unsystematically reversed from trial to trial until a pair was first responded to correctly, after which the pair directionality remained constant on all trials. This condition produced less errors per pair than an average value taken from a standard unindirectional procedure and a bidirectional condition , wherein the stimulus and response term order of all pairs was unsystematically varied throughout.

In a study by Brown, Battig and Pearlstein (1965), facilitation was found when new second and third letters were added to an original single-letter stimulus term for a given pair immediately following the attainment of a specified performance criterion (either one or three successive correct responses). Of course, in all these studies it is quite possible that facilitation resulted from cues to the learning difficulty of pairs (or their adequacy of encoding) rather than their "state of learning". However, such a distinction is merely a semantic quibble, as to all intents and purposes, they would be equivalent in so far as telling the subject where to direct his effort. However, none of these results can be taken as evidence that subjects actively group items of equal difficulty together during on-going learning.

In an experiment by Battig and Bernstein (1965), subjects learned a 12-item list which was either homogenous (all item pairs of equal learning difficulty) or heterogenous (items included both words, and CVC's of minimal association value). Subjects were then given 12 individual cards, each containing one of the pairs, and asked to arrange them into groups on any basis they could. Results suggested that each subject tended to group the cards on the basis of his own difficulty in learning them, although this effect was more pronounced in the heterogenous list condition. Battig (196) interpreted this result as evidence that subjects employ an active grouping strategy during on-going paired-associate learning.

This interpretation, however, is suspect. The fact that subjects are demonstratably able to group items on a basis of difficulty does not necessarily mean that they actively employ this ability during learning. In fact, as Herriot (1974) has pointed out (see 1.25) it has not even been established that such a strategy is employed in free-recall, where the evidence is somewhat more convincing. However, these studies do show that subjects become aware of the learning or encod-difficulty of items when learning lists, and that furthermore they can make use of cues to such difficulties during on-going learning. However, such information may just indicate to them where to direct their effort, since there is no way that, in general, encoding paired associates in groups would be of value at test.

4.3. The Hypotheses

The various hypotheses to be advanced to account for the beneficial effects upon memory performance derived from the spacing of presentations in a paired-associate task must all be made in terms of a general theory of paired-associate forgetting. Thus, in the following section, the theory tentatively outlined in section 1.33 will be restated in the light of the results discussed earlier in this chapter. This theory will thus provide a general conceptual background against which to examine the various hypotheses.

4.31 A theory of paired-associate forgetting

Although the theory to be stated is principally an encoding theory, it makes a number of implicit assumptions regarding short-term memory (STM) which should first of all be stated. Short-term retention effects are seen as possibly reflecting three different underlying factors. The first of these is some kind of echoic sensory store or structure in which traces are seen to decay over time. Secondly, an active conscious rehearsal process may be responsible for short-term retention. A distinction is made between "passive" or "rote" rehearsal, by which the nominal stimulus or a number of nominal stimuli may be cycled through attention by sub-vocalisation, and "active" rehearsal, which operates to hold functional and nominal aspects of a stimulus in attention whilst encoding takes place. The rote rehearsal process is clearly of limited capacity, because of the nature of echoic decay, whilst the active , or encoding, rehearsal process is seen as being limited to some extent by the assumption that it is an attentional, real time process. Thirdly, short-term retention effects may reflect memory proper (which will be called LTM on the grounds that it is responsible for long-term performance) to the extent that certain episodic and articulatory-acoustic - phonemic aspects of an encoded stimulus may be subject to an enormous amount of interference and hence rapid decay. For the sake of brevity, all these potential processes will be grouped together and called short-term memory (STM).

In a typical paired-associate memory procedure, the subject is presented once only with each pair, and the following sequence of events is hypothesised to occur. Firstly, some perceptual response is made to the nominal stimulus, and this perceptual response will to some extent determine the functional, or encoded form, of the stimulus. Since it is thought that much of the difficulty that occurs in paired-associate memory stems from an inability to recognise the stimulus under test or to discriminate it from other, similarly encoded stimuli, attention will principally be focussed on stimulus encoding. However, it should be borne in mind that a functional encoding of the response must also be made, and furthermore, an associative encoding which will link it with functional aspects of the appropriate stimulus.

It is possible that certain functional aspects of the stimulus are related to the response fairly early on in the encoding process, since there is some evidence that stimulus encoding is determined to some extent by the response. Although it is not known exactly where in the sequence association codes are formed, it is fairly certain that a second stage of stimulus encoding takes place, wherein encodings of previously-presented stimuli that are similar to that of the current stimulus become available and are taken into account in completing the encoding of the current stimulus. It is postulated that these encodings are "cued" both by semantic aspects of the current stimulus, and by episodic aspects of the current presentation event which contribute to the current perceived functional stimulus. Consequently, the subject may find it necessary to elaborate, or even change, his current stimulus encoding in order to discriminate it from previously-presented stimuli. However, it is postulated that in all probability, these previous encodings are not modified or elaborated, since if they were, one would expect far smaller retroactive interference effects relative to proactive interference effects than actually occur. Thus, at test, when a perceptual response is made to the nominal stimulus that is ambiguous, in that it shares functional aspects with several stimuli, then the stimulus will be identified as that whose encoding possesses the most similar episodic features to the functional stimulus at test, all other things being equal.

At first glance, this theory appears to contradict the general finding that stimulus encoding during on-going paired-associate learning is reductive in nature. However, it should be borne in mind that in a list-learning situation, pairs are presented repeatedly in randomized order, so that the subject will rapidly gain some appreciation of the total stimulus set which he has to discriminate, which after all comprises a relatively small number of items. During a memory task, however, the subject has to discriminate each stimulus from a potentially very large set of not-yet-seen stimuli. Thus the list-learning process of "focussing in" on a specific preferred version of the stimulus may not be typical of memory tasks in general. Furthermore, codings might well be elaborated in terms of episodic cues that point to the preferred semantic encoding that these list-learning studies have isolated; clearly, it would be extremely difficult to detect the episodic features by which basic semantic encodings may be elaborated.

Although this theory is by no means complete, it adequately describes the effects of prior and posterior activity on paired-associate memory performance, and is certainly consistent with the current state of knowledge concerning stimulus encoding. It will therefore suffice to serve as a conceptual basis for the hypotheses now to be advanced.

4.32 Consolidation

In its most simple form, the consolidation hypothesis states that a memory trace is able to gain strength in some way every moment it remains in memory. In other words, every moment that a memory trace has failed to decay increases its subsequent resistance to decay. Close examination of retention curves often reveals that the rate at which performance declines itself declines as a function of the retention interval (e.g. Wickelgren, 1973) and this is often taken as evidence of consolidation. However, it should be pointed out that retention curves are almost always obtained by averaging data across many subjects, or by averaging data for a single subject across many observations taken across a period of time. Thus, if forgetting rates varied between subjects, or within a subject over a period of time, one would expect to find a decline of forgetting rate with retention interval, since with a longer retention interval more points arising from "good" subjects, or each subject's better part of the session will contribute to the data. In addition, it is quite possible that there are various levels at which material can be encoded; some items will decay rapidly, others more slowly. Again, the longer the retention interval, the greater the contribution of slowly-decaying items to the data, and the slower the observed rate of forgetting. It would be almost impossible to design an experiment to discriminate between consolidation and a hypothesis based upon a sampling distribution of forgetting rates.

In addition to these arguments, the results of Bjork and Allen (1970) discussed in section 3.33 place such a consolidation hypothesis beyond consideration. However, an alternative form of the consolidation hypothesis has been advanced by Atkinson and Shiffrin (1968) and a mathematical model based on this hypothesis has been applied with great success to the paired-associate experiments of Atkinson, Brelsford and Shiffrin (1967) and Brelsford, Shiffrin and Atkinson (1968) described in Chapter One. Stated simply, the hypothesis claims that items entering the memory system from a sensory store are placed in a fixed capacity "rehearsal buffer". Items in the buffer remain there until displaced by the entry of new items. A long-term memory trace is assumed to be built while an item remains in the buffer; the longer the stay in the buffer, the greater the long-term trace strength. Once an item has left the buffer, then its long-term trace is assumed to decay in some way.

Thus, this theory would explain the improvement in performance on paired-associates with spaced presentations in terms of the hypothesis that items continue to be learned during the presentations of successive later items regardless of the state of learning of these items (since they may still remain in the rehearsal buffer after their own presentations trial has ceased). Thus, if on average an item remained in the buffer for say, ten trials, then with an interpresentation spacings of less than ten, items would not receive their full complement of processing. They would be optimally processed with an interpresentation spacing of ten trials, and would receive maximal processing but would also suffer from decay with an interpresentation spacing in excess of ten trials. The theory also predicts an interaction of interpresentation and retention interval of the kind found by Peterson, Hillner and Saltzman (1962) (see 1.44 and Figure 9), since at short retention intervals, performance will be enhanced by recall from the rehearsal buffer, and an item is more likely to be in the buffer after two massed, rather than two spaced presentations.

There are a number of objections to this buffer theory as a general explanation of the effects of the spacing of presentations in paired-associate memory. In the first place, it should be pointed out that the studies that the theory was designed to explain all involved an extremely slow rate of presentation (11-second anticipation trials) which might well allow the subject ample time to rehearse previously-presented items, even whilst processing the current item. Furthermore, in these studies short retention intervals were far more frequent than long ones (another factor that might lead the subject to employ a shared-rehearsal strategy), and in addition, the material employed (two-digit number stimuli and alphabetical responses) would not be easily amenable to deeper forms of encoding. There is considerable doubt that shared cyclic or sequential rehearsal would be a significant factor in such studies as that of Peterson, Wampler, Kirkpatrick and Saltzman (1963) described in 1.44 when a 2-second presentation and test rate was employed, and furthermore, stimulus material consisted of highly encodable common words. Nevertheless a beneficial effect of the spacing of presentations is found in such studies. Furthermore, most of the studies involving spacing in paired-associate memory have employed a CPA procedure during which the material to be learned is constantly refreshed, and experimental sessions are relatively long. It is unlikely that a shared rehearsal strategy would be maintained for very long in such situations. Even during the learning of relatively short free-recall lists, (incidentally, a situation in which shared, sequential rehearsal is an even more feasible strategy than in paired-associate procedures) it has been found that subjects are unable or unwilling to maintain a high rate of rehearsal from the beginning to the end of the list, and that indeed the frequency of rehearsal declines monotonically with list position (Rundus, Loftus and Atkinson 1970).

4.33 Multiple Encoding

The multiple encoding hypothesis as applied to the spacing of presentations in paired-associate memory is essentially identical to that propounded by Melton (1970) to explain spaced presentation effects in free-recall (see 3.14). The hypothesis basically states that as the interpresentation interval increases, so the contexts in which the successive presentations occur become less correlated, increasing the probability that different encodings of the item are formed on each presentation. Consequently, on a subsequent test, there will be a greater chance that the perceived functional stimulus is recognised as one of these encoded forms, and thus the response will materialise. In other words, the hypothesis states that the spacing of presentations indirectly serves to increase the number of potential retrieval routes to the response.

Such a hypothesis appears to conflict with the observation that in list-learning paired-associate studies, subjects tend to "focus in" on one particular encoding of the stimulus during on-going learning (See 4.12). However, there is even more convincing evidence that the hypothesis is false, which comes from a recent study by Schwartz (1975), who employed lists consisting of 16 pairs, the stimulus and response elements of which were letter bigrams. For each bigram pair, there were two corresponding word pairs. The stimulus word of each pair began with the two letters constituting the corresponding stimulus bigram, and a similar relationship existed between the response words and the corresponding response bigram. Furthermore, the word pairs were selected from word association norms; the response word of each pair was one of the six most common normative responses to its stimulus word. An example of one of Schwartz's bigram and corresponding word pairs is:- AR-LE, arm-leg, arrive-leave.

Four conditions were tested in an unmixed list, 2 x 2 factorial design. The two factors were presentation (massed VS distributed) and coding (varied VS constant). In the massed presentation condition, each of the 16 bigram pairs was presented twice in succession, whilst in the spaced condition, the list of 16 pairs was presented once, and then repeated in the same order, so that there were always 15 presentations of other bigram pairs intruding between the two successive presentations of any particular bigram pair. Pairs were presented visually at a 4-second rate. On each presentation trial, the display constituted a bigram pair, beneath which appeared one of its corresponding word pairs. In the constant coding conditions, the same word pair appeared on each of the two presentations of the bigram pair whilst in the varied coding condition, a different corresponding word pair accompanied the bigram pair on each presentation.

Subjects were instructed to read aloud both the bigram and the word pairs, and to use the words to help remember the bigram pairings; they were also informed that paired-associate memory of the bigram pairs would subsequently be tested. After the 32 exposures which constituted the presentation of the lists, subjects were required to perform a short distracting task (number reading) in order to remove short term recency effects from performance, and were then given a retention test. This consisted of a sheet containing all 16 bigram stimuli against each of which subjects were required to write down the appropriate bigram response. Once this had been completed, subjects were asked to write down the corresponding word pair to each bigram (or one such pair in the case of varied encoding subjects).

It was found that, given a correct bigram response, the overall probability of correctly producing a corresponding word pair was 0.984, which strongly suggests that subjects were making use of the particular semantic encodings (i.e. the word pairs) that the experimenter had gone to such great pains to provide.

In terms of bigram response performance, it was found that in both varied and constant coding conditions there was a significant spacing effect, with distributed presentations leading to superior performance; furthermore, there was no significant interaction of spacing with coding. Of course, Schwartz confounded spacing with sequential effects in this study, but even so, coding did not differentially affect performance differences with spacing. However, the most damaging result for the multiple encoding hypothesis lies in the finding that, in both spacing conditions, constant coding produced superior performance to varied coding. In other words, in a given spacing condition, it was better to practice the same encoding twice than to form two different ones.

Although Schwartz claimed to have demonstrated the superiority of spaced to massed presentations in an unmixed list study, thereby casting considerable doubt upon a shared rehearsal hypothesis as well, it has already been pointed out that spacing in this study was confounded with sequential factors, rendering such a strong interpretation of the results doubtful. Nevertheless, the result that the varied coding condition produced inferior performance clearly renders the multiple encoding hypothesis untenable.

4.34 Differential Encoding

The differential encoding hypothesis can have a number of equally feasible, alternative underlying rationales. Stated simply, it claims that if an item is presented again after a short-interpresentation interval then that item is likely to be encoded less efficiently that it would have been with a longer presentation interval. Three distinct and separate positions may be adopted. In the first place, differential encoding may be regarded as a passive process, whereby at short interpresentation intervals, "bad" or interference-prone encodings that occurred on the first presentation might survive sufficiently on re-presentation to be employed again; hence at shorter interpresentation intervals, bad first presentation codes will survive because the subject merely employs them again on the second presentation. A more active view may be taken of the subject's role, and it may be postulated that if a bad code survives to the second presentation, then the subject believes he has adequately processed the item, and uses his time either to process something else, or to take a rest from processing. Finally, it may be postulated that at long interpresentation intervals, the subject is generally less confident in any surviving first presentation encodings, and is consequently motivated to improve them. Again, when a first presentation encoding survives in which the subject has confidence, he may either devote his time during the second presentation to processing other items, or he merely may rest and do nothing. It is proposed to set aside the question as to what the subject does if for some reason he doesn't process a currently presented item, although as has been pointed out (see 1.43) there is very little evidence that other items are processed during the virtually useless representation trials on short-spaced items.

Evidence from paired-associate retention curves (see 1.32 and Figure 1) suggests that there are only two types of code; essentially stable long-term codes and essentially rapidly-decaying short-term ones. Consequently, the problem in the first two hypotheses is to find some way of determining how long a rapidly-decaying short-term encoding survives to the extent that it will be reproduced on a re-presentation. There are three possible positions; firstly, a "bad" code will carry over onto the second presentation as long as it will support recall; secondly, such an encoding will carry over as long as the stimulus encoding survives, since this will tend to result in the same association being formed; and thirdly, such an encoding will carry over a longer interval than it will support retention, and a shorter interval than it will support stimulus recognition.

It may be quite feasible to discriminate between these three positions experimentally, but no obvious method based on behavioural data suggests itself as a way of discriminating between the two rationales, namely, that a bad code may be maintained or reproduced on a re-presentation either because the subject just can't help reproducing the encoding, or because he doesn't know enough about the encoding to decide that its inadequate. Both rationales predict that once a sufficient interpresentation interval has elapsed, the bad encodings will not be maintained, either because the subject just can't help thinking of a new one, or because he recognises the bad encoding for what it is, and actively tries to find a new encoding.

The third rationale would predict on ability on the part of the subject to improve an even quite stable encodings at a sufficient interpresentation interval, so that the long-term forgetting rate after a second presentation should be slower than the long-term forgetting rate following a single presentation, and should furthermore decline as a function of interpresentation interval.

4.4 Summary

A number of alternative hypotheses concerning the effect of spacing presentations in paired-associate memory have been proposed. It is considered that in general, differential encoding hypotheses offer the simplest and most general explanations of the effect. In particular, it has been argued that a consolidation hypothesis based upon a serial shared rehearsal process may only apply to situations in which a slow rate of presentation is paired with a preponderance of short retention intervals, whilst a multiple encoding hypothesis is almost certainly erroneous.

A number of discriminable rationales for a differential encoding theory have been isolated as being equally feasible explanations of the spacing effect, as follows:-

1) On the hypothesis that some aspects of inadequate first presentation encodings may somehow survive until a representation, leading to the maintenance of the inadequate code, it is possible that
   a) Such an encoding will be maintained if it can support recall on the second presentation.
   b) Such an encoding will be maintained if the original stimulus encoding survives until the second presentation
   c) Such an encoding will be maintained if the second presentation occurs sometime after the encoding has ceased to support recall, but the encoding may not be maintained over all interpresentation intervals at which stimulus recognition will occur. In particular, at long interpresentation intervals, the surviving stimulus encoding may no longer evoke the original association encoding, or the subject may recognise the inadequacy of the original association encoding.
2) Alternatively, the subject may be dissatisfied with his first

presentation encoding at long interpresentation intervals, even though such an encoding may still support recall, and he may consequently be motivated to improve it.

The following experiments were performed in order to attempt to discriminate between these hypotheses.

CHAPTER FIVE

THE PRESENT EXPERIMENTS

The three experiments reported here were designed to provide data upon the basis of which to discriminate between the various hypotheses outlined at the end of the previous chapter.

5.1 Experiment 1.

It has been suggested earlier (see 1.44) that a particularly sensitive indication of the effect of interpresentation spacing may be provided by an examination of performance at final test conditional upon performance on a test immediately preceding the second of two presentations. Although Young (1966) included such conditions as part of his more complex study, unfortunately it appeared that he tested insufficient replicates of these conditions to observe really stable conditional performance effects. Experiment 1. was designed to correct this omission.

5.11 Method (Exp. 1.)

Subjects The nine subjects employed in this study were undergraduate and postgraduate students at Stirling University, who were paid a small fee for their participation in the experiment. All nine subjects were experimentally naive.

Materials. The stimulus materials employed in this study were selected at random from a stimulus pool of 886 common monosyllabic English words of 3 - 4 letters (the pool may be found in Appendix 1). Responses were the integers 1 - 15.

Apparatus The lists of material were prepared in paper tape form on Stirling University's Elliott 4130 computer. These tapes were interpreted on a standard teletype machine, which had a cardboard mask fitted to it so that only one line of print was visible. The rate at which the display in the "window" of the mask was updated was controlled by having a teletype punch runouts for the desired period (whilst doing this, the machine carriage remains stationary). A line-feed character on the paper tape served to up-date the visible display.

Because of the great volume of noise generated by the teletype the entire sequence output on the teletype was filmed using standard videotape equipment. This process incidentally allowed the experimenter to check each list for errors before actually presenting it to subjects. During the experiment, the videotape was played back over a closed-circuit television screen, giving a display about 1" x 6". Although subjects sat some distance away from the screen, they all reported that they could read the display without difficulty.

Procedure. A study-test CPA paradigm was employed. Six lists were prepared using an "interleaving" procedure as described in section 1.31. Each critical item received two presentations (or study trials). A test trial (T_1) occurred immediately prior to the second presentation (P_2), and a final test trial (T_2) always occurred after an interval of 8 intruding trials on other items following P_2. The interval between the first presentation (P_1) and the first test trial (T_1) was varied according to condition; ten interpresentation spacings were employed, namely 0,1,2,3,4,5,6,8,12, or 16 intruding trials on other items. A typical schedule may be depicted as follows:-

$$P_1 \underset{i \text{ trials}}{\ldots\ldots\ldots} T_1 P_2 \underset{8 \text{ trials}}{\ldots\ldots\ldots} T_2$$

The interpresentation interval i takes values as described above, in the range 0-16.

Each of the six lists comprised ten overlapping blocks. Each of the ten spacing conditions occurred once in each block, in a randomly-determined order. Trials towards the end of one block overlapped slightly with trials at the beginning of the next block. This procedure ensured that the various spacing conditions would be evenly distributed through the list. The stimulus word for each item was selected at random (without replacement) from the stimulus pool. This meant that each critical stimulus occurred on only four trials; two study trials and two test trials. Each stimulus word was randomly paired with a response in the range of 1-15. The same response appeared with the word on each of its two study trials. Where vacant list positions were left by the interleaving process, dummy "filler" pairs were presented. Such dummies comprised a randomly-selected stimulus word paired at random with a response integer in the range 1 - 15. No dummy stimulus was presented more than once. Thus, no word was employed more than once in a list. Furthermore, different words were used in each of the six lists.

Each subject was tested on all six lists. Due to practical difficulties encountered in arranging individual sessions, it was necessary to test the subjects as a group. This unfortunately meant that all subjects were tested in the same order on all six lists. The subjects were instructed to read to themselves everything that they saw on the display, and to respond to test trials by writing down the appropriate response (guessing if necessary) on a prepared response sheet. The response sheets employed comprised rows of ten boxes, in which subjects were to write their responses. After every ten test trials, the experimenter cued the subjects to begin a new row of their response sheets. This procedure was adopted so that if a subject accidentally omitted a response, only one row of the response sheet (involving ten responses) would be lost. Before the experiment proper, subjects were given a short practice session after which they were allowed to ask questions about any points in the instructions that they didn't fully understand. The subjects were then tested on each of the six lists in turn; each list lasted for about 20 minutes, and there was a five-minute break between successive lists. All the lists were presented at a rate of 1 (study or test) trial every 2 seconds.

5.12 Results of Exp. 1.

The items occurring in the first and last block of each list were omitted from the analysis; the first block served as a "primacy buffer" and short practice session, whilst the final block was omitted because it contained large numbers of singly-presented filler items, and was thus untypical. This meant that there would be 48 pairs contributing to each condition; it seemed reasonable to hope that the relatively large sample of material, plus idiosyncratic subject effects would more than compensate for any systematic biases that might be introduced into the data by the fact that all subjects were tested on the same lists in the same order.

It should also be mentioned that any subject item which had a missing response was omitted from the data. This meant that in general, the total numbers of observations varied from condition to condition, and in addition, that subjects did not contribute equal numbers of observations to each condition. It should be borne in mind when following the analysis that the interval i between P_1 and T_1 is both the T_1 retention interval, and the effective interpresentation interval (since P_2 always followed immediately after T_1). Furthermore the interval between P_2 and the final test T_2 was always 8 intruding (study or test) trials on other items.

A useful notation that will be employed throughout is to represent an error on a test trial by the symbol W (for "wrong") and a correct response by the symbol C (for "correct"). In addition, a number may be subscripted to indicate the trial to which the symbol refers. For example, C_1 represents a correct response on T_1, whilst W_1W_2 represents the response sequence "wrong on both T_1 and T_2". Consequently the data from this study may be expressed as proportions of subject item responses falling into the various response categories.

The overall results of Exp. 1. are summarized in Table8. The proportions were calculated across subjects and lists. The symbol n refers to the total number of observations in each spacing condition. Clearly the proportions of items falling into the categories (W_1C_2), (C_1C_2), (W_1W_2) and (C_1C_2) must sum to unity in any particular condition. The T_1 and T_2 performance scores were obtained by adding the proportions

TABLE 8

Results of Experiment 1.

P_1-T_1 LAG	n	$Pr(W_1C_2)$	$Pr(C_1C_2)$	$Pr(W_1W_2)$	$Pr(C_1W_2)$
0	412	.005	.456	.066	.473
1	413	.145	.380	.262	.213
2	420	.219	.321	.357	.102
3	418	.196	.311	.426	.067
4	417	.158	.391	.362	.089
5	420	.138	.383	.424	.055
6	417	.132	.441	.357	,070
8	412	.202	.396	.357	.046
12	413	.179	.392	.378	.051
16	414	.201	.333	.406	.060

P_1-T_1 LAG	$Pr(C_1)$	$Pr(C_2)$	$Pr(C_2/C_1)$	$Pr(C_2/W_1)$
0	.930	.461	.491	.069
1	.593	.525	.641	.357
2	.424	.541	.758	.380
3	.378	.507	.823	.315
4	.480	.549	.815	.304
5	.438	.521	.875	.246
6	.511	.573	.863	.270
8	.442	.597	.896	.361
12	.443	.571	.885	.322
16	.394	.534	.847	.331

P_1-T_1 LAG	Perseverations
0	.370
1	.185
2	.293
3	.320
4	.272
5	.298
6	.268
8	.259
12	.295
16	.369

$Pr(C_1C_2)$ and $Pr(C_1W_2)$ and the proportions $Pr(C_1C_2)$ and $Pr(W_1C_2)$ respectively, whilst the conditional proportions were computed in the usual way; e.g. $Pr(C_2/C_1) = Pr(C_1C_2) / Pr(C_1)$. In addition, perseveration errors are listed for the ten spacing conditions; these are merely the proportions of (W_1W_2) items on which the same incorrect response integer occurred on both T_1 and T_2. When examining these data, it should be borne in mind that the probability of making a correct response by chance (or guessing) is just the inverse of the number of response alternatives; i.e. 1/15 or .067.

Two methods of analysis were adopted. In the first place, in order to gain a rough impression of the trends present in the data, the various performance scores for each condition were computed for individual subjects, and the resulting proportions were subjected to a subjects x conditions analysis of variance. Essentially, the data may be summarized by the three statistics $Pr(C_1)$, $Pr(C_2/C_1)$ and $Pr(C_2/W_1)$, although the proportion $Pr(C_2)$ was also analysed in this way because of its obvious interest. It should be stressed that $Pr(C_2)$ merely measures the proportion of correct responses on T_2 regardless of performance on T_1. Therefore, the conditional proportions $Pr(C_2/C_1)$ and $Pr(C_2/W_1)$ will serve to provide additional information on the relationship of $Pr(C_2)$ with interpresentation spacing i, since performance on T_1 will give some insight as to the state of a particular item when its second presentation P_2 occurs, as P_2 immediately succeeds T_1 in all conditions.

The analyses of variance are all presented in Table. 9.

TABLE 9

Analysis of Variance by Subjects for Exp. 1.

1. Pr(C_1)

Source	SOS	DF1	VE	DF2	F
S	2.6826	8	0.3353		
T_L	0.5691	1	0.5691	8	58.699***
T_Q	0.4343	1	0.4343	8	14.612 **
T_H	1.1404	7	0.1629	56	29.916***
SXT_L	0.0776	8	0.0097		
SXT_Q	0.2378	8	0.0297		
SXT_H	0.3050	56	0.0055		
TOTAL	5.4467	89			

2. Pr(C_2)

Source	SOS	DF1	VE	DF2	F
S	3.5672	8	0.4459		
T_L	0.0289	1	0.0289	8	10.869 *
T_Q	0.0550	1	0.0550	8	14.057 **
T_H	0.0361	7	0.0052	56	1.244 N.S.
SXT_L	0.0213	8	0.0027		
SXT_Q	0.0313	8	0.0039		
SXT_H	0.2320	56	0.0041		
TOTAL	3.9718	89			

3. Pr(C_2/C_1)

Source	SOS	DF1	VE	DF2	F
S	2.3081	8	0.2885		
T_L	0.5490	1	0.5490	8	54.801 ***
T_Q	0.4080	1	0.4080	8	28.925 ***
T_H	0.1301	7	0.0186	56	1.925 N.S.
SXT_L	0.0801	8	0.0100		
SXT_Q	0.1129	8	0.0141		
SXT_H	0.5408	56	0.0097		
TOTAL	4.1291	89			

TABLE 9 continued..

