AN EXPERIMENTAL STUDY OF MNEMONIC DEVICES IN VERBAL LEARNING

by P.S. Delin, M.A. (Oxon)

Thesis presented for the degree of Doctor of Philosophy at the University of Adelaide, 1969

TABLE OF CONTENTS

Page No.

Page No.

SUMMARY

The object of this thesis was to study the performance of subjects using mnemonics to learn verbal material, with a view to discovering what sort of relationship might obtain between this sort of learning and the kinds of learning normally exhibited in serial learning experiments.

Six experiments were carried out, using first-year psychology students as subjects, and English nouns as experimental material. The mnemonic technique primarily investigated involved the use of bizarre imagery to connect adjacent pairs of items in serial lists.

In experiment 1 subjects having mnemonic instructions, normal serial anticipation instructions, and free recall instructions and conditions, were compared under two recall conditions, serial anticipation, and free recall. Presentation time was generous (20 seconds per item) and financial reward was used. For both modes of recall mnemonic subjects made less errors (on the single recall trial), than free input subjects, while subjects with standard instructions made most errors. Mnemonic subjects showed no serial position error distribution,

and tended to adhere to serial order recall even under free recall conditions.

In experiment 2 subjects with and without mnemonic instruction learned lists by serial anticipation at an 11 second rate of presentation, to a criterion of one perfect anticipation of the list. In terms of both errors and trials to criterion mnemonic subjects performed better. On the earlier anticipation trials mnemonic subjects showed longer latency to respond than did control subjects. Mnemonic subjects made more of their errors in the form of omissions.

In experiment 3 large groups of subjects were required to provide descriptions of bizarre images designed to link three word-pairs having high, medium, and low inter-item association. 15 weeks later they were asked to try to recall the response terms of the pairs, given the stimulus terms. In the interval the image descriptions had been rated by judges on the extent to which they fulfilled the instructions. These ratings proved to be different for the three word-pairs, and related to correct recall.

An attempt was made in experiment 4 to break down the mnemonic instructions into their component sub-instructions. Groups of subjects received subsets of these instructions ranging, in seven stages, from no mnemonic instruction to the complete set of mnemonic instructions. In addition some subjects received recall instructions and others did not. Each subject received four anticipation trials at a 7 second rate, and returned after 5, 10, or 15 weeks for four relearning trials. Performance on both learning and retention proved to be related to the degree of completeness of the instructions received. On learning, though not on retention, the proportion of the subjects showing more errors of omission than of commission increased with the completeness of the instructions. There was some reason to think that the subjects receiving the full mnemonic instruction had difficulty in carrying it out under the conditions of the experiment.

In experiment 5 subjects with and without mnemonic instruction learned lists of one of two different lengths at a 6 second rate of presentation over four trials. All subjects returned after 6 weeks for four relearning trials. The results showed no interaction between list length and mnemonic instructions, for either learning or retention. There was little sign that mnemonic instructions benefited retention independently from their effect on learning. For both long and short lists the serial position error curve was flatter for mnemonic subjects than for control subjects.

In experiment 6 subjects with serial anticipation instructions learned lists having either high or low meaningfulness and either high or low rated inter-item association. Presentation was at a 4 second rate, and learning was carried to a criterion of one error-free anticipation of the list. Inter-item association proved to have a more powerful positive effect upon learning performance than did meaningfulness. Between groups differences in error type and distribution, and in the subjective reports given by the subjects, were consistent with the hypothesis that differences in list structure affect performance at least partly through their effect upon the learning techniques used by the subjects.

Overall conclusions were that mnemonic instructions improve performance in the learning of serial lists of concrete nouns, and that this improvement is, at least in part, independent of the subjects awareness that he will be required to recall the material, and dependent upon his carrying out of the instructions. Persistent qualitative differences in performance and subjective reports between mnemonic and non-mnemonic subjects, together with similar differences produced by variation in the structure of the material, suggest that there may be a number of modes of learning at the disposal of the subject, which are called into operation by such factors as the nature of the material, the mode of presentation, and the instructions given by the experimenter.

One-trial learning, the nature of the functional stimulus in serial learning, meaningfulness, and mediation were among the topics discussed and reviewed.

DECLARATION

I hereby declare that this thesis does not to my knowledge contain material written or published by another, except where I have so indicated in the text.

I further declare that I have not previously submitted any of the theoretical or experimental material included in this thesis in either partial or full fulfilment of the requirements of any other degree.

Signed (P.S. Delin)

Date 15th February 1969

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Dr. F.R. Dalziel and Professor M.A. Jeeves, for the time they have spent with me, and for their many guiding and motivating suggestions.

I am also most grateful to Dr. A.H. Winefield and Professor S.H. Lovibond for the help and encouragement which they have been ever willing to give me, and to Professor A.T. Welford, who kindly undertook the unpleasant task of reading the manuscript in draft form, and made many helpful suggestions.

Thanks are due also to Mrs. M. Blaber, who transformed my untidy jottings into a neat typescript, and to the rest of my colleagues at Adelaide, for their willingness to listen, in triumph and adversity.

Finally I should like to thank my wife, Jane, without whose patience and forbearance this work would not have been possible.

CHAPTER 1.

Introduction and Description of the Problem.

The study of human memory for verbal material is one of the oldest fields of psychology, and one of the most popular. That this should be so is hardly surprising when one considers the high proportion of human behaviour which is dependent on the prior learning of such material.

Although its obvious relevance to a specifically human psychology has ensured a vast outpouring of both theoretical and experimental work, this area is fraught, as a result of its complexity, with all sorts of research dangers. Not the least of these is the danger, which it shares with many areas of social psychology, that the researcher may find himself talking in terms not amenable to experimental handling.

Perhaps as a response to these dangers two major research tendencies have been evident. The first has been the tendency to restrict study to areas in which it seemed that there was more possibility of controlling all the variables, and the second has been the tendency to formulate theories and to design experiments in such terms that a close relationship can be maintained with the learning work that has been done on animals. Postman (1964, p.358) establishes himself as an exemplar of this approach when he says, "Nevertheless, the formal model used in the analysis of behavior transcends differences among species in capacities and skills".

These tendencies are both understandable and in some ways laudable, but they may also have had unfortunate side-effects. The first of them has resulted in the exclusion from consideration of a number of phenomena which may have much intrinsic interest; such phenomena as eidetic imagery, the effects of mnemonic devices, and the occurrence of memory prodigies. The second tendency may have resulted in the assumption of a much closer correspondence than is really the case between the laboratory learning performance of animals and the real-life performance of human beings. Certainly it has left us unable to say how close this correspondence is.

In spite of the difficulties both practical and social attendant upon such efforts some psychologists have attempted to investigate human learning of "real life" materials such as stories and prose in such a way as to bring out the specifically human aspects of such learning. Notable among these have been Bartlett (1932) and a number who followed in his footsteps, for instance Wees and Line (1937), and Tresselt and Spragg (1941). More recently, Chomsky (1959) has attacked what he sees as being an excessively reductionist approach to linguistic behaviour on the part of Skinner (1957). Gagné (1965) distinguishes eight different kinds of learning in human beings, and makes it clear that he feels that at least one of these kinds of learning is specifically human. Although some of the results, particularly those obtained by using Bartlett's method of serial reproduction, are sufficiently reproducible to warrant their inclusion in the practical class sections of psychology courses, the theoretical problems they raise are daunting enough to confirm most researchers in their determination to work with simpler and more artificial materials and experimental situations.

The effects of this avoidance of difficult areas may be unfortunate. Research may be focussed upon a biassed sample of human behaviour, and theory may reflect this bias. On the other hand it could be that an examination of the more neglected phenomena of human learning would produce nothing that could not be fitted in with modern theories of learning.

A superficial look at everyday learning does suggest some differences from what we are used to observing in the laboratory. In the first place the emphasis on repetition which is so general in laboratory experiments is by no means as obvious outside the laboratory. Indeed, there are many circumstances in which we become quite irritated if people do not remember things which they have been told only once. Secondly there are many people who seem to learn very much more, and to retain it better, than might have been expected to have been even remotely possible on the basis of laboratory experience. Thus Luria (1960) describes the Russian prodigy, Shereshevskii, as retaining perfectly, for periods of up to twenty years, material which had only been presented once. Indeed, he failed to find an instance of Shereshevskii's forgetting anything.

The notion that learning may take place in one trial is not by any means a new one. Guthrie (1935) says in effect that all learning is one-trial learning. Within the field of verbal learning the problem was first brought into prominence by Rock (1957).

Rock used a paired associate task in which items which were not anticipated correctly on a particular trial were replaced on the next trial from a pool of items. He found no advantage for a control group which simply learned the original list to criterion. This result suggested that repeating the items did not facilitate learning to produce the correct response. Rock's experiment was attacked (e.g. Williams, 1961), for methodological reasons, and, in spite of amplificatory experiments by Rock and associates (e.g. Rock and Heimer, 1959; Rock and Steinfeld, 1963), independent discovery of similar phenomena (e.g. Estes, 1960), and numberous experiments by other authors using some variant of Rock's technique (e.g. Lockhead, 1961; Clark, Lansford and Dallenbach, 1960) neither the factual nor the theoretical status of one-trial learning is yet clear (Postman, 1963). Variables which seem to be relevant to the appearance or non-appearance of the one-trial learning effect are type of material (Restle, 1965), and presentation time (Lockhead, 1961). Since Rock's 1957 paper the question has tended to shift from whether learning ever occurs in one trial to whether learning is all-or-none or incremental.

Some theorists may still doubt the laboratory occurrence of one-trial learning, and it may be equally difficult to prove that such learning takes place as a regular feature of everyday life, but there is one type of learning to which the concept of repeated presentation seems quite alien. The learning in question is that making conscious use of mnemonic devices.