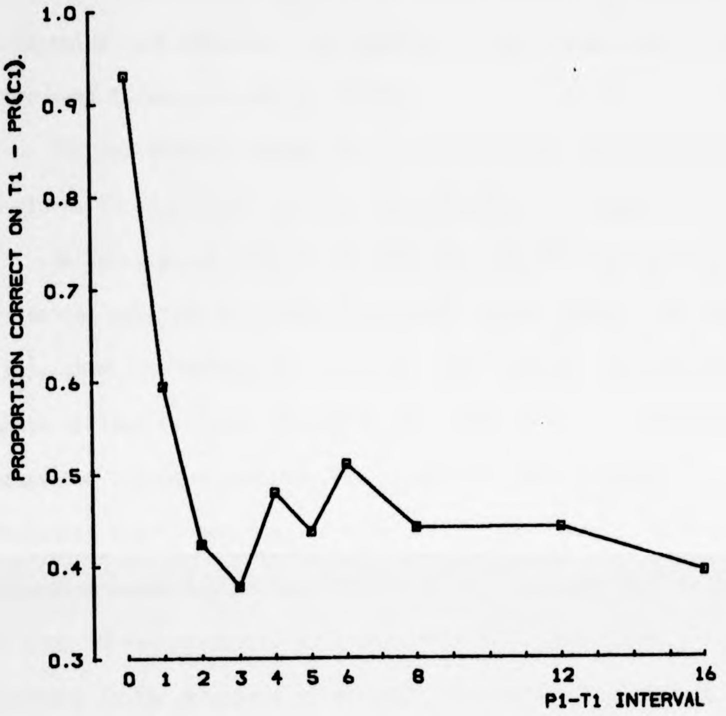
4. $Pr(C_2/W_1)$

Source	SCS	DF1	VE	DF2	F
S	1.6221	8	0.2038		
T_L	0.1170	1	0.1170	8	9.441 *
T_Q	0.1083	1	0.1083	8	8.484 *
T_H	0.6267	7	0.0895	56	9.093 **
SXT_L	0.0991	8	0.0124		
SXT_Q	0.1021	8	0.0128		
SXT_H	0.6195	56	0.0111		
TOTAL	3.2946	89			

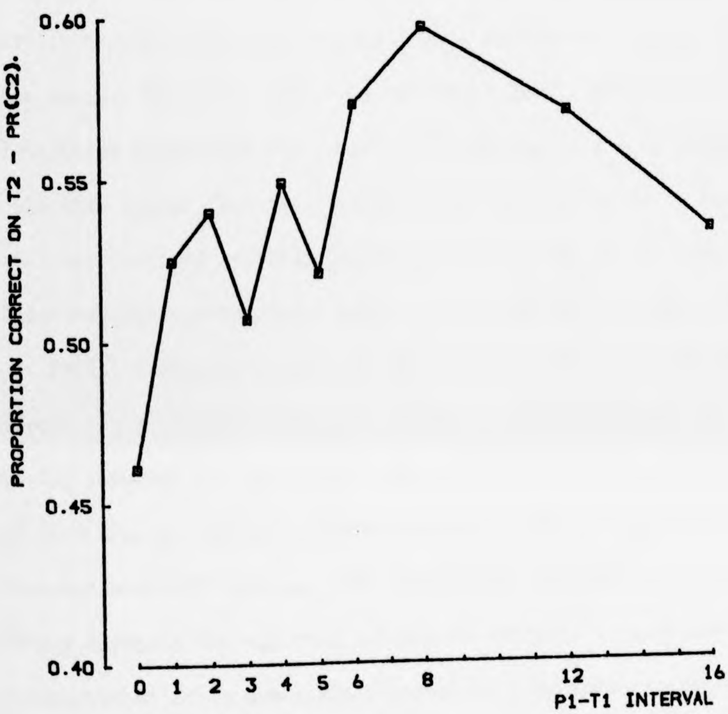
The analyses were performed on the raw proportions observed for individual subjects, in order to retain the meaningfulness of the linear and quadratic components of the spacing interval. Thus, T_L represents the linear effect of the P_1-T_1 interval i on the particular statistic, T_Q the quadratic effect, and T_H the higher - order or residual effects. There are obviously strong theoretical objections to submitting raw proportions to analysis of variance, not the least of which concerns the enormous heterogeneity of single plot variances which will result, However, it should be pointed out that the within-subject proportions were generally based on different numbers of observations in any case, and no appropriate transformation exists in this case. Furthermore, it should be restated that these analyses are only intended as a rough guide; more acceptable statistical methods will also be applied. In addition to the analyses of variance by subjects, similar analyses of the various proportions computed by lists across subjects were also carried out, and these yielded almost identical results to the analyses reported, although unfortunately

FIGURE 14

Proportions of correct responses on T_1 and T_2 as a function of interpresentation interval in Exp. 1.



PROPORTION CORRECT ON T1 - PR(C1).
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0 1 2 3 4 5 6 8 12 16
P1-T1 INTERVAL



PROPORTION CORRECT ON T2 - PR(C2).
0.60
0.55
0.50
0.45
0.40
0 1 2 3 4 5 6 8 12 16
P1-T1 INTERVAL

there were insufficient observ tions to carry out a conditions x lists x subjects analysis, and considering the many reservations concerning such analysis, and exercise of this kind would certainly have been rather a waste of effort.

The performance scores on T_1 and T_2, $Pr(C_1)$ and $Pr(C_2)$, are depicted as a function of the P_1-T_1 interval i in figure 14.

In the case of $Pr(C_1)$, the spacing i may be regarded as a retention interval separating the first presentation of an item (P_1), from its subsequent test (T_1). The curve in the figure is quite similar to those typically found for retention of a singly-presented paired-associate (See Figure 1,) and it appears to display the usual short- and long-term retention components. Short-term retention seems to disappear at retention intervals of 2 or more trials. These observations are underlined by the analysis of variance However, in the analysis of variance, there is a significant effect of T_H; this may reflect the apparent "noise" in the retention curve between intervals of 2 and 8 trials. Furthermore, the long-term portion of the curve a pears if anything to recover slightly over this range. When the last eight values of $Pr(C_2)$ were compared, a significant difference was found ($\chi^2 = 21.72$, df = 7, p =.003). It would thus appear that the long-term portion of the curve is not stable, and this effect is probably due to the inadequacy of the basic design in confounding testing order with material and with conditions.

$Pr(C_2)$ certainly appears to display a spaced presentations effect. In examining $Pr(C_2)$, it should be remembered that the spacing interval is essentially the interpresentation interval, and that $Pr(C_2)$ reflects performance on T_2, which always follows P_2 at an interval of 8 trials. The appropriate analysis of variance clearly supports the apparent pattern of improvement with spacing; an improvement up to a maximum followed by a subsequent decline. However, there was no significant difference between the values of $Pr(C_2)$ at interpresentation spacings of 2 or more ($\chi^2 = 10.42$, df = 7,

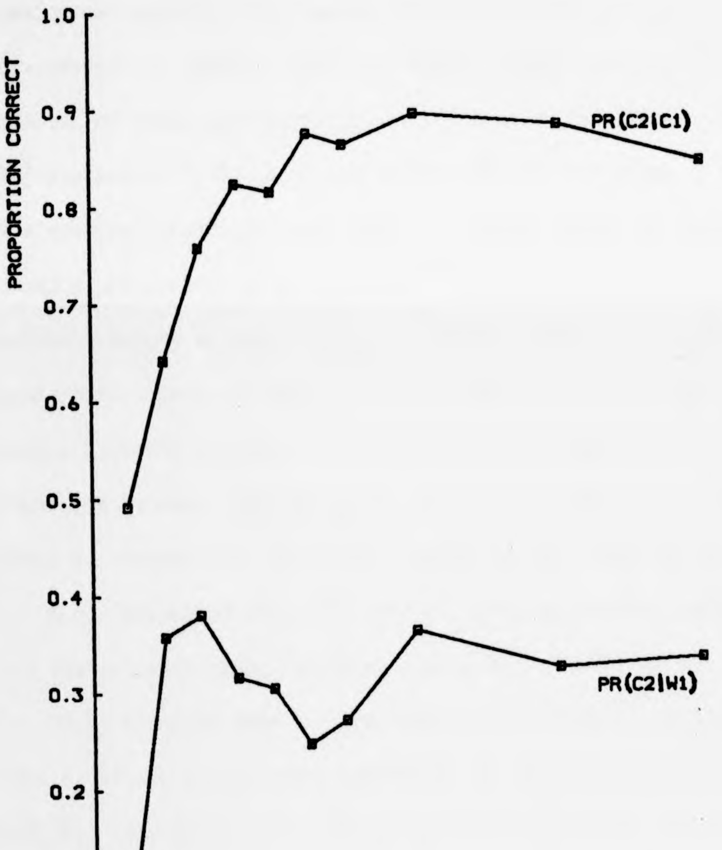
$p = .167$). It was also found that the apparent decline over lags of 8, 12 and 16 was not significant ($\chi^2 = 3.41$, df = 2, $p = .182$). It is, of course, quite likely that $Pr(C_2)$ is being affected by the "noise" detected in $Pr(C_1)$; if there is a regular relationship between $Pr(C_2)$ and $Pr(C_1)$, then variation in $Pr(C_1)$ will be reflected by $Pr(C_2)$, although the variation may no longer reach significance since it may be confounded with more systematic effects.

A clearer picture may be obtained from an examination of the relationship of the conditional proportions $Pr(C2/C_1)$ and $Pr(C_2/W_1)$ with $P_1 - T_1$ interval depicted in Figure 15. Certainly $Pr(C_2/C_1)$ appears to exhibit an extremely regular relationship with spacing and again this observation is emphasised by the analysis of variance. Now, on the hypothesis that the improvement found in long-term recall with the spacing of presentations results from the forgetting of short-retention items that would otherwise be poorly processed on the second presentation for some reason (cf 4.34), then on the basis of the $Pr(C_1)$ function, one would predict that $Pr(C_2/C_1)$ should reach its upper asymptote at the same time that short term retention effects disappear, that is, at a spacing of 2 trials. However, the improvement in $Pr(C_2/C_1)$ appears to be maintained up to a spacing of around 8 trials, and indeed the last eight values (at spacings of 2 or more) were found to differ significantly ($\chi^2 = 9.92$, df = 7, $p = .006$). Consequently, these results markedly conflict with the short-term forgetting hypothesis.

However, it may be argued that the continued improvement of $Pr(C_2/C_1)$ at spacing intervals in excess of 2 could well result from subject differences, in that a greater proportion of the observations contributing to the statistic at long interpresentation intervals will come from the better or more competent subjects. However, by the same argument, even a within-subject comparrison may prove misleading, since the subject may experience both positive and negative transfer across even a single session, and consequently will contribute more

FIGURE 15

Performance on T_2 conditional on performance on T_1 as a function of interpresentation interval in Exp. 1.



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PR(C2|C1)
PR(C2|W1)
0
1
2
3
4
5
6
8
12
16
P1-T1 INTERVAL

points to $Pr(C_2/C_1)$ at long intervals when he is going through a relatively "good patch".

Furthermore, a sampling argument could also be applied to $Pr(C_2/W_1)$; namely that more relatively poor subjects will contribute to this statistic at short intervals. An examination of Figure 15 certainly suggests that $Pr(C_2/W_1)$ might show some improvement with spacing and again the analysis of variance emphasises this suspicion. Furthermore, it appears that if an error is made on T_1 at a retention interval of zero, then performance following an immediate re-presentation (P_2) appears on T_2 to be no better than chance; the value of $Pr(C_2/W_1)$ at a spacing interval of zero was 2/29 = .069, whilst the theoretical guessing probability is 1/15 = .067. This result is surprising, as one would expect an error on T_1 immediately after P_1 to result from inattention. However, even if the subject failed to see P_1 for some reason, he would certainly have attended to the display during T_1 (otherwise he would have failed to respond) and so there is little reason to suspect that the subject would fail to attend to P_2.

A comparison of $Pr(C_2/W_1)$ over all spacing intervals showed that the values differed significantly ($\chi^2 = 24.06$, df = 9, p = .0042) although when the low value at a spacing of zero was omitted, it was found that a comparison of the remaining values of $Pr(C_2/W_1)$ only just reached significance ($\chi^2 = 15.61$, df = 8, p = .048). However, there is no obvious systematic relationship of $Pr(C_2/W_1)$ with P_1-T_1 spacing of the kind predicted by a sampling hypothesis. Rather an examination of Figure 15 suggests a certain amount of "noise" in the response, and this again could well result from shortcomings in the design.

It was also noticed that $Pr(C_2/W_1)$ appeared to be generally lower than performance following a first presentation and a subsequent test at lag 8 would suggest. If a wrong response on T_1 indicated merely that the item was unlearned, one would expect $Pr(C_2/W_1)$ to equal $Pr(C_1)$ at a retention interval of 8 trials. However, a subject -

sampling hypothesis would predict poorer performance on $Pr(C_2/W_1)$, as would an interference hypothesis.

The interference hypothesis will be considered first. If some errors result from, say, stimulus mis-recognition, or confusion, then clearly performance on an item on which an error was previously made should be inferior to that on a new item at the same retention interval. It is tempting to regard perseveration errors as a measure of interference, although if perseveration scores are added to the corresponding values of $Pr(C_2/W_1)$ the resulting values are far higher than long-term retention at T_1. It is suggested that perseveration scores may be particularly contaminated by the adoption by subjects of a "guesssing number" strategy, by which they have a particular number that they always employ when they have no knowledge of the appropriate response. Pure guessing would then result in a far higher than chance rate of perseverations.

It seems that the only way of resolving this dispute is to test the subject sampling hypothesis directly by comparing $Pr(C_2/W_1)$ with $Pr(C_1)$at a retention interval of 8 within subjects. It was decided to compare the pooled values of $Pr(C_2/W_1)$ at spacings of 8, 12, and 16 trials with $Pr(C_1)$ at a retention lag of 8 for each individual subject. This procedure would do much to remove sampling effects from "bad patches" that each subject might have gone through, since these would mainly have a deleterious effect on $Pr(C_2/W_1)$ at short spacings. The nine 1-tailed significance levels obtained were combined to yield a χ^2 of 37.75 with 18 df ($p < .005$). Consequently, it was concluded that $Pr(C_2/W_1)$ was inferior to $Pr(C_1)$ at lag 8 within subjects, and that on the whole, the interference hypothesis would be most likely to account for this result.

Although not entirely convincing, this argument may be turned around and applied to the observations made earlier concerning $Pr(C_2/C_1)$. In other words, if subject sampling doesn't account for the low values of $Pr(C_2/W_1)$, then there is little support for the hypothesis that it does account for the continued improvement of $Pr(C_2/C_1)$ beyond those spacings at which short-term retention effects disappear.

Finally, a comparison of perseveration errors across the ten spacing conditions proved non-significant ($\chi^2 = 13.83$, df = 9, p = .128) and so it was concluded that there was little evidence that perseveration errors were dependent on the P_1-T_1 spacing interval. This result supports the argument advanced earlier that perseveration errors may result at least in part from a "guessing number" strategy.

5.13 Conclusions (Exp. I)

The main conclusions drawn from this study may be summarized as follows. $Pr(C_1)$ certainly exhibited a rapidly-decaying short-term component which had disappeared by a retention interval of 2 trials, although subsequent performance was found to vary significantly in an unsystematic way. Although $Pr(C_2)$ exhibited a significant improvement at non-zero spacings over a spacing interval of zero, no further spacings effects could be isolated. However, $Pr(C_2/C_1)$ appeared to exhibit a much more stable relationship with spacing, and certainly continued to improve beyond the range of short-term retention. This was interpreted as evidence against Greeno's (1970a) version of the processing attenuation hypothesis or equivalent short-term memory explanations of the spacing effect (cf 4.34), although it is clear that short-retention items must receive little benefit from a re-presentation. It was also found that performance at T_2 following an error on T_1 was inferior to performance at a similar retention interval on a new item, and it was concluded that the evidence marginally supported an interference explanation.

In conclusion, it appears that the uncontrolled "noise" resulting from design faults was mainly restricted to the marginal performance measures $Pr(C_1)$ and $Pr(C_2)$, and to some extent to $Pr(C_2/C_1)$. Clearly, these observations would all be affected by specific contextual and semantic relationships that may have been present in the six lists. However, $Pr(C_2/C_1)$ was extremely stable, and exhibited a remarkably orderly relationship with spacing. This suggests that the majority of encoded items were relatively unaffected by specific contextual factors, and this in turn suggests that such items were quite elaborately and deeply encoded. Finally, it seems that the analyses of variance did not make much contribution to the overall interpretation of the data, because of the strong reservations held about accepting their results.

5.2. Experiment II

It was suggested in section 4.32 that recall performance following two presentations of a paired associate may be directly related to the rate of short-term forgetting of stimulus recognition. In other words, it may be that even with a stable, relatively deep associative encoding, performance at final test may suffer if confusable or inadequate stimulus encodings survive at the second presentation. Such a hypothesis would predict that recall performance at final test would continue to improve with interpresentation spacing until that spacing was sufficiently long to ensure that short-term stimulus recognition components were effectively "wiped out" on P_2. This experiment was designed to examine such a hypothesis directly.

5.21 Method of Exp 11

Subjects The subjects were 18 undergraduate students at Stirling University who opted to act as experimental subjects in partial fulfillment of the practical requirements of their introductory psychology course. All subjects were experimentally naive.

Materials The stimulus and response materials employed in this study were identical to those employed in the previous experiment. Stimuli were selected randomly from the common word pool in Appendix 1, whilst responses were randomly selected integers in the range 1.-15.

Apparatus The apparatus employed in this experiment was identical to that used in Exp. 1.

Procedure Critical items were assigned schedules similar to those employed in the previous study; thus

$$P_1 \underset{i \text{ trials}}{\ldots\ldots} T_1 \; P_2 \underset{8 \text{ trials}}{\ldots\ldots\ldots} T_2$$

However, only 5 spacing conditions were employed, with the spacing interval i taking values of 0,4,8,12 or 16 intruding trials on other items. An additional condition was also included, in which a stimulus that had never occurred before was tested. This condition allowed the estimation of the probability of a false recognition; i.e. the subject identifying a stimulus word as "old", when it was, in fact new.

The procedure employed in this study was almost identical to that of Exp. 1, with three important exceptions. In the first place, subjects were required to make two responses on each test trial. They were first of all required to indicate whether they thought they had seen the stimulus under test earlier in the session (by writing the letter "O" for "old" on the response sheet) or not (by writing "N" for "new"), and then a normal recall response was required. Again subjects were instructed to guess if necessary. A recall response was required even if the subject had thought that he had not previously seen the stimulus under test, and had consequently made a "new" recognition response.

In the second place, to allow extra time for the additional response to be made on test trials, the lists were presented at a 3-second rate. Finally, only three ten-block lists were employed in this study as compared with six in Exp. 1. The 18 subjects were tested in small groups of between 2 and 5 individuals; in other words all the subjects in a group were tested simultaneously, which meant that they all saw the three lists in the same order. Although this list order was varied from group to group, the unequal group sizes (caused by the frequent failure of students who had "signed up" for the experiment to actually attend) meant that list order was not perfectly counterbalanced across the subject sample.

It should be pointed out that in fact, Exp.II was not really completed. The original intention had been to test six groups each of five subjects, counterbalancing list order between these groups. However, the experiment had reached its present stage of completion at the end of the academic year, which meanst that the subject pool had "dried up" for the summer. However, an examination of the extant results at this stage strongly suggested that very little additional information, if any, would become available even if the design had been completed. Consequently, the experiment was terminated at the present stage, and the results of the incompleted design are presented below.

5.22 Results of Exp.II

Once again, the items occurring in the first and final (tenth) block of each list were omitted from the analysis for the reasons outlined in section 5.12, which meant that only 24 different pairs contributed to the results of each spacing condition, and of the false recognition condition. Furthermore, any subject item which had an omitted response was discarded from the analysis, so that, in general subjects did not contribute equal numbers of observations to each condition.

On each test trial, four response categories were defined in an analagous way to the two categories in Exp.I; in addition to the "correct/wrong" or C/W recall categories, the symbols "O" and "N" were employed to describe the corresponding recognition response. Of course, a response of "O" to an item in one of the five spacing conditions would have been correct, and a "new" response to such an item would be wrong. On the other hand, the opposite would be true of such responses to items in the "false recognition" condition; an "old" response would be incorrect. On each test trial in the spacing conditions, the subjects two responses (recognition followed by recall) could fall into one of four categories; OW, OC, NW or NC. Four such categories on T_1 taken in conjunction with four such categories on T_2 gives a total of 16 possible response categories by which to summarize performance on each item (i.e. $O_1W_1C_2W_2$, $O_1W_1O_2C_2$, $O_1W_1N_2W_2$,...., $N_1C_1N_2C_2$). The overall proportions of the total number of subject items (n) falling into these sixteen categories are presented in Appendix 2. Of course, such a table is almost impossible to interpret, and more detailed breakdowns of the data will be presented at the appropriate points in the following discussion. When examining these results it should be borne in mind that the probability of making a correct recall response by chance is just 1/15 or .067, and that analysis of false recognition items yielded a false recognition rate, P("old/new"), of 95/528, or .180.

It is proposed firstly to deal separately with recognition and recall performance. The relevant data are summarized in Table 10. It was decided to omit analysis of variance on these data, following the somewhat disappointing results of such a procedure in the previous study. Recognition performance was generally of such a high

TABLE 10

Recognition and recall performance in Exp II

P_1-T_1 LAG	n	$Pr(O_1)$	$Pr(O_2)$	$Pr(O_2/O_1)$	$Pr(O_2/N_1)$	
0	423	.981	.979	.983	.750	$(\frac{1}{4})$
4	415	.918	.993	.995	.971	
8	422	.879	.979	.978	.980	
12	419	.862	.995	.994	1.000	
16	422	.855	.991	.989	1.000	

P_1-T_1 LAG	n	$Pr(C_1)$	$Pr(C_2)$	$Pr(C_2/C_1)$	$Pr(C_2/W_1)$
0	423	.903	.482	.513	.195
4	415	.451	.533	.840	.281
8	422	.408	.507	.814	.296
12	419	.401	.487	.827	.259
16	422	.374	.533	.823	.360

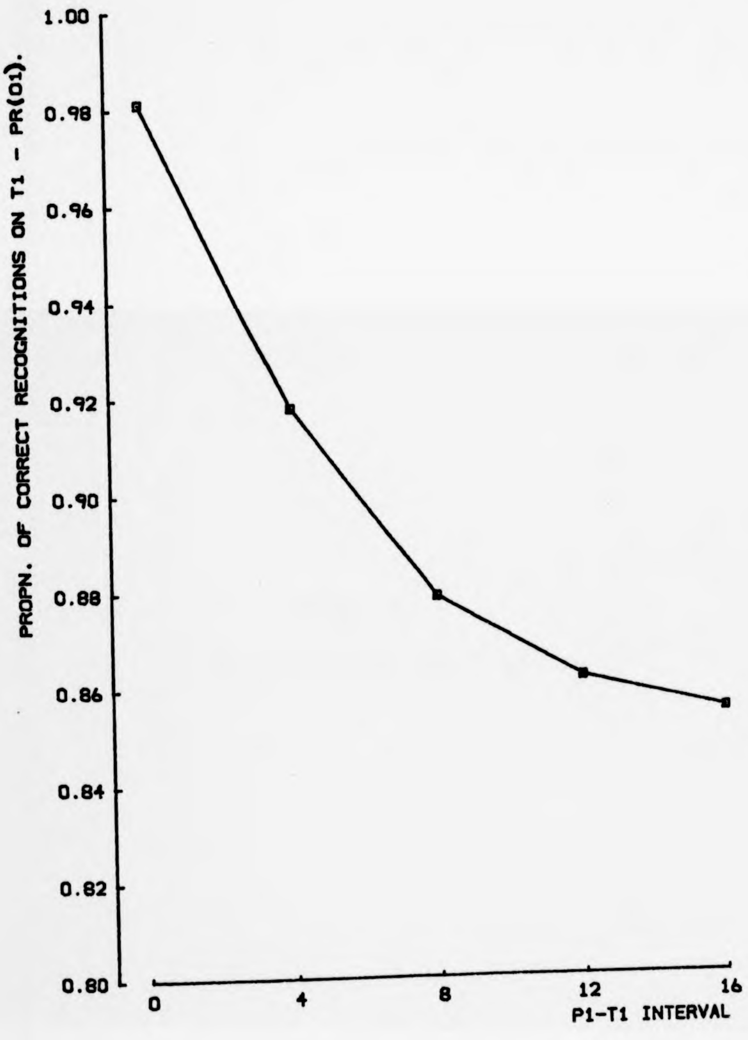
level that valid tests of the effects of P_1-T_1 spacing were impossible to contruct, owing to the very small numbers of scores falling into the "new" categories on T_1 and T_2. However, a comparison of the values of $Pr(O_1)$ proved highly significant ($\chi^2 = 50.21$, df = 4, $p < .0001$), and an examination of Figure 16 suggests that short term recognition effects may still be effective up to retention intervals of 12 or more; there is not clear division of performance into long-and short-term components, although it is quite possible that very little additional decline would have been found at retention intervals in excess of 16 trials, so that the entire portion of the curve tested in this study might lie in the short-term recognition region. There were insufficient data to statistically compare the values of $Pr(C_2)$ at the various spacing intervals, and this was also true of $Pr_2(O_2/O_1)$ and $Pr(O_2/N_1)$, but an inspection of the data in Table 10 suggests that these scores were not particularly affected by the spacing variable. Thus, there is little evidence that recognition on T_2 was affected by the spacing interval between P_1 and P_2.

Analysis of the recall data, however, proved less disappointing. Recall at T_1 shows the usual short- and long-term retention components, (See Figure 17) and $Pr(C_1)$ was found to differ significantly as a function of P_1-T_1 lag (χ^2=336.00, df = 4, $p < .0001$), although no differences were found in recall on T_1 at lags of 4, 8, 12 and 16 ($\chi^2 = 5.18$, df = 3, $p = .159$). These results strongly suggest that short-term retention had disappeared by a retention interval of 4 intruding trials, and that, furthermore, long-term forgetting was at a negligible level.

Recall performance at T_2, $Pr(C_2)$, rather disappointingly failed to exhibit a spacing effect ($\chi^2 = 3.94$, df = 4, $p = .419$), and this may well have been due to systematic but uncontrolled effects of material or list order, since material and spacing were completely confounded in this study, and list order was not completely counter-

FIGURE 16

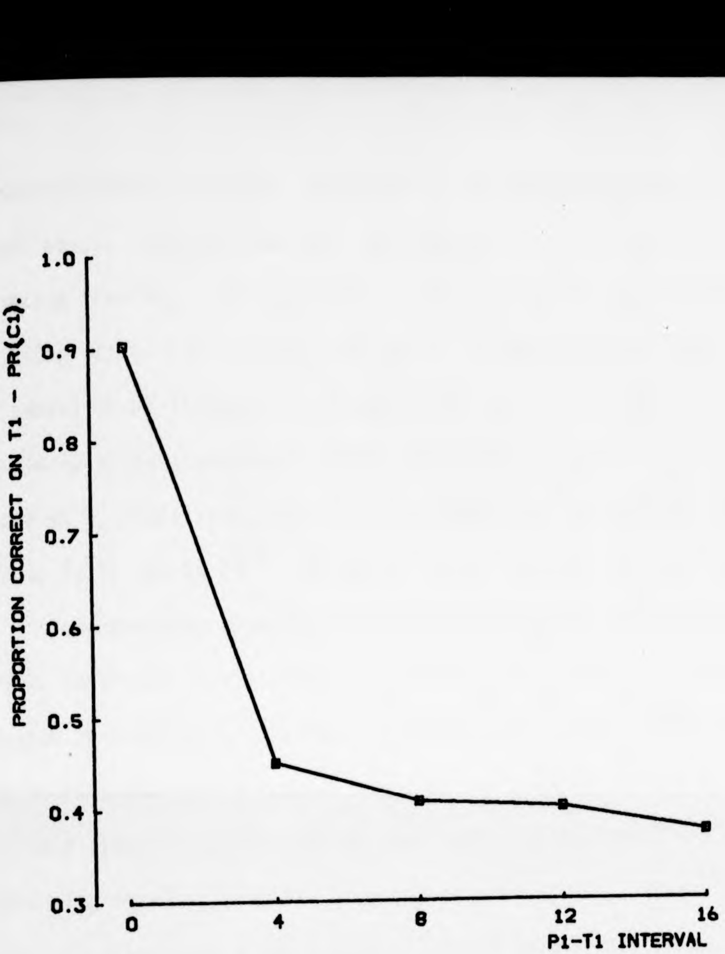
Recognition at T_1 as a function of retention interval in Exp. II.

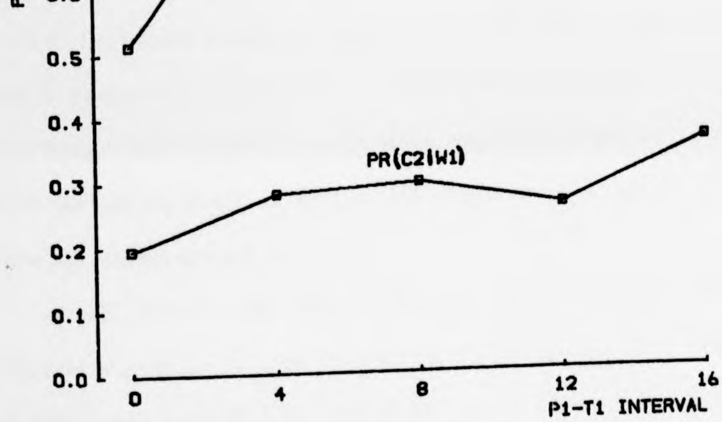


PROPN. OF CORRECT RECOGNITIONS ON T1 - PR(O1).
1.00
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.84
0.82
0.80
0
4
8
12
16
P1-T1 INTERVAL

FIGURE 17.

Paired-associate recall in Exp.II


PROPORTION CORRECT ON T1 - PR(C1).
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
4
8
12
16
P1-T1 INTERVAL


PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PR(C2|C1)
PR(C2|W1)
0
4
8
12
16
P1-T1 INTERVAL

balanced across subjects. However, it was found that, as in the last study, $Pr(C_2/C_1)$ showed a far more orderly relationship with spacing (See Fig. 17), and furthermore exhibited a significant spacing effect ($\chi^2 = 118.14$, df= 4, $p < .0001$). However, this effect appeared to be limited to an improvement up to a spacing of 4 trials or more, since there was no significant difference in recall performance at T_2 conditional upon a correct recall at T_1 at P_1-T_1 lags of 4, 8, 12 and 16 ($\chi^2 = 0.43$, df = 3, $p = .935$).

Consequently, it appears that in this experiment, insufficient P_1-T_1 intervals were included to fix the spacing effect with any degree of certainty. All that can be stated is that performance at T_2 conditional upon a correct response at T_1 improved with spacing up to a limit which occurred at some P_1-T_1 lag between 0 and 4 intruding trials. In contrast to the results of Exp.I these findings are consistent with a short-term retention hypothesis, in that improvement with spacing appears to reach its upper asymptote at the same point that short-term retention reaches its lower asymptote. However, this conflict may well be only apparent; were P_1-T_1 spacings of 1, 2 and 3 trials included, it is quite likely in the light of the results of Exp.1. that $Pr(C_2/C_1)$ would have shown a continued improvement beyond the range of short-term retention at T_1. However, it should be pointed out that these data appear to conflict markedly with a short-term stimulus recognition hypothesis; an examination of the T_1 recognition curve (Fig. 16) would suggest under such a hypothesis that T_2 recall performance would show continued improvement over the entire range of spacing intervals included in the study. This was certainly not the case.