The conscious application of mnemonic devices has a long recorded history. Cicero (in De Oratore, II lxxxvi) described himself as having learned his speeches by associating a series of key ideas with a sequence of locations in his house and grounds, and credits the Greek poet Simeonedes with the discovery of "the art of memory". It is interesting to note that Shereshevskii seems spontaneously to have arrived at a similar system. It has been suggested that the expressions, "In the first place", and "In the second place", derive from this system (Weinland, 1957).

It seems that the spatial organisation of material proved, for many people, to be inferior to some kind of logical ordering, the most obvious of these being the attaching of the items to a sequence of numbers. Numbers are also, as items, difficult to remember, so it is not surprising that many systems should have been devised to translate numbers into letters or words, both to facilitate the remembering of digit sequences, and to enable the numbers to be attached to items of different kinds. Leibnitz, in his attempt to create a universal language, made use of a number alphabet invented by

Pierre Herigone (in his book "Cursus Mathematici", 1634) while another of their contemporaries, Stanislaus Mink von Wenaussheim, was responsible for the number alphabet which is used in most commercial memory courses today. Using this type of system one may turn numbers into object names, or into phrases, so that items of other kinds may be remembered by associating them with the resulting key items.

As an example of the simplest type of number-to-object translation system one might take the rhyming system. For instance, if a person wishing to remember a list of words, or a speech, knows in advance a rhyme of the form, 'One is a gun. Two is a shoe. Three is a tree,,,,,', he can associate the first word of his list, or the first point of his speech, with a gun, the second with a shoe, and so on.

Miller, Gallanter, and Pribram (1960) describe a small demonstration of this system, using 'One is a gun', etc. They say (p.134), "The antagonistic attitude of experimental psychologists towards mnemonic devices is even more violent than their attitude towards their subject's word associations; mnemonic devices are immoral tricks suitable only for evil gypsies and stage magicians. As a result of this attitude almost nothing is known by psychologists about the remarkable feats of memory that are so easily performed when you have a Plan ready in advance". One of the aims of this thesis is to do something towards filling this lacuna.

The invention of digit translation systems may be seen as a development of the location systems from the organisational point of view. There have also been developments in techniques for forming associations.

Most of the memory systems commercially available (e.g. Lorayne, 1963; Weinland, 1957; Furst, 1954) emphasise that one should connect the items to the number-objects by making, not merely an association, but an image; and not merely an image, but one which is as absurd, bizarre or fantastic as possible. Some suggest that even where there exists a perfectly natural association, as between "gun" and "bullet", one should reject this in favour of a bizarre one.

Where earlier systems usually referred specifically to "Visualising", most modern systems seem to be referring to "Imagining". Modern systems often also suggest that the connections which one constructs should be vivid and active, and that one should use as many sensory modalities in their construction as possible.

Systems of the type described above are used by most entertainers who display feats of memory. A typical performance would run as follows. The performer is blindfolded. An assistant writes the numbers from 1 to 25 on a blackboard, and asks members of the audience to call out the names of objects. These are written as selected against the sequence of numbers, the assistant saying them aloud as he writes them. The items tend to be written up at an average of about one every six or seven seconds. Once the list is written up members of the audience may call out either a number, when the performer will name the appropriate item, or one of the items, when the performer will give its number in the list.

This type of performance becomes even more impressive when one considers that the performer may learn several lists each day using the same number-items. Furthermore the system has been used to learn lists of several hundred items at the same rate of presentation, and with the same error-free recall.

A simpler system also contained in most commercial memory courses makes use of the bizarre associations but not of the number-digit translation system. Here the subject simply makes his association between each item and the previous item in the list. Thus the list is learned as a series of independent bizarre associations which chain the items together. This system, which will be referred to as the "Link" system, is the one with which we shall be primarily concerned.

Although there has been little actual research on mnemonics a number of psychologists have referred to them. Thus H.B. Reed (1918) discusses associative aids in relation to learning and retention. I.M.L. Hunter (1956) classifies mnemonic systems, and makes an evaluation of their properties.

Rock (1957) sees mnemonics as playing a major role in associative learning. He too expresses concern over the lack of attention psychologists have shown in relation to them. "Most subjects succeed in learning only a few pairs per trial, and many do so with the aid of some mnemonic device. The theoretical significance of the widespread use of such devices in rote learning experiments has not been sufficiently emphasised in the past. The successful use of such devices may mean that an idea suddenly occurs to S which enables him to link two items then and there; it has, to some extent, the character of insightful learning". (pp.191-192).

There are, then, three related reasons for studying mnemonics. The first is that this is an area of learning behaviour about which we know very little, although there is reason to think that mnemonic learning is very widespread. The second is that we need to know whether, at the upper end of the continuum of mnemonic performance, the results really are as inconsistent with laboratory findings on rote learning as they appear to be. Both of these reasons could be subsumed under the rubric of checking to see that we have not been basing our theories on a biassed sample of human learning behaviour. Thirdly, certain researchers, particularly Rock (1957), and Restle (1965) have made use of the notion of mnemonics in discussing one-trial learning as it has appeared in recent laboratory experiments. Our studies should leave us in a better position to say how consistent their ideas are with the facts of mnemonic performance.

CHAPTER 2.

Discussion of the literature on mnemonics.

Although a number of authors mention, describe, or give examples of mnemonics, it is only recently that the climate of opinion in the area of verbal learning has changed in such a direction as to make it likely that laboratory experiments on them would take place. The previous attitude of psychologists is represented in the words of Woodworth and Schlosberg (1954, p.708). They describe attempts on the part of experimenters to set up conditions such that it is not possible for subjects to use mnemonics. They suggest that mnemonics are undesirable because they "make the learning task less uniform and introduce variability and unreliability into the quantitative results. Besides, the experimenter wants to study the formation of new associations, not the subject's clever utilisation of old ones".

Researchers might at any stage have questioned the desirability of excluding from consideration a range of implicitly accepted phenomena on the grounds that they might make the results less tidy, but it was not possible to counter the concluding argument until it was realised that the subject's utilisation of old associations is just what we are concerned with, or at least that we are not in a position to make a distinction between the formation of new associations and the utilisation of old ones.

Hunter (1956) has attempted to classify mnemonic systems. He uses three categories; visual symbol, digit-letter, and successive comparison. Visual symbol systems include the "topological" system described above (p.6), and attributed by Cicero to Simeonides, and Henry Herdson's (17th century) system of associating the first item with a candle, or any 1-shaped object, the second with a swan, or any 2-shaped object, and so on. The defining characteristic of a visual symbol system is that sequence is represented by a succession of predetermined visual symbols with which the items to be remembered are associated.

Digit-letter systems are systems for translating sequences of digits into sequences of letters, usually consonants, which may then be transformed into words. Hunter gives as an example of the use of this type of system Brayshaw's metrical rhymes. These were a series of ingenious mnemonics which made use of rhyme, metre, wit, and a digit-letter code, covering facts relevant to most school subjects of the period (mid-19th century).

The successive comparison method is described by Hunter as being so simple as hardly to merit the name "system". It consists in comparing the first item with the second, and making use of any association technique to combine them into a single unit, then doing the same with the second and third, the third and fourth, and so on.

Hunter remarks on the infrequency with which professional mnemonists make use of either the digit-letter, or the successive comparison methods, which is odd since most memory-courses suggest a combination of all three methods as a basis of the kind of performance described on p.9 of this thesis.

Hunter attempts to account for the success of mnemonics in terms of factors important in ordinary memory; the factors of organisation, observation, and repetition. He subdivides organisation into assimilation-to-schemata, and recoding, under which heading he includes such devices as rhyming and acrostics. He concludes by suggesting that mnemonics have little relevance to real-life problems of memorisation, and by making a distinction between the sort of outstanding memory performance produced by mnemonists, and that produced by memory prodigies like Macaulay, who, he suggests, had "a native capacity for rapid and permanent learning which is out of the ordinary" (p.96).

The primary emphasis in Hunter's article is on the similarities between mnemonic and non-mnemonic learning. He does, however, mention several aspects of mnemonic technique and performance which might be considered suggestive of differences between mnemonic learning and verbal learning as it has traditionally been studied in the laboratory. Thus he refers to the fact that mnemonists using visual symbol methods report no interference between lists, although this situation corresponds to an S1.-R1./S1.-R2. interference paradigm. Again, he points out, and attests to, the effectiveness of the rule of not comparing more than two items at a time; a rule quite foreign to the picture of rote learning as usually observed in the laboratory. Similarly he mentions the suggestion frequently made in memory courses that a mediating association should be as bizarre as possible; a suggestion which looks inconsistent with the usual view of memory as a process of fitting items into pre-existent schemata.

The Link system, with which we shall mainly be concerned, does not fit cleanly into Hunter's scheme, as it has the form of his successive comparisons system but the associations between items are made in a way similar to, though not identical with, that which he describes in relation to the visual symbol approach. The subject is in fact asked to use as many sensory modalities as possible in his image, and not just to make a visual image. Indeed, the word "Imagining" would better describe what he is asked to produce than does the word "Image".

There is some reason to think that it is worth making a distinction between this type of system and those making use of rhymes and acrostics. It is characteristic of the latter that the mnemonic tends to become "fixed", so that the person using it cannot remember the material without making use of the mnemonic. Thus few people can remember how many days there are in a particular month without making use of the rhyme which starts, "Thirty days hath September". On the other hand it is frequently remarked by users of the Link system that it is remarkable how quickly one ceases to use the mnemonics if one repeatedly recalls the material. Weinland (1957) mentions the "fixation" of the mnemonic as a specific criticism of the acrostic approach (p.104). On the other hand Zinchenko (1959) gave his subjects either a "cognitive" or a "mnemonic" orientation, and found that as the learning process proceeded the mnemonic subjects became more similar to the cognitive ones, both in performance and in their introspections.

The earliest piece of research in the literature which is both fairly extensive and specifically concerned with imagery mnemonics is that of Wallace, Turner, and Perkins (1957). This was an Army Signals Engineering Laboratory project carried out with a view to assessing the practical limits of human memory storage capacity.