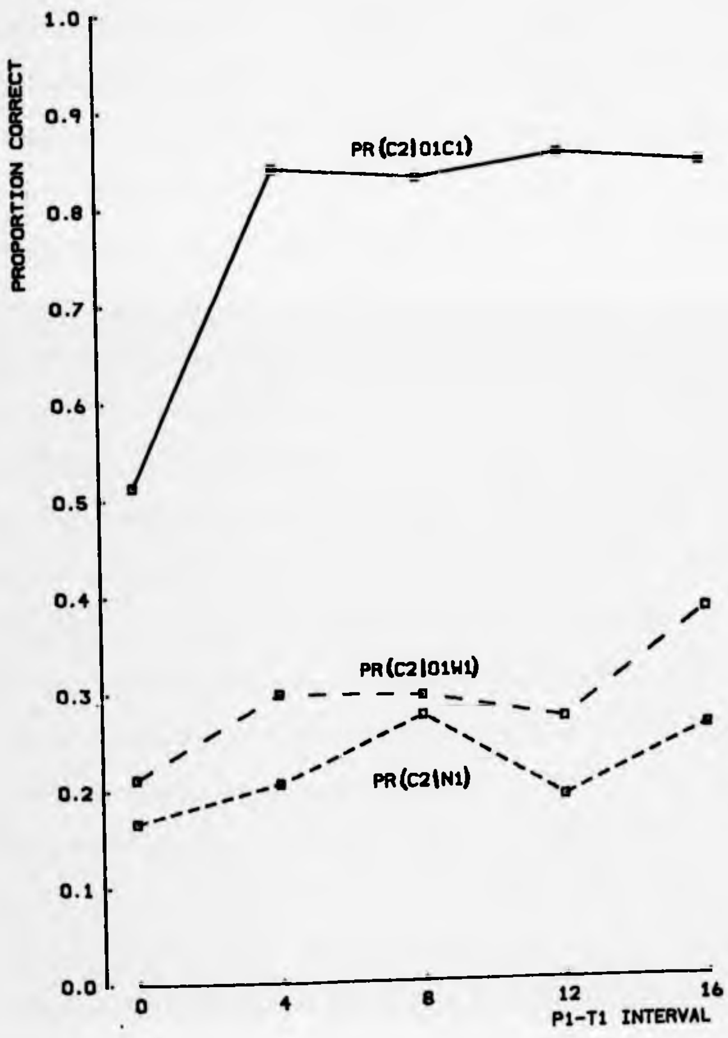
Recall performance at T_2 conditional upon a recall error at T_1 followed a similar pattern to that in the previous study, although in this case, $Pr(C_2/W_1)$ did not appear to be affected by P_1-T_1 spacing ($\chi^2 = 9.08$, df = 4, $p = .05$.) However, once again overall recall performance at T_2, which followed eight trials after P_2, was poorer following two presentations with a recall error on T_1 (.296) than performance at T_1 at a similar retention interval of eight trials (.408); ($z = 4.03$, $p = .001$, 1 - tailed).

So far, recognition and recall performance have been treated separately; consequently, it is now necessary to examine the effects of stimulus recognition upon recall performance. Although insufficient "new" responses were made on both T_1 and T_2 to allow a meaningful examination of recall performance conditional upon a stimulus recognition failure on the same trial, it was possible to produce a meaningful estimate of this performance measure by aggregating $Pr(C/N)$ across P_1-T_1 lag on both T_1 and T_2, to yield a value of $Pr(C/N) = 20/239 = .084$. Thus, given a "new" response on either T_1 or T_2, the proportion of correct recall responses on the same trial was only .084; this value did not differ significantly from the theoretical chance recall level of 1/15 ($z = .967$, $p = .28$, 2-tailed). This result is in accordance with earlier findings that recall performance on a particular test trial is no better than chance if the subject fails to recognize the stimulus on that test trial (see section 4.13).

Recall performance on T_1 given a correct recognition on T_1 (i.e. $Pr(C_1/C_1)$) certainly showed a significant decline with retention interval ($\chi^2 = 290.84$, df = 4, $p < .0001$), although, as with $Pr(C_1)$, this decline appeared to be limited to a rapid short-term decay effect, since the observed values of this statistic did not differ at P_1-T_1 lags of 4, 8, 12 and 16 ($\chi^2 = 2.88$, df = 3, $p = .410$). Because recognition at T_1 was at such a high level, $Pr(C_1/O_1)$ was very similar to $Pr(O_1)$ at all P_1-T_1 lags; however, this result is still of interest because it firmly establishes that recall performance declines much more rapidly than recognition performance as a function of retention interval. The values of $Pr(C_2/O_2)$ were virtually identical to those of $Pr(C_2)$ at all P_1-T_1 intervals, again as a result of the high level of recognition performance on T_2, so that consequently

FIGURE 18

Recall performance on T_2 conditional upon non-recognition, and recognition and recall performance on T_1, in Exp. II.



PROPORTION CORRECT
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PR (C2|O1C1)
PR (C2|O1W1)
PR (C2|N1)
0
4
8
12
16
P1-T1 INTERVAL

the same observations apply in that this statistic did not exhibit a spacing effect.

Recall performance da a on T_2 conditional upon both recognition and recall performance on T_1 are presented in Figure 18. Again, because recognition on T_2 was almost perfect, it was not considered necessary to examine both recognition and recall on T_2 conditional upon total performance on T_1. Furthermore, since there is little evidence that subjects were performing better than chance on a test trial iven a stimulus recognition failure on the same trial, it seems reasonable only to consider $Pr(C_2/N_1)$ irrespective of recall performance on T_1; in any case, there were insufficient N_1W_1 observations to meaningfully examine $Pr(C_2/N_1W_1)$ as a function of $P_1 - T_1$ spacing.

Because of the very high level of recognition performance on T_1 there is hardly any difference between the values of $Pr(C_2/O_1C_1)$ and $Pr(C_2/C_1)$; consequently, it appears that the $Pr(C_2/C_1)$ function is almost perfectly accounted for by the $Pr(C_2/O_1C_1)$ curve. On the other hand, $Pr(C_2/W_1)$ results from the functions $Pr(C_2/O_1W_1)$ and $Pr(C_2/N_1)$, and it is clear from Figure 18 that the former of these two curves lies entirely above the latter, Furthermore, both these functions appear to lie entirely below the value of $Pr(C_1)$ at a retention interval of 8; that is, a value of .408. It was found that neither $Pr(C_2/O_1W_1)$ nor $Pr(C_2/N_1)$ showed a spacing effect (χ^2=8.35, df = 4, p = .08; χ^2 = 1.68, df = 4, p = .79, respectively), and when the values of these proportions were estimated across spacing intervals and compared, it was found that the overall value of $Pr(C_2/O_1W_1)$ differed significantly from that of $Pr(C_2/N_1)$, (z=2.29, p=.022, 2-tailed). In addition, both overall values were found to be significantly lower than the value of $Pr(C_1)$ at a retention interval of 8; for $Pr(C_2/O_1W_1)$, a z value of 3.38 resulted for this comparison ($p < .001$, 1-tailed), whilst the value for the corresponding comparison involving $Pr(C_2/N_1)$ yielded z = 4.59, ($p < .001$, 1-tailed).

These results confirm an extend the findings of Exp.1.; not only was $Pr(C_2/W_1)$ found to be significantly inferior to $Pr(C_1)$ at a retention interval of 8, but both underlying components of this conditional recall performance measure, $Pr(C_2/C_1W_1)$ and $Pr(C_2/W_1)$ were found to take values that were significantly inferior to that of $Pr(C_1)$ at lag 8. This means that following a successful recognition with a recall error on T_1, subsequent performance following an immediate presentation, P_2, and a retention interval of 8 trials was significantly inferior to performance on a brand new item that had received just one presentation, P_1, and was tested at a similar lag of 8 trials. This was also true of recall performance on T_2 following a recognition error on T_1, and furthermore, a recognition error on T_1 was found to be more deleterious to performance on T_2 than a recall error. The results of Exp.1. suggested that this depression in performance could not be accounted for in terms of subject differences (i.e. an error at T_1 suggesting that that particular subject item was more likely to have resulted from a less able subject), so that the relatively poor performance on T_2 following just a recall error on T_1 in this study might well result from interference or response competition and confusion. The even poorer performance on T_2 following a recognition error on T_1 could be explained in terms of the encoding variability hypothesis; those stimulus items not recognised on T_1 are seen as more fractionable, less well integrated, and hence more variably encoded. Consequently, such items are much more likely to be mis-recognised, or recognised in terms of stimulus features not associated with the appropriate response, on T_2 (see 4.13).

5.23 Summary (Exp. II)

The recall performance results of Exp. II are on the whole consistent with those of Exp. I, although a smaller sample of inter-presentation spacings were tested, and the spacing at which recall on T_2 reached its upper asymptote was "missed". This result was obviously disappointing. However, several of the results of Exp. I were confirmed, principally the finding that the spacing effect appears to operate only on those items that are relatively well encoded on P_2 (i.e. those that are correctly recognised and recalled on the immediately preceding test trial, T_1.) However, the recognition results were also disappointing, in that recognition was at a very high level throughout the study, although there is evidence that interpresentation spacing effects on recall performance on T_2 reached their upper asymptote far earlier than short-term stimulus recognition components on P_2 had completely decayed; this finding conflicts with the hypothesis that the spacing effect is caused by the maintenance of encodings with poor stimulus components on P_2.

On the whole, recognition did not appear to offer the slightest explanation of the spacing effect. However, it is possible that because of the nature of the stimulus material (i.e. common words) stimulus recognition was dependent upon the encoded stimulus aspects that were not employed in associative encodings, since it is feasible that, for example, all levels of encoding assisted stimulus recognition (auditory, episodic and semantic) whilst say, only semantic aspects of stimuli were employed in associative encodings. In other words, additional cues may have been available to aid recognition which would not materially benefit recall.

Finally, it should be pointed out that despite certain obvious inadequacies in the (incompleted) experimental design, the data from this study exhibited surprisingly regular relationships with spacing; consequently, it appears that the partial counterbalancing of list order in this study was successful in removing some of the uncontrolled "noise" that was present in the results of Exp. I.

5.3 Experiment III

As had just been pointed out, it was felt that the results of Exp II left one or two important questions to be answered. In the first place, the recognition data might well have resulted from elaborate stimulus encodings only part of which were employed in association codings, and secondly, because of the small range of spacings tested, the experiment appeared to "miss" the point at which the improvement in recall performance at T_2 with interpresentation spacing reached its upper asymptote. Experiment III was designed to clarify these two points.

5.31 Method of Exp III

Subjects The 25 subjects employed in this study were undergraduate and postgraduate student volunteers from the Psychology Department at Nottingham University. All subjects claimed to be experimentally naive.

Materials The stimuli employed in this study were selected at random from a pool of 106 consonant-vowel-consonant (CVC) trigrams in the range 30-40 Archer (1960). The stimulus pool (see Appendix 1) was made up in such a way as to ensure that the first two letters of each CVC were unique, and as far as possible, each initial consonant occurred equally often in the pool. Responses were randomly-selected integers in the range 1 - 5. It was hoped that with more fractionable, low-M stimuli, there would be agreater likelihood that the stimulus aspects employed in recognition would exhibit a high degree of correlation with those employed in associative codes.

Apparatus The CPA study-test lists employed in this study were prepared in paper-tape form on the Nottingham University Psychology Department's Elliot 903 computer. These tapes were then read into the department's PDP-11 computer, which controlled the real-time durations of both study and test trials, and output the material on a GT40 display console. Subjects sat in a small darkened booth in front of the console, and responded where appropriate by pressing keys. Responses were recorded, stored, and subsequently output by the PDP-11.

__Procedure__. The basic paradigm remained the CPA study - test procedure employed in the previous two studies. However, each subject in the current experiment was tested on only one list, and a separate list was made up for each subject. Again, lists were composed of ten overlapping blocks, each of which contained one exemplar of each experimental condition. There were five double presentation conditions, again employing the schedule,

$$P_1 \underset{i \text{ trials}}{\ldots\ldots} T_1 P_2 \underset{8 \text{ trials}}{\ldots\ldots} T_2$$

but in this study the P_1-T_1 interval i comprised 0,2,4,6 or 8 intruding trials on other items, in order to reduce the likelihood of "missing" the optimal P_1-T_1 spacing. Furthermore there was an additional "false recognition" condition, wherein items received a single unreinforced test trial.

Slightly more realistic filler items were constructed in this study to occupy list positions left vacant by the random interleaving process. Each filler item was presented once, and a filler was tested on a subsequent vacant list position if it occurred between X and 8 trials after the filler's presentation trial, where X took values 1,2, 3 and 4 each with probability ¼. However, responses to fillers were not recorded. Within each list, each paired-associate item (i.e. double presentation, false recognition or filler) comprised a CVC stimulus randomly selected without replacement from the pool in Appendix 1, paired with a randomly selected integer in the range 1-5. The interleaving order was varied randomly from list to list.

Instructions were similar to those in the previous study, so that on each test trial, subjects were required to make two responses; a recognition response (O or N) followed by a recall response in the range 1-5 guessing where necessary. However, there was one major difference between this experiment and Exp. II. Study trials were of 2.7 second's duration whilst the duration of each test trial was determined by the subject, in that the trial was terminated only when the subject had made his second (i.e. recall) response. During study trials, a stimulus response pair appeared together on the GT40 screen, whilst on a test trial, the word "TEST" was displayed followed by the CVC currently under test. In an effort to pace subjects, the word "TEST" began to flash on and off 2.7 seconds after the trial onset if both responses had not been made, to cue the subjects to hurry up and finish responding. There was a blank period of 0.3 seconds between the offset of each trial and the onset of the next one. Upon completion of the session, each subject in addition was given a short , informal postexperimental interview in order to ascertain his reactions to the task.

5.32 Results (Exp. III)

Once again, the first and final (i.e. tenth) blocks of each list were discarded, so that only eight items per condition were analysied for each subject. However, the procedure adopted ensured that subjects could not possibly omit responses, so that consequently, there were in all 8 x 25, or 200 observations contributing to each condition. The theoretical chance level of currect recall in this study is simply 1/5 or .2 (there were five response alternatives) whilst the false recognition rate Pr("O"/New) was found to be 47/200, or .235. This value is somewhat higher than that recorded in Exp II (.180) which suggests that the relatively low-M CVC's employed in this study were prone to more encoding variability, and were hence more likely to be mis-recognised, than the common words employed in the earlier study. The data from this study are fully summarised in Appendix 3.

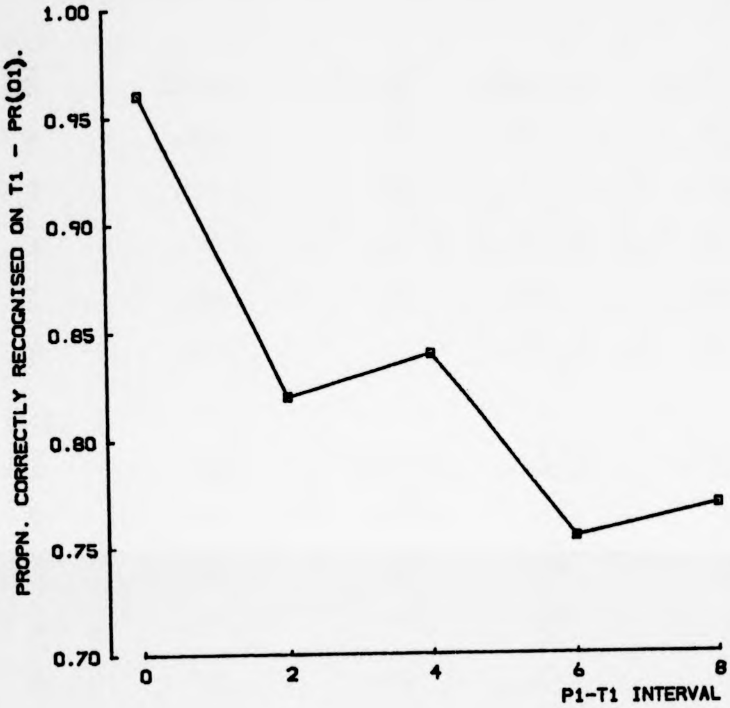
Recognition and recall data from this study are presented in Table 11. The recognition data are also depicted in Figure. 19. It is apparent from the figure that recognition performance at T_1, $Pr(O_1)$, exhibits a rapid short-term decline from a P_1-T_1 interval of zero to an interval of 2 trials, and possibly a somewhat slower subsequent decline with larger retention intervals. This is borne our by

FI URE 19

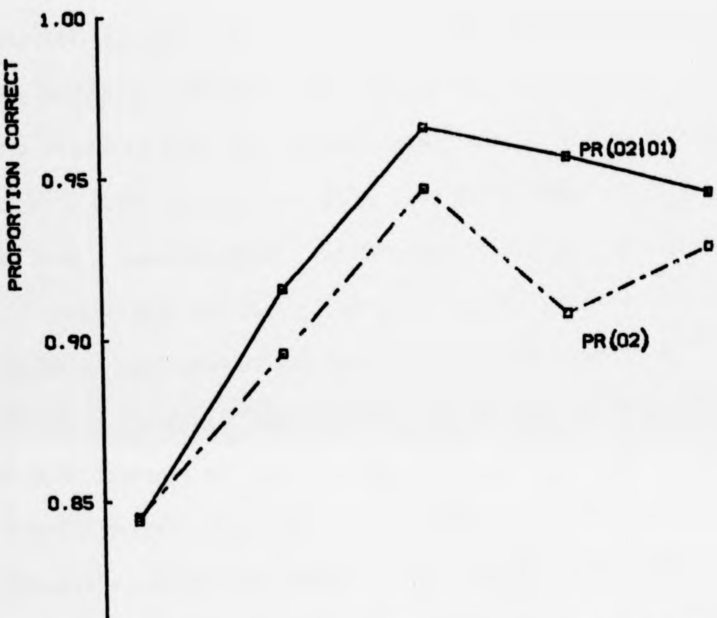
Stimulus recognition in Exp. III

FIGURE 19

Stimulus recognition in Exp. III



1.00
0.95
0.90
0.85
0.80
0.75
0.70
PROPN. CORRECTLY RECOGNISED ON T1 - PR(O1).
0
2
4
6
8
P1-T1 INTERVAL



1.00
0.95
0.90
0.85
0.80
PROPORTION CORRECT
PR(O2|O1)
PR(O2)
0
2
4
6
8
P1-T1 INTERVAL

TABLE 11

Recognition and recall data for Exp. III

P_1-T_1 LAG	$Pr(O_1)$	$Pr(O_2)$	$Pr(O_2/O_1)$	$Pr(O_2/N_1)$
0	.960	.845	.844	.875
2	.820	.895	.915	.806
4	.840	.945	.964	.844
6	.755	.905	.954	.755
8	.770	.925	.942	.870

P_1-T_1 LAG	$Pr(C_1)$	$Pr(C_2)$	$Pr(C_2/C_1)$	$Pr(C_2/N_1)$
0	.950	.365	.363	.400
2	.500	.450	.540	.360
4	.425	.365	.541	.235
6	.370	.430	.622	.317
8	.315	.380	.571	.292

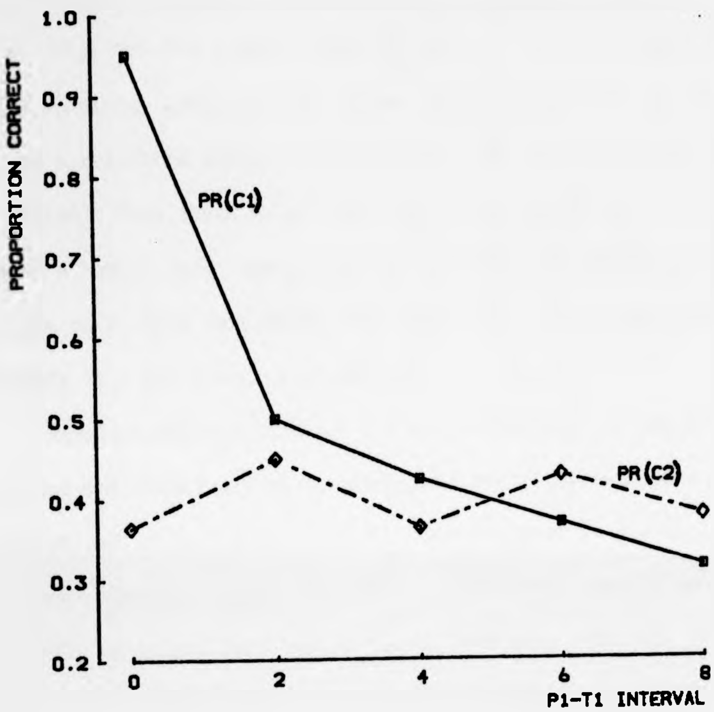
Note: n = 200 observations in each condition

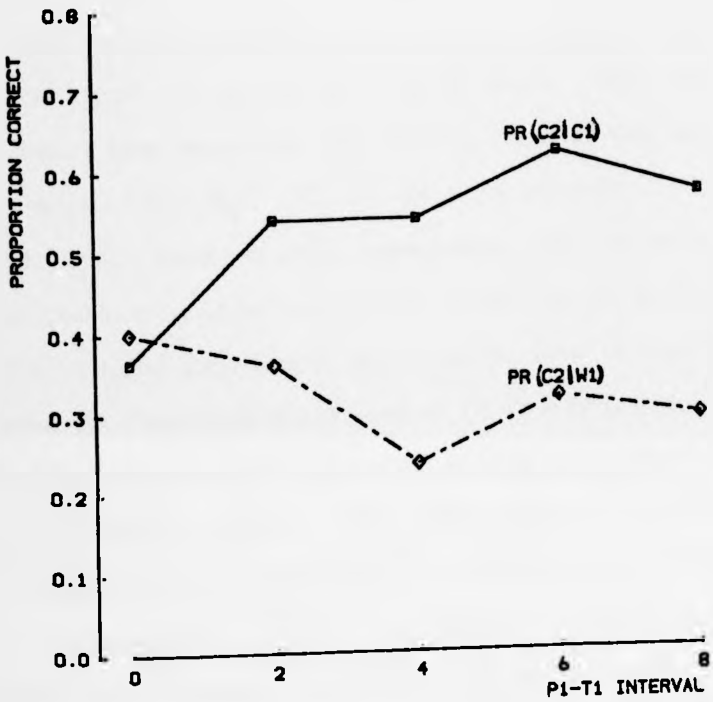
statistical analysis; a comparison of all five values of $Pr(O_1)$ was highly significant ($\chi^2 = 37.13$, df = 4, $p < .0001$) whilst the values at lags of 2 or more trials did not differ significantly ($\chi^2 = 6.00$, df = 3, p = .112). Both $Pr(O_2)$ and $Pr(O_2/O_1)$ appear to show a spacing effect. In the case of $Pr(O_2)$, a comparison of all five values was significant ($\chi^2 = 12.97$, df = 4, p = .011) whilst a comparison of the last four values was not ($\chi^2 = 8.9$, df = 3, p = .273). This suggests that recognition on T_2 improved with an increase of P_1-T_1 spacing from zero to two trials, but that thereafter, performance was not affected by interpresentation spacing. Comparable results were observed for $Pr(O_2/O_1)$; a comparison of all five values was significant ($\chi^2 = 23.10$, df= 4, $p < .001$) whilst the last four values did not differ significantly ($\chi^2 = 4.23$, df = 3, p = .238). These results are consistent with the hypothesis that the improvement in recognition performance at T_2 with P_1- T_1 spacing results from the decay of short-term recognition components which would otherwise be maintained at P_2. However, since a P_1-T_1 lag of 1 was not included in the study, it is quite possible that the coincidence of the upper asymptote of the spacing effect with the final decay of short-term recognition on T_1 is merely apparent. However, in the light of the small spacing intervals required to achieve optimal recognition on T_2, it should be suspected that similarly disappointing observations will result from the recall data. There were insufficient observations to statistically compare the five values of $Pr(O_2/N_1)$, but the overall value of this proportion computed across lag, was .819, which exceeded the value of $Pr(C_1)$ at a retention interval of 8 trials (.770), so there is no evidence that items are not recognised on T_1 are more difficult to encode than brand new items.

The recall data for this study are presented in Figure 20. Recall performance at T_1, $Pr(C_1)$, clearly shows a rapid decline as the P_1-T_1 interval increases from 0 to 2, and a subsequent slower decline. Statistical comparisons of all five, and the last four, values of $Pr(C_1)$ were both significant ($\chi^2 = 206.94$, df = 4, $p < .001$; $\chi^2 = 15.57$, df = 3, $p = .001$) so that both long-and short-term components exhibit significant forgetting with increasing retention intervals, in contrast to the results of the two previous studies. $Pr(C_2)$ was disappointingly unaffected by P_1-T_1 spacing ($\chi^2 = 5.20$, df = 4, $p = .267$), although a comparison of the five values of $Pr(C_2/C_1)$ was significant ($\chi^2 = 20.92$, df, = 4, $p = .0003$). When the values of $Pr(C_2/C_1)$ were compared at P_1-T_1 intervals of 2,4,6 and 8 no significant differences were found ($\chi^2 = 1.43$, df = 3, $p = .699$) so that, as expected, after an initial improvement in recall performance at T_2 conditional upon correct recall at T_1 from a $P_1 - T_1$ interval of zero to an interval of two trials, there was no evidence of further improvement with subsequent increases in interpresentation spacing.

FIGURE 20

Paired-associate recall in Exp. III


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
PROPORTION CORRECT
PR(C1)
PR(C2)
0
2
4
6
8
P1-T1 INTERVAL

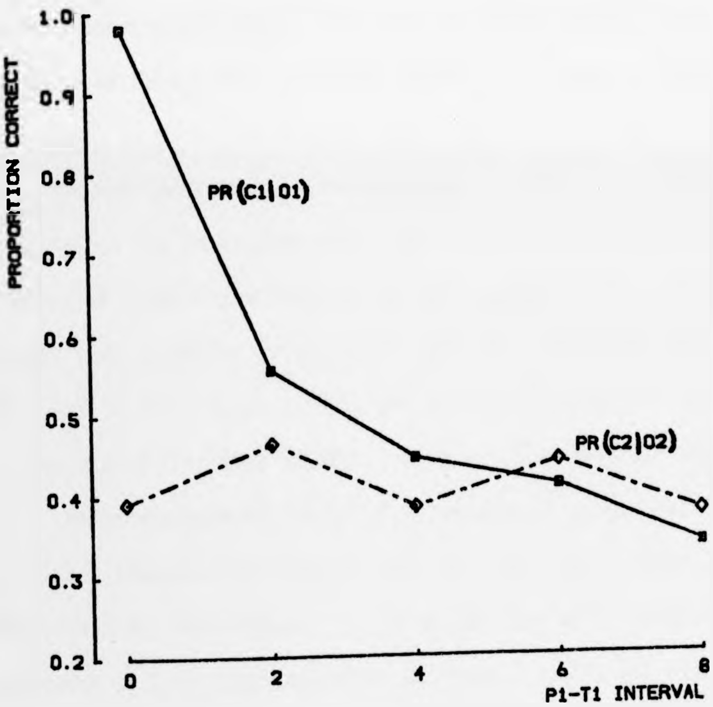

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PROPORTION CORRECT
PR(C2|C1)
PR(C2|W1)
0
2
4
6
8
P1-T1 INTERVAL

Recall performance on T_2 conditional upon a recall error at T_1, $Pr(C_2/W_1)$ was not affected by P_1-T_1 spacing (χ^2=4.73, df = 4, p = .316), and the overall value of $Pr(C_2/W_1)$ computed across the five spacings, .301, did not differ significantly from recall at T_1 at a retention interval of 8 trials, .315 (z = .360, p = .361, 2-tailed). Thus there is no evidence in this study that items upon which a recall error was made on T_1 were any more difficult to encode on P_2 than were brand new items on P_1. This result conflicts sharply with the findings of Exp II..

Recall conditional upon a correct recognition on the same trial did not materially differ from marginal recall performance. The proportions $Pr(C_1/O_1)$ and $Pr(C_2/C_2)$ are depicted in Figure 21, and a comparison with Figure 20 confirms that overall recall can be almost entirely explained in terms of recall following a correct recognition on the same trial coupled with guessing following a recognition error on the same trial. Certainly, $Pr(C_1/N_1)$ failed to exhibit any relationship with P_1-T_1 spacing (χ^2 = .64, df = 3, p = .887). There were insufficient observations to permit a meaningful comparison of the values of $Pr(C_2/N_2)$ at various P_1-T_1 spacing. Thus, overall values of these proportions were computed across spacing intervals. The value of $Pr(C_1/N_1)$, .257, did not differ significantly from that of $Pr(C_2/N_2)$, which was .268. Consequently, these two estimates were pooled to yield an overall value of P(C/N on the same trial) of .261. Somewhat surprisingly, this value was found to significantly exceed the theoretical chance level of 1/5 (z = 2.53, p < .01, 1-tailed). This result suggests that given a recognition error, recall performance on the same trial was better than chance, in direct conflict with the results of the previous study, and those outlined in section 4.13. It is suggested that this result may have arisen from the relatively small number of response alternatives employed in this study. It is possible that subjects had some idea of which responses they had been presented with most often in the recent past, so that on failing

FIGURE 21

Recall performance following a correct recognition on the same test trial in Exp. III



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
PROPORTION CORRECT
PR (C1|01)
PR (C2|02)
0
2
4
6
8
P1-T1 INTERVAL

to recognise a stimulus, they may have guessed away from this set. Such a strategy might well serve to boost the guessing level, and the failure to find any effect of retention interval on $P(C_1/N_1)$ is certainly consistent with such a hypothesis.

Since $Pr(C_2/O_2)$ and $Pr(C_2)$ were so similar at each P_1-T_1 spacing, it was considered sufficent to examine only recall performance at T_2 conditional upon both recognition and recall performance on T_1. $Pr(C_2/N_1)$ did not exhibit a P_1-T_1 spacing effect ($\chi^2 = 2.56$, df = 3 $p = .461$), and moreover, the overall value of $Pr(C_2/N_1)$ computed across the five spacing conditions .275, did not differ significantly from the value of $Pr(C_1)$ at a retention interval of 8 trials, .315 ($z = .84$). This result supports the earlier conclusion that a recognition failure at T_1 did not indicate that the encoding of a pair on a subsequent presentation was in any way more difficult or less adequate than the encoding of a brand new pair on its first presentation. Similar results were found for $Pr(C_2/O_1W_1)$; there was no spacing effect ($\chi^2 = 3.74$, df = 3, $p = .290$), and the overall value of .319 clearly did not differ from that of $Pr(C_1)$ at a similar retention interval, .315. An examination of $Pr(C_2/O_1C_1)$ revealed that this statistic was almost identical in form to $Pr(C_2/C_1)$, in that a comparison of all five values was significant ($\chi^2 = 26.99$, df = 4, p .001), whilst the values at P_1-T_1 spacings of 2, 4, 6 and 8 trials did not differ significantly ($\chi^2 = 3.44$, df = 3, $p = .329$). Again, an increase in P_1-T_1 spacing from 0 to 2 trials was beneficial, whereas subsequent increases produced no additional benefit.