Six subjects were used. The material consisted of words which were within the vocabularies of the subjects. Some phrases were used, and many names; e.g. "Adam and Eve", "Fagin", "Danny Kaye". The material was presented in the form of paired-associates for one trial only, followed by a recall trial with the order unchanged.

The mnemonic instructions were very simple, the subject being instructed to form a mental image connecting each pair. Initially the subjects were self-paced, but later a 20 sec. time limit was imposed, and each subject also carried out "Speed trials", at presentation times as short as 3 sec.

List length ranged from 25 to 700 pairs, and with lists as long as 500 pairs subjects produced practically error free performance. With 700 items subjects recalled about 660, but it should be taken into account that they would in this case have been working for about $4\frac{1}{2}$ hours, so one might reasonably expect fatigue to have affected their performance.

With lists of 25 items subjects achieved perfect recall using presentation times as low as 3 seconds, but with lists of 50 they could not perform perfectly with less than 5 seconds per item. Clearly although list-length may not play the same role in this sort of learning as it does in ordinary rote learning, it does have some effect.

The authors say little about retention, except to remark that after two or three days subjects could usually recall about 90 items out of a list of 100. It is perhaps unfortunate that they did not have more data on retention. The uniformly high performance on immediate recall may have been obscuring a large amount of variability in individual performances which retention studies would have made apparent, and which might, through error analysis, have given important information about the mnemonic process.

The authors attempt to analyse the task in informational terms. Their analysis makes a number of assumptions which are almost certainly contrary to fact, but which should lead to a highly conservative estimate of the amount of information stored. Thus it is assumed that the subject's possible responses are limited to the list of task words, or of a given list from which they are drawn, as they would be if the item list consisted of, for instance, the names of all the playing cards, or the letters of the alphabet. It is also assumed that all errors are equally likely to occur. Different sorts of errors are not distinguished on this equal-probability assumption.

On this basis the subject who stores 664 items correctly out of a list of 700 is described as having stored 6968.8 bits, and as having lost 604.4. The authors say the fastest information storage rate observed was that achieved by a subject learning 25 items at the rate of one every 3 seconds, which they calculate to represent a rate of 0.5 bits per second. This is incorrect, the actual rate being 1.1 bits per second, but even this is a surprisingly low rate of data processing.

Wallace, Turner, and Perkins conclude that if there is an upper storage limit it is higher than 700 words. They are unable to decide whether the increase in errors with lists longer than 500 words represents the beginning of a levelling-off process.

The concept of a upper limit to human memory storage capacity is one which is difficult to interpret in a way which is meaningful for verbal learning theorists. We do, after all, learn and retain vastly greater amounts of material than those with which Wallace, Turner, and Perkins were concerned. But they were orienting their research around a practical rather than a theoretical question; the question of how big a short-term memory task has to become before it becomes necessary to supplement the human operator's memory with some sort of machine aid.

Smith and Noble (1965) carried out what they described as a quantitative test of the validity of claims made by Furst (1954) for his "Hook" method of memorising. The Hook method consists in using bizarre imagery to associate the items of a list with a series of key items derived from the serial numbers of the items by the use of a digit-letter code.

Experimental and control subjects were matched on a seven point learning ability scale by the use of three tests, spelling, practice list learning performance, and performance on the Reasoning subtest of the Primary Mental Abilities battery (Science Research Assoc., 1958). Experimental subjects were given a 1-hr. lecture-demonstration of the mnemonic technique. They were then sent away to practise the key item list 15 minutes per day for four days. All subjects were then given 20 serial anticipation trials on one of three 10-item CVC lists at a four-second rate of presentation. The CVC lists were of high, medium, and low meaningfulness. After 24 hours all subjects were given 10 relearning trials on the same list.

Furst's method produced significant improvement in recall after 24 hours, and in relearning, but the differences in learning failed to reach significance. The biggest differences were for subjects learning the medium meaningfulness list, and for the low ability subjects. The authors conclude that Furst's method has, "limited usefulness during the acquisition of a serial verbal list, and perhaps a significant facilitating effect on retention, provided the material to be recalled is of medium or low meaningfulness, but no efficacy for remembering highly meaningful materials learned under the present conditions" (p.123).

In fact there were other findings in the study which suggest that this is a summary which does not do full justice to the differences between the experimental and control groups. The best of the three matching variables used was the practice list score. This correlated with learning performance at rho=.76 for the controls, but at only rho=.38 for the experimental subjects, a difference which is significant (p .05). This would have been an interesting finding even if there had been no difference in performance between experimental and control subjects. At the very least it suggests that the instructions were leading the experimental subjects to do something different from what they would have done without instructions.

Again, when the subjects had completed their acquisition trials, they were asked to respond to the serial numbers of the items in 5 different random orders. All experimental subjects were able to do this, but few of the controls. This is in fact the task that Furst's technique is designed for but the experimenters did not analyse this result any further, and did not in fact make any further reference to it.

Irrespective of these comments there is reason to suspect some anomalies in the data that was obtained in the experiment. A diagram on p. 129 presents the percentage of correct responses for each of the 6 groups, plotted against the ordinal number of the trials. From this diagram one can extract such information as, for instance, that the mean subject learning the high meaningfulness list made about 8 errors on his 8th trial. Since these high meaningfulness CVCs were all words it looked as if these subjects had found the task much more difficult than seemed reasonable. Running 8 subjects under conditions very similar to those used by Smith and Noble in their high meaningfulness control group the present writer found that they reached a criterion of one error-free run in 1, 2, 3, 4, 4, 7, 8, and 15 trials respectively, and even this last subject never made more than two errors after the 6th trial. For these subjects the mean percentage of correct responses on the 4th presentation was higher than that produced by Smith and Noble on the 9th trial. Since the performance of subjects in the other control groups appears to be similarly depressed in relation to expectations based on the learning speed of other psychology students under similar conditions it seems unwise to regard Smith and Noble's data as representative.

Olton (1966) investigated the differential effects of a mnemonic on learning and retention. He presented his 5th grade subjects with sentences linking the stimuli to the responses (e.g. The COW chased the red BALL). These sentences were presented before learning, before recall or at both times. He attempted to equate learning in his experimental and control subjects by giving the former six learning trials, while the latter received nine, having arrived at this differential on the basis of pilot studies. A successive probability analysis indicated that this procedure was successful.

The results suggested that when degree of learning was equated the mnemonic did not improve retention. A recall test after one week gave a marginally significant difference in favour of the control subjects. Neither presentation nor representation of the mnemonic just prior to recall had any effect on retention.

It is difficult to interpret this study without knowing what difference it makes to a subject's use of a mnemonic to present it to him rather than allowing him to construct it. The present writer carried out a small study in which he provided subjects with descriptions of imagery mnemonics appropriate to the material. These subjects learned faster than the control subjects, but slower than subjects who constructed their own mnemonics. However, it is possible that this treatment affects imagery mnemonics more than it affects mnemonics of the type used by Olton, and that its effects on retention are different from its effects on learning.

Wood (1966) carried out a series of five experiments on imagery mnemonics. In his first experiment he used seven groups, four of which were given a series of key items, and three of which were not. Of his key-item groups, one was told to use bizarre images, and to rehearse previous images, one was told to use bizarre images, and not to rehearse previous images, one was told to use "common" (that is, "mundane") images, and one was told to use verbal mediators, an example of which might be the connecting of the stimulus 'rope' to the response 'shoe' by evoking the mediating link-word 'lace'. Of the remaining three groups one was told to make images of the items, one to use bizarre images to link the successive items, and one was given standard free recall instructions. This last group served as a control.

The main findings were that all the groups given key items and the group told to make bizarre linking images performed better than the control group, but that instructions as to type of imagery and rehearsal made no difference.

In the second experiment subjects using key items and bizarre imagery were compared with controls at 5 sec. and 2 sec. presentation rates. There was an interaction between method and presentation time such that the differences between the groups in favour of the experimental subjects were greater at the slower rate of presentation.

In all of the remaining experiments groups using key items and bizarre images were compared with groups using key items and verbal mediation. In a negative transfer paradigm, with a high interference list, with a list designed to facilitate mediation, and with a series of lists of words of varying degrees of abstractness no significant differences were found. Wood notes that the presence of peg-words (our 'key items') appears to facilitate performance as compared with a free recall control group, but also notes that a more appropriate control group would have been given standard paired associate instructions. He suspects nevertheless that it is not just the presence of the peg-words, but also the strategy that the subjects use, which is responsible for the facilitation of performance. In another experiment he finds a significant learning-to-learn effect with peg-words, and points out that this is consistent with previous findings in relation to paired associate learning (e.g. Postman & Schwartz, 1964).

The finding that instructions to link successive items by making bizarre images facilitates performance is interpreted by Wood as support for an item-associative view of serial learning, and he shows that the subjects so instructed tend more than the free recall control subjects to adhere in recall to the order in which the items were presented. He does not, however, attribute any facilitative effect to the imagery aspect of the instructions, and in fact does not find evidence in any of his experiments that imagery has a facilitative effect. He does, on the other hand, suggest that it is possible that his subjects did not, or could not, respond to this aspect of the instructions. Uncertainty about this vital aspect of his experiments makes it difficult to arrive at any clear evaluation of their import.

It is possible that the presentation times used by Wood were too short to enable naive subjects to make effective use of the instructions given. Seibel, Lockhart, and Taschman (1967), on the other hand, allowed their subjects to spend as long as they liked building a mnemonic link for each pair in a list of 100 paired associates. In spite of the fact that they only used 8 subjects in each group their experimental subjects performed significantly better than their control subjects. It is, however, worthy of note that their subjects took approximately 1 minute to learn each item which they subsequently recalled correctly. The longest presentation-time Wood used was 5 seconds per item.

Bugelski, Kidd, and Segman (1968) investigated the effectiveness of the, "One is a bun, Two is a shoe" system described in Chapter 1 (p.7). They used three different presentation times: 2 seconds, 4 seconds, and 8 seconds, and had, at each of these rates of presentation, two control groups, of which one had been taught the number-rhymes, but had not been told how to use them, and the other had not been told about them. Recall was tested by reading to the subjects at a 4-second rate a randomly ordered sequence of the numbers from 1 to 10, and requiring the subjects to name the items which had originally been associated with those numbers, 2 points were given for a correct response, and 1 for an intralist intrusion. The maximum possible score was thus 20. All subjects had initially learned a list without mnemonic instructions, but otherwise under the same conditions.

The results showed that the mnemonic instructions did not facilitate performance at the 2-second rate of presentation, but did so at the 4 and 8-second rates. At the 8-second rate most experimental subjects made perfect scores. Bugelski et al estimate that it takes most subjects between 4 and 8 seconds to implement the instructions. The control group which had been taught the number rhymes did not perform any better than the other control group, and the authors suggest there is some reason to think this pre-training actually interfered with their performance to some degree. Analysis of the results on the first list (before mnemonic instructions had been given) showed that presentation time per se was not very potent in its effect upon performance. There was no significant difference between the 2-second and 4-second groups, and the 8-second group, while better than the others, was not strikingly better. Bugelski et al recognise that they have not proved that imagery (as distinct from other types of mediator) was responsible for the improved performance shown by the experimental subjects. However, they point out that the subjects made it evident that they ascribed their success to the use of images. They also point out that the control subjects made attempts to use other types of mediator, but were, to judge from their performance, unsuccessful.

Although the experimental literature on mnemonics is still rather sparse, some generalisations seem to be justified. Positive findings appear to be associated with the use of words rather than nonsense syllables, and with presentation times per item of 4 seconds or longer. When these conditions apply unselected subjects seem to have little difficulty in using either the link system or systems involving key-words. In general the subjects as well as the experimenters appear to be surprised by the levels of performance achieved.

The research directly oriented towards mnemonics need not necessarily be our only source of information on the subject, for there is another field which is growing very fast in importance, and which is so closely related to the one under discussion as probably to be co-extensive with it. The field in question involves the experimental control and analysis of the conceptual manipulations and cognitive and perceptual reorganisation which a subject in a verbal learning task consciously applies to the material in order to facilitate his learning of it. The literature in this area will be considered in Chapter 7.

CHAPTER 3.

Methodological and theoretical considerations for research on mnemonics. I General.

In his review of the current state of the field of verbal learning Mandler (1967) places great stress on the growing awareness among researchers of the active role that the subject plays in the learning process. This trend is attested to by work on coding (e.g. Underwood and Erlebacher, 1965), and on natural language mediation (e.g. Montague, Adams, and Kiess, 1966; Rundquist, 1965), as well as by work on mnemonics such as that here reported.

General Research Strategy.

Researchers interested in the effects upon performance of the subject's manipulations of the material presented have used a number of different research strategies. Perhaps the simplest strategy is that used by Martin, Boersma and Cox (1965). Essentially this consists in asking subjects to learn the material and then obtaining information from them as to the techniques which they used in learning it. This information is then related to indices of performance. This approach is not suitable to the investigation of a highly specific learning technique unless one can specify a type of material or mode of presentation which will elicit the technique in question with some degree of reliability.

Probably the most sophisticated strategy is that used by Underwood and Erlebacher (1965). Their subjects learned material which was difficult to learn without coding, and which was most amenable to coding in a particular way. Pretraining was used to set the subjects to use the appropriate coding techniques, and error analysis to check that the subjects had done so. Coding of the type studied by Underwood and Erlebacher offers far fewer dimensions for individual variation than does the type of mnemonic considered here. It is also the case that we do not have a clear enough theoretical understanding of the link system to guide us in the selection or construction of the stimulus material in the way required by this research strategy.

A third approach, and the one which has been used most generally in the area of research under discussion, is to give the subjects instructions as to the type of techniques which they are to use in learning the material. Workers who have used this method include Wallace, Turner and Perkins (1957), Rundquist and Farley (1964), and McNulty (1966). The method involves some only partially justified assumptions and some difficulties of interpretation, but nevertheless appears to be the most practical if not the only practicable way to deal with the subject matter of this thesis.

The first assumption of the method is that the learning techniques under discussion are ones which the subjects are capable of consciously applying. Informal observation of people trying the techniques for the first time suggests that this is an assumption which is in some degree justified. Another assumption is that it is possible to motivate subjects to follow the instructions. Again informal evidence in the form of expressed interest on the part of subjects suggests some degree of justification. A third assumption is that the instructions will be interpreted similarly by all of the subjects. This assumption appears rather more questionable. One can, of course, keep some sort of a check on these factors by supplementing the instructions with examples and by asking questions after the test-session. Error analysis too will provide some degree of check. But it is inevitable with this approach that, in the event of negative results being obtained, the researcher will have some difficulty in deciding whether they were due to the falsity of his experimental hypothesis, to a failure to get his instructions implemented, or to contamination of the control groups by subjects who are adopting similar techniques to those specified in the instructions to the experimental groups.

Material to be used in lists.

In the research that has been carried out relating to the active role of the subject engaged in a verbal learning task there has been a tendency for experimenters to use "traditional" verbal learning materials, and this has sometimes been the case even when the rationale of the experiment seemed to demand the use of more "natural" materials. Thus Smith and Noble (1965) used CVC triads although their experiment was explicitly intended to test the claims of Bruno Furst's memory system; a system intended for use with lists of unconnected object names.

In the research aimed at investigating natural language mediation most experimenters have been to some degree interested in demonstrating the occurrence of processes overlooked in the more traditional memory experiments. It is not therefore surprising that they have wanted to use the same sorts of material as those used by previous verbal learning experimenters. At the same time the term 'natural language mediation' does seem to imply some sort of connection with real life memorisation, and it therefore seems unfortunate that we have so little data which involves real life material. Even paired adjectives, as used by Rundquist (1965), and Rundquist and Farley (1964), impose some restrictions on the kinds of mediator easily used by the subject, while the use of CVC triads, as by Groninger (1966), imposes so many restrictions that very little generality may be assumed for his results.

Considerable doubt has been expressed recently, for instance by Deese (1961), as to the advantages of so-called 'nonsense' materials. But even if these doubts were unjustified the use of such materials in the work on mnemonics and on natural language mediation would seem odd, since these are just the sorts of phenomena that the materials in question were designed to prevent.

Our basic concern is in any case with the link system as applied to the material to which it is most appropriate. This material consists of unrelated object names, and its use has the advantages of being consistent with the above discussion, and of facilitating comparison with the most relevant previous studies, for example, that of Wallace, Turner and Perkins (1957).

Paired Associates versus Serial Learning.

Since Bousfield's (1953) paper on associative clustering there has been a rapid growth of interest in free recall learning, and there is some reason to think that the free recall situation introduces less artificiality into the learning process than does either the paired-associate or the serial learning paradigm. Unfortunately, however, to present material and test learning in the free recall mode, using link system instructions, would obscure some of the most interesting attributes of the system. For instance, users of the system claim that ordering errors do not occur when material has been learned in this way. It would in any case be difficult to devise adequate instructions for use with the link method in the free recall mode.

Paired-associate tasks appear more often in the recent verbal learning literature than do serial ones. One reason for this is the fact that with paired-associates it is possible to keep some degree of separate control of stimulus and response factors, since the stimulus item as presented may be treated as the nominal stimulus. Starting with the work of Mattocks (reported in Underwood and Schulz, 1960), some evidence has accrued, as reported by Underwood (1963), which suggests that the functional stimulus is usually some fraction of the nominal stimulus.

The situation in serial learning is by no means as clear. There is no specifiable nominal stimulus, and the evidence does not support the obvious hypothesis that the functional stimulus is, for each item, the previous item. In their important review of the matter Jensen and Rohwer (1965) find evidence against both this hypothesis and the hypothesis that the serial position of the item is the functional stimulus. Jensen's own hypothesis, that a serial list is learned as a single complex response, has not been tested, and it is difficult to see how it could be tested. Whatever the outcome of this theoretical conundrum may be, it provides us with a clear reason for studying serial learning rather than paired associates. The one finding which emerges clearly from experimentation on the nature of the functional stimulus in serial learning is that the previous item does not usually fulfil this role. The link system instructions clearly suggest that the subject is to attempt to associate each item with the previous one, and people using the system with apparent success seem to think that they are doing this. If we can demonstrate that they are learning in this way, we shall have found one difference between mnemonic and non-mnemonic learning.

Other reasons for studying serial learning rather than paired associates are that people making claims for the system usually make these claims in relation to the former mode of presentation rather than the latter, and that, as in the case of free recall, a paired associate approach would make it difficult to test the claim that ordering errors are eliminated by the use of the link system. It is also worthy of consideration that although the paired associate task looks to be theoretically simpler than the serial learning one, part of this apparent simplicity may be illusory. The mere fact that item input time goes up with the length of the list suggests that individual pairings are not learned independently as a naive approach might anticipate, while experiments showing that a prior ordering of the stimulus items in the paired associate task produces a serial position effect (Ebenholtz, 1966) suggest a complex internal organisation to the task itself.

Naive versus trained subjects.

One way of avoiding some of the problems of the instructions approach would be to use mnemonic subjects who had had sufficient training that one could be reasonably sure that they had understood the instructions and reached asymptote in their ability to apply them. However, for the purposes of the present research this approach would introduce as many problems as it solved.

In the first place there are practical limits on the length of time for which subjects are available. If, as in the case of Wallace, Turner, and Perkins (1957), only a few subjects had been required, volunteers could have been used. But most of the work for this thesis was likely to take the form of group comparisons, using statistical analysis, and hence much larger numbers. The only subjects available in sufficient numbers were first year psychology students, and these were not usually available for longer than about two hours each.

Secondly, even if subjects had been available in sufficient numbers, and for long enough, there are objections to the use of pretraining to asymptote. The most cogent of these again reflects the comparative nature of the research. If control subjects were given equivalent amounts of pretraining, in terms either of time or of number of lists learned, the assumption that the pretraining was in fact equivalent for the control and experimental groups might easily be unwarranted; especially if there is a real difference in process between the groups. Thus one would hardly maintain that two hours, or two miles, of training at driving a car was in any important way equivalent to the same amount of training at riding a bicycle.