An examination of perseverations in this study was interesting. The proportions of items to which an incorrect recall had been made on both T_1 and T_2 which were also perseverations were examined as a function of the two recognition responses. In the case of two correct recognitions, i.e. $O_1W_1C_2W_2$ items, perseverations were found to exhibit a significant lag effect, and to increase with P_1-T_1

spacing. This result suggests that as the retention interval increases after a first presentation, the probability of a perseveration given an error also increases. However, it is not clear whether this indicates an interference effect, or merely reflects a "guessing number" strategy, since the highest proportion of perseverations given two recall errors and two correct recognitions was 34/66, or .515, whilst the overall proportion of perseverations given two recall errors and two recognition errors was 9/17 or .53! In the latter case, there is considerable evidence that subjects were guessing responses on both test trials, so that the high level of perseverations clearly suggests an underlying "guessing number" strategy of the kind postulated in section 5.13. Thus, it is quite likely that at long retention intervals, given correct stimulus recognitions, a greater proportion of recall errors were pure guesses than at shorter intervals.

5.33 Summary (Exp. III)

Clearly, Exp III failed to fulfill the function for which it was designed, since the range over which T_2 performance improved with interpresentation spacing was, if anything, smaller than that observed in Exp II. In addition, there were almost certainly insufficient observations to make full use of the recognition data, and more particularly of T_2 recall performance conditional upon T_2 recognition and T_1 recognition and recall performance.

However, it was interesting to note that in this study, poor performance on T_1 did not deleteriously affect subsequent performance on T_2. This suggests that in the previous two studies, although inadequate encoding on entry to P_2 clearly was not always rectified by that presentation, the encodings which resulted from P_2 were inadequate only in that they did not permit recall in the short fixed period allowed for test trials in these studies. Items to which T_1 errors were made in the current study would be if anything more inadequately encoded than such items in the first two studies, since much more time was potentially available to respond on T_1 in this experiment. Despite this, the deleterious effects of these T_1 errors were totally "wiped out" by allowing ample recall time on T_2. These results may well be interpreted as evidence that recodin s on P_2 following an error on P_1 are somewhat elaborative, perhaps of the form "the response is not the erroneous one just made, but etc., " in which case the search time required to retrieve the response on T_2 might well be excessive if T_2 is of a small, fixed duration.

5.4 Conclusions

Taken together, the results of Exps. I and II strongly suggest that the improvement at final test with increasing interpresentation spacing may be partially accounted for by the relatively small effect of the second presentation upon items only in short-term memory at the onset of the second presentation. However, there is considerable evidence in Exp. I that the improvement is maintained across inter-presentation spacings far in excess of those sufficient to "wipe out" short-term retention effects at P_2. However, in both Exps. II and III. there was little evidence to support the hypothesis that more slowly decaying short-term stimulus recognition components were responsible for this.

All three experiments suggest that the spacing effect can be accounted for only in terms of those items that are successfully responded to just prior to the second presentation, that is, on T_1. Therefore any hypothesis based upon some kind of strength theory (where there is no sharp division between those items to which correct responses are made on T_1 and those on which errors are made) appears to be untenable, since there is very little evidence that T_2 performance on items on which errors are made at T_1 is affected in any regular way by interpresentation spacing.

A comparison of Exps. II and III with Exp. I suggests that the introduction of a recognition probe on test trials may well affect the subject's task perceptions to such an extent that he changes his encoding strategy altogether, since neither of the latter studies exhibited improvements in T_2 performance over anything like the range of interpresentation spacings found to have an increasingly beneficial effect in Exp.I. This strategy change may well have involved the employment of more elaborate stimulus encodings in order to improve performance on the stimulus recognition component of the task. However, it is far from clear as to how such a strategy change would operate to reduce the range of the spacing effect.

There is very little evidence to suggest that subjects were employing a shared cyclic rehearsal strategy in these studies similar to that proposed by Atkinson and Shiffrin (1968) and outlined in 4.32. In the post-experimental interview at the conclusion of Exp III, subjects were specifically questioned on this point, and the results were illuminating. Out of a total of 25 subjects, only one subject claimed to have exclusively employed a strategy that involved the rehearsal of previously-presented items during current study and test trials. A second subject admitted to employing such a strategy "occasionally"; in particular, previously-presented items were rehearsed during study trials on pairs that the subject had decided were very difficult to memorize. In other words, this subject employed a strategy of essentially ignoring certain pairs that he found difficult. Two other subjects admitted that they had employed a shared rehearsal strategy during the early part of the session, but had subsequently given it up (as too difficult to maintain) in favour of the procedure followed by the vast majority of subjects; that is, the concentration of effort and attention upon each item as it occurred on a study or test trial. Consequently, it is most unlikely that shared rehearsal contributed significantly to the results to Exp. III; at least it is certain that subjects were not conscious of employing such a procedure.

Although there is no direct evidence on this matter in the first two experiments, it seems likely that were shared cyclic rehearsal to occur, then it would be more probable in Exp. III than in the first two studies. In the first place, the subject-pacing of test-trials in this study meant that the rate of occurrence of trials was at least partially under to subject's control, and such a procedure would surely be less disruptive of rehearsal processes than the fixed rate method used in Exps. I and II. Secondly, the relatively low-M CVC stimuli employed in Exp III would be more difficult to encode than the common words employed in the first two studies, and so a shared rehearsal strategy might well have been employed as an alternative to a deep encoding strategy. These two arguments both imply that shared cyclic rehearsal would be even less likely in Exps. I and II than in Exp III. Furthermore, since subjects in the first two studies were tested over a far longer period than those in the final experiment, there is a strong possibility that even had subjects initially employed shared rehearsal, they would have given it up relatively early during testing, so that the bulk of their results would not be affected by it. This argument applies to Exp. I in particular, when each subject was tested on six lists, and it should be pointed out in addition that the rate of occurrence of trials in this study was extremely rapid, and would therefore be extremely disruptive to rehearsal processes. In conclusion, then, it appears most unlikely that shared rehearsal occurred to any significant extent in all three studies; this is not only evidence against a shared cyclic rehearsal interpretation, but also renders unlikely a shared processing hypothesis of the type postulated by Greeno (1967, see 1.43). Although shared rehearsal was almost certainly a significant factor in the Brelsford et.al. study, due to the extremely slow presentation rates that were employed, it thus appears that such a process does not offer an explanation of the spaced presentations effect for relatively rapidly presented material.

A number of observations made in Exps. I and II were not replicated in Exp III, namely those concerning T_2 recall performance on items that were incorrectly responded to on T_1. There is considerable evidence in the first two studies that the second presentation of such items was less effective than a first presentation of a brand new item. The failure to find a similar effect in Exp III suggests t at this result is somehow tied up with the fixed duration test trials employed in Exps. I and II, in that items that were not responded to correctly on T_1 appeared to be, in general, more difficult to encode for subsequent rapid recall on T_2. It is quite possible that an elaborative encoding is employed on P_2 when a correctly recognised stimulus still results in a recall error on T_1 of the form "the response is not X but Y". Such an encoding would be decoded on a subject - paced T_2, whilst on a fixed duration T_2, there is a strong possibility that the subject often only has time to decode the item to the extent that "the response is not X". The resultant response would thus be a guess from the remaining response alternatives, and in particular, it would not be a perseveration. This hypothesis is consistent with the results of Exp. III which suggested that perseverations very much reflected a "guessing number" strategy, and would consequently occur only on those items that were either not recognised, or whose responses had not been adequately encoded in the first place, on both test trials.

However, such a hypothesis is called into question by the even poorer recall at T_2 of items that were not recognised on T_1, as compared with those that were recognised but incorrectly recalled on T_1, in Exp. II. It is quite possible that t e association encoding in this case is of the elaborative type described above, since even a guess on T_1 may serve to establish an inappropriate response pairing which is corrected on P_2. In addition, it is postulated that so much time is employed on T_2 in fulfilling the stimulus recognition requirement (since that particular stimulus was originally difficult to recognised on T_1) that even less time is available for the retrieval of the appropriate response. Such a hypothesis is con istent with the res l s of Exp. II, and furthermore, would predict no reduction in T_2 recall following a T_1 recognition error relative to a brand-new item if test trials were subject paced. This prediction is borne out by the results of Exp III.

There is one somewhat startling result from Exp. I that deserves further comment, which is that items incorrectly recalled at T_1 at a retention interval of zero were recalled at chance level on T_2.

It is difficult to offer a really convincing explanation of this result, since an attentional hypothesis clearly does not apply. If items that are incorrectly recalled at a retention interval of zero are just those that were not attended on P_1, then performance following P_2 should be equivalent to that on new items at T_1 at a retention interval equal to the P_2-T_2 spacing, or 8 trials. This is clearly not the case. Several explanations of the result are possible; for example, items incorrectly responded to on T_1 at a zero retention lag may be items that were totally mis-perceived on T_1, were again mis-perceived on P_2, but were correctly perceived on T_2. Whatever rationale is accepted, it appears that there are some items that are extremely difficult to encode uniquely for rapid recall at test. However, there is some evidence that these encoding difficulties can be overcome if P_2 is sufficiently long, since T_2 recall performance on such items was not significantly different to T_2 recall performance on items to which T_1 errors were made at non-zero retention intervals in Exp. II. This observation is at least consistent with a mis-perception hypothesis, since with a longer P_2, the subject may have time to realise that he has mis-perceived the stimulus, and may consequently have some time in which to produce an appropriate encoding. However, a mis-perception hypothesis is not completely satisfactory, since it implies that the subject perceives the stimulus correctly on P_1 and T_2; this should produce better-than-chance recall on T_2. An alternative, and perhaps more attractive hypothesis is that certain stimuli are for some reason already associated with a particular, inappropriate response, Such a pre-existant association might well enormously interfere with any attempt to produce a new encoding. However, it is still most surprising that recall performance on T_2 given an error on T_1 does not show some kind of systematic improvement with the P_1 and T_1 spacing, since in general, the shorter the P_1-T_1 interval, the more "difficult" on average the item to which a T_1 error is made, and the poorer the performance on T_2. It should be pointed out, though, that in Exp. I, the $Pr(C_2/W_1)$ curve may well reflect such an improvement with spacing (See Figure 15), since overall, the value of this statistic appears somewhat lower at P_1-T_1 spacings of 0 to 6 than at spacings of 8 to 16.

In conclusion, it is clear that Exp. I exhibits the most interesting results, in that this study clearly discriminates between short-term retention at T_1 (and consequently at P_2) and the extent to which increasing P_1-T_1 spacing benefits performance at T_2. However, there are still a number of questions to be answered. In particular it is not clear whether the apparently continued improvement of $Pr(C_2/C_1)$ with interpresentation spacings from zero up to about 8 trials reflects a reduction in the long-term forgetting rate of items re-presented at longer lags in comparison with that of singly-presented items, or whether the improvement reflects an increase in the benefit derived from the second presentation by those items that were incorrectly recalled on T_1 or correctly guessed on that trial. In other words, does P_2 principally operate to improve the encodings of those items that are already adequately encoded, or to produce more adequate encodings of those items that were inadequately encoded upon re-presentation? An attempt is made to examine these points in the next chapter.

CHAPTER SIX

SOME THEORETICAL ANALYSES

As was pointed out at the conclusion of the preceding chapter, it is only by a detailed theoretical analysis that the effect of a second presentation may be deduced. In particular, the specific question to be answered concerns whether a re-presentation serves to improve the encoding of already adequate encoded items, or to improve the encoding of inadequately encoded items, or both. In addition, such an analysis will be of great value in indicating which of the above factors is the main contributer to the spacing effect. There are many mathematical models of human memory in existance which could be employed in such an analysis, so that the first task is clearly to select the most appropriate formulation for testing the hypotheses of interest.

6.1 Models of Human Memory.

Generally speaking, there are three extant classes of models that have some chance to account for the effects of spacing. Although each of these classes embraces a wide range of specific models, it will be sufficient for the purpose of this thesis to deal briefly with the three general classes only.

6.11 Stimulus sampling theory

The first class of models derives from stimulus sampling theory, and is based on the idea of stimulus fluctuation (Estes, 1955; Izawa, 1971). These models assume that an item to be remembered together with the context in which is is presented may be described in terms of a set of "stimulus elements". At any one time, some of these elements are assumed to be available to the subject, whilst the remainder are not available. Over a period of interfering activity, each stimulus element is assumed to move at random, or fluctuate, between the available and unavailable sets.

The models as applied to paired associates generally assume that during presentation, all the available stimulus elements are sampled, and each may become associated with the response with some fixed probability.

The presence of one or more response-associated stimulus elements in the available set at test is assumed to lead to a correct response with probability 1. Consequently, if two presentations of an item are given, then if the second presentation occurs soon after the first, the stimulus items in the available set will mainly be those that were associated with the response on the first presentation, so that the second presentation will be of little value. However, as the interpresentation interval increases, the available set of stimulus elements at the second presentation will become less and less likely to include many of the already-conditioned elements, so that the second presentation will be effective in producing a large number of new stimulus element-response associations. Consequently, on a subsequent test, there will be a greater likelihood that an associated element is in the available set.

There are a number of objections to such a model. The first objection is made on broadly psychological grounds, in that models of this nature clearly assume that paired-associate encoding is elaborative in nature, and that, in particular, paired-associate learning reflects the association encoding of a wide range of stimulus attributes or elements. This supposition is in direct conflict with the results outlined in 4.12 which show that during on-going paired-associate learning, subjects tend to "home in" on a specific,preferred version of the functional stimulus, so that learning is partly a consequence of the stabilisation of the available set of stimulus attributes. Secondly, fluctuation models certainly imply that the spacing effect is dependent upon multiple contextual encoding, and this is also at variance with the extant data (e.g. see 4.33). Thirdly, sampling models are somewhat inflexible, since they postulate a very definite relationship (via the stimulus elements falling into the available set) between retention following a single presentation and the effectiveness of a second presentation. In particular, performance following a second presentation should be inversely related to performance just prior to the second presentation. This interpretation is precisely one of the hypotheses to be tested! Consequently, sampling models will be of little value in producing acceptable tests of the hypotheses of interest. Finally, it should be pointed out that sampling models would predict that a second presentation would have an optimal effect if a recall error were made just prior to it. This prediction is, to say the least, somewhat at variance with the results of Exps. I and II.

6.12 Multiprocess and consolidation models

The second class of models are the multiprocess "buffer" models of the type proposed by Atkinson and Shiffrin (1968). These models essentially assume that there is a fixed-capacity rehearsal buffer which operates both as a short-term store, and as an attentional device whereby material is encoded into a form suitable for long-term storage. The long-term storage of items is assumed to lead to imperfect retrieval due mainly to the traditional processes of associative interference. The probability that an item is represented in the long-term store is assumed to be an increasing function of the length of time that it resides in the rehearsal buffer before it is displaced (by the entry of a new item), and a decreasing function of the length of time since the item was so displaced from the buffer. Thus, these models in general interpret the spacing effect as resulting from the fact that items may continue to reside in the rehearsal buffer (and consequently increase their long-term storage probability) during the presentation and testing of other items which constitues the interpresentation interval. This notion constitutes a mechanism through which memory traces consolidate over time.

There is considerable controversy as to whether, in fact, it is possible to consciously process material whilst attending to the presentation and testing of other items, and indeed there is some evidence that shared rehearsal did not contribute greatly to the results of the three experiments in the preceding Chapter. Furthermore, models of this kind place a very specific interpretation upon the spacing effect, so that once again, it is difficult to see how they could provide a framework for testing the hypotheses of interest. It should also be noted that during the period over which items are assumed to be "consolidated", they are resident in the rehearsal buffer, whence they are assumed to be recalled perfectly. There is clearly very little evidence of such a consolidation process in the single-presentation retention curve of Exp.I (see Fig.14). Finally, it should be pointed out that the models of Atkinson and Shiffrin were intractable to the extent that predictions had to be generated using monte-carlo methods, so that enormous practical difficulties would be expected if models of this type were adopted.

6.13 Markovian Models.

Models of the third class are generally known as Markovian models, and in their simplest form they assume that an item may be held in one of three states: a "naive" state (i.e. not in memory), a short-term retention state, and a permanent, long-term memory state. The individual members of this class are defined by their various assumptions about transition probabilities from one state to another, and include the original models of Atkinson and Crothers (1964) and Greeno (1967), the more generalised model of Bjork (1966), and the modified version of Bjork's model proposed by Rumelhart (1967) and called by him GFT (General Forgetting Theory).

A Markovian approach offers several advantages. In the first place, such models are not based upon mechanisms which relate very specifically to any particular psychological theory, as are the stimulus sampling and buffer types of model. However, Markovian models do provide a flexible framework within which hypotheses relating to specific psychological processes may be tested. For example, additional states may be added to the model, in such a way that the transition probabilities between these states represent the desired psychological process. A statistical test may then be constructed to determine whether the inclusion of these additional parameters significantly improves the fit of the model to the data (see 6.32).

Perhaps the most pronounced advantage offered by a Markovian model lies in the fact that the postulated states relate in a very direct and obvious way to the observed aspects of the data. This means that, first of all, the states of the model may be operationally defined, so that the long-term and short-term retention states of the model may be taken as the smallest number of states necessary to predict the observed relationship between retention interval and recall performance. This does not necessarily imply the acceptance of a theory of memory that postulates a dichotomy into a specific short-term and a specific long-term memory store, but rather reflects the observation that such a dichotomy in the model is necessary in order to predict the broader aspects of the extant data. In the second place, the direct relationship between the postulated states of the model and the predictions of the model may be particularly useful in suggesting improvements and modifications to the model in the eventuality that it does not successfully characterise the data.

Finally, a Markovian approach appears to be unique in providing a framework within which to evaluate the hypotheses outlined at the beginning of this chapter; in other words, does a second presentation operate to retard the rate of forgetting of already quite adequately encoded items, or simply to improve the encodings of those items that were not adequately encoded on their first presentation? Markovian models will provide a method of determining which of these mechanisms best accounts for the spacing effect.

Consequently, it is proposed to employ a Markovian approach in the following theoretical analyses, in the expectation that the application of such models will provide a far more precise summary of the data of Experiment I than that available as a result of the preliminary analyses carried out in the previous chapter. It is hoped that this information may prove to be of great value in the evaluation of the various psychological hypotheses advanced to explain the effects of the spacing of paired-associate study trials upon subsequent recall performance.

6.2 Numerical and Statistical Methods

The analyses to be reported in the remainder of the Chapter were carried out solely upon data from Experiment I, since this was the only study that included a sufficient range of interpresentation spacings to effectively discriminate between the range of interpresentation intervals over which short-term retention effects survive until P_2, and that range over which performance at final test apparently continues to improve. Independent analyses of $Pr(C_1)$, $Pr(C_2/C_1)$ and $Pr(C_2/W_1)$ clearly cannot convincingly provide an appreciation of the extent to which each of these performance measures determines the relationship of $Pr(C_2)$ with the interpresentation interval. Before proceding to describe the theoretical analyses in detail, however, it is proposed to give a summary of the numerical and statistical methods which they employed.

6.21 Numerical Methods.

Since the data essentially takes the form of the frequencies of observations which were observed to fall into the various response categories, the appropriate goodness-of-fit statistic is clearly Chi-squared. Thus, a minimum Chi-squared technique was employed to fit the various models to the data (i.e. to produce the "best" estimates of the various parameters of each model). It was not found possible to minimise χ^2 analytically, due to the complex relationship between χ^2 and the various parameters, and to the relatively large numbers of parameters involved (between four and nine).

Consequently, minimisation in all the analyses was accomplished by the use of Fortran computer programmes which employed the NAG subroutine E04CAF. This subroutine minimises a function of several variables by an iterative procedure based upon a direct search, non-gradient method developed by Powell (1964).

6.22 Statistical Methods

The statistical techniques employed were based on a result in the works of Neyman (1949). Suppose it is desired to fit a mathematical model to an array of frequency data with n degrees of freedom, and that there is a vector of the theoretical parameters of the model (p_1, $p_2,\ldots,p_s$) where $s \leq n$, such that the model predicts the probabilities of observations falling into each category of the data array as a function of this parameter vector. Suppose, in addition, there is a special case of the model with q ($<$ s) parameters, ($p_1,\ldots,p_q$) where $p^*_{q+1} = p_{q+1},\ldots, p^*_s = p_s$, and $p^*_{q+1},\ldots, p^*_s$ are either constants, or functions of $(p_1,\ldots,p_q)$. In other words, the special case defines a "sub-model" of the original model. Then, if fitting the full model yields a minimum χ^2 with n - s degrees of freedom of m_1, and fitting the special case yields a minimum χ^2 with n - q degrees of freedom of m_2, Neyman (1949) showed that m_2-m_1 has a χ^2 distribution with s - q degrees of freedom under the null hypothesis that the improvement in fit from allowing $p_{q+1},\ldots, p_s$ to vary freely is due merely to capitalising on chance.

Thus, if the null hypothesis is accepted, nothing is gained by allowing $p_{q+1},\ldots, p_s$ to vary freely, so that they can just as well be left at their a-priori values of $p^*_{q+1},\ldots, p^*_s$. This result is of obvious applicability in comparing the fits of models involving additional parameters which define the spacing mechanisms described by the various hypothesis outlined earlier with that of a model which does not include these extra parameters.

6.3 Fits to performance on T_1

Because the $Pr(C_1)$ curve resulting from Exp.I did not show a regular, monotonic decline with retention interval as was expected from previous data (see Figures 14 and 1), it was anticipated that difficulties would occur in fitting a Markovian forgetting model to this aspect of the data. Two alternative formulations were proposed, each involving four parameters, and their fits to T_1 performance were compared.

6.31 Model 1

It will be remembered that in Exp.I, so far as T_1 performance is concerned, each item received a schedule which may be depicted as

$$P_1 \underset{i \text{ trials}}{\cdots\cdots\cdots} T_1$$

where the retention interval i took values of 0,1,2,3,4,5,6,8,12 or 16 trials. Model 1 was constructed in order to predict recall performance at T_1 as a function of the retention interval i.

The model includes 3 states:

L - Long-term retention state

S - Short-term retention state

N - Naive (unlearned) state

The probability of a correct response on a test trial given that an item currently resides in one of the three states of the model is represented by the vector $\underline{R}'$, where

$$\underline{R}' = \begin{array}{ccc} L & S & N \\ (\ 1, & 1, & g\ ) \end{array}$$

and g is the probability that the subject correctly guesses the correct response. The results of Exp.II suggested that this probability did not differ from the theoretical guessing probability (see 5.22), so that as there were 15 response alternatives in Exp.I, g was set to 1/15, and consequently g did not constitute an effective parameter of the model.

The effect of the study trial P_1 may be represented by the vector of probabilities $\underline{P}$, where

$$\underline{P} = \begin{array}{ccc} L & S & N \\ (\ a\ , & (1-a)c, & (1-a)(1-c)\ ) \end{array}$$

Finally, the effect of an intruding trial on another item, or a "forgetting" trial, is described by the matrix $\underline{F}$, where

$$\underline{F} = \begin{array}{c} \\ L \\ S \\ N \end{array} \begin{array}{c} \begin{array}{ccc} L & S & N \end{array} \\ \begin{bmatrix} p & 0 & 1-p \\ 0 & s & 1-s \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

It was assumed that the test trial T_1 did not affect the state in which items were held.

The probability of a correct response on T_1 at a retention interval of i is predicted by the following equation

$$P_i(C_1) = \underline{P}\ \underline{F}^i\ \underline{R}$$

or, alternatively, by

$$P_i(C_1) = B(i) + (1-B(i))g$$

where $B(i) = ap^i + c(1-a)s^i$

Thus, the model interprets performance at T_1 as resulting from long-retention items, that are retained in state L with a relatively high probability p on each forgetting trial, partially from short-retention items (especially at short retention intervals) which are retained with a relatively low probability s on each forgetting trial, and from chance-level guessing to items that have been forgotten, or were never learned in the first place.

6.32 Model 2

It is clear from the above formulation that Model 1 predicts non-zero short-term retention effects at retention intervals of two or more trials, whereas examination of Figure 14 suggests that short-term retention has been essentially "wiped out" by a retention interval of two or more trials. This might well cause Model 1 to yield a low estimate of the parameter s, in order to compensate for the over-prediction of short-term retention effects at moderate retention intervals, resulting in an underprediction of $Pr(C_1)$ at short retention intervals. Although no monotonic forgetting model could possibly hope to satisfactorily explain the $Pr(C_1)$ data of Exp. I, it was thought that a superior fit might ensue from a model that predicts no short-term retention at all at retention intervals of 2 or more trials. Model 2 makes such a prediction.

The model includes 4 states:

L - long-term retention state

S_1, S_2 } Short-term retention states

N - Naive (unlearned) state

The probability of a correct response on T_1 is represented by the vector $\underline{R}'$, where

$$\begin{array}{cccccc} & & L & S_1 & S_2 & N \\ \underline{R}' & = & (1, & 1, & 1, & g) \end{array}$$

and $g = 1/15$ as before. The effect of the initial presentation P_1 is represented by the vector $\underline{P}$, where

$$\begin{array}{cccccc} & & L & S_1 & S_2 & N \\ \underline{P} & = & (a, & (1-a)c, & 0, & (1-a)(1-c)) \end{array}$$

Finally, the effect of an intruding trial involving another item, or a forgetting trial, is represented by the matrix $\underline{F}$, where

$$\underline{F} = \begin{array}{c} \\ L \\ S_1 \\ S_2 \\ N \end{array} \begin{array}{c} \begin{array}{cccc} L & S_1 & S_2 & N \end{array} \\ \begin{bmatrix} p & 0 & 0 & 1-p \\ 0 & 0 & s & 1-s \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Again, it was assumed that the test-trial, T_1, had no effect upon the state in which an item resided.

The probability of a correct response at a retention interval of i trials is predicted to be

$$Pi(C_1) = \underline{P}\ \underline{F}^i\ \underline{R}$$

or alternatively

$$Pi(C_1) = B(i) + (1-B(i))g$$

$$\text{where } B(i) = \begin{cases} a + (1-a)c, & \text{if } i = 0 \\ ap + c(1-a)s, & \text{if } i = 1 \\ ap^i, & \text{if } i \geqslant 2 \end{cases}$$

The splitting of the short-term retention state into the two states S_1 and S_2 was really a mathematical device employed to preserve the Markovian aspects of the model, in that all four parameters of the model, c, a, p and s remain constant over trials.

6.33 The fits of Models 1 and 2 to $Pr(C_1)$

For each model, the predicted value of $Pi(C_1)$ was multiplied by the total number of observations at a retention interval of i, to yield a prediction of the number of observations falling into the C_1

category. When subtracted from the total number of observations, this yielded the predicted number of observations falling into the W_1 category (i.e. the sum of categories W_1C_2 and W_1W_2), and the value of χ^2 was consequently computed as $\sum (O-E)^2/E$, where O and E refer respectively to the observed and predicted frequencies in each of the 20 categories.

TABLE 12

Results of fitting models 1 and 2 to retention data at T_1 from Exp. I

(i) Observed and Predicted values of $Pr(C_1)$

P_1-T_1 LAG; i	Observed	Predicted Model 1	Predicted Model 2
0	.930	.931	.930
1	.593	.560	.593
2	.424	.473	.447
3	.378	.451	.445
4	.480	.444	.443
5	.438	.442	.442
6	.511	.440	.440
8	.442	.437	.437
12	.443	.430	.431
16	.394	.424	.425

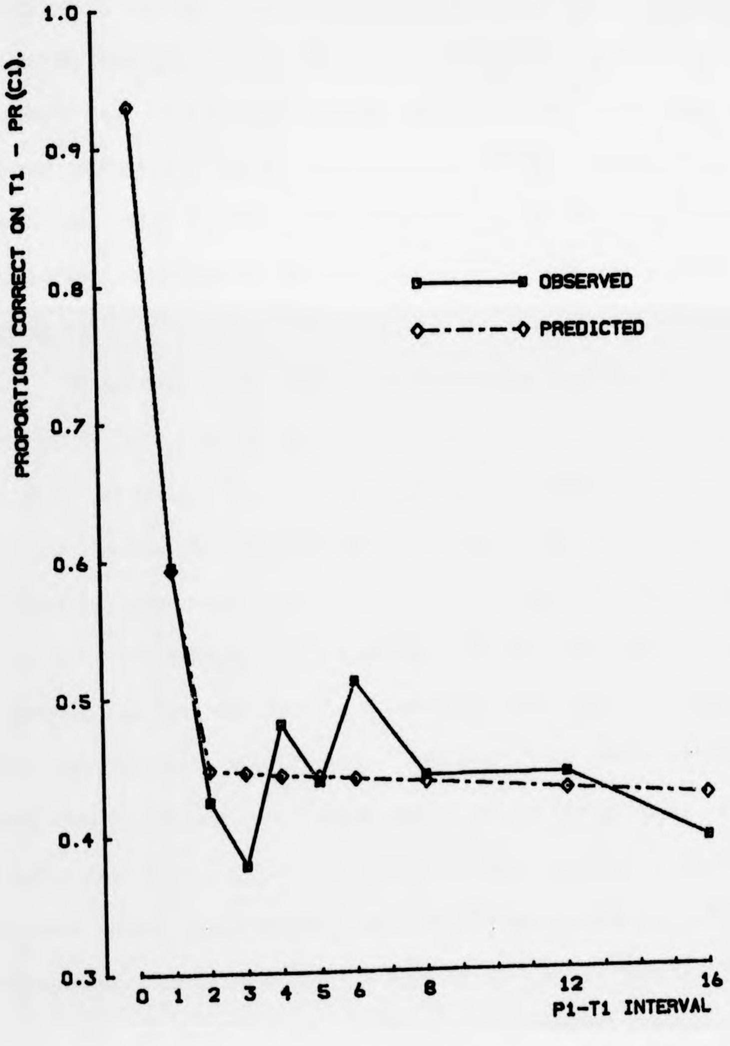
(ii) Parameter estimates and goodness of fit

	Parameter Estimates c	a	p	s	Minimum Chi-squared	d.f.	prob.
Model 1	.874	.410	.996	.234	27.274	6	<.001
Model 2	.872	.410	.996	.302	21.067	6	<.001

The data array comprises 10 degrees of freedom, whilst the fitted models each allowed 4 parameters (c,a,p and s) to vary freely, yielding 6 degrees of freedom for the minimum chi-squared statistics. Details of the fits of the two models are presented in Table 12. Although Model 2 did result in a reduction of 6.207 in the minimum chi-squared value, nevertheless this value was still highly significant, which suggests that the fit of Model 2 still left much to be desired. However, a close examination of the observed and predicted values in the table show that Model 2 did result in an improvement in fit at retention

FIGURE 22

Observed values of $Pr(C_1)$ from Exp.I with the values predicted by Model 2.