Thirdly, both the results of Wallace, Turner, and Perkins, and those inferable from informal reports of the work of mnemonists suggest that the trained mnemonist makes very few errors of any kind. But in the present research it was hoped that analysis of the differences in type and distribution of errors between experimental and control subjects would be an important source of information.

Discussion of some suggested explanations of mnemonic superiority.

Although it has yet to be demonstrated that, under satisfactorily controlled conditions, subjects using the link system really do perform better than do control subjects, there does exist sufficient informal evidence to make this a very reasonable expectation. Furthermore, from the point of view of the planning of future research, it seems worthwhile to consider what sorts of explanation might be put forward should such positive results arise.

Let us then assume that a comparison has been made between subjects with and without mnemonic instructions, and, given the same presentation conditions, the instructed subjects have learned the list in significantly fewer presentations than did the controls. What explanations of this might one entertain, and what implications would they have for further research? It is worth consideration that the effect might result from the combination of a number of mechanisms. One possibility is that, as was observed in the well-known Hawthorn experiment (see Roethlisberger and Dixon, 1940), the experimental subjects operate with higher morale and expectations as a result of the extra attention which they receive from the experimenter. Acceptance of this suggestion would depend on a number of considerations. Firstly it would depend on the size of the difference between the two groups. Secondly it would depend on the existence of other evidence relevant to the question of differential motivation of the groups. Thirdly one would have to consider whether there were between groups differences in type of distribution of errors, since it would be difficult for an explanation in motivational terms to account for such differences. The subjective reports of the subjects would also be relevant.

Further experiments might examine the effect on performance of changing the instructions in order to produce variations in the subject's expectations of their effectiveness.

A similar suggestion would be that the situation produced differences in achievement motivation between the subjects. Error analysis and subjective reports would be relevant to this suggestion too, as would any additional indicators of motivational strength. One might study the effects on performance of a financial incentive, though it is worth pointing out that undergraduate subjects in verbal learning tasks usually seem to be very highly motivated, even without mnemonic instructions.

It could be that the mnemonic subjects are receiving additional reward from some aspect of the task, for instance, from its növelty (Berlyne, 1960). The implications of this suggestion are similar to those of the other motivational suggestions. One might try the effect of a financial incentive offered to the controls in order to balance out such extra rewards for the experimental subjects.

A different kind of suggestion is that the instructions force the subjects to attend to the items more fully, or for longer, than is the case with the control subjects. Apart from the usual considerations of whether all of the aspects of the differences between experimental and control subjects are accounted for by this hypothesis, this is a rather difficult one to test. One possibility might be to test the thresholds of experimental and control subjects for response to stimuli extraneous to the learning task.

A related suggestion is that although the mnemonic subject may need less trials, his total presentation time is the same as that required by the control subject. This explanation would not apply to our imaginary experiment, since all subjects would have had the same presentation time, but it might well be worth entertaining in relation to some other studies, for instance, that of Wallace, Turner, and Perkins (1957). The wide applicability of the rule that total time taken to learn an item is constant, irrespective of the length of the presentation time on each trial (Bugelski, 1962), raises the question of whether the subjects of Wallace, Turner, and Perkins were, when they were taking more than 20 seconds to learn items, doing so any more quickly, in terms of total presentation time, than control subjects would have done at a two second presentation rate. It is, however, worth noting that these same subjects, after some practice, were to learn lists of similar items at rates of about 5 seconds per item, which is better than one would have expected control subjects to do, in terms of total presentation time per item.

One way of arriving at hypotheses to account for mnemonic superiority is to postulate that the instructions that have been arrived at by the various mnemonic 'authorities' do have direct relevance to the performance the subject produces, and to examine the instructions, component by component, attempting to assess what the effect of the components might be. Thus, in the version of the link system taught by Bruno Furst, it is suggested that the subject deals with the items in overlapping consecutive pairs, and that he makes images connecting the two items in each pair. It is suggested that these images should be vivid, active, and bizarre, and that they should invoke as many sensory modalities as possible.

As remarked above, these instructions explicitly suggest that the subject should use each item as the stimulus to which the next item is the response. This alone, if carried out by the subject, would distinguish mnemonic learning from rote learning as usually studied in the laboratory. Secondly, if the instructions are carried out, the information as to the correct ordering of the items will have been neatly encoded in the image sequence. At any given point in the recall procedure the only item which the system is likely to bring to mind is the 'next' one. Now from the point of view of information analysis this represents a considerable gain over a method of storage in which ordering information was stored to any extent separately from information about what the items were. For while information about the nature of the items increases as a linear function of the number of items in the list, information about their order increases as the factorial of list length. If this is part of the explanation of mnemonic superiority we should find that for mnemonic subjects increasing the length of the list should not increase item-input-time as much as it does for control subjects. Although Wallace, Turner and Perkins did not have any control subjects it is worth noting here that their subjects learned lists containing hundreds of items at rates not very different from those at which they learned lists of 25 items.

Consideration of instruction components may suggest also all or any of the alternative hypotheses that were considered above. Thus it would be difficult to carry out the instructions without attending to the items in a way which is not forced on the control subjects. The experimental subjects may indeed be forced to attend to each item for longer. It would also seem quite reasonable to suggest that the creation and contemplation of vivid, bizarre, active images is rewarding.

Another hypothesis which is particularly relevant to the multi-modality instruction is that the mnemonic instruction forces the subject to set up multiple connections between pairs of items, as against a relatively simple response tendency induced in the control subjects. Thus if the subject envisages a pig that is made into a piano he may set up associations between the smell of the pig and the smell of varnish, the movement of his fingers on the keys and the bucking about of the pig, the sound of the piano and the squeals and grunts of the pig, and so on.

Another source of hypotheses is the subjective reports of the people applying the mnemonic instructions. These tend, on the whole, to suggest that the subject feels as if he is doing something very different from what he would normally do when asked to learn a list. Consider comments like, "It doesn't feel like learning at all", and, "It's like magic". Many subjects emphasise a curious quasi-perceptual aspect of the method. One subject expressed this by saying, "Usually I remember things because they fit in with each other. This way they stand out instead". This kind of comment might give rise to the suggestion that there are two different kinds of memory, one of which has hitherto received the exclusive attention of psychologists. It is the case, after all, that although we may remember a name apparently because it is very common, we may also remember it apparently because it is unusual. It may indeed often seem as if we have forgotten the name because it was common. If the link method really does work by means of images, it may be even less surprising to find that this contrast element is present. An image is easily interpreted as a quasi-perception, and perception is very dependent on contrast effects. Similarly, perception is very dependent on changes in the stimulus situation, and this could be an explanation for the presence of the activity component in the mnemonic instruction. Consistent with these remarks, and with what most mnemonic subjects seem to think they are doing, would be the following model of the mnemonic process. In the input phase the subject does not make a conscious attempt to memorise anything. He simply makes his image according to the instructions. This image, being highly complex, containing numerous highly integrated connections between the items, and perhaps having a reward attached to its construction, is a fairly robust psychological unit.

In the recall phase the subject starts with one of the items of the pair and asks himself what images he made which incorporated it. Unless the item in question is the first or the last item the subject actually made two images which included it, but one of these has already been used to recall the item itself. In the case of any fairly common or familiar object there will be a whole field of possible images which might spring to the subject's mind at this stage, but the one he made in learning the list has at least two features which are likely to make it more salient. Firstly it is more recent, and secondly its bizarreness and vividness will make it contrast with any other images which might arise in the mind of the subject. If the right image does recur it may have lost some detail through forgetting, but because of the multiplicity of the connections between them the two main objects are likely to have been retained. All the subject has now to do is to recognise the second of these two objects.

This model clearly makes use of terms and concepts which would be unacceptable to most theorists in the field of verbal learning, but might nevertheless prove to be a useful basis for discussion, and perhaps for prediction too. It is in any case not inconsistent with the simultaneous operation of some of the other mechanisms suggested above.

The above model does suggest some pitfalls into which the less experienced mnemonist may fall, and the data often give the appearance of showing examples of these pitfalls in operation. Thus the subject may not make an image at all, and may in consequence not be able to make any response at a particular point in the recall of the list. For mnemonic subjects the gaps in their responses on the first recall run do appear to show some correspondence with their admissions that they were unable to make an image for particular item pairs. Again, the subject may introduce some extra-list object into his image and, in the recall phase, misrecognise this as the list object. This too does seem to occur. The subject may, on his first list, misunderstand the instructions, and only connect alternate pairs of items. This should lead to an alternation of correct with incorrect, mainly null, responses in the recall phase of the task. It did in fact prove possible to identify, from the result sheets, the subjects who had made this particular error.

Whatever model or theory of the operation of mnemonics we may decide to test, it must be emphasized that we cannot expect that we will have in any of our subject groups an uncontaminated sample on which to test it. Some subjects will fail to carry out the instructions entirely or in part. Subjects who do carry out the instructions may not do so for all items. Subjects who are not instructed to do so may use mnemonic techniques having all degrees of similarity to those under test. It may even be that the nature of the material to some extent determines the use of mnemonic techniques by subjects, even without instructions or specific previous training. Analysis and interpretation of the results of experiments will therefore have to take account of these contaminating influences, and of the possibility that some of the sub-phenomena under consideration may occur spontaneously, in the absence of mnemonic instruction.

CHAPTER 4.

Methodological and theoretical considerations for research on mnemonics <u>II</u>. One-trial learning.

One of the most striking aspects of mnemonic performance as reported in the experimental literature and as demonstrated by professional mnemonists is the absence of the emphasis on repetition which is so salient in the literature on rote learning. Mnemonists typically give the impression that they do not need more than one presentation of the material.

One-trial learning has been the topic of a major controversy within the field of verbal learning in recent years, since the publication of Rock's (1957) classic study. No clear conclusion has emerged from the studies that have been done since, and one reason for this is that there is not just one one-trial hypothesis, but a number of alternatives, and the different experimenters and discussants have often been at cross-purposes. Clearly the question as to whether subjects ever produce correct responses after one presentation must be answered in the affirmative, since, even in conventional rote learning experiments, some correct responses are usually produced after the first presentation. Equally clearly learning is not always one-trial in this sense. Because of the vacuity of both of these propositions it soon became evident that the real issue was not one-trial versus many-trial, but all-or-none versus incremental learning.

Rock makes a distinction between the formation and the strengthening of an association, and suggests that the former is an all-or-none process. He is explicit in assigning a major role in this process to mnemonic devices, though he does not necessarily, if at all, have in mind the kinds of device we are considering here. His work has mainly been criticised on methodological grounds, and particularly on the grounds that the exchanging of unlearned items for new ones is a form of stimulus selection, and will result in the experimental subjects learning a list which is easier than that learned by the control subjects (Postman, 1963). Several experimenters (e.g. Williams, 1961; Underwood, Rehula, and Keppel, 1962), have produced evidence that such item selection does in fact occur under the conditions of Rock's study. On the other hand it is not clear (Rock and Steinfeld, 1963) that such stimulus selection does entirely account for the findings.

Another approach has been that of Estes (1960, 1961). Estes defines learning in terms of the probability of a correct response, and reinforcement in terms of the paired presentation of the stimulus and the response. In what he termed his "miniature experiment", the subjects were presented with the stimulus and response items (a reinforced presentation) followed by two test presentations of the stimulus item alone, the response being required of the subject. This procedure has become known as the RTT paradigm.

The strength of a learned association between a stimulus and a response can, according to Estes, be represented by the probability that the stimulus will, when presented, elicit the response. On the incremental view this probability can, for a particular subject and a particular item-pair, take any value between 0.0 and 1.0. On the all-or-none view 0.0 and 1.0 are the only values this probability can take. In Estes' experiment the proportion of correct responses on the first test-trial can then be taken either as being an estimate of the average strength of learning per subject per item, or as being an estimate of the average proportion of subjects who have learned the average item perfectly.

The second test-trial should make it possible to test these two alternatives. If the all-or-none view is correct items which were recalled correctly on the first trial should be recalled correctly again, while those not recalled the first time should not be recalled the second time. The incremental hypothesis implies that the probability that a particular subject will recall a particular item correctly the second time is equal to the proportion of correct responses made overall on the first trial irrespective of whether the particular subject recalled that particular item correctly the first time or not. Estes found that 71% of items correct on the first test-trial were correct on the second test-trial, while only 9% of items not correct on the first test-trial were correct on the second. He interpreted his result as favouring the all-or-none view.

However, as Postman (1963) points out, Estes assumed equality of difficulty in his items, and equality of learning ability in his subjects. He also assumed that weakly-learned items would be forgotten at the same rate as strongly-learned ones, and that correct responses are not strengthened by their occurrence on the first trial. Consideration of any of these possibilities makes the incremental view more tenable in the face of Estes' data. Postman (1963) also produces some experimental evidence which appears inconsistent with the position of Estes, by finding that the probability of producing a correct response on the second test trial after an incorrect response on the first test trial is a function of the number of prior reinforcements, in Estes' rather odd sense of this term (i.e. prior presentations of the stimulus item followed by the response item).

Underwood and Keppel (1962) also produce data which appear to be inconsistent with the all-or-none position. They also suggest that the data originally produced by Estes (1960), and confirmed by Estes, Hopkins, and Crothers (1960), can be accounted for by an incremental theory of the Hullian type, having separate concepts of associative strength and of performance, together with some such notion as that of "oscillation" of the "reaction threshold". Although Postman (1963) makes use of the data obtained by Underwood and Keppel (1962), he does not seem to have found their theoretical position congenial, with its implication of a learning model with more than one stage, for he consistently treats the notions of strength of association and degree of learning as equivalent. Miller (1963), in his reply to Postman's paper, points out that it is possible to envisage multi-stage models of learning, and to ask in relation to each stage whether it is incremental or all-or-none in its operation.

Although the all-or-none versus incremental learning controversy still continues, it has now to some degree changed its character, and seems to have become an aspect of multi-stage theories of verbal learning (see e.g. Restle, 1965). The stimulation of the development of such theories is probably the most important outcome of the controversy, since once it is recognised that the learning process may have more than one stage, or facet, it becomes possible to ask other questions about these facets than whether they are all-or-none or incremental. It also becomes possible to distinguish between, and make distinct predictions about, tasks which prior to this development would not have been in distinct categories at all. Thus, once one has become aware, as did Underwood and Schulz (1960), that there is a distinction to be made between response learning and associative learning, it becomes possible to ask which if either of these is all-or-none, whether the same forgetting principles apply to them, whether reinforcement principles apply in a similar way to both of them, and so on.

To return to a consideration of mnemonics, it is clear that no amount of evidence as to the one-trial nature of mnemonic learning would constitute evidence that it was an all-or-none process, in Postman's sense. It might even be maintained that the process carried out by the successful mnemonist represents a rapid overlearning, by an incremental process, of an association between the stimulus and response items. On the other hand, it is the case that the type of learning under consideration is almost pure associative learning, in the sense of Underwood and Schulz (1960), and it is in relation to this kind of learning that the all-or-none hypothesis is best supported.

There are, however, some conceptual problems associated with dealing with mnemonic devices, and there is some temptation to step outside of the non-introspective and non-purposive vocabulary created by the psychologists of the last five decades. Thus as long as a theorist such as Postman is thinking of association in terms of the strength of some kind of internal connection built up between two items it clearly admits of degrees of strength. But once the notion is raised of conscious, deliberate, connections created between the two items, then the subject either does make them or he doesn't. And this notion is in fact raised by the mere use of the term 'mnemonic devices'. In this sense there clearly is at least one aspect of mnemonic performance which is all-or-none. Subjects do say of a given pair of items that they either did or did not make an association between them. Now if the rules which the subject has been given for the formation of such associations are good ones, in the sense that if correctly carried out they will mediate the recall of the items over the time interval involved in the experiment, and if the subject does carry out the rules, then in the sense we have been discussing, the process is all-or-none.

This frame of reference suggests other all-or-none questions. Thus one might ask whether the creation of an association which does not mediate the recall of the response item facilitates the future creation of an association which will mediate a response. One may also ask whether the item pairs which the subject did not associate at all on the first presentation are subsequently easier to associate than they were the first time. If the answers to either or both of these questions are in the negative, then it would seem that there are additional senses in which the mnemonic process is all-or-none.

Clearly questions like that of whether the subject did or did not make an association are difficult to ask with any assurance of achieving reliable replies, but it is important to distinguish between questions which are scientifically meaningless, and those which are merely experimentally difficult to answer. And there is some reason to think that the climate of opinion in psychology is at present moving in such a direction as to place the sort of question raised in the above discussion in the latter category. The study by Mattocks reported in Underwood and Schulz (1960); Farber's (1963) article entitled "The things people say to themselves", and the wealth of recent studies on natural language mediators (e.g. Bugelski, 1962; Dean and Martin, 1966), making use of the subjects' reports on their thought processes, all suggest an increased tendency to ask this sort of question.

CHAPTER 5.

Methodological and theoretical considerations for research on mnemonics. III The functional stimulus in serial learning.

When the subject makes a response in a serial learning task it seems reasonable, given an orientation towards S-R theory, to ask just what it is that he is responding to. Since a high proportion of early work in the field of verbal learning made use of serial tasks, and since most of this work took place during an era in which S-R theory was at its most influential, it is strange that this question was hardly ever asked. Indeed, as late as 1952 McGeoch and Irion were able to treat it as self-evident (pp.89-90) that the subject learns to produce each item as the response to the previous one, so that each item is first a response, and then a stimulus. Since 1952 a large amount of research suggests that this assumption is seldom if ever warranted. But even prior to this date there had been evidence which ran counter to the assumption of McGeoch and Irion. Primoff (1938) found that when subjects were required to learn a paired-associate list in which the items were used twice, once as stimuli, and once as responses, these 'double function' lists took about twice as many trials to learn as did ordinary paired-associate lists. Primoff's hypothesis was that backward associations were interfering with the occurrence of the correct responses, and he presented evidence that this was at least part of the answer. However, there is ample evidence that a serial task is easier than a paired-associate task of equal length, and not more difficult, as Primoff's result might suggest. When a paired-associate list is presented in constant order it is usually, though not always (see Martin and Saltz, 1963), easier than when presented in the usual varied order, and this might be explained in part by Underwood's (1963) observation that the subject tends under these conditions to learn the responses as a serial list.

Although the results of Primoff, and the confirmatory results of Young (1961), are not crucial to the hypothesis that the previous item is the functional stimulus in serial learning (referred to by Young as the 'specificity hypothesis'), they do make it clear that it is worth considering alternative hypotheses. Indeed, as early as 1920, Woodworth and Poffenberger suggested one alternative, namely that the subject associates each item with its serial position in the list.

The work of Mattocks (see Underwood and Schulz, 1960; Underwood, 1963), on the nature of the functional stimulus in paired-associate learning also carried implications relevant to serial learning, and a number of experimenters have since addressed themselves to this problem. There has been some tendency for the assumption to be made that whatever answer is arrived at will apply to all serial learning, and it is important to note that this assumption has neither theoretical nor empirical grounds to support it.

Many of the experiments in this area have involved transfer from serial learning to paired-associates, or vice-versa. Transfer is usually measured in terms of the percentage reduction in the errors made in learning one list as a result of having previously learned another. The primary intention has been to test variants of the specificity hypothesis. In its simplest form this hypothesis implies that transfer should be more or less perfect both from a serial task to a paired-associate task containing the same associations and from a paired-associate task to a serial task containing the same associations. In practice the results in these two situations have been quite disparate. Generally a moderate degree of transfer has been found from paired associates to serial learning. Primoff (1938) found 35%, and Young (1959) found 55% transfer. These levels of transfer, although fairly high, are not as high as one would expect if the specificity hypothesis were correct.