PROPORTION CORRECT ON T1 - PR(C1).
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
OBSERVED
PREDICTED
0 1 2 3 4 5 6 8 12 16
P1-T1 INTERVAL

intervals of 1 and 2 trials, so that the hypothesis that short-term retention disappears after a retention interval of 2 or more trials does appear justified, since Model 1 clearly under-predicted $Pr(C_1)$ at a retention interval of 1 trial to compensate for its over-prediction at a retention interval of 2 trials.

The observed values of $Pr(C_1)$ are presented along with the predicted values from Model 2 in Figure 22. The figure suggests that the model predicts the data as well as any monotonic forgetting model is likely to. It therefore appears that the noise in the data curve between retention intervals of 2 and 8 is the cause of the enormously significant value of the minimum Chi-squared, and this observation only serves to underline the fears expressed at the beginning of section 6.3.

Of course, in the absence of convincing evidence that the apparent recovery in $Pr(C_1)$ is due to anything more than shortcomings in the design of Experiment I, there remains no alternative but to attempt to complete the desired analysis within the framework of Model 2. Fortunately, some methods do exist whereby meaningful results may be obtained. Before going on to describe these methods, however, it should be pointed out that the results obtained in this section do have some slight psychological significance. Although it is fairly certain that short-term forgetting is shomehow caused by the displacement of items by later ones (see 1.23), on the basis of these results, it appears that the least recent item currently held in STM is the one most likely to be displaced. In other words, it appears that short-term retention arising from an echoic sensory store and/or "passive" rehearsal has a capacity of about two nominal items, and that the displacement of attention caused by an intruding trial has a fair chance of "wiping out" the item attended on the previous trial, and is almost certain to displace the item attended on the previous trial but one. This observation is a last consistent with the supposed sequential properties of rote rehearsal and echoic storage.

6.4. Further analyses involving Model 2

Of course, Model 2 as presented in the preceding section, whilst clearly superior to Model 1 in accounting for the relationship of $Pr(C_1)$ with retention interval, is clearly not able to offer an explanation of all the aspects of the results of Exp.I. Consequently, the model must undergo a process of augmentation and modification before it can be expected to provide a satisfactory tool for the evaluation of the two hypotheses advance as alternative explanations of the spacing effect.

6.41 Model 2 augmented

In its most complete form, Model 2 comprises 8 states:-

$$\left.\begin{array}{l}\mathrm{LU}\\ \mathrm{LI}\\ \mathrm{LD}\end{array}\right\}\ \text{long-term retention states}$$

$$\left.\begin{array}{l}S_1\\ S_2\end{array}\right\}\ \text{short-term retention states}$$

$$\left.\begin{array}{l}\mathrm{Ns}\\ \mathrm{Nf}\\ \mathrm{Nx}\end{array}\right\}\ \text{"naive" or unlearned states.}$$

The probability of a correct response on any test-trial (either T_1 or T_2) may be represented by the vector $\underline{R}'$, where

$$\begin{array}{cccccccccc} & & \mathrm{LU} & \mathrm{LI} & \mathrm{LD} & S_1 & S_2 & \mathrm{Ns} & \mathrm{Nf} & \mathrm{Nx} \\ \underline{R}' & = (& 1, & 1, & 1, & 1, & 1, & g, & g, & g\ ) \end{array}$$

and where $g = 1/15$ as before. As before also, the model assumes that the state in which an item resides is unchanged by a test trial. The effect of a P_1 (the first presentation of an item) is represented by the vector $\underline{P}_1$, where

$$\begin{array}{cccccccccc} & & \mathrm{LU} & \mathrm{LI} & \mathrm{LD} & S_1 & S_2 & \mathrm{Ns} & \mathrm{Nf} & \mathrm{Nx} \\ \underline{P}_1 & = (& 0, & a_0, & 0, & (1-a_0)c, & 0, & 0, & 0, & (1-a_0)(1-c)) \end{array}$$

Thus, on its first presentation, an item enters one of the three states L_1, S_1 or N_x. The parameter c clearly corresponds to the parameter c in the earlier formulation, whilst a_0 corresponds to the original parameter a. The effect of a second presentation, P_2 is represented by the matrix $\underline{P}_2$, where

$$\underline{P}_2 = \begin{array}{c|cccccccc|} & LU & LI & LD & S_1 & S_2 & Ns & Nf & Nx \\ \hline LU & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LI & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ LD & a_3 & 1-a_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ S_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ S_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ Ns & 0 & a_2a_1 & 0 & 1-a_2a_1 & 0 & 0 & 0 & 0 \\ Nf & 0 & a_1 & 0 & 1-a_1 & 0 & 0 & 0 & 0 \\ Nx & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{array}$$

Thus, items in states LU, LI and S_1 remain there on a P_2. An item in state LD enters LU with probability a_3, or LI with probability $1-a_3$. Items in states S_2 and Nx automatically enter S_1, whilst an item in state Ns or Nf may enter either LI or S_1.

The effect of an intruding trial on another item, or a forgetting trial, may be represented by the matrix $\underline{F}$, where

$$\underline{F} = \begin{array}{c|cccccccc|} & LU & LI & LD & S_1 & S_2 & Ns & Nf & Nx \\ \hline LU & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ LI & 0 & pr & p(1-r) & 0 & 0 & 0 & 1-p & 0 \\ LD & 0 & 0 & p & 0 & 0 & 0 & 1-p & 0 \\ S_1 & 0 & 0 & 0 & 0 & s & (1-s)q & (1-s)(1-q) & 0 \\ S_2 & 0 & 0 & 0 & 0 & 0 & q & 1-q & 0 \\ Ns & 0 & 0 & 0 & 0 & 0 & q & 1-q & 0 \\ Nf & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ Nx & 0 & 0 & 0 & 0 & 0 & 0 & 1-q & q \end{array}$$

Before going on to discuss the completed version of Model 2 in detail, it may be helpful to remind the reader of the form taken by the presentation-test schedules applied to items in Exp.I. These schedules may be depicted as follows:

$$P_1 \underbrace{\ldots\ldots}_{i\ \text{trials}} T_1 P_2 \underbrace{\ldots\ldots}_{8\ \text{trials}} T_2$$

Thus, each particular item received an initial presentation, P_1, and was subsequently tested at T_1 after an interval of i = 0,1,2,3,4,5,6, 8, 12 or 16 trials involving other items. A second presentation, P_2, always followed immediately after T_1, so that performance at T_1 would serve as some indication as to the state of an item on its entry to P_2; in particular, an error on T_1 would imply that the item entered P_2 in one of the three "naive" states, as these are the only states in which recall errors can occur at test. Furthermore, the P_1-T_1 interval i may be thought of both as the T_1 retention interval, and the P_1-P_2 interpresentation interval. Finally, a second test trial, T_2, always occurred after an interval of 8 trials involving other items (i.e. "forgetting" trials) following P_2.

An examination of the forgetting matrix $\underline{F}$ will reveal that items in the long-term state LU are never forgotten. Thus, the state LU may be thought of as a "unique" state, in which items have received a unique long-term encoding which is not prone to the usual interference effects. Items in states LI and LD may be forgotten into what is essentially a "forget" state, Nf, and indeed, state LI corresponds exactly with state L in the earlier formulation of the model. The state is, however, split into two components, LI and LD. in such a way that given that an item remains in state L on a forget trial, it will remain in state LI with probability r, and will enter state LD with probability 1-r. Items in states LI and LD may be thought of as being fairly deeply encoded, but still prone to a certain amount of interference, so that they may eventually be forgotten into Nf. A re-examination of $\underline{P}_2$ will reveal that items which are in LI on a subsequent presentation will remain there; in other words, their original encoding is maintained, whilst there is a non-zero probability that items in LD will enter LU, or in other words, have their encoding improved on a P_2. Thus, the three long-term states, LU, LI and LD, together with their various transition probabilities, represent a mechanism whereby the improvement with T_2 performance with interpresentation spacing is seen as a consequence of the increasing probability that items that are relatively well-encoded at P_1 (and enter LI) enter LU on P_2 from LD.

This mechanism may have several underlying psychological rationales. LI may be though of as an "immediate" long-term retention state, wherein items are immediately recalled, so that the subject has no reason to suspect that his encoding is not perfectly adequate, and simply maintains it on P_2. Recall from the "delayed" long-term state LD may be delayed to the extent that the subject has to initiate an effortful search process in order to retrieve the correct response, and may therefore be motivated to alter or otherwise improve his encoding. Alternatively, with short interpresentation spacings, it is probable that the two presentation trials have many episodic aspects in common, whilst with a longer interpresentation interval, P_1 and P_2 may share very few episodic cues. Thus, if P_2 follows shortly after P_1, an item that has been deeply encoded on P_1 (and enters LI) will simply maintain its encoding on P_2, and will remain in LI, either because there are relatively few new episodic components available at P_2 that can be used to elaborate and enrich the encoding, or because the P_1 and P_2 episodes are so similar that the subject doesn't realise on P_2 that there are alternative, superior encodings of the pair. With a longer interpresentation interval, there is a far greater probability that the episodic aspects of P_1 and P_2 are vastly different, so that a superior encoding may result on P_2, either because there are many new episodic attributes available on P_2 which may be employed to elaborate the original encoding, or because the two presentation episodes are so different that the subject just can't help thinking of new, improved encodings on P_2. Thus, LD may be thought of as representing these items that were deeply encoded on P_1, but whose encodings share relatively few episodic aspects with the current trial.

Of course, the state LU may be thought of as a "unique" long-term state representing those items whose encoding is so unique as to be unaffected by interference from other items, so that these items will never be forgotten. It will be noticed, however, that items cannot enter LU on their first presentation, $\underline{P_1}$. Rather than place a psychological interpresentation on state LU, it is perhaps more sound to conceive of LU as representing a convenient mathematical way of expressing the notion that the rate of decay of deeply encoded items may be reduced by a second presentation which occurs at a sufficiently long interval after the first presentation. Since items enter LU with probability a_3 (which is likely to be less than unity) the overall effect of the state will be to reflect such a decline in forgetting rate. It is certainly very doubtful that any item is so well encoded that it will never be forgotten. This particular formulation was chosen to represent the hypothesis of a reduction in the long-term decay rate because it was the most efficient way of expressing the idea; only two additional parameters r and a_3 are necessary to specify the process.

The two states S_1 and S_2 are identical to the corresponding states in the earlier formulation of the model, and require no further explanation; they merely reflect the short-term retention effects which are clearly identifyable in the $Pr(C_1)$ function. It should be pointed out, though, that the model predicts that short-retention items will remain in the short-retention state S_1 following a second presentation, P_2. This assumption is justified by previous results (e.g. Greeno, 1964; see 1.43) and by the results of Exp.I. Performance at T_2 following two massed presentations (.461) was hardly any better than performance following a single presentation at a similar retention interval of 8 trials (.442), so that there is little evidence that short-retention items received any benefit from the second presentation.

It will be noticed that the original naive state N has been split into three states, Ns, Nf and Nx. An examination of the matrix $\underline{F}$ will confirm that all items which are prone to forgetting are eventually forgotten into state Nf, the "forget" state. On a subsequent presentation, P_2, items in the "forget" state may enter LI with probability a_1, or S_1 with probability $1-a_1$. The state was postulated in order to predict the observation that recall of items at T_2 (which followed 8 trials after P_2) was poorer if an error was made on T_1 (so that the item would be in Nf on entry to P_2) than recall of a brand-new item of T_1 at a retention interval of 8 trials (see 5.12). Consequently, it was expected that the value of a_1 would be estimated to be less than that of a_0, the probability that a new item enters LI on its first presentation.

It will be noticed that the matrix $\underline{P}_2$ makes no provision for items in the naive state Nf to remain in a naive state, in contrast to the vector $\underline{P}_1$, which does allow new items to enter a naive or guessing state on initial presentation. Since $Pr(C_1)$ was less than perfect at a retention interval of zero, it is clear that some new items remain unlearned on P_1. However, since the interval between P_2 and T_2 was always 8 forgetting trials, it is impossible to discriminate between a formulation that allows some of the items that do not enter a long-term state of P_2 to enter either S_1 or a naive state, and a formulation that only allows such items to enter S_1. Items which enter S_1 at P_2 will certainly be in a naive state by the time T_2 occurs, so that effectively, T_2 performance on items that entered P_2 in a naive state would only depend on the probability that such items entered LI on the second presentation, P_2. Consequently, the simpler formulation was chosen, since this would lead to somewhat simpler predicted values, and would thus save valuable computer time when finding the best fit of the model by an iterative procedure.

It will be seen from $\underline{P}_1$ that items that do not enter a retention state on their first presentation enter the naive state Nx, whilst $\underline{P}_2$ ensures that items still resident in Nx on their second presentation automatically enter S_1, whence they will have entered a guessing state with probability 1 by the time that T_2 occurs (i.e. 8 trials after P_2; S_1 items are automatically forgotten into a guessing state after 2 or more forgetting trials). Thus, in particular, the model predicts that items which are incorrectly recalled on T_1 at a P_1-T_1 interval of zero (and thus must be in state Nx) will be recalled at chance level on T_2. This result, of course, is precisely what is observed in Exp.I. On a forget trial, F, items in Nx are seen to remain there with probability q, and to enter the "forget" state, Nf, with probability 1-q. Thus, as the interpresentation interval increases, the proportion of items to which T_1 errors are made that are in Nf (as opposed to Nx) increases, so that $Pr(C_2/W_1)$ is predicted to increase from the guessing level observed at an interpresentation interval of zero.

The postulated forgetting process from Nx to Nf, in conjunction with the assumptions concerning the third naive state Ns, constitutes the mechanism by which the alternative spacing hypothesis is represented by the model. This hypothesis states that the improvement in recall on T_2 with increasing interpresentation spacing results from a process whereby items that were not adequately encoded on P_1 (and therefore entered S_1 or Nx) will for some reason be poorly processed on P_2 if aspects of the original poor encoding survive until P_2 even though they may no longer support recall. Thus, items that enter state S_1 on their first presentation may, over a period of forgetting trials, enter state Ns, a naive state. Items in Ns on entry to P_2 have a probability if a_2a_1 of entering LI (which is thus less than or equal to the probability a_1 that Nf items enter LI on a second presentation). Forgetting occurs from Ns to Nf at the same rate as forgetting from Nx to Nf, so that if q takes a value greater than zero, the model predicts that $Pr(C_2/W_1)$ increases with interpresentation interval from an initial chance level at an interval of zero, and furthermore, this improvement will be maintained over interpresentation intervals in excess of 2 trials. It was not clear in the preliminary analysis of Exp.I whether this was, in fact, the case, so that the proposed theoretical analysis offers the only method of evaluating this hypothesis. Furthermore, it should be pointed out that the proposed mechanism is also capable of predicting the observed continued increase in the value of $Pr(C_2/C_1)$ with interpresentation spacings in excess of 2 trials, since T_2 performance on items that were correctly guessed on T_1 will show continued improvement over such a range.

The mechanism represented by the state Ns may be given the following psychological interpretation. It may be that episodic aspects of the first presentation trial P_1 to some extent determine the subject's perceptual response to the presented pair. In particular, this perceptual response may be such that the encoding of the pair which results will not subsequently support recall. Of course, an inadequately encoded pair may survive for a short period because of echoic memory and passive rehearsal effects, and it has already been pointed out that an item held in such a short-retention state will not be more adequately encoded on P_2. However, the current mechanism states that even if such a pair has left the short-term system, it may enter Ns, which will result in P_2 failing to fulfill its maximum encoding potential. It is postulated that with short interpresentation spacings, P_2 may share a number of episodic aspects with P_1, and that these aspects may operate to produce the same perceptual response to the pair that resulted in the original inadequate encoding made on P_1, so that consequently, there is some likelihood that the original inadequate encoding may recur. The decline in this probability with increasing interpresentation spacing is represented in the model by the forgetting which may occur from Ns to Nf, although it should be pointed out that if a_1 is less than a_0 on a P_2, then this will mean that the probability that an original inadequate encoding survives on P_2 never reaches zero.

There are definitely probelms in similarly accounting for the mechanism represented by forgetting from Nx to Nf. It has already been pointed out that the state Nx was postulated in order to predict the chance level of $Pr(C_2/W_1)$ at an interpresentation interval of zero, and it was argued in section 5.4 that this result may have been caused by stimulus items to which T_1 errors are made at a zero retention lag possessing pre-existent associations with inappropriate responses. The pairing of such a stimulus, pre-associated with response X, with response Y on P_1 may result in an association encoding which states that "the response X is incorrect, and the appropriate response is Y". With only 2 seconds to respond to the stimulus on T_1, it is possible that even at a retention interval of zero, the subject does not have sufficient time to unravel the encoding beyond "the response X is incorrect...". The current formulation regards an inadequate encoding of this type as being affected in a similar way by episodic cues as other types of inadequate encoding, although because it is possible that the original pre-existent inappropriate pairing is based on fairly "deep", semantic aspects of the stimulus, the model predicts that the survival of episodic cues from P_1 to P_2 will automatically result in the original extended encoding being reproduced on P_2.

There are, however, a number of feasible alternative mechanisms. For example, transition from Nx to Nf may occur at a different (possibly more rapid) rate to transition from Ns to Nf on a forgetting trial. This mechanism was rejected on the grounds that it would require the postulation of an additional parameter. Alternatively, forgetting may occur from Nx to Nf via Ns; thus at short non-zero interpresentation intervals, items that entered Nx on P_1 might enter LI on P_2 by reason of their having moved into Nf during the retention interval, so that the maintenance of episodic cues from P_1 to P_2 would not automatically imply the recurrence of the original poor encoding on P_2. This hypothesis would also require the postulation of an additional parameter to reflect the rate of transition from Nx to Ns on a forget trial, although this might conceivably be set to unity. However, it should be pointed out that the proportion of items to which T_1 errors are made at non-zero retention intervals, that had entered Nx on P_1, would be relatively small, so that this more complex hypothesis would only have the slightest marginal effect on the predictions of the model. Thus, the formulation as presented was adopted, even though it might tend to underpredict $Pr(C_2/W_1)$ at short, non-zero interpresentation spacings.

Operationally speaking, the current formulations appears quite sound, and will allow the comparison of three hypothesis, which may be defined as follows:-

H_1 : the spacing effect results merely from the fact that short-retention items that were otherwise inadequately encoded on P_1 will be inadequately encoded on P_2.

H_2 : the spacing effect results from the fact that items that were adequately encoded at P_1 may receive enhanced encodings on P_2, and that the probability of such an enhancement on P_2 increases with interpresentation spacing, provided the original encoding survives until P_2.

H_3 : the spacing effect results from the fact that items that were inadequately encoded on P_1 may receive the same inadequate encoding on P_2 even after short-term retention has ceased to support recall. The probability that the original inadequate encoding will survive in this way decreases with interpresentation spacing.

It will be clear on an examination of the model that the additional mechanism representing H_2 can be removed from the model by applying the restriction that r = 1 and/or $a_3 = 0$. Similarly, the mechanism that represents H_3 can be removed from the model by applying the restriction that q = 0 and $a_2 = 1$. The full model possesses nine parameters : a_o, a_1, a_2, a_3, c,p,r,s, and q (since g is postulated to be 1/15), so that versions of the model representing the following hypotheses may be fixed to the data:

Version I	(H_1):	c, a_o, a_1, p and s free; r=1, a_3=0; q=0, a_2=1.
Version II	$(H_1 \cup H_2)$:	c, a_o, a_1, p, s, r and a_3 free; q = 0, a_2 = 1
Version III	$(H_1 \cup H_3)$:	c, a_o, a_1, p,s,q and a_2 free; r =1, a_3 = 0
Version IV	$(H_1 \cup H_2 \cup H_3)$:	c, a_o, a_1, p, s, q, a_2, r and a_3 all free.

The improvement in goodness of fit that results from including an additional hypothesis (i.e. freeing two additional parameters) may be tested by using the result of Neyman (1949) described in section 6.22, which states that the difference between the two appropriate minimum Chi-squareds it itself distributed as a Chi-squared with 2 degrees of freedom.

6.42 Derivations of Model 2.

It is immediately apparent that the proportion of correct responses on T_1, $Pr(C_1)$, is predicted by the equation

$$Pi(C_1) = \underline{P}_1 \underline{F}^i \underline{R}$$

where i represents the interpresentation spacing, and that the proportion of correct responses on T_2, $Pr(C_2)$ is predicted by the equation

$$Pi(C_2) = \underline{P}_1 \underline{F}^i \underline{P}_2 \underline{F}^\delta \underline{R}$$

However, these equations are not particularly enlightening, and, furthermore, they do not constitute predictions which are sufficient for all aspects of the data.

An alternative formulation for $Pi(C_1)$ is given by

$$Pi(C_1) = B(i) + (1-B(i))g$$

$$\text{where } B(i) = \begin{cases} a_o + (1-a_o)c, & \text{if } i = 0 \\ a_o p + (1-a_o)cs, & \text{if } i = 1 \\ a_o p^i & \text{if } i \geq 2 \end{cases}$$

This formulation is identical to that presented in section 6.32 for the simple version of Model 2, since a_o in the augmented version corresponds to the parameter a in the simple version. In order to generate further predictions, it will first be necessary to define a number of different probabilities. The following represent the probabilities that an item is in each of the states of the model following an initial presentation P_1, and i subsequent forgetting trials:

$$Q_{LI}(i) = a_o p^i r^i$$

$$Q_{LD}(i) = a_o p^i (1-r^i)$$

$$Q_S(i) = \begin{cases} c(1-a_o) & \text{if } i = 0 \\ c(1-a_o)s & \text{if } i = 1 \\ 0 & \text{if } i \geq 2 \end{cases}$$

$$Q_{Ns}(i) = \begin{cases} 0 & \text{if } i = 0 \\ c(1-a_o)(1-s)q & \text{if } i = 1 \\ c(1-a_o)(sq^{i-1} + (1-s)q^i) & \text{if } i \geq 2 \end{cases}$$

$$Q_{Nx}(i) = (1-c)(1-a_o)q^i$$

and finally,

$$Q_{Nf}(i) = 1 - Q_{LI}(i) - Q_{LD}(i) - Q_S(i) - Q_{Ns}(i) - Q_{Nx}(i)$$

It will be noticed that since T_1 recall is perfect in both states S_1 and S_2, and that, furthermore, items in both states automatically enter S_1 on a P_2, these two states have been combined to give a single probability $Q_S(i)$, which represents the appropriate probability that an item is either in state S_1 or in state S_2. Also, since no items can enter LU on a P_1, the appropriate probability for this state is zero.

On exit from a second presentation, P_2, items can only be in one of three states, namely LU, LI or S_1. The resultant probabilities of a correct response on T_2 (which always follows P_2 after 8 forgetting trials) are therefore

1, if the item left P_2 in state LU

A, if the item left P_2 in state LI

(where $A = p^8 + (1-p^8)g$)

g, if the item left P_2 in state S_1

It is now possible to generate the predicted values of $Pr(C_1C_2)$ and $Pr(W_1C_2)$ for each value of i, as follows

$$Pi(C_1C_2) = a_3 Q_{LD} + A(Q_{LI} + (1-a_3)\,Q_{LD} + ga_1(Q_{Nf} + a_2\,Q_{Ns}))$$
$$+\, g\,(Q_S + g\,((1-a_1)\,Q_{Nf} + (1-a_1a_2)\,Q_{Ns} + Q_{Nx}))$$

$$Pi(W_1C_2) = (1-g)(Aa_1(Q_{Nf} + a_2 Q_{Ns}) + g\,((1-a_1)\,Q_{Nf} + (1-a_1a_2)\,Q_{Ns} + Q_{Nx}))$$

Since $Pi(C_1)$ has already been derived, it is now possible to predict $Pr(C_1W_2)$ and $Pr(W_1W_2)$ by

$$Pi\,(C_1W_2) = Pi\,(C_1) - Pi\,(C_1C_2)$$
$$Pi\,(W_1W_2) = 1 - Pi(C_1) - Pi\,(W_1C_2)$$

and the two conditional probabilities by

$$Pi(C_2/C_1) = Pi\,(C_1C_2) \,/\, Pi(C_1)\,.$$
$$Pi\,(C_2/W_1) = Pi\,(W_1C_2)/(1-Pi(C_1)).$$

6.43 Fits to the conditional performance scores

It will be recalled that the fit of Model 2 to $Pr(C_1)$ yielded an extremely unsatisfactory minimum χ^2, and that this was interpreted as a consequence of the unsystematic variation in the observed values of $Pr(C_1)$ at retention intervals between 2 and 6 trials (see 6.33). Consequently, it was thought that an attempt to fit the various versions to all the data simultaneously (i.e. to the four joint proportions

$Pr(W_1C_2)$, $Pr(C_1C_2)$, $Pr(C_1W_1)$, and $Pr(W_1W_2)$) would not yield a satisfactory minimum χ^2 in any case, due to this unsytematic variation.

However, it was noticed that all four versions of the model outlined at the end of section 6.41 were identical insofar as their predictions concerning $Pr(C_1)$. In fact, $Pi(C_1)$ depends only on the parameters c, a_o, p and s. It was therefore decided to try an approach which involved fitting the model to $Pr(C_1)$ to yield estimates of these four parameters; these values were then carried forward, and the remaining parameters of the model were estimated by fitting the model simultaneously to $Pr(C_2/C_1)$ and $Pr(C_2/W_1)$.

Of course, in each case, the fit to $Pr(C_1)$ yielded identical results to those already described in section 6.33 for the simpler version of Model 2. The observed minimum Chi-squared was $\chi_1^2 = 21.067$ on 6 degrees of freedom, and the resultant parameter estimates were c = .872, a_o = .410, p = .996 and s = .302 (cf Table 12). In the further fitting procedure to be described below, these four parameters were constrained to take these estimated values.

In order to provide a clear description of the subsequent fitting procedure, it will be necessary to introduce some new notation. Let Ni be the total number of observations in Exp.I at an interpresentation interval of i trials, and let $ni(W_1C_2)$, $ni(C_1C_2)$, $ni(W_1W_2)$ and $ni(C_1W_2)$ be the corresponding numbers of observations falling into the four response categories.

Then, $ni(C_1) = ni(C_1C_2) + ni(C_1W_2)$

and $ni(W_1) = ni(W_1C_2) + ni(W_1W_2)$

are merely the observed numbers of items that were correct and wrong respectively on T_1 at a P_1-T_1 lag of i.

The procedure basically took the form of minimising

$$\chi_t^2 = \chi_2^2 + \chi_3^2$$

where

$$\chi_2^2 = \sum_i \frac{(ni(C_1)\,Pi(C_2/C_1) - ni(C_1C_2))^2}{ni(C_1)\,Pi(C_2/C_1)} + \frac{(ni(C_1)(1-Pi(C_2/C_1)) - ni(C_1W_2))^2}{ni(C_1)(1-Pi(C_2/C_1))}$$

and

$$\chi_3^2 = \sum_i \frac{(ni(W_1)Pi(C_2/W_1)-ni(W_1C_2))^2}{ni(W_1)Pi(C_2/W_1)} + \frac{(ni(W_1)(1-Pi(C_2/W_1))-ni(W_1W_2))^2}{ni(W_1)(1-Pi(C_2/W_1))}$$

Consequently, it can be seen that χ_t^2 was a measure of the goodness-of-fit of the model to the two conditional probabilities $Pr(C_2/C_1)$ and $Pr(C_2/W_1)$ which did not depend upon the fit of the model to $Pr(C_1)$, since χ_2^2 is calculated by restricting attention only to those observations that were correct on T_1, and χ_3^2 similarly restricts attention to those observations that included a T_1 error. In other words, χ_t^2 would not be inflated by the relatively poor fit of the model to $Pr(C_1)$, except to the extent that it was based on a procedure that fixed the values of c, a_o, p and s to their best estimates from a consideration of $Pr(C_1)$. It therefore appeared likely that this procedure would produce a reasonable fit to the conditional performance scores, and would therefore constitute an acceptable method of discriminating between the four versions of the model outlined at the end of section 6.41.

The minimum χ_t^2 that results from such a method involves fitting the model to a data array with 20 degrees of freedom since i takes 10 separate values. Thus, for any particular version of the model, its degrees of freedom will be 20 less the number of parameters allowed to vary freely (and it should be remembered that in computing χ_t^2, the probabilities c, a_o, p and s do not constitute effective parameters of the model). Although it is possible to examine the values of the two components of χ_t^2, that is, χ_2^2 and χ_3^2, these are unfortunately not distributed as a Chi-squared; for example, it is not possible to deduce their degrees of freedom, since, say, the free parameter a_1 contributes both to the value of χ_2^2 and to that of χ_3^2.

Finally, an overall value of Chi-squared for each version of the model was computed upon completion of the two fitting procedures, which measured the fit of the model to the four joint proportions $Pr(W_1C_2)$, $Pr(C_1C_2)$, $Pr(W_1W_2)$ and $Pr(C_1W_2)$, by use of the formula:

TABLE 13

Results of fitting Model 2 to conditional performance data from Exp. I.