Studies on transfer from serial learning to paired-associates have, on the other hand, generally found that little or no transfer took place. Thus Young (1959) found 8% transfer which was not statistically significant. Even when the serial list was over-learned Young (1962) did not find a significant degree of positive transfer. In an important review of the problem Jensen and Rohwer (1965) conclude that "the literature contains no bona fide demonstrations of significant positive transfer in the serial to paired-associate situation" (p.65).

The specificity hypothesis having found so little support some researchers have considered a modification of it, the compound stimulus hypothesis, according to which the functional stimulus consists of two or more of the items immediately preceding the response item. Results in relation to this hypothesis too have been equivocal. Thus Young (1962) found a small amount of negative transfer when the stimulus in the paired-associate task consisted of two successive items from a previously learned serial list. Horowitz and Izawa (1963), on the other hand, found some degree of support for this hypothesis when the items did not have associates in common.

Another hypothesis which has received some attention is the serial-position hypothesis initially suggested by Woodworth and Poffenberger (1920). In general this hypothesis has been tested by using as an experimental list some arrangement of a list which has been learned serially by the subjects. Thus if a list has been learned serially and is then presented in the reverse order the items central in the list will have had their order disturbed less than those at the ends of the list, and therefore, according to the serial-position hypothesis, there should be more positive transfer on these items. Young, Patterson, and Benson (1963) found some support for the hypothesis from this type of experiment, while Ebenholtz (1963) found that as long as the original items retained their first-list order, positive transfer was obtained, even when new items were interspersed with them. Though this does not represent unequivocal support for the serial-position hypothesis, it is clearly easier to account for under that hypothesis than under either the specificity or compound stimuli hypotheses.

A number of researchers have suggested, on the basis of the analysis of sequential changes in the behaviour of the subjects through the learning process, or of the distribution of errors through the lists, that more than one of the above hypotheses may apply to the learning of a single list. Thus Ebenholtz (1963) found specificity operating in the middle of the list, but position at the ends. On the other hand, Young (1962), and Young and Clark (1964), had results which suggested serial position as the basis of the functional stimulus in the middle of the list, but the specific prior item at the ends. Battig, Brown and Schild (1964) find the functional stimulus to be complex in the middle of the list but simpler at the ends, and especially at the beginning. They suggest that the question of which of the possible types of functional stimulus will be used will depend on the level of difficulty of the list, and go on to argue that with low meaningful material of the type they use (consonant bigrams) high intralist interference caused by stimulus and response generalisation should be expected to lead to a decrease in the applicability of simple functional stimuli.

Jensen (1962, a,b) suggests that there may be no functional stimulus in serial learning, but that the list may be learned as a single integrated response, like the trill of a pianist. Underwood (1961) had suggested

(1963, p.43), that although the positive transfer that had been obtained in the paired-associates to serial learning paradigm showed that subjects could learn a serial list as a series of paired-associates there was reason to think that this was not natural to the subjects. On the basis of this suggestion Jensen (1962,b) set out to discover whether the amount of transfer from paired-associate to serial learning was a function of the degree to which the experimenter made it easy for the subject to retain the paired-associate learning set during the serial task. The findings being positive Jensen concluded that Underwood's suggestion was correct. Furthermore, given that the evidence for the serial-position hypothesis was equivocal, the displacement of the specificity hypothesis and its variants seemed to Jensen to exhaust the possible candidates for the role of functional stimulus in serial learning. Jensen's own integrated response hypothesis attempts to avoid the problem by suggesting that, rather than the individual items being stimuli to which a response is to be made, they are reinforcers of elements in a total response (the recall of the entire list) which is integrated by some central process.

This hypothesis was restated in the context of a general review of the functional stimulus problem by Jensen and Rohwer (1965), but even after a three year delay they do not arrive at any suggestions as to how their hypothesis might be tested, or indeed, what consequences its truth would have, apart from the consequence that experimenters looking for a functional stimulus will be unsuccessful. Although it is by no means a clear cut argument against their position it is worth pointing out that other hypotheses about the nature of the functional stimulus do not entirely lack experimental support. On the contrary, there are experiments lending some support to each of the hypotheses, a situation which might be coped with by suggesting, not that they are all wrong, but rather that there are circumstances in which each of them is right. Certainly this would seem a suggestion preferable to the putting forward of an alternative hypothesis as experimentally intractable as that of Jensen.

The one conclusion upon which most of the researchers in this area seem to agree is that the functional stimulus is not, in general, the previous item, or any part of the previous item. However as we pointed out in Chapter 3 (p.37) the mnemonic instructions which are under consideration in this dissertation seem to require the subject to produce each item as a response to the previous one, and at least some of the subjects seem to think that they are following these instructions. If this turns out to be the case we may be considering one sort of serial learning in which the previous item really is the functional stimulus.

Nor is this necessarily just a matter of subjects responding to a highly artificial situation, for it is now widely recognised that, given meaningful material, and enough time, a high proportion of subjects will make use of mnemonic devices of one type or another without waiting to be instructed to do so. Many of these devices, as described by the subjects, seem to exploit similarities between adjacent items, or pre-existent associations betwen them. It may then be the case that laboratory experiments on rote learning have tended to under-estimate the complexity of the problem of the nature of the functional stimulus in serial learning. Its nature may depend on a number of situational variables. Thus it may depend on the nature of the material; the mode of presentation used; and individual characteristics of the subjects, quite apart from any modifications in the behaviour of the subjects that the experimenter may deliberately produce (e.g. by instruction). A corollary of this suggestion is that the reason why researchers have failed to find much evidence for the use of the previous item as the functional stimulus in serial learning is that they have been operating on a biassed sample of learning, in terms of the materials used and the conditions of presentation of those materials.

CHAPTER 6.

Methodological and theoretical considerations for research on mnemonics. IV Meaningfulness.

From the earliest days of experimental psychology it has been recognised that performance in a verbal learning task is related to the meaningfulness of the material (to the subject). Ebbinghaus (1885) gave recognition to this variable in two different ways. Firstly, he invented and used the CVC nonsense syllable, with a view to preventing the material from being meaningful to the subject. Secondly, he carried out an experiment in which he showed that 80-syllable stanzas of the poem "Don Juan" could be learned to a criterion of one perfect reproduction with fewer presentations than were needed with lists of 12 nonsense syllables.

Since the work of Ebbinghaus the development of the concept of meaningfulness has continued in several different directions, until there are today a large number of different approaches to the topic; approaches which over-lap, include, and diverge from each other in a highly complex pattern. However meaningfulness is measured, it usually seems to be relevant to performance in a learning task, and even now there are aspects of the lay view of what makes material meaningful which have yet to be adequately formalised, but which will also probably turn out to be highly relevant to performance.

One distinction which is not usually clearly made, but which is clearly suggested by a survey of the literature, is between what might be called 'item meaningfulness', and what might be called 'connective meaningfulness'. The former is represented by the measurement of the capacity of individual items to evoke associations (e.g. Glaze, 1928; Noble, 1952a). The latter is illustrated by Ebbinghaus's Don Juan experiment and by the work of Miller and Selfridge (1950) on different orders of statistical approximation to passages of connected English. Begging the question a little one might say that item meaningfulness is concerned with the potential of particular items for becoming associated in learning with other items unspecified, whereas connective meaningfulness relates to the potential of particular pairs or sequences of items for becoming associated in that pairing or in that sequence. Bearing in mind recent work suggesting that serial learning is at least sometimes non-associative, one might add to these variables one which might be called 'learnability' which may turn out simply to reflect the frequency with which the subject has experienced that item, or others similar to it, or may, at least, turn out to have little connection with 'meaning' except through causally neutral correlation with variables which do have such a connection. An example of such correlation would be a negative correlation between word length and meaningfulness in the sense of association value. Few would suggest that being short made words meaningful, or that being meaningful made words short, but high frequency may, through quite independent causal mechanisms, lead both to high word association values and to shortness.

Another point about the above analysis is that there would probably be some point in specifying that where the term 'item' is used, it means, in some sense, 'functional item'. Thus it may be that when the subject is given English words, these are already sufficiently integrated to be items from his point of view, whereas with material such as Witmer's (1935) CCC syllables he has to learn each sequence of consonants, as well as the list, so there is a sense in which his 'items' are the consonants and not the syllables.

Another distinction which has a tendency to become obscured in the literature is that between the measurement of meaning, as exemplified, for instance, in the work of Osgood (1952) and of Osgood, Succi, and Tannenbaum (1957), and the measurement of meaningfulness, as in the work of Noble (1952a) and Noble, Stockwell and Prior (1957). Osgood seems to think that the sort of analysis of meaning which emerges from the application of his semantic differential approach is basically antithetical to that which emerges from Noble's analysis. Noble (1963, pp. 96-99) effectively replies to the attacks of Osgood, Succi, and Tannenbaum (1957) by pointing out their misinterpretations of his earlier work, and by pointing out also that some of their criticisms would apply equally to their own approach. In fact there appears to be little theoretical conflict between the views of Osgood and of Noble. Osgood et al (1957) pointed out (p.17) one solution to the confrontation, when they said of the index of item meaningfulness "m'" developed by Noble (1952a), "Noble's m might be identified as meaningfulness rather than meaning, or better, simply the association value of the stimulus, since this is actually what he is measuring". However, they had already made sure of an unsympathetic reading by Noble and his collaborators by characterising his approach (p.16) as being, "...... as simple as it is ludicrous". Although it is becoming clear that the approach of Osgood has identified some very important factors in the structure of connotative meaning (Osgood et al make it clear (p.231) that they do not regard their model of meaning as exhaustive) it remains true that two words obviously very different in meaning can have very similar profiles on the semantic differential. Similarly, though its clear relationship to learnability makes Noble's m a useful measure of meaningfulness, it is becoming clear that there are some aspects of item-meaningfulness (e.g. concreteness) which it does not estimate very well.