(i) Parameter estimates and goodness-of-fit

Estimated (Constrained) Parameters

Version	a_1	a_2	q	a_3	r
I	.265	(1.000)	(.000)	(.000)	(1.000)
II	.265	(1.000)	(.000)	.000	1.000
III	.265	1.000	.000	(.000)	(1.000)
IV	.265	1.000	.000	.000	1.000

Version	χ^2_2	χ^2_3	χ^2_t	dft	χ^2_0	dfo
I	110.094	17.945	128.039	19	146.564	25
II	110.094	17.945	128.039	17	146.564	23
III	110.094	17.945	128.039	17	146.564	23
IV	110.094	17.945	128.039	15	146.564	21

(ii) Observed and predicted values.

	$Pr(C_2/C_1)$		$Pr(C_2/W_1)$	
i	obs	Pred	Obs	Pred
0	.491	.465	.069	.067
1	.641	.701	.357	.306
2	.758	.911	.330	.306
3	.823	.911	.315	.306
4	.815	.910	.304	.306
5	.875	.910	.246	.306
6	.864	.910	.270	.306
8	.896	.909	.361	.306
12	.885	.907	.322	.306
16	.847	.906	.331	.306

$$\chi^2_0 = \sum_i \frac{(Ni.Pi(W_1C_1)-ni(W_1C_1))^2}{Ni\ Pi(W_1C_1)} + \ldots + \frac{(Ni.Pi(W_1W_2)-ni(W_1W_2))^2}{Ni\ Pi\ (W_1W_2)}$$

For each version of the model, the degrees of freedom were simply 30 less the total number of fitted parameters (including c, a_o, p and s) of the model.

Details of the fitting procedure are presented in Table 13. It is clear from the table that this particular approach was a failure. Not only was the minimum value of χ^2_t for Version I highly significant beyond all conventional values, but the freeing of the parameters

TABLE 13

Results of fitting Model 2 to conditional performance data from Exp. I.

(i) Parameter estimates and goodness-of-fit

Estimated (Constrained) Parameters

Version	a_1	a_2	q	a_3	r
I	.265	(1.000)	(.000)	(.000)	(1.000)
II	.265	(1.000)	(.000)	.000	1.000
III	.265	1.000	.000	(.000)	(1.000)
IV	.265	1.000	.000	.000	1.000

Version	χ_2^2	χ_3^2	χ_t^2	dft	χ_0^2	dfo
I	110.094	17.945	128.039	19	146.564	25
II	110.094	17.945	128.039	17	146.564	23
III	110.094	17.945	128.039	17	146.564	23
IV	110.094	17.945	128.039	15	146.564	21

(ii) Observed and predicted values.

	$Pr(C_2/C_1)$		$Pr(C_2/W_1)$	
i	obs	Pred	Obs	Pred
0	.491	.465	.069	.067
1	.641	.701	.357	.306
2	.758	.911	.330	.306
3	.823	.911	.315	.306
4	.815	.910	.304	.306
5	.875	.910	.246	.306
6	.864	.910	.270	.306
8	.896	.909	.361	.306
12	.885	.907	.322	.306
16	.847	.906	.331	.306

$$\chi_0^2 = \sum_i \frac{(Ni.Pi(W_1C_1)-ni(W_1C_1))^2}{Ni\ Pi(W_1C_1)} + \ldots + \frac{(Ni.Pi(W_1W_2)-ni(W_1W_2))^2}{Ni\ Pi\ (W_1W_2)}$$

For each version of the model, the degrees of freedom were simply 30 less the total number of fitted parameters (including c, a_0, p and s) of the model.

Details of the fitting procedure are presented in Table 13. It is clear from the table that this particular approach was a failure. Not only was the minimum value of χ_t^2 for Version I highly significant beyond all conventional values, but the freeing of the parameters defining the additional spacing processes resulted in absolutely no overall improvement in fit. Some idea of the cause of these disappointing results may be gleaned from an inspection of the predictions of the model. Clearly, when $q = 0$ and $a_2 = 1$, the model predicts that $Pr(C_2/W_1)$ will be a constant for non-zero values of i, but the minimum Chi-squared prediction of .306 appears somewhat low (the pooled value of $Pr(C_2/W_1)$ across all non zero value of i was .320). It is also apparent that $Pr(C_2/C_1)$ is consistently overpredicted at all non-zero values of i. Thus, this procedure has resulted in the lowest possible predicted value of $Pi(C_2/W_1)$ consistent with the data, since $Pi(C_2/C_1)$ involves a component equal to $Pi(C_2/W_1)$ which results from those items that were correctly guessed on T_1.

The consistent overprediction of $Pr(C_2/C_1)$, even with $a_3=0$, suggests that either the model is in error (so that adequately encoded items at P_2 should possess a non-zero probability of being forgotten on P_2, which is patently absured) or as is far more likel , the initial fit to $Pr(C_1)$ resulted in an overestimate of the parameter p. It may be that a lower value of p coupled with a somewhat higher value of a_0 would result in almost as good a fit to $Pr(C_1)$, and in a far superior fit to the conditional probabilities.

6.44 Simultaneous fits to all the data

The results discussed in the preceding section suggest that there is no alternative but to accept the inflation of the overall goodness of fit resulting from the unexplained variation of $Pr(C_1)$, and to proceed by fitting each version of Model 2 simultaneously to the four observed proportions $Pr(W_1C_2)$, $Pr(C_1C_2)$, $Pr(W_1W_2)$, $Pr(C_1W_2)$. Consequently, all the unconstrained parameters in each version of the model were estimated simultaneously by a process of minimising the overall Chi-squared, χ^2_0. Parameter estimates and goodnes-of-fit statistics resulting from this procedure are presented in Table. 14.

It is apparent from the table that Versions I and III of the model, and Versions II and IV resulted in identical fits, and indeed

TABLE 14

Results of the fits of Model 2 to the joint data of Exp. 1.

Estimated (Constrained) parameters

Version	c	a_0	a_1	p	S	a_2	q	a_3	r
I	.860	.438	.306	.986	.312	(1.000)	(.000)	(.000)	(1.000)
II	.856	.449	.316	.982	.297	(1.000)	(.000)	1.000	.953
III	.860	.438	.306	.986	.312	1.000	.000	(.000)	(1.000)
IV	.856	.449	.316	.982	.297	1.000	.000	1.000	.953

Version	χ^2_1	χ^2_2	χ^2_3	χ^2_0	df	prob
I	25.253	25.500	15.600	65.520	25	<.001
II	28.881	17.890	15.600	61.777	23	<.001
III	25.253	25.500	15.600	65.520	23	<.001
IV	28.881	17.890	15.600	61.777	21	<.001

it can be seen that even when allowed to vary freely, the parameters a_2 and q were estimated to be 1 and 0 respectively. Thus, the inclusion of the mechanism representing H_3 produced absolutely no improvement in the fit. However, an improvement in the fit of the model was observed when a_3 and r were allowed to vary freely. Hence, the inclusion of the mechanism whereby long-retention items may have their encoding still further improved by a P_2 occurring after a sufficient interpresentation interval resulted in a reduction in the observed value of the minimum χ^2_0 from 65.520 to 61.777. Thus, applying the result described in section 6.22, the difference between these two values (3.743) is distributed as a Chi-squared with 2 degrees of freedom. This does not represent a significant improvement in fit ($p > .1$).

Again, these results are disappointing since neither additional mechanism appears to account for the apparent continued improvement of T_2 performance with interpresentation spacing. Values of χ^2_1, χ^2_2 and χ^2_3 were computed as before, and although these values cannot be meaningfully tested, they do yield some useful information. A comparison of versions I and II of the model suggests that the inclusion of the mechanism representing H_2 substantially improved the fit of $Pi(C_2/C_1)$ (see χ^2_2), did not affect the fit of $Pi(C_2/W_1)$ (see χ^2_3), but resulted in a worsening of the fit of $Pi(C_1)$ (see χ^2_1). Taken in conjunction with the not unexpectedly highly significant value of the minimum overall Chi-squareds, these results suggest that these straightforward fits of the model may be unduly influenced by the noise present in the $Pr(C_1)$ data, in that there clearly appears to be some "trade off" between the fit of the model to $Pr(C_1)$, and its fit to the conditional probabilities. The predictions arising from these two fits are not particularly interesting in themselves, especially in the light of the results to be discussed below, but for completeness, they are presented in Appendix 4.

Although the attempt made to "get around" the noisiness of $Pr(C_1)$ described in section 6.43 was unsuccessful, it is clear that some attempt must be made to solve this problem. A glance at the equation for $Pi(C_1)$ on page 249 will confirm that for values of the retention interval i of 2 or more trials,

$$Pi(C_1) = a_o p^i + (1-a_o p^i)g$$

so that

$$a_o = (Pi(C_1)-g)/(1-g)p^i$$

It was proposed to ensure the fit of the model to $Pr(C_1)$ by replacing the single parameter a_o by the 10-vector $(A_o(i))$, where $A_o(0) = A_o(1)=a_o$, and

$$A_o(i) = (Pr(C_1)-g)/(1-g)p^i \text{ for } i \geqslant 2.$$

This would ensure a perfect fit of the model to $Pr(C_1)$ at all but the first two points (when i = 0 and i = 1), and would also go some way towards explaining the unsystematic variation in $Pr(C_1)$ as reflecting different average P_1 encoding probabilities for the items that were assigned to the various spacing conditions. A fit to the four joint probabilities under this scheme would involve the estimation of an additional 8 parameters (i.e. the $A_0(i)$ for $i \geq 2$), although under the proposed scheme these would not be minimum Chi-squared estimates. However, it was thought that the additional computing time necessary to fit the four versions of the modified model (with between 13 and 17 free parameters) would prove prohibitive, and in all probability would result only in a marginal improvement in overall fit.

It is not proposed to describe in detail the modifications to the equations giving the predictions of Model 2 with the additional eight parameters, since these are both obvious and straightforward. Suffice it to say that χ_0^2 was minimised under all four versions of the modified model (which will henceforth be called Model 2a) and the results are summarized in Table 15. It should be immediately pointed out that

TABLE 15

Results of fitting Model 2a to the joint data of Exp. I.

Estimated (constrained) parameters.

Version	C	a_o	a_1	p	s	a_2	q	a_3	r
I	.859	.442	.308	.985	.307	(1.000)	(.000)	(.000)	(1.000)
II	.835	.517	.369	.963	.220	(1.000)	(.000)	.761	.761

Version	Ao(2)	Ao(3)	Ao(4)	Ao(5)	Ao(6)	Ao(8)	Ao(12)	Ao(16)
I	.394	.349	.470	.429	.521	.453	.484	.446
II	.412	.373	.514	.479	.595	.542	.632	.638

Version	χ_1^2	χ_2^2	χ_3^2	χ_0^2	df	prob
I	.580	20.837	15.598	37.038	17	<.01
II	.911	9.263	15.598	25.789	15	<.05

the fit of version III was identical to that of version I, and the fit of version IV to that of version II, and in both cases q was freely estimated to be zero. Thus, once again, the addition of the mechanism representing the hypothesis H_3 produced no improvement in fit whatsoever.

The fit of version I of Model 2a yielded a reduction in the minimum value of χ_0^2 of 28.482 in comparison with version I of

Model 2, which on 8 degrees of freedom represents a significant improvement ($p < .01$). Thus, the freeing of the eight parameters Ao(2),..., Ao(16) from their constrained value of a_0 in Model 2 yielded a significant improvement in fit. The mean value of these eight parameters was found to be .443, which is very near the estimated value of a_0 (.442); this result is consistent with the hypothesis that the unsystematic variation in $Pr(C_1)$ was primarily due to sampling differences in the way that long-term encodable material was assigned to condition. A comparison of the respective values of χ^2_2 from the fits of version I of Model 2 and version I of Model 2a (see Tables 14 and 15) suggests that freeing $A_0(2),\ldots,A_0(16)$ did not greatly improve the fit of version I to $Pr(C_2/C_1)$, whilst a comparison of the corresponding values of χ^2_3 reveals hardly any difference in the respective fits to $Pr(C_2/W_1)$. In other words, the bulk of the observed improvement in the fit of version I of Model 2a over that of the same version of Model 2 appears to have resulted from the enormous improvement in fit to $Pr(C_1)$.

Similar comparisons of version II of Model 2a with version II of Model 2 reveal that freeing the eight parameters Ao(2),..., Ao(16) from their constrained value of a_0 yielded a significant improvement in χ^2_0 ($\chi^2 = 35.988$, df = 8, $p < .01$), and the mean value of the eight free parameters $A_0(2),\ldots, A_0(16)$ was found to be .523, which again did not differ significantly from the estimated value of a_0 (.517). This latter finding is also consistent with the hypothesis that the unsystematic variation observed in the relationship between $Pr(C_1)$ and the retention interval i was primarily due to differences in the sampling probabilities that a long-term encodable item would be assigned to any particular condition. A comparison of the two values of χ^2_2 suggests that in the case of version II, the freeing of $A_0(2),\ldots,A_0(16)$ contributed substantially to the improvement of the fit of the model to $Pr(C_2/C_1)$, whilst the two values of χ^2_3 again hardly differed. Thus, the freeing of $A_0(2),\ldots, A_0(16)$ in version II of the Model 2a served to improve the fits of the model both to $Pr(C_1)$ and to $Pr(C_2/C_1)$.

Comparisons will now be made within Model 2a, between version I and version II. It is apparent from an examination of Table 15 that both minimum overall Chi-squared values were significant, so that in general it may be said that neither version of Model 2a provides a completely satisfactory explanation of the data of Exp I. However, this result is not too disheartening, since an examination of χ^2_1, χ^2_2 and χ^2_3 suggests that the basic inadequacy of the model lies in its continued failure to provide a good fit to $Pr(C_2/W_1)$, since χ^2_3 has remained largely unchanged over all versions of the model so far examined. This point will be discussed more fully a little later. Although version II of Model 2a appears to predict $Pr(C_2)$ a little less well than version I, as a comparison of the respective values of χ^2_1 reveals, version II clearly predicts $Pr(C_2/C_1)$ far better than version I (cf χ^2_2). Consequently, the improvement in overall fit observed with version II which is reflected in a reduction in χ^2_0 of 11.249, clearly results from a superior fit to $Pr(C_2/C_1)$. With two degrees of freedom, the observed improvement in fit is highly significant ($p < .01$). Since $Pi(C_2/C_1)$ is affected only slightly by $Pi(C_2/W_1)$ (to the extent that, for the relatively small proportion of items correct at T_1 that are correctly guessed, $Pi(C_2/C_1)$ will be equal to $Pi(C_2/W_1)$) these results may be taken as fairly reliable evidence that version II of Model 2a provides a far superior account of $Pr(C_2/C_1)$ than does version I. Versions III and IV, which include the alternative spacing mechanism, do not provide the slightest improvement in the accounts of the spacing effect represented respectively by versions I and II.

Data from Exp. I together with the corresponding values predicted by versions I and II of Model 2a are presented in Table 16. The data are presented in terms of $Pr(C_1)$, $Pr(C_2/C_1)$ and $Pr(C_2/W_1)$ rather than the four joint proportions $Pr(W_1C_2), \ldots, Pr(C_1, W_2)$ that were actually employed in the minimum Chi-squared procedure, since both the observed and predicted values of the joint proportions may be recovered from the components presented. Furthermore, it is felt that this particular way

TABLE 16

Data from Exp I and the corresponding values predicted by Versions I and II of Model 2a

	$Pr(C_1)$			$Pr(C_2)$		
i	obs	Pred I	Pred II	Obs	Pred I	Pred II
0	.930	.926	.926	.461	.433	.425
1	.595	.611	.615	.525	.533	.539
2	.424	.424	.424	.540	.541	.519
3	.378	.378	.378	.507	.513	.502
4	.480	.480	.480	.549	.575	.569
5	.438	.438	.438	.521	.549	.550
6	.511	.511	.511	.573	.594	.600
8	.442	.442	.442	.597	.552	.563
12	.443	.443	.443	.571	.552	.569
16	.394	.394	.394	.534	.552	.538

	$Pr(C_2/C_1)$			$Pr(C_2/W_1)$		
i	Obs	Pred I	Pred II	Obs	Pred I	Pred II
0	.491	.462	.454	.069	.067	.067
1	.641	.669	.675	.357	.322	.322
2	.758	.838	.786	.380	.322	.322
3	.823	.826	.798	.315	.322	.322
4	.815	.849	.837	.304	.322	.322
5	.875	.841	.843	.246	.322	.322
6	.864	.855	.867	.270	.322	.322
8	.896	.842	.867	.361	.322	.322
12	.885	.842	.880	.322	.322	.322
16	.847	.831	.872	.331	.322	.322

of presenting the results is more meaningful, in that it provides a clearer intuitive picture of the relationship between the data and the various theories. In addition, the observed and predicted values of $Pr(C_2)$ are presented, since this is the conventional way of depicting the spacing effect.

The values of $Pr(C_2/W_1)$ predicted by both versions of Model 2a are identical, and follow a pattern with interpresentation spacing that is very much as expected given that in each version, q was fixed at 0 and a_2 at 1. In other words, the conditional probability is predicted at chance level for i = 0, and thereafter is predicted to take a constant value (.322). The failure of versions III and IV to produce any improvement in fit strongly suggests that this formulation of the relationship of $Pr(C_2/W_1)$ with interpresentation interval is superior to a hypothesis that the statistic shows a gradual improvement with spacings in excess of zero. It will be recalled that in the formulation of Model 2a, some misgivings were expressed in regard to the assumptions regarding the transition probability on an F-trial from Nx to Nf (see 6.41). In particular, it was thought that the specific formulation adopted might lead to an underprediction of $Pr(C_2/W_1)$ at short non-zero interpresentation intervals; this might consequently have lead to an underestimation of the parameter q.

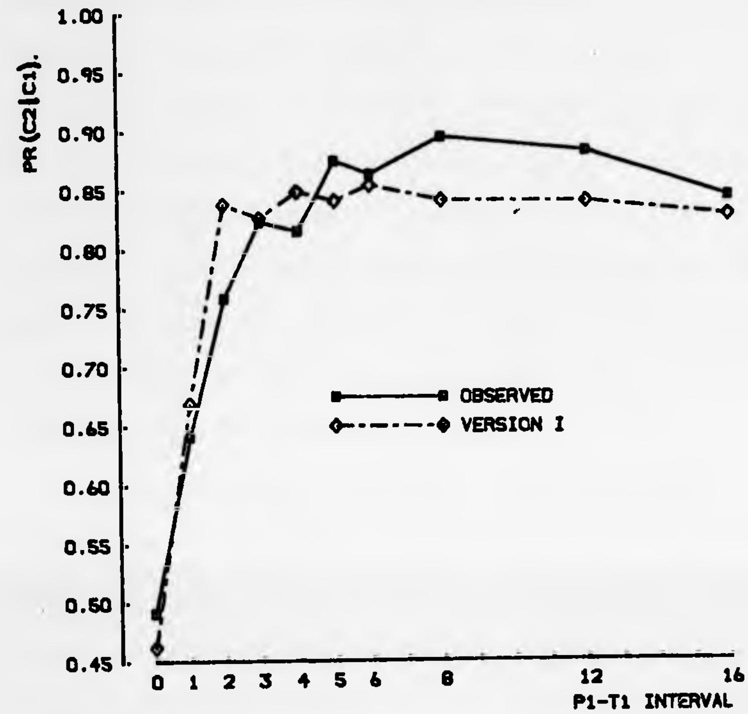
This objection was checked by fitting the four versions of a slightly modified form of the model; in this case, the probability of an F-trial transition from Nx to Nf was set to unity, so that even with non zero values of q, $Pr(C_2/W_1)$ would be predicted to increase sharply from chance at an interpresentation interval of zero, and would thereafter show a far more gradual improvement with interpresentation spacing than would be predicted by Model 2a. Of course, versions I and II of this modified form would be identical to the corresponding versions of Model 2a. It was found, however, that in fitting both versions III and IV of the modified model, q was again freely estimated to be zero. Thus, even with a mechanism representing a far less pronounced rate of improvement in $Pr(C_2/W_1)$ with interpresentation spacing, no improvement in fit was found over versions I and II of Model 2a. This result is conclusive, and strongly suggests that the observed variation of $Pr(C_2/W_1)$ with non-zero interpresentation interval found in Exp.I (see 5.12) could not be accounted for in terms of a gradual increase across non-zero spacings.

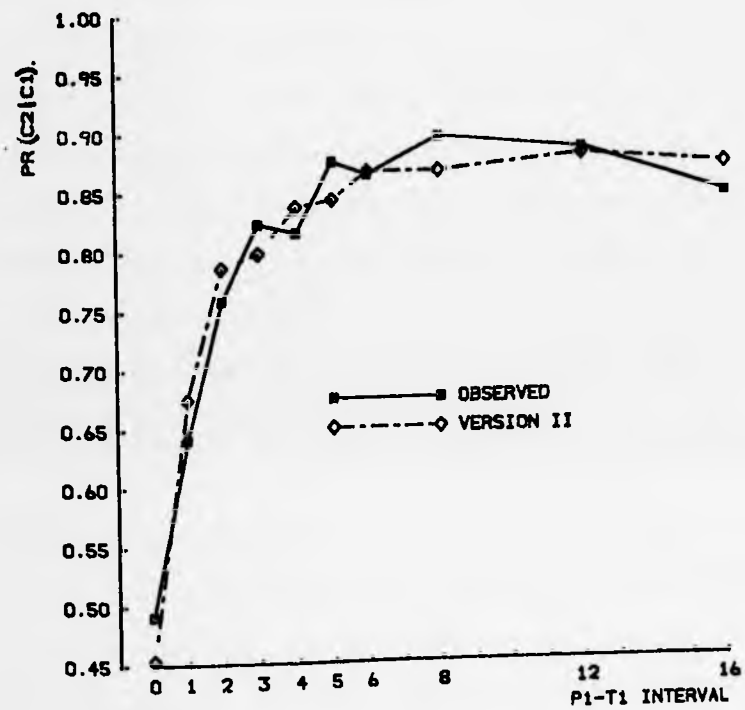
Consequently, the failure of Model 2a to satisfactorily predict $Pr(C_2/W_1)$ is not regarded as very important, in that this statistic clearly does not contribute in any meaningful way to the spacing effect, and after all, the main objective of these analyses was precisely to investigate the effects of interpresentation spacing on T_2 recall. It is suggested that the variation in the values of $Pr(C_2/W_1)$ may be due to differences in the sampling probabilities that an item to which an error was made on T_1 in each particular spacing condition was an item that would have been recalled at a retention interval of zero (and was therefore encodable on P_2) as opposed to an item of the type that would lead to a T_1 error at a retention interval of zero. It is quite possible that these latter items are always difficult to encode adequately on P_2. Although it is, in theory, quite possible to construct a model to this effect, such an exercise appears hardly worthwhile, since it would very probably involve the postulation of additional parameters, and in any case would contribute very little to our understanding of the spacing effect. In other words, even were a model constructed which was able to adequately explain the observed variation in $Pr(C_2/W_1)$, it is certain that the predicted variation would not be an increasing function of the interpresentation spacing. Moreover, such a model would still predict the spacing effect in terms a mechanism involving of long-retention items, although its predictions of $Pr(C_2/C_1)$ would differ slightly from those of the current model to the extent that a relatively small proportion of items that were correct on T_1 would be guesses, and would produce T_2 recall equal to that of items that were errors on T_1 at the same retention interval. However, these small variations would be essentially "swamped" by the systematic predicted improvement in T_2 recall with interpresentation spacing for those items that were correct on T_1 by reason of their having been recalled from memory proper, as opposed to having been correctly guessed.

It has already been pointed out that the improvement in overall fit observed when fitting version II of Model 2a stems entirely from its improved fit to $Pr(C_2/C_1)$. The values presented in Table 16 are

FIGURE 23.

Observed values of $Pr(C_2/C_1)$ from Exp. 1. and the corr sponding values predicted by versions I and II of Model 2a.


PR(C2|C1).
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0 1 2 3 4 5 6 8 12 16
P1-T1 INTERVAL
OBSERVED
VERSION I


PR(C2|C1).
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0 1 2 3 4 5 6 8 12 16
P1-T1 INTERVAL
OBSERVED
VERSION II

presented in graphical form in Figure 23, and an examination of the figures only serves to emphasise this observation; the fit of version II to $Pr(C_2/C_1)$ is clearly superior to that of version I. Thus, at first sight, it appears reasonable to conclude that the mechanism represented by version II (which accounts for a continued improvement in $Pr(C_2/C_1)$ over interpresentation intervals in excess of 2 trials in terms of a process whereby a second presentation becomes increasingly effective in improving the resistance to decay of already adequately encoded items) gives a perfectly satisfactory account of the relationship of $Pr(C_2/C_1)$ with interpresentation spacing.

However, it should be pointed out that in fitting Model 2a, there is room for considerable variation in the long-term retention parameter p. This is because the model was fitted in such a way as to ensure a perfect fit to $Pr(C_1)$ at intervals or 2 or more trials. Thus, p could take almost any value, provided that it was large enough to ensure that the estimated values of $A_o(i)$ did not exceed unity, and still produce an excellent fit to $Pr(C_1)$. In addition, identical values of $Pi(C_2/W_1)$ could result with any value of p that was sufficiently large to allow a_1 to assume a value of less than 1 in compensating for the new value of p in order to produce identical predictions of $Pr(C_2/W_1)$ to those involving the current value of p. Thus, in fitting Model 2a, the value of p is estimated almost entirely in terms of the fit of the model to $Pr(C_2/C_1)$. Furthermore, the value of p estimated by version II of Model 2a (.963) is the lowest value observed of such an estimate.

Thus, if the long-term retention parameter (and hence the long term forgetting rate following a first encoding) can take any value within a fairly broad range without affecting $Pi(C_1)$ and $Pi(C_2/W_1)$, then how can one state with any certainty that a re-presentation serves to reduce this forgetting rate? For example, the additional mechanism represented by version II could equally well be interpreted as serving to reduce the long-term retention parameter from 1 to .963 on a re-presentation at short-lag, whilst maintaining it at 1 on a re-presentation at long lag (since an initial long-term retention parameter of 1 could still give identical predictions of $Pr(C_2/W_1)$ if a_1 were estimated to be lower, and identical predictions of $Pr(C_1)$ if a_0 and $Ao(i)$ were estimated to be smaller).

There is clearly no way of resolving this issue from a consideration of the model; however, the alternative interpretation of the spacing effect described above may be rejected on psychological grounds. At first sight, it appears patently absured to suggest that an additional presentation of an item that is currently recallable from memory could in any way operate to increase the subsequent decay rate of that item. However, it is just possible that subjects for some reason believed that, following a second presentation at short lag, an item would not be subsequently tested, and they may therefore have treated the second presentation as an implicit "forget" instruction (see 1.33). Nevertheless, there appears to be no convincing explanation as to why this should be the case, let along why such an effect would be more likely at shorter interpresentation spacings. The fact that each subject in Exp I. was tested on 600 items suggests that even were such a misapprehension concerning the subsequent testing of a re-presented item initially present, then there would be ample opportunity for subjects to realise the falseness of such a supposition.

Consequently, it appears safe to conclude that the superior fit of version II of Model 2a really does reflect a continued improvement with interpresentation spacing in the effectiveness of a second presentation in reducing the subsequent forgetting rate of those items that are recallable from memory at the time of the re-presentation. Therefore, the fits of Model 2a have satisfactorily fulfilled the main objectives of a theoretical analysis, in that they have successfully discriminated between the various spacing hypotheses.

6.5. Discussion

The results of the theoretcal analyses described above may be summarised as follows:

(i) The observed variation of $Pr(C_1)$ with retention intervals of 2 or more trials was well accounted for by a sampling hypothesis which states that the proportions of items that could be relatively easily memorised in a single presentation varied between the spacing conditions. This hypothesis was relevant because the material assigned to each spacing condition was identical for each subject in Exp. I.

(ii) There was no evidence whatsoever that $Pr(C_2/W_1)$ increased with interpresentation spacings in excess of zero. Thus, the apparent improvement in T_2 performance with interpresentation spacings in excess of zero was not found to result from an enhancement with spacing of the effectiveness of a second presentation on items that could not be recalled at the time of the second presentation.

(iii) Consequently, it can be concluded that the spacing effect results from an improvement in T_2 performance with interpresentation spacing on those items that could be correctly recalled from memory at the time of the second presentation, i.e. on T_1. The observed relationship of $Pr(C_2/C_1)$ with interpresentation spacing was partially accounted for by the hypothesis that items that can only be recalled from a short-term retention state at the time of their second presentation receive no benefit from their second presentation. A significantly superior account was provided by the additional hypothesis that with increasing interpresentation intervals, a second presentation may be increasingly effective in reducing the subsequent mean forgetting rates of items that can be recalled from memory at the time of the second presentation.

A discussion of the psychological implications of these results is reserved until the next chapter. However, it does appear on the whole that Model 2a has proved successful in establishing that the spacing effect operates over interpresentation intervals in excess of those that will be sufficient to "wipe out" short-term retention effects on the second presentation, that furthermore the spacing effect operates only on those items that can be correctly recalled at re-presentation, and that finally, the spacing effect operates via a reduction in the subsequent forgetting rate of such items on their second presentation.

It should be pointed out that the models devloped in this chapter are not intended to provde a full account of all the spacing results outlined in section 1.44. Instead, they were intended to produce a clarification of the results of Exp. I with reference to the specific hypotheses outlined at the conclusion of Chapter Five. In this, they have been successful.

However, it is doubtful if Model 2 could account for the interaction between the interpresentation and retention interval described by Peterson, Hillner and Saltzman (1962; see Fig. 9), or for the commutativity of two interpresentation intervals in a three-presentation schedule (Bjork and Abramowitz, 1968; See 1.44). This is due to the fact that the model was principally designed in order to answer specific questions concerning the results of Exp. I. In particular, the retention interval between P_2 and T_2 was always 8 trials in this study, and would therefore be sufficiently long to "wipe out" short-term retention effects at T_2. Consequently, the assumptions made by the model concerning the effect of a P_2 were fairly simple, since it was only concerned with whether an item left P_2 in a long-retention state or not. In order to predict the results of Peterson et al. a Markovian model might well need to include some kind of enhanced short-term state wherein short-term recall might be possible over, say, four trials or so. The introduction of a third presentation into the presentation-test schedule might well provide the model with additional information upon which to base an estimation of the LU retention parameter. In the absence of such information, this was set to unity.

However, these comments are somewhat o f the point. Essentially, the Markovian models employed were intended to provide a description of the data, no a complete explanation, and although their basic structure of long- and short-retention states and naive states may be justified by pointing to the "obvio s" parallels wit the extant data, they have obvious s ortcomings when taken to represent, in anything but the most superficial way, underlying psychological mechanisms. For example, the fundamental Markovian assumptions of constant transition probabilities will probably fail to stand up to detailed scrutiny. Although the dichotomy into short- and long-retention states may stand up at a superficial level, it is extremely likely that some items categorised as long-retention items are easier to remember than others. Thus, when data is pooled over many subjects, many sessions, or even many subject-sessions, it is extremely likely that items of varied "retention probabilities" will all be placed by the model in, say, the long-term state. Consequently, the long-term retention probability estimated by the model will be too high at short retention intervals and too low at long ones. Thus, at best, Markovian models can only provide an adequate description of the data. Fortunately, in this case, this is a l that was required.

Finally, it should be re-emphasised that the application of more complex, psychologically-based models was neither warranted by current theory (See 6.1), nor by the current data, which was clearly too "noisy" to give much h pe of a meaningful analysi . Fortunately, it proved possible to account for a great deal of the noise, but it should also be pointed out that in addition to the faulty design of Exp I, which meant that the allocation of material to conditions was the same for all subjects, there is also a strong possibility that the trials which constituted the "forgetting" or intruding trials in each condition differed somewhat in composition. For example, it is highly probable that intruding presentation and test trials on other items have different effects on retention (see 1.34), and with the limited number of lists employed in Exp I., it is possible that the overall ratio of intruding presentations to intruding tests differed significantly from one spacing condition to another. However, it is also possible that even with the best possible design, these ratios may differ significantly, since the list order produced by the "random" interleaving procedure may well be determined to some extent by the precise presentation-test schedules that are being employed. Thus, it is not certain that the interleaving procedure really is as random as all that. In addition, it may not just be intruding presentation and test trials that differ in their effect on the retention of a given item; for example, second presentations may differ in effect from first presentations, and the interfering effect of a second trial on an item may depend on the interval between it and the item's first trial, or on the difficulty of the subject's task on that trial.

These points emphasise not only the specific effects that may have contributed to the data in the present study; many of these problems may occur in virtually any memory experiment. The inference to be drawn is that sophisticated "psychological" models may be just as misleading as the somewhat more naive Markovian models when fitted to virtually any set of memory data. Consequently, it is likely that the present analyses are no more and no less reliable than analyses based upon mathematical models in general.

CHAPTER SEVEN

CONCLUSIONS

The results which have emerged from this study broadly speaking fall into two categories; those which are specifically concerned with the spacing effect in paired-associate memory, and those which have implications concerning paired-associate memorising in general. It is proposed to deal firstly with the latter category.

One point to emerge from the results of the experiments which involved stimulus recognition testing (i.e. Expts. II and III) was that it generally appeared that stimulus recognition performance alone could not account for all, or even for the greater proportion of, paired-associate forgetting. Certainly it was found in Exp. II that response recall was no better than chance following a stimulus recognition failure on the same test trial, in agreement with the previous studies reported in section 4.13. A slightly different result was found in Exp. III; although response recall performance following a stimulus recognition failure on the same test trial was far poorer than that following a correct stimulus recognition, it was significantly above chance level. However, recall performance following a recognition error was not affected by retention interval, nor by the number of presentations, so that it was concluded that some kind of guessing strategy might have served to inflate the level of recall performance above the theoretical chance level, since a "true" memory process would be expected to be affected by such factors.

Therefore, it was found that some recall errors could be explained in terms of stimulus forgetting. However, in Exp II, it was also found that stimulus recognition was at such a high level that the overall recall data were almost identical to the recall data conditional upon correct stimulus recognition on the same test trial. Although a great proportion of recall errors in this study clearly resulted from a failure to produce an adequate association encoding in the first place, it was found that the proportion of recall errors given a correct stimulus recognition on T_1 exhibited the usual relationship with retention interval; a short-term component of performance followed by a more gradual decline in performance with retention interval (see 5.22). In other words, even after short-term effects had been "wiped out", response recall performance appeared to decline more rapidly with retention interval than did stimulus recognition.

Similar results were found in Exp.III (see 5.32 and Figure 21). However, in this case, the results are less convincing, since the relatively high false recognition rate in this study (.235) implies that stimulus recognition performance at T_1 in fact declined more rapidly than an examination of $Pr(O_1)$ would suggest (see Figure 19). The data are further complicated by the fact that the theoretical chance probability of a correct recall response given a false stimulus recognition was .2 (since there were 5 possible responses). Therefore, it was decided to estimate, at each retention interval, the probability of a successful response retrieval (corrected for chance) at T_1 conditional upon a correct "true" stimulus recognition at T_1 (corrected for false recognitions). The resultant estimates were .986, .47[illegible], .326 .293 and .190 at retention intervals of 0,2,4,6 and 8 intruding trials respectively. Clearly, the corrected data are extremely striking, and strongly suggest that, even after short-term retention effects have been "wiped out", the rate of decay of response associations is greater than that of stimulus recognition codes. It has thus been established that there is considerable evidence to support the conclusion that a great deal of long-term paired-associate response forgetting cannot be explained in terms of long-term stimulus forgetting.

These observations are of obvious relevance to any theory of paired-associate encoding, and certainly suggest that the hypotheses outlined in sections 1.33 and 4.31 require some modification, in that they place too much emphasis on selective stimulus encoding. Consequently, the following formulation is proposed. The subject's task in encoding a paired associate is seen principally as his having to produce an association encoding that will be uniquely linked to the current stimulus and to no other. Thus, on the first presentation of a pair, it is postulated that the subject attempts to find some feature of the stimulus which will be unique to that stimulus, and to incorporate this feature in a response-evoking association code. In deciding which stimulus aspect or aspects are best used in the association encoding, the subject may well have access to previous encodings which have certain aspects in common with the current presentation episode. In particular, it is possible that the subject may recognize that certain semantic aspects of the current stimulus have been previously utilised in a different association encoding involving a different pair, and may therefore be motivated to seek alternative aspects of the current stimulus to use in the current association coding. These features may not be particularly easy to incorporate into an association code; in other words, a similar previous encoding may have a deleterious effect on the current encoding, and in this way, proactive inhibition effects may be explained. In addition, it is postulated that the subject only takes previous encodings into account if he recognises them as pertaining to other pairs currently to be remembered, and this must surely occur via the mediation of episodic aspects common to the presentation of the previously encoded pair and the current presentation episode. Furthermore, it is not necessary to assume that confusable previous encodings receive additional processing during the current presentation trial; indeed, it is quite probable that they do not. The postulated dependence of proactive inhibition effects on rapidly-decaying episodic cues is certainly consistent with the small size of the primacy effect in PA probe studies described in section 1.33.

Retroactive inhibition is seen as a consequence of the subject's failure to successfully take into account a confusable prior encoding in the way described above; consequently, subsequently presented confusable pairs will compete with the encoding of the current pair at recall, especially since they will probably share more episodic aspects with the test trial. However, when the retention interval becomes extremely long, these episodic cues are less likely to be present at test, so that the dominance of the more recently presented pair would tend to disappear. This is, of course, a widely observed phenomenon described by classical interference theory as the "spontaneous recovery" of the older association.

It has already been explained in some detail how this type of formulation handles the results of "cued forgetting" studies (in section 1.33). However, the current formulation differs from that presented earlier in that is postulates that stimulus aspects which are not selected for use in the association encoding may still be employed as stimulus recognition cues. In other words, stimulus recognition performance is not seen as being based only on those stimulus features that are employed in the association encoding. Indeed, it may be that the "non-associated" stimulus aspects are combined into an encoding which "points" to those aspects employed in the association coding. This hypothesis would certainly go some way towards explaining how subjects "focus in" on a preferred functional stimulus during on-going paired-associate learning (see 4.12). Finally, this hypothesis offers an explanation of how association codes decay more rapidly than stimulus recognition codes; sufficient stimulus features may survive to guarantee correct recognition, but these may not be sufficiently well-integrated into an encoding which "points" to the functional aspects of the stimulus required to evoke the association encoding, and hence the response. Consequently, an inappropriate stimulus aspect may be used to initiate the search process, and hence an inappropriate response may result. It should, perhaps, be stressed that although "non-associated" stimulus features may serve to point to the associated aspects in their presence, they are not seen as sufficiently well-integrated in general to support recall of the associated aspects in their absence; in other words, "non-preferred" stimulus features will not generally evoke the association encoding by themselves, so will not be sufficient for recall.

The theory outlined above admits several processes by which additional practice in the form of re-presentations of the pairs to be remembered may prove beneficial to subsequent response recall performance almost independently of stimulus recognition performance. In the first place, they may allow extra processing time during which the subject can somehow enhance or improve the encoding of the non-preferred stimulus features to "point" to the preferred functional encoding of the stimulus. Such a process may only marginally improve subsequent stimulus recognition performance. Secondly, further presentations may be effective in allowing the subject to enhance, deepen, elaborate or otherwise improve the association part of his encoding. In contrast, it is also possible that the subject may wish to change the entire encoding, especially if it is thought to be inadequate, or insufficiently unique, on a subsequent presentation. This might involve the selection of a new "preferred" functional stimulus version, the formation of a new "pointer" code, and the formation of a new association encoding. It certainly appears that in learning a finite PA list, the subject may well vary his "preferred" stimulus until he is satisfied that he has a unique formulation, following which the "pointer" code is refined, and only with overlearning are additional stimulus features employed to elaborate the association encoding (see 4.12). Finally, of course, a re-presentation almost certainly gives the subject another chance of producing an association encoding of those pairs that were not adequately coded on previous presentations.

Two other results have implications concerning paired-associate memorizing in general. Firstly, it was found that following an error just prior to the second presentation of a pair, recall performance at final test was poorer than recall of a singly-presented item at a similar retention interval in those experiments which employed a fixed duration test trial. This depression in recall performance to items on which an error had been made did not appear in Exp.III, in which test-trial duration was subject-determined. This suggests that if an encoding is inadequate on the onset of a re-presentation and the subject is aware of this inadequacy, then he uses the re-presentation to improve and probably elaborate his encoding of the pair. This conclusion is based on the hypothesis that more elaborate encodings take longer to decode during search and retrieval at test, so that with test trials of a short fixed duration, those items which are encoded more elaborately are disadvantaged.

Secondly, it was found in Exp.III that the proportion of double recall errors that were perseverations (i.e. with the same inappropriate response being made on each of the two test trials) following a correct recognition response on each test trial, was no higher than the proportion of perseverations when a recognition error was made on each test trial. Since in the latter case it is likely that subjects were guessing on each test trial, the greater than chance perseveration error level found was interpreted as resulting from a "guessing number" strategy, whereby some subjects may always have guessed the same response if they had no idea as to the appropriate response on any test trial. Consequently, when recall errors followed correct stimulus recognition on each test trial, it appears that not only was the second recall response unlikely to be a pure guess, but that if anything, subjects were able to avoid perseverations in this situation. The implication of this observation is that subjects could often remember on a second test trial that their first test trial response had been wrong. This in turn suggests that on a re-presentation which followed a correct stimulus recognition but a recall error, subjects may have elaborated their association encoding in such a way as to incorporate the original inappropriate resonse with the additional information that this response is wrong, for example "the response is not X but Y". Even if the extended encoding were not able to produce Y at subsequent test, it might be sufficient to enable the subject to avoid responding X again, so that he could guess away from this clearly inappropriate response.

Of course, it is possible that two parallel codings are formed, in such a way that the second encoding employes a different functional version of the stimulus; this seems quite likely, since the original encoding was prone to interference or competition. In this case, the old encoding would be "tagged" as inappropriate, so that if the original functional stimulus were employed to initiate search on a subsequent test, the subject would realise he was on the wrong track, and he would therefore be able to go back to the stimulus again and try to find the appropriate functional version. Both these formulations could explain the observed depression in T_2 recall performance following a recall error prior to a re-presentation on a schedule involving fixed duration test trials.

It is now proposed to deal with those results which concern the spacing effect. It was only in Exp.I that T_2 performance appeared to show a continued improvement with interpresentation spacings in excess of those required to "wipe out" short-term retention effects from P_1 to P_2. It is possible that the inclusion of a stimulus recognition test in Expts. II and III somehow altered subjects' encoding strategies (perhaps towards an increased emphasis on stimulus encoding) in some way that precluded the normal spacing effect. Nevertheless, the results of these studies suggested that recall performance at T_2 was not affected in any systematic way by stimulus recognition performance alone at T_1. In particular, there was no evidence that T_2 recall performance was depressed if the second presentation of an item followed a "short-term" stimulus recognition. Thus, the results of these studies are not consistent with the hypothesis that the spacing effect is caused by the maintenance of encodings with inadequate stimulus components on a second presentation at short interpresentation intervals. On the other hand the results of Exp.I suggested that the continued improvement in T_2 performance with interpresentation spacing was confined mainly to those items that were correctly recalled just prior to the second presentation.

A detailed Markovian analysis of Exp.I confirmed this suspicion.

There was no evidence whatsoever that T_2 performance on items to which a T_1 recall error had been made showed any systematic improvement with increasing non-zero interpresentation intervals. Although it appeared that items correctly recalled just prior to a second presentation at very short interpresentation intervals received very little benefit from the second presentation, it was found in addition that T_2 recall on such pairs continued to improve significantly with longer interpresentation intervals. Thus, the spacing effect appeared to be confined only to those items that could be recalled from memory proper just prior to their second presentation, and furthermore, it was found to operate over a range of interpresentation intervals in excess of that which would be predicted solely on the hypothesis that short-retention items received no benefit from a re-presentation. Finally, there was considerable evidence that the second presentation operated to increase the resistance to forgetting of those items that were already adequately encoded at its onset, and that it became increasingly effective as the interpresentation interval increased. Although the Markovian model predicted some limit in the interpresentation interval beyond which the effect of a re-presentation would decline (due to the fact that with long interpresentation intervals, there would be significant long-term forgetting, so that an increasing proportion of correct T_1 responses would be guesses) there was no evidence that this limit fell in the range of interpresentation intervals employed in Exp. I (i.e. 0-16 trials).

These results are clearly consistent with some kind of differential encoding hypothesis, since the second presentation almost certainly operates to produce a superior encoding of those items already adequately encoded at its onset. There are a number of psycholgical rationales that could underly such a hypothesis. For example, there is an "active" hypothesis, which would claim that as the interpresentation interval increases, so the subject's confidence in a recall-supporting code at P_2 declines, so that he consequently becomes more and more motivated to "think up" improvements or modifications to his current encoding. Alternatively, a more "passive" view may be taken of the subject's role. For example, it may be that the subject's encoding of a pair is to some extent determined by episodic cues or aspects available during the presentation sequence. Thus, at short interpresentation intervals, there will be a great likelihood that the episodic cues available at the second presentation are very similar to those present on the first presentation, so that the subject just can't help thinking of the same encoding that he used before. On the other hand, after a long interpresentation interval, the second presentation may occur in so different an episodic context to the first that the subject just can't help thinking of new ways of encoding the pair that may be employed to elaborate his original association encoding, or to replace his original encoding altogether by a better one. Although this passive hypothesis may appear at first sight to correspond with a "multiple retrieval route" position, it should be stressed that this is not the case. In other words, it emphasisesthe differential aspects of second presentation encoding, but does not necessarily imply that a second presentation encoding in any way provides an alternative independent retrieval route to that provided by the original encoding. On the contrary, it assumes that the original encoding is available at re-presentation, and will therefore play some part in determining the form of the improved encoding.

Unfortunately, it is not possible to discriminate between these hypotheses on the basis of the current data, and further experimentation will be necessary. An experiment involving a recall test and a confidence rating response would seem to offer the most fruitful approach, and if this failed to explain the spacing effect, it could be followed up by a study in which the context of each presentation of a critical item (in terms of the pairs preceding it) was systematically varied in an attempt to manipulate the episodic aspects of each presentation sequence. Finally, it should be stressed that the present experiments do not yield any information concerning the way in which an original encoding may be improved. Several feasible mechanisms have already been advanced as part of the paired-associate encoding theory discussed earlier in this chapter. However, this problem is not likely to be solved until we have a far greater understanding of paired-associate encoding than we have at present, and it is certainly a very difficult area of research.

To summarize, then, the present results strongly suggest that the spacing effect in paired-associate memory results from an increase in the effectiveness of a second presentation with increased inter-presentation spacing in improving the encoding of those pairs that are already moderately well encoded when re-presented. Although this conclusion supports some kind of differential encoding hypothesis, it is not at present possible to make any kind of inference about the nature of the differential encoding which occurs. However, this position is not too different from the current state of knowledge concerning spacing effects in Brown-Peterson memory (see section 3.33), and it is quite possible that a similar underlying rationale applies to the spacing effects observed both in paired-associate and Brown-Peterson memory. In other words, although the nature of the encodings employed in the two paradigms may differ considerably, it is still possible that the reason why a second presentation increases in effectiveness with spacing is common to both areas.

BIBLIOGRAPHY

* (Abbreviations: JVLVB = Journal of Verbal Learning and Verbal Behaviour.
JEP = Journal of Experimental Psychology)

AARONSON, D (1967): Temporal factors in perception and short-term memory. Psycholigical Bulletin, 67, 130-144.

ADAMS, J.A. (1952): Warm-up decrement in performance on the pursuit rotor. American Journal of Psychology, 65, 404-414.

ARCHER, E.J. (1960): A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs, 74, (10, whole No. 497).

ATKINSON, R.C., BRELSFORD, J.W., and SHIFFRIN, R.M. (1967): Multiprocess models for memory with applications to a continuous presentation task. Journal of Mathematical Psychology, 4, 277-300.

ATKINSON, R.C., and CROTHERS, E.J. (1964): A comparison of paired-associate learning models having different acquisition and retention axioms. Journal of Mathematical Psychology. 1, 285-315.

ATKINSON, R.C., and SHIFFRIN, R.M. (1968): 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 II, New York, Academic Press.

AVERBACH, E., and CORIELL, A.S. (1961) : Short-term memory in vision. Bell System Technical Journal, 40, 309-328.

BADDELEY, A.D., (1966a): Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. Quarterly Journal of Experimental Psychology, 18, 362-365.

BADDELEY, A.D., (1966b) : The influence of acoustic and semantic similarity on long-term memory for word sequences. Quarterly Journal of Experimental Psychology, 18, 302-309.

BADDELEY, A.D., and PATTERSON, K. (1971): The relation between short-term and long-term memory. British Medical Bulletin, 27, 237-242.

BADDELEY, A. D., SCOTT, D., DRYNAN, R., and SMITH, J. C. (1969): Short-term memory and the limited capacity hypothesis. British Journal of Psychology, 60, 51-55.

BARNES, J. M., and UNDERWOOD, B. J. (1959) : "Fate" of first-list associations in transfer-theory. J.E.P. 58, 97-105.

BATTIG, W. F. (1966) : Evidence for coding processes in "rote" paired-associate learning. JVLVB, 5, 177-181.

BATTIG, W. F., and BERNSTEIN, D.A.(1965) : Intralist heterogeneity and interpair grouping in paired-associate learning. Paper read at the Psychonomic Society, Niagara Falls, Ont.

BATTIG, W. F., BROWN, S.C., and NELSON, D. (1963) : Constant vs. Varied serial order in paired-associate learning. Psychological Reports, 12, 695-721

BELL, H. M. (1942) : Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. Psychological Monographs, 54, No. 1.

BERNBACH, H. A. (1967a) : Stimulus learning and recognition in paired-associate learning. J.E.P, 75, 513-519.

BERNBACH,H. A., (1967b) : The effects of labels in short-term memory for colours with nursery school children. Psychonomic Science, 7, 149-150.

BJORK, R. A. (1966) : Learning and short-term retention of paired-associates in relation to specific sequences of interpresentation intervals. Technical Report No. 106. Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, Ca.

BJORK, R. A. (1970) : Positive forgetting : the noninterference of items intentionally forgotten. JVLVB, 9, 255-268.

BJORK, R. A., and ABRAMOWITZ, R. L. (1968) : The optimality and commutativity of successive interpresentation intervals in short-term memory. Paper presented at the meeting of the Midwest Psychological Association, Chicago, May '68.

BJORK, R. A. and ALLEN, T. W. (1970): The spacing effect : consolidation or differential encoding? JVLVB, 9, 567-572.

BOUSFIELD, W. A. (1953) : The occurrence of clustering in the recall of randomly-arranged associates. Journal of General Psychology, 49, 229-240.

BOUSFIELD, W. A., PUFF, C.R., and COWAN, T.M. (1964) : The development of constances in sequential organisation during repeated free recall. JVLVB, 3, 489-495.

BOWER, G.H. (1970): Mental imagery and associative learning. In Gregg, L (ed), Cognition in Learning and Memory, New York, Wiley.

BRELSFORD, J.W. Jnr., SHIFFRIN, R.M., and ATKINSON, R.C. (1968) : Multiple reinforcement effects in short-term memory. British Journal of Mathematical and Statistical Psychology, 21, 1-19.

BROADBENT, D.E. (1958) : Perception and Communication, London, Pergamon.

BROWN, J. A. (19 4) : The nature of set to learn and of intra- material interferences in immediate memory. Quarterly Journal of Experimental Psychology, 6, 141-148.

BROWN, S.C., and BATTIG, W.F. (1962) : Partial serial position constancy in paired-associate learning. JVLVB, 1, 42-47.

BROWN, S. C., BATTIG, W. F., and PEARLSTEIN, R. (1965): Effect of successive additions of stimulus elements on paired-associate learning. JEP, 70, 87-93.

BUMSTEAD, A.P., (1940) : Distribution of effort in memorizing prose and poetry. American Journal of Psychology, 53, 423-427.

CALFEE, R. C., (1968): Interpresentation effects in paired-associate learning. JVLVB, 7, 1030-1036.

CALVIN, J. S. (1939) : The effects of regularity of stimulation in conditioned-response learning. Psychological Bulletin, 36, 596.

CIEUTAT, V.J., STOCKWELL, F.E., and NOBLE, C.E., (1958): The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. JEP, 56, 193-202.

COHEN, B. H., (1966): Some-or-none characteristics of coding behaviour. JVLVB, 5, 182-187.

COLLINS, A. M., and QUILLIAN, M. R. (1972) : Experiments on semantic memory and language comprehension. In. Gregg. L. (ed.), Cognition in Learning and Memory, New York, Wiley.

COOK, S. A. (1928) : The effect of various temporal arrangments of practice on the mastery of an animal maze of moderate complexity. Archives of Psychology, New York, No.98.

COOK, T. W. (1944): Factors in massed and distributed practice. JEP 34, 325-334.

COOPER, E. H., and PANTLE,A. J. (1967): The total time hypothesis in verbal learning. Psychological Bulletin, 68, 221-234.

CRAIK, F.I.M., (1968) : Two components in free recall. JVLVB, 7, 996-1004.

CRAIK, F.I.M., (1971): Primary memory. British Medical Bulletin, 27, 232-236.

CRAIK, F.I.M., and LOCKART. R.S. (1972): Levels of processing: a framework for memory search. JVLVB, 11,671-684.

CROWDER, R.G., and MORTON, J. (1969): Precategorical acoustic storage (PAS). Perception and Psychophysics, 5, 365-373.

DEESE, J. (1957) : Serial organization in the recall of disconnected items. Psychological Reports, 3, 577-582.

DORÉ, L.R., and HILGARD, E. R. (1937): Spaced practice and the maturation hypothesis. Journal of Psychology, 4, 245-259.

EBBINGHAUS, H. (1885) : Über das Gedächtris : Untersuchungen zur experimentallen Psychologie . Leipzig, Dunker and Hublot.

ELIAS, C. S., and PERFETTI, C. A. (1973) : Encoding task and recognition memory : the importance of semantic coding. JEP, 99,151-156.

EPSTEIN, B. (1949): Immediate and retention effects of interpolated rest periods on learning performance. Teach. Coll. Cont. Educ. No. 949.

ERICKSEN, S. C., (1942) : Variability of attack in massed and distributed practice. JEP, 31, 339-345.

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

ESTES, W. K., (1960) : Learning theory and the new "mental chemistry". Psychological Review, 67, 207-223.

ESTES, W. K. (1967) : Reinforcement in human learning. Technical Report No. 125, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, Ca.

GARDINER, J. M., (1974): Levels of processing in word recognition and subsequent free recall. JEP,102, 101-105

GARRETT, H. E. (1940) : Variability in learning under massed and spaced practice. JEP,26, 547-567.

GARTMAN, L. M., and JOHNSON, N. F., (1972): Massed v.s. distributed repetition of homographs : a test of the differential encoding hypothesis JVLVB, 11, 801-808.

GLANZER, M. (1972): Storage mechanisms in recall. In Bower G.H. (Ed.), The Psychology of Learning and Motivation, Vol. 5. New York, Academic Press.

GLANZER, M., and CUNITZ, A. R., (1966) : Two storage mechanisms in free recall JVLVB, 5, 351-360.

GLANZER, M., GIANUTSOS, R., and DUBIN, S. (1969): The removal of items from short-term storage. JVLVB, 8, 435-7.

GLANZER M., and MEINZER,A., (1967): The effects of intra-list activity on free recall JVLVB, 6, 928-935.

GLAZE, J. A., (1928) : The association value of nonsense syllables. Journal of Genetic Psychology, 35, 225-269.

GREENO, J.G. (1964): Paired-associate learning with massed and distributed repetitions of items. JEP, 67, 286-295.

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

GREENO, J.G. (1969): A cognitive interpretation of negative transfer and forgetting of paired associates. Human Performance Centre. Memorandum Report No. 9. University of Michigan.

GREENO, J. G., (1970a) : Conservation of information processing capacity in paired-associate memorizing. JVLVB, 9, 581-586.

GREENO, J.G. (1970b) : How associations are memorized. In D.A. Norman (ed). Models of Human Memory, New York, Academic Press.

GUTHRIE, E. R., (1935): The Psychology of Learning, New York, Harper.

HEBB , D. O. (1949): The Organisation of Behaviour, New York, Wiley.

HELLYER, S. (1962) : Frequency of stimulus presentation and short-term decrement in recall. JEP, 64, 650.

HERRIOT, P. (1974) : Attributes of Memory. London, Methuen.

HINTZMAN, D.L. (1969a): Apparent frequency as a function of frequency and the spacing of repetitions. JEP, 80, 139-145.

HINTZMAN, D.L. (1969b) : Recognition time : effects of recency, frequency, and the spacing of repetitions. JEP, 79, 192-194.

HINTZMAN, D.L. (1970) : Effects of repetition and exposure duration on memory. JEP, 83, 435-44.

HINTZMAN, D.L., and BLOCK, R.A. (1971) : Repetition and Memory: evidence for a multiple-trace hypothesis. JEP, 88, 297-306.

HINTZMAN, D.L., BLOCK, R.A. and INSKEEP, N. R. (1972) : Memory for mode of input. JVLVB, 11, 741-749.

HINTZMAN, D.L., and SUMMERS, J.F., (1973); Long-term visual traces of visually presented words. Bulletin of the Psychonomic Society, 1, 325-327.

HOVLAND, C. I. (1938) : Experimental studies in rote - learning theory. III. Distribution of practice with varying speeds of syllable presentation. JEP, 23, 172-190.

HOVLAND, C.I. (1939) : Experimental studies in rote-learning theory. V. Comparison of distribution of practice in serial and paired-associate learning. JEP, 25, 622-633.

HOVLAND, C. I. (1940) : Experimental Studies in rote-learning theory. VII.

Distribution of practice with varying lengths of list. JEP, 27, 271-284.

HOVLAND, C. I. (1949): Experimental studies in rote-learning theory. VIII. Distributed practice of paired-associates with varying rates of presentation. JEP, 39, 714-718.

HOWE, M.J.A. (1967): Consolidation in short term memory as a function of rehearsal Psychonomic Science, 7, 355-356.

HULL, C. L. (1943) : Principles of Behaviour. New York, Appleton-Century - Crofts.

HUMPHREYS, L. G. (1940): Distributed practice in the development of the conditioned eyelid reaction. Journal of General Psychology, 22, 379-385.

IZAWA, C. (1966): Reinforcement - test sequences in paired-associate learning. Psychological Reports. 18, 879-919.

IZAWA, C. (1967).: Function of test trials in paired-associate learning. JEP, 75, 194-209.

IZAWA, C. (1968): Effects of successive reinforcements, tests and neutral trials in paired-associate learning. In Proceedings of the 76th Annual Convention of the American Psycholgical Association. Washington, D.C., APA, 33-34.

IZAWA, C. (1969): Long sequences of successive tests in paired-associate acquisition. In Proceedings of the 77th annual convention of the American Psychological Association. Washington D.C., APA, 57-58.

IZAWA, C. (1970): Reinforcement-test-blank acquisition programming under the unmixed list design in paired-associate learning. Psychonomic Science, 19, 75-77.

IZAWA, C. (1971): The test trial potentiating model. Journal of Mathematical Psycholcgy, 8, 200-224.

JAMES, C.T. and GREENO J.G. (1967): Stimulus selection at different stages of paired-associate learning. JEP, 74, 75-83.

JAMES, W. (1890): The Principles of Psychology, New York, Henry Holt & Co.

JENKINS, J.J., and RUSSELL, W.A. (1952): Associative clustering during recall. Journal of Abnormal and Social Psychology, 47, 818-821.

JOHNSON, N.F. (1970): The role of chunking and organisation in the process of recall. In Bower, G.H. (Ed). The Psychology of Learning and Motivations, Vol. 4. New York, Academic Press.

JOHNSON, N.F. (1972): Organisation and the concepts of a memory code. In Melton, A.W. and Martin, E. (Eds). Coding Processes in Human Memory. Washington D.C., Winston.

JOST, A. (1897): Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen. Zeitschrift fur Psychologie, 14, 436-472.

KELLER, L., THOMPSON W.J., TWEEDY, J.R. And ATKINSON, R.C., (1967): Effects of reinforcement interval on paired-associate learning. JEP, 73, 268-277.

KEPPEL, G. (1968): Retroactive and proactive inhibition. In Dixon, T.R. and Horton D.L. (Eds). Verbal Behaviour and General Behaviour Theory. New Jersey, Prentice-Hall.

KEPPEL, G. and UNDERWOOD, B.J., (1962) : Proactive inhibition in short-term retention of single items. JVLVB, 1, 153-161.

KIENTZLE, M.J. (1946) : Properties of learning curves under varied distribution of practice. JEP, 36, 187-211.

KIENTZLE, M.J. (1949): Ability patterns under distributed practice JEP, 39, 532-537.

KIMBLE, G.A., and SHATEL, R.B., (1952): The relationship between two kinds of inhibition and the amount of practice. JEP, 44, 355-359.

KINTSCH, W. (1966) : Recognition learning as a function of the length of the retention interval and changes in the retention interval. Journal of Mathematical Psychology, 3, 412-433.

KINTSCH, W. (1970a): Learning, Memory and Conceptual Processes. New York, Wiley.

KINTSCH, W. (1970b) : Models for free recall and recognition. In Norman, D.A. (Ed). Models of Human Memory. New York Academic Press.

KINTSCH, W., and BUSCHKE, H. (1969): Homophones and synonyms in short-term memory. JEP, 80, 403-407.

KOPPENAAL, R.J. (1963).: Time changes in the strength of A-B, A-C lists; spontaneous recovery? JVLVB, 2, 310-319.

LANDAUER, T. K. (1962) : Rate of implicit speech. Perceptual Motor Skills, 15, 646.

LANDAUER, T.K. (1969): Reinforcement as consolidation. Psychological Review, 76, 82-96

LOESS, H. (1967). Short-term memory, word class, and sequence of items. JEP, 74, 556-561.

LOESS, H. (1968). Short-term memory and item similarity. JVLVB. 7, 87-92.

LOESS, H. and WAUGH, N.C. (1967): Short-term memory and inter-trial interval. JVLVB, 6, 455-460.

LORENTE de NÓ, R. (1938): Analysis of the activity of the chains of internuncial neurons. Journal of Neurophysiology, 1, 207-244.

LORGE, I. (1930): Influence of regularly interpolated time intervals upon subsequent learning. Teach. Coll. Contr. Educ, No. 438.

LYON, D.O., (1914): The relation of length of material to time taken for learning and the optimum distribution of time. Journal of Educational Psychology, 5, 1-9, 85-91, 155-163.

LYON, D.O. (1917) : Memory and the Learning Process. Baltimore, Warwick and York.

MADIGAN, S.W. (1969): Intraserial repetition and coding processes in free recall. JVLVB, 8, 828-835.

MANDLER, G. (1967): Verbal learning. In New Directions in Psychology, III. New York; Holt, Rinehart and Winston.

MANDLER, G, (1972) : Organisation and recognition. In Tulving, E. and Donaldson W.A. (Eds). Organisation of Memory, New York. Academic Press.

MARTIN, E. (1966): Stimulus pronounceability in aural paired-associate learning JVLVB, 5, 18-22.

MARTIN, E. (1967a) : Relation between stimulus recognition and paired-associate learning. JEP, 74, 500-505.

MARTIN, E. (1967b): Responses to stimuli in verbal learning. Human Performance Centre Memorandum Report No. 3. University of Michigan.

MARTIN, E. (1967c): Stimulus recognition in aural paired-associate learning JVLVB, 6, 272-276.

MARTIN, E. (1968a): Recognition and correct responding mediated by first letter of trigram stimuli. JVLVB, 7, 703-706.

MARTIN, E. (1968b): Stimulus meaningfulness and paired-associate transfer : an encoding variability hypothesis. Psychological Review. 75, 421-441.

MARTIN, E. (1971): Verbal learning theory and independent retrieval phenomena. Psycholigical Review, 78, 314-332.

MARTIN, E. (1972): Stimulus recoding in learning and transfer. In Melton A.W. and Martin E. (Eds.) Coding Processes in Human Memory, Washington D.C., Winston.

MARTIN E, and CAREY, S.T. (1971) : Retroaction, recovery and stimulus meaningfulness in the A-B, A-Br paradigm. American Journal of Psychology, 84, 123-133.

MELTON, A.W. (1963): Implications of short-term memory for a general theory of memory. JVLVB, 2, 1-21.

MELTON, A.W. (1970): The situation with respect to the spacing of repetitions and memory. JVLVB, 9, 596-606.

MELTON, A.W., REICHER, G.M. and SHULMAN, H.G. (1966): A distributed practice effect on probability of recall in free recall of words. Paper presented at the Psychonomic Society, St. Louis. Oct. '66.

MELTON, A.W. and SHULMAN, H.G. (1967): Further studies of a distributed practice effect on probability of recall in free recall.

Paper presented at the Psychonomic Society, Chicago, Oct '67.

MERRYMAN, C. T., and MERRYMAN, S. S. (1971): Stimulus encoding in the A-B', AX-B and the A-Br', AX-B paradigms. JVLVB, 10,681-685.

MILLER, G. A. (1956): The magical number seven plus or minus two: some limits on our capacity for processing information. Psychological Review, 63. 81-97.

MILNER, B. (1967): Amnesia following operation of the temporal lobes. In Zangwill, C.L. and Whitty, C.M.W. (eds) Amnesia, London, Butterworth.

MURDOCK, B. B. Jnr. (1963a) : Short-term retention of single paired associates. JEP , 65, 433-443.

MURDOCK, B. B. Jnr. (1963b): Interpolated recall in short term memory. JEP, 66, 525-532.

MURDOCK, B. B. Jnr. (1963c): Short term memory and paired-associate learning. JVLVB. 2, 320-328.

MURDOCK, B. B. Jnr. (1964) : Proactive inhibition in short-term memory. JEP, 68, 184-189.

NEISSER, U. (1967): Cognitive Psychology. New York, Appleton-Century-Crofts.

NEYMAN, J. (1949): Contribution to the theory of the χ^2 test. In Neyman J. (Ed). Proceedings of the Berkeley Symposium on Mathematical Statistics and Prbability. Berkeley, University of California Press.

NOBLE, C. E., (1952): An analysis of meaning. Psychological Review, 59, 421-430.

OLSON, G.M. (1969): Learning and retention in a continuous recognition task. JEP, 81, 381-384.

PAIVIO, A. (1969): Mental imagery in associative learning and memory. Psychological Review, 76, 241-263.

PAVLOV, I.P., (1927): Condition Reflexes : An Investigation of the Physiological Activity of the Cerebral Cortex. London, Oxford University Press.

PECHSTEIN, L. A. (1921), Massed vs. distributed effort in learning. Journal of Educational Psychology, 12, 92-97.

PETERSON, L. R., (1963): Immediate memory : data and theory. In Cofer, C.N. and Musgrave, B.S. (Eds) Verbal Behaviour and Learning : Problems and Processes. New York, McGraw-Hill.

PETERSON, L. R. and BREWER, C. L. (1963) : Confirmation, correction and contiguity. JVLVB, 1, 365-371.

PETERSON, L. R., HILLNER, K., and SALTZMAN, D. (1962): Supplementary report : time between pairings and short-term retention. JEP, 64, 550-551.

PETERSON, L. R. and PETERSON, M. J. (1959): Short-term retention of individual verbal items. JEP, 58, 193-198.

PETERSON, L. R. and PETERSON, M. J. (1962): Minimal paired-associate learning. JEP, 63, 521-527.

PETERSON, L. R. SALTZMAN, D., HILLNER, K. and LAND, V. (1962): Recency and frequency in paired-associate learning. JEP, 63, 396-403.

PETERSON L. R., WAMPLER R., KIRKPATRICK, M., and SALTZMAN, D., (1963). Effect of spacing presentations on retention of a paired associate over short intervals. JEP, 66, 206-209.

POLLATSEK, A. (1969): Rehearsal, interference and spacing of practice in short-term memory. Technical Report, Human Performance Centre, University of Michigan.

POSTMAN, L., and GREENBLOOM, R., (1967) : Conditions of cue selection in the acquisition of paired-associate lists. JEP 73, 91-100.

POSTMAN, L. and PHILLIPS, L. W. (1965): Short-term temporal changes in free recall. Quarterly Journal of Experimental Psychology 17, 132-138.

POSTMAN, L., and STARK, K. (1971): Encoding variability and transfer. American Journal of Psychology, 84, 461-472.

POSTMAN, L. STARK, K, and FRASER, J., (1968): Temporal changes in interference. JVLVB, 7, 672-694.

POTTS, G.R. (1969): Distance from a massed double presentation as a factor in paired-associate list learning. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May '69.

POWELL, M.J.D. (1964): An efficient method of finding the minimum of a function of several variables without calculating derivatives. Computer Journal, 7, 155-162.

PYLE, W.H. (1928): The Psychology of Learning : An advanced Text in Educational Psychology. Baltimore, Warwick and York.

RAFFEL, G.(1936): Two determinants of the effects of primacy. American Journal of Psychology, 48, 654-657.

RAYMOND, B. (1969): Short-term storage and long-term storage in free recall. JVLVB. 8, 567-574.

REITMAN, W.R., MALIN, J.T., BJORK, R.A. and HIGMAN, B. (1973): Strategy control and directed forgetting. JVLVB, 12, 140-149.

RICHARDSON, J. (1972): Encoding and stimulus selection in paired-associate verbal learning. In Melton A.W., and Martin, E. (Eds) Coding Processes in Human Memory, Washington D.C., Winston.

RUMELHART, D.E. (1967): The effects of interpresentation interval on performance in a continuous paired-associate task. Technical Report No. 116, Institute for Mathematical Studies in the Social Sciences, Stanford University, Standord, Ca.

RUNDUS, D. (1971): Analysis of rehearsal processes in free recall. JEP, 89, 63-77.

RUNDUS, D., LOFTUS, G.R. and ATKINSON, R.C., (1970): Immediate free recall and three-week delayed recognition. JVLVB, 9, 684-688.

SCHEIDER, N.C., and HOUSTON, J.P., (1969): Retroactive inhibition, cue selection, and degree of learning. American Journal of Psychology, 82, 276-279.

SCHILD, M.E. and BATTIG, W.F. (1966): Directionality in paired-associate learning. JVLVB, 5, 42-49.

SCHULZ, R.W., and MARTIN. E. (1964): Aural paired-associate learning: stimulus familiarisation, response familiarisation and pronounceability. JVLVB, 3, 139-145.

SCHWARTZ, M. (1975): The effect of constant vs. varied encoding and massed vs. distributed presentations on recall of paired associates. Memory and Cognition, 3, 390-394.

SHAFFER, W.O., and SHIFFRIN, R.M. (1972): Rehearsal and storage of visual information. JEP, 92, 292-296.

SHALLICE, T. (1967): Paper presented at the NATO Advanced Study Institute on Short-term Memory, Cambridge.

SHALLICE, T. and WARRINGTON, E.K. (1970): Independent functioning of verbal memory stores; a neuropsychological study. Quarterly Journal of Experimental Psychology, 22, 261-273.

SHEPARD, R.N., (1963): Comments on Professor Underwood's paper. In Cofer, C.N., and Musgrave, B.S. (Eds) Verbal Behaviour and Learning, New York, McGraw-Hill.

SHEPARD, R.N., and TEGHTSOONIAN, M. (1961): Retention of information under conditions approaching a steady state. JEP, 62, 302-309.

SHULMAN, H.G. (1971): Similarity effects in short term memory. Psychological Bulletin, 75, 399-415.

SKINNER, B.F. (1938): The Behaviour of Organisms, New York, Appleton-Century-Crofts.

SNODDY, G. (1945): Evidence for a universal shock factor in learning. JEP, 35, 403-417.

SPERLING, G. (1960): The information available in brief visual presentations. Psychological Monogrpahs, 74, No. 11.

STRAND, B.Z. (1971): Further investigation of retroactive inhibition in categorized free recall. JEP, 87, 198-201.

THOMPSON, D.M. and TULVING, E. (1970): Associative encoding and retrieval: weak and strong cues. JEP, 86, 255-262.

THORNDIKE, E.L. and LORGE, I. (1944): The Teachers Word Book of 30,000 Words. New York, Columbia University Press.

TOLMAN, E.C. (1932): Purposive Behaviour in Animals and Men. New York, Century.

TRAVIS, R.C., (1937): The effect of the length of the rest period on motor learning. Journal of Psychology, 3, 189-194.

TREISMAN, A.M. (1964): Monitoring and storage of irrelevant messages in selective attention. JVLVB, 3, 449-459.

TULVING, E. (1962): Subjective organisation in free recall of unrelated words. Psychological Review, 69, 344-354.

TULVING, E. (1964): Intratrial and intratrial retention: notes towards a theory of free-recall verbal learning. Psychological Review, 71, 219-237.

TULVING, E. (1966): Subjective organisation and effects of repetition in multi-trial free-recall learning. JVLVB, 5, 193-197.

TULVING, E. (1967): The effects of presentation and recall of material in free-recall learning. JVLVB 6, 175-184.

TULVING, E. (1968): Theoretical issues in free recall. In Dixon, T.R. and Horton, D.L. (Eds) Verbal Behaviour and General Behaviour Theory, New Jersey, Prentice-Hall.

TULVING, E., and ARBUCKLE, T.Y. (1963): Sources of intertrial interference in immediate recall of paired associates. JVLVB, 1, 321-324.

TULVING, E., and ARBUCKLE, T.Y. (1966): Input and output interference in short-term associative memory. JEP, 72, 145-150.

TULVING, E., and PEARLSTONE, Z. (1966): Availability versus accessibility of information in memory for words. JVLVB, 5, 381-391.

TULVING, E., and PSOTKA, J. (1971): Retroactive inhibition in free recall: inaccessibility of information available in the memory store. JEP, 87, 1-8.

UNDERWOOD, B.J. (1963): Stimulus selection in verbal learning. In Cofer C.N. and Musgrave B.S. (Eds) Verbal Behaviour and Learning, New York, McGraw-Hill.

UNDERWOOD, B.J. (1969): Some correlates of item repetition in free-recall learning. JVLVB, 8, 83-94.

UNDERWOOD, B.J. (1970): A breakdown of the Total-Time Law in free-recall learning. JVLVB, 9, 573-580.

UNDERWOOD, B.J., and POSTMAN, L.(1960): Extra - experimental sources of interference in forgetting. Psychological Review, 67, 73-95.

UNDERWOOD, B.J. and SCHULZ, R.W. (1960): Meaningfulness and Verbal Learning. Chicago, Lippincott.

WARDEN, C.J. (1923): The distribution of practice in animal learning. Comparative Psychological Monographs, No. 3.

WAUGH, N.C. (1963): Immediate memory as a function of repetition. JVLVB, 2, 107-112.

WAUGH,N.C. (1967): Presentation time and free recall. JEP 73, 39-44.

WAUGH ,N.C. (1970): On the effective duration of a repeated word. JVLVB, 9, 587-595.

WAUGH, N.C. and NORMAN,D.A. (1965): Primary memory. Psychological Review, 72, 89-104.

WAUGH, N.C. and NORMAN, D.A. (1968): The measure of interference in primary memory. JVLVB, 7, 617 - 626.

WEAVER,G.E., McCANN, R.L. and WEHR,R.J., (1970): Stimulus meaningfulness, transfer and retroactive inhibition in the A-B, A-C paradigm. JEP, 85, 255-257.

WICKELGREN, W.A. (1970): Multitrace strength theory. In Norman D.A. (Ed) Models of Human Memory, New York, Academic Press.

WICKELGREN, W.A. (1973): The long and the short of memory, Psychological Bulletin ,80, 425 - 438.

WITMER, L.R. (1935): The association value of three-place consonant syllables. Journal of Genetic Psychology, 47, 337-360.

WOOD, G. (1972): Organisational processes and free recall. In Tulving, E, and Donaldson, W.A. (Eds) Organisation of Memory, New York, Academic Press.

YOUNG,J.L. (1966): Effects of intervals between reinforcements and test trials in paired-associate learning. Technical Report No.101.

Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, Ca.

YOUNG, J.L. (1971): Reinforcement-test intervals in paired-associate learning. Journal of Mathematical Psychology, 8, 58-80.

YOUTZ, A.C. (1941): An experimental evaluation of Jost's Laws. Psychological Monographs, 53, No. 1.

YUM, [illegible].S. (1931): An experimental test of the law of association. JEP 14, 68-82.

APPENDICES.

APPENDIX 1

(i) The pool of 886 common words from which the stimuli in Exps I and II were selected.

ACHE	ACT	ADD	AGE	AIM	AIR	ALE	ANT	APE	ARM
ART	ASK	ASS	AXE	BACK	BAD	BAG	BAIT	BALL	BAN
BAR	BARK	BARN	BASE	BAT	BAY	BEAD	BEAK	BEAT	BEAN
BED	BEE	BEER	BEG	BELL	BEND	BEST	BET	BIAS	BID
BIDE	BIG	BILL	BIN	BIT	BITE	BLIP	BLOT	BLOW	BLUR
BOAR	BOG	BOMB	BONE	BOOK	BOOT	BORN	BOSS	BOW	BOWL
BOX	BOY	BRAG	BRAN	BRAT	BREW	BRIM	BUD	BUG	BUN
BUS	BUY	CAB	CAGE	CAKE	CALL	CAMP	CANE	CAP	CAPE
CAR	CARE	CASE	CAT	CELL	CHAP	CHAT	CHEF	CHEW	CHIN
CHIP	CHOP	CHUM	CLAM	CLAN	CLAP	CLAW	CLIP	CLOD	CLOG
CLOT	CLUB	CLUE	COAL	COD	CODE	COG	COIL	COIN	COMB
CONE	COPE	CORE	COST	COT	COVE	COW	COY	CRAB	CRAG
CREW	CRIB	CROP	CROW	CUB	CUBE	CUE	CUFF	CUP	CURE
CUT	CUTE	DAB	DAFT	DAM	DAMP	DANK	DARE	DARK	DARN
DATE	DAY	DAZE	DEAF	DEAL	DEAN	DECK	DEED	DEEP	DEER
DEFT	DELL	DEN	DENT	DESK	DEW	DIAL	DICE	DIE	DIG
DIM	DIN	DINE	DIP	DIRE	DIRT	DOCK	DOE	DOG	DOLL
DOME	DOOR	DOSE	DOT	DOVE	DRAB	DRAG	DRIP	DROP	DRUG
DRUM	DUCK	DUEL	DUET	DULL	DUMB	DUMP	DUNE	DUPE	DUST
EAR	EBB	EDGE	EEL	EGG	ELM	END	ERR	EVE	EWE
EYE	FACE	FACT	FADE	FAIL	FAIR	FAKE	FALL	FAME	FAN
FAR	FARM	FAST	FAT	FATE	FEAR	FEAT	FEE	FEEL	FEN
FERN	FEND	FIG	FILL	FILM	FIN	FINE	FIR	FIRE	FIST
FIT	FLAG	FLAP	FLAT	FLAW	FLAY	FLEA	FLIP	FLIT	FLOG
FLOP	FLOW	FLY	FOAL	FOAM	FOE	FOG	FOIL	FOLD	FOLK
FOOD	FOOL	FOOT	FOWL	FOX	FRAY	FREE	FRET	FROG	FRY
FUEL	FUME	FUN	FUR	FUSS	GAG	GAIN	GAP	GAPE	GARB
GAS	GATE	GAY	GAZE	GEAR	GEM	GERM	GET	GIFT	GIN
GIRL	GLAD	GLEN	GLIB	GOAL	GOAT	GOLF	GORE	GRAB	GRID
GRIM	GRIN	GRIP	GRIT	GROW	GRUB	GULL	GUM	GUN	GUT

HACK	HAG	HAIL	HAIR	HALL	HALT	HAM	HAND	HARD	HARM
HAT	HATE	HAVE	HAY	HAZE	HEAD	HEAL	HEAP	HEAR	HEAT
HEED	HELP	HEMP	HEN	HEW	HIDE	HIKE	HILL	HINT	HIP
HIRE	HIT	HIVE	HOE	HOG	HOLE	HOME	HOOD	HOOK	HOOP
HOOT	HOP	HOPE	HORN	HOSE	HOST	HOT	HOUR	HUB	HUG
HUGE	HULL	HUM	HUMP	HUNT	HUT	ICE	ILL	IMP	INK
JAB	JADE	JAM	JAR	JAW	JAY	JAZZ	JEST	JET	JIB
JIBE	JIG	JOB	JOG	JOT	JOY	JUG	JUT	JUTE	KEEL
KEEP	KEEN	KEY	KICK	KILL	KILT	KILN	KIT	LACE	LACK
LAD	LAG	LAIR	LAKE	LAME	LAMP	LANE	LAP	LARD	LASS
LAST	LATE	LAW	LAY	LEAD	LEAF	LEAK	LEAN	LEAP	LEG
LEST	LET	LIAR	LICK	LID	LIE	LIFE	LIFT	LIKE	LIME
LINE	LIP	LIST	LIVE	LOAD	LOAD	LOB	LOCK	LOG	LOOK
LOOM	LOOP	LOOT	LOPE	LORD	LOSS	LOT	LOVE	LOW	LUCK
LULL	LURE	MACE	MACK	MAD	MAID	MAIL	MAIM	MAKE	MALE
MAN	MAP	MARK	MAT	MATE	MAUL	MAZE	MEAD	MEAL	MEAN
MEET	MEEK	MEND	MESS	MICE	MILE	MILK	MILL	MILT	MIND
MINE	MINT	MIRE	MISS	MIST	MOAN	MOB	MOCK	MOLE	MOP
MOPE	MORE	MOSS	MOST	MOTE	MOW	MUCK	MUD	MUG	MULE
MURK	MUTE	NAG	NAIL	NAME	NAP	NAPE	NAVE	NEAR	NEAT
NECK	NEED	NEST	NET	NEW	NEWT	NIB	NICE	NICK	NIP
NOD	NOTE	NUDE	NUT	NUN	OAF	OAR	OAT	ODD	ODE
OIL	OLD	OPT	ORB	OWE	OWL	OWN	PACE	PACK	PACT
PAD	PAGE	PAIL	PAIN	PALL	PALM	PAN	PANT	PARK	PART
PASS	PAST	PAT	PATH	PAVE	PAW	PAY	PEA	PEAK	PEAR
PEAT	PECK	PEEL	PEER	PEG	PELT	PEN	PERT	PEST	PET
PEW	PICK	PIE	PIG	PIKE	PILE	PILL	PIN	PINE	PINT
PIP	PIPE	PIT	PITH	PLAN	PLAY	PLOD	PLOT	PLOY	PLUG
PLUM	POD	POLE	POOR	PORK	PORT	POSE	POST	POT	PRAM
PREY	PRIM	PROD	PROP	PROW	PUB	PUFF	PULL	PUN	PUNT
PUT	QUAY	QUIP	QUIT	RACK	RACE	RAFT	RAG	RAGE	RAID
RAIL	RAIN	RAKE	RAM	RANT	RAP	RARE	RAT	RATE	RAVE

RAW	RAY	READ	REAL	REAP	REAR	REEF	REEK	REND	RENT
REST	RIB	RICE	RICH	RICK	RIDE	RIG	RIM	RIND	RIP
RIPE	RISE	ROAD	ROAM	ROAR	ROB	ROBE	ROCK	ROD	ROLL
ROMP	ROOF	ROOK	ROOM	ROOT	ROPE	ROSE	ROT	ROVE	ROW
RUB	RUG	RULE	RUM	RUN	RUNT	RUSE	RUST	RUT	SACK
SAD	SAFE	SAG	SAGE	SAIL	SALT	SAME	SAND	SANE	SAP
SAVE	SAW	SAY	SEAL	SEAM	SEAT	SEE	SEED	SEEK	SEER
SELL	SEND	SET	SEW	SHAM	SHED	SHIN	SHIP	SHOE	SHOP
SHOT	SHOW	SHUN	SICK	SIDE	SILL	SIN	SIP	SIT	SITE
SLAB	SLAM	SLAP	SLAY	SLED	SLIM	SLIP	SLIT	SLOB	SLOG
SLOP	SLOT	SLOW	SLUG	SLUR	SNAG	SNAP	SNIP	SNOB	SNOW
SNUB	SNUG	SOAK	SOAP	SOAR	SOB	SOCK	SOFT	SOIL	SOLE
SOME	SON	SOON	SOP	SOT	SOUL	SOW	STAB	STAG	STAY
STEM	STEP	STEW	STIR	STOP	STOW	STUB	STUD	STUN	STY
SUCH	SUCK	SUIT	SUM	SUN	SUP	SWAB	SWAN	SWAP	SWAT
SWAY	SWIM	TAB	TACK	TAG	TAIL	TAKE	TALL	TAME	TAMP
TAN	TANK	TAP	TAPE	TAR	TASK	TEA	TELL	TEND	TENT
TERM	TERN	TEST	THAW	THIN	TICK	TIDE	TIE	TILE	TILL
TIME	TIN	TINT	TIP	TOAD	TOE	TOIL	TOLL	TOMB	TON
TONE	TOOL	TOP	TOSS	TOUR	TOW	TOY	TRAP	TRAY	TRIM
TRIP	TUB	TUG	TUNE	TURN	TYPE	TYRE	URN	VALE	VAN
VANE	VASE	VAST	VAT	VEAL	VEIL	VEND	VENT	VEST	VIEW
VINE	VOW	WAD	WADE	WAFT	WAG	WAGE	WAIL	WAIT	WAKE
WALK	WALL	WAN	WANT	WAR	WART	WASP	WAVE	WAX	WAY
WEAK	WEAL	WEAR	WEB	WED	WELL	WELT	WEND	WET	WHIM
WHIP	WICK	WIDE	WIFE	WIG	WILT	WINE	WIPE	WIRE	WISE
WIT	WOE	WOOD	WOOL	WORD	WORN				

(ii) The pool of 106 30% - 40% Archer (1960) CVC trigrams from which stimuli in Exp. III were selected.

BAZ	BEX	BIY	BOV	BUP	BYH	CAJ	CEB	CIQ	COH
CUX	CYF	DAQ	DEQ	DIW	DOJ	DUY	DYV	FAQ	FEG
FIK	FOZ	FUH	FYS	GAK	GEV	GIS	GOK	GUB	GYD
JAF	JEC	JIZ	JOW	JUK	JYL	KAX	KEF	KIG	KOH

KUC	KYC	MAQ	MEP	MIJ	MOF	MUQ	MYH	NAQ	NEZ
NIR	NOF	NUK	NYG	PAZ	PEX	PIW	POB	PUY	PYZ
QAR	QEN	QIX	QOY	QUR	QYL	SAJ	SEH	SIY	SOZ
SUW	SYV	TAH	TEF	TIV	TOZ	TUH	TYX	VAK	VEY
VIB	VOP	VUM	VYL	WAB	WEC	WIQ	WOX	WUM	WYS
ZAB	ZET	ZIR	ZOD	ZUP	ZYG	REH	RIQ	ROJ	RUY
RYS	YAG	YEM	YIK	YOL	YUN.				

APPENDIX 2

Overall proportions of items falling into the 16 responses categories in Exp. II (Note: There were n observations in each spacing condition).

P_1-T_1 LAG	O_1W_1: O_2W_2	O_2C_2	N_2W_2	N_2C_2	n
0	.057	.017	.005	.000	423
4	.328	.142	.005	.000	415
8	.325	.142	.014	.000	422
12	.341	.129	.005	.000	419
16	.299	.187	.007	.002	422

P_1-T_1 LAG	O_1C_1: O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.428	.463	.012	.000
4	.070	.373	.000	.000
8	.062	.332	.005	.000
12	.055	.332	.000	.000
16	.055	.306	.000	.000

P_1-T_1 LAG	N_1W_1: O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.012	.002	.005	.000
4	.060	.012	.002	.000
8	.078	.033	.000	.000
12	.098	.026	.000	.000
16	.095	.036	.000	.000

P_1-T_1 LAG	N_1C_1: O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.000	.000	.000	.000
4	.002	.005	.000	.000
8	.007	.000	.002	.000
12	.014	.000	.000	.000
16	.012	.002	.000	.000

APPENDIX 3

Overall proportions of items falling into the 16 response categories in Exp. III. (Note: there were n = 200 observations in each spacing condition).

O_1W_1

P_1-T_1 LAG	O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.005	.010	.005	.000
2	.200	.130	.030	.005
4	.345	.115	.005	.000
6	.270	.150	.020	,005
8	.330	.145	.020	.015

O_1C_1

P_1-T_1 LAG	O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.485	.310	.110	.035
2	.185	.235	.025	.010
4	.135	.215	.025	.000
6	.100	.200	.000	.010
8	.090	.160	.005	.005

N_1W_1

P_1-T_1 LAG	O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.015	.010	.005	.000
2	.075	.040	,015	.005
4	.075	.020	.015	.000
6	.105	.035	.035	.010
8	.120	.040	.015	.000

N_1C_1

P_1-T_1 LAG	O_2W_2	O_2C_2	N_2W_2	N_2C_2
0	.010	.000	.000	.000
2	.020	.010	.000	.015
4	.025	.015	.010	.000
6	.030	.015	.010	.005
8	.035	.005	.005	.010

APPENDIX 4

Data from Exp. I and the corresponding values predicted by versions I and II of Model 2.

	$Pr(C_1)$			$Pr(C_2)$		
i	Obs	Pred I	Pred II	Obs	Pred I	Pred II
0	.930	.926	.926	.461	.432	.429
1	.593	.611	.609	.525	.534	.532
2	.424	.464	.471	.540	.569	.566
3	.378	.459	.464	.507	.565	.564
4	.480	.453	.457	.549	.562	.561
5	.438	.448	.450	.521	.559	.559
6	.511	.443	.443	.573	.555	.557
8	.442	.432	.429	.597	.549	.552
12	.443	.412	.404	.571	.537	.541
16	.394	.394	.380	.534	.525	.530

	$Pr(C_2/C_1)$			$Pr(C_2/W_1)$		
i	Obs	Pred I	Pred II	Obs	Pred I	Pred II
0	.491	.461	.458	.069	.067	.067
1	.641	.669	.667	.357	.322	.322
2	.758	.854	.840	.380	.322	.322
3	.823	.853	.843	.315	.322	.322
4	.815	.851	.847	.304	.322	.322
5	.875	.850	.850	.246	.322	.322
6	.864	.849	.852	.270	.322	.322
8	.896	.847	.857	.361	.322	.322
12	.885	.842	.865	.322	.322	.322
16	.847	.838	.869	.331	.322	.322