Most of the work relating meaningfulness to learning has been in terms of either item meaningfulness or connective meaningfulness, either separately or in combination. Although attempted quantifications of item meaningfulness began much earlier, with Glaze (1928), Hull (1933), Krueger (1934) and Witmer (1935), it was not until the work of Noble (1952) a, b that this sort of meaningfulness began to be studied in any depth separately from connective meaningfulness. Deese (1952) wrote, "But meaningful material, almost by definition, contains a lot of short phrases which go together" (p.166), a statement which does not acknowledge the existence of item meaningfulness in our sense. Since then, however, most work has been concerned with item meaningfulness.

The earliest attempts to measure meaningfulness of individual items followed, in principle, at least, Glaze's (1928) concept of 'Association Value', according to which each item was assigned a number representing the proportion of subjects who produced or reported associations for the item within a fixed time. Following Glaze, Hull (1933) and Krueger (1934) produced such values for CVC syllables, while Witmer (1935) did the same for CCC syllables. Archer (1960) has restandardised these association-values for all the CVC syllables.

Noble (1952,a) introduced the concept of a 'Production Value' which he denoted by "m", this being the average number of separate associations to an item produced by subjects within a specific time period. He used two-syllable words or nonsense words. Mandler (1955) carried out a similar procedure for some of the CVC syllables. Later Noble, Stockwell, and Prior (1957), introduced the notion of 'Rated Value' or "m'", obtained by having subjects rate on a five point scale the number of associations the items evoked in them. Noble (1963) applied this technique to dissyllabic nouns. Noble has in a number of studies shown that the relationships between rated and production values are very close, ranging between r = .7 and r = .9.

In addition Underwood and Schulz (1960), in reviewing these different methods of measuring meaningfulness, point out that the same methods, applied to different populations at different points in time have typically resulted in remarkably similar values. High correlations between m and three other variables, subject ratings of learning speed, rated familiarity, and pronunciability, led Underwood and Schulz to suggest that a single variable underlies all of these measures, and on the basis of their own studies they concluded that this variable is frequency of previous experience. Noble (1963) insists on a clear separation of the notions of familiarity and of meaningfulness in the sense of high associate-availability, pointing out that although meaningfulness is a sufficient condition for availability the reverse is not the case.

The findings on the relations between meaningfulness and learning have been reviewed by Underwood and Schulz (1960), Noble (1963), and Kausler (1966, pp. 182-191). These findings, insofar as they are relevant to our purposes, may be summarised as follows. Serial learning is facilitated as a negatively accelerated (increasing) function of meaningfulness (e.g. Noble, 1952,b). Some evidence suggests that the serial position curve flattens out with increased meaningfulness (e.g. Braun and Heyman, 1958). Meaningful material is retained better, at least for retention intervals up to seven days (e.g. Dowling and Braun, 1957).

For paired-associates the situation is rather more complicated. With equal m for St. and R. items the facilitative effect of increasing m seems much the same as in serial learning (e.g. Noble and McNeely, 1957). However, when m is varied independently for St. and R. items it appears that the facilitative effect is much greater on the response side than on the stimulus side. This generalisation still needs to be treated with some caution, as a few studies (e.g. Mandler and Campbell, 1957), have given results in the opposite direction.

With regard to the separate roles of m and familiarity in the paired-associate situation, there is some support for the Underwood and Schulz stage analysis view that familiarity has the main effect on the response-learning stage, after which m is the chief facilitative factor in the associative phase. Against this must be set the fact that studies by Winzen (1921) confirmed by Gannon and Noble (1961) show that stimulus pre-familiarisation has more facilitative effect on performance than does response pre-familiarisation.

Although interest is in the process of shifting away from nonsense material towards words there has not yet been enough research done with such highly meaningful material to warrant an assumption that all of the above-described relations continue to hold at such high levels of meaningfulness. Much of our information about relative meaningfulness of different words tends to revolve about the stimulus words of the Kent-Rosanoff test, and the norms for associations to these words produced by Russell and Jenkins (1954). The other major source of indirect information about meaningfulness of words is the Thorndike-Lorge word count (1944). A result of this is that we do not have the same sort of systematic meaningfulness tabulations for particular types of verbal material that we have for nonsense materials. On the other hand there seems no a priori reason for expecting the findings about meaningfulness for nonsense materials to cease to apply when the material is relatively highly meaningful.

One recent development which may be highly relevant to mnemonics has been a revival of interest in what we have called connective meaningfulness. Deese (1959), using lists of words from the Minnesota (Russell and Jenkins, 1954) norms, tabulated for each word in the list the percentage frequency with which each other word was given in the norms as a free association to it. He then treated the mean of the totals of these percentages for all the words in the list as an index of inter-item associative strength for the list as a whole. Using eighteen different lists, and a free recall procedure, he found a correlation of .88 between this index and recall scores for the lists. Deese (1961) argues that commonality of particular extra-list intrusions between different subjects suggests that where inter-item associations influence learning they do so automatically, and without any deliberate selection on the part of the subject. In view of the evidence that now exists for selection and editing on the part of the subject, at least in other types of learning task, this seems a somewhat extreme position, and it is hard not to agree with Cofer (1961) when, in replying to Deese, he says, (p.35), "At any rate, I am not convinced that the high correlations Deese has reported between measures of associative strength, amount of recall, and intrusions are necessarily inconsistent with the editing of associations".

Deese's inter-item associative index is a measure of the cohesiveness of the list as an associative cluster. It is a not unnatural extension of this procedure to make use of data from associative norms to construct indices of associative cohesiveness between pairs of items, either in terms of the percentage of subjects who gave each item as a free association to the other, or in terms of the extent to which the two items had overlapping associative hierarchies. A number of researchers (e.g. Amster, 1967; Garskof and Houston, 1963), have published such inter-item association norms. Alternatively, bearing in mind the high correlations typically found between direct measures and ratings for different types of item meaningfulness, one might make use of such ratings for connective meaningfulness (as in experiments 3 and 6 of this thesis).

In relation to serial learning, in spite of the evidence against inter-item association as a major element in acquisition in this type of task, it would be surprising if the subject did not make use of such associations if they were maximised in the construction of the list. A study by Wearing and Montague (1967) is highly suggestive here. They found that two lists were still not equally difficult to learn when all of the variables usually associated with item-meaningfulness had been controlled. They attributed the difference in difficulty to a between-lists difference in inter-item association. They advance as evidence for this interpretation the fact that when subjects were given opportunities to report natural language mediators many more were reported for the easier list. Although the task they used was not a serial one it may be relevant in this connection to note that Abramczyk and Bousfield (1967) found that sequential ordering in repeated free recall was a positive function of inter-item association.

The addition of words to the list of acceptable materials to be used in verbal learning has also resulted in the consideration of dimensions of variation related to meaningfulness which were not relevant to the materials previously used. Two such dimensions are abstractness-concreteness and capacity to evoke imagery (imageability). The work of Paivio and his associates (e.g. Paivio, 1963, 1965; Paivio and Yarmey, 1966; Yarmey and Thomas, 1966), suggests that these variables may, for material consisting of English nouns, relate to performance in ways similar to the ways in which meaningfulness relates to performance with nonsense materials. Thus Dukes and Bastion (1966) find that when m is held constant concrete words are learned faster than are abstract words. Paivio (1965) also finds concrete nouns to be easier to learn than abstract nouns, and that there is a difference in the facilitative effect of concreteness when applied to the St. and the R. terms in paired-associate learning. This difference is, however, in the opposite direction from that usually found when meaningfulness is the independent variable. Concrete-abstract learning is easier than abstract-concrete. This finding was confirmed by Paivio and Yarmey (1966).

Some recent studies suggest that when the material consists of concrete nouns imageability has a much stronger facilitative effect than does meaningfulness per se. Thus Paivio, Yuille, and Smythe (1966) find that imageability and concreteness have their effect primarily on the stimulus side, that meaningfulness has a facilitative effect on both sides with abstract words, but that when the words are concrete meaningfulness has little effect on either term. Similarly Paivio and Yuille (1967) found that imageability was a better predictor of performance with concrete nouns than was meaningfulness.

Although the information in this area is still too sparse for detailed theorising, some general suggestions may be worthwhile. Subjects in experiments with meaningful materials often make comments about the sorts of conscious mediation in which they were engaging, and it may be that the material determines to some extent what kind of mediation the subject will use, and also what kind of mediation will be most effective. Yarmey and Thomas (1966) found that for concrete nouns setting the subjects to use verbal chaining mediation interfered with their learning, whereas with abstract nouns setting them to use imagery mediation interfered with performance. It may also be that the material determines to some degree whether the uninstructed subject will use mediation at all. Thus, with nonsense material the construction of mediators may, for most subjects, be so difficult and inefficient that it is easier to rely on rote, unmediated, and perhaps, position learning, whereas the combination 'egg-spoon', even if it occurs in a serial list, has sufficient pre-existing associative strength that it would be surprising if any subject did not make use of it.

CHAPTER 7.

Methodological and theoretical considerations for research on mnemonics. V Mediation and natural language mediation.

Mandler (1967), in discussing the changes which have taken place in the field of verbal learning in recent years, points to an increasing awareness of the fact that the subject in a learning experiment cannot be realistically treated as responding in a purely passive way to the stimuli presented by the experimenter; that he must be seen as operating, often in a highly individual way, on the stimulus material. This movement away from the anti-introspection, frequently "black box", approach, is detectable in many fields of psychology. Thus the main conclusion of Farber's (1963) article in the American Psychologist is that it is a great mistake not to ask one's subjects what they thought the experiment was about, and what they were trying to do. Associated with this movement is a resurgence of interest in topics such as Imagery, Creativity, and what is now known as "Natural Language Mediation".

To say that the subject does not respond in a purely passive way to the stimuli presented to him is to imply that some form of internal activity is interposed between stimulus and response. One such form of activity is the construction of mediators or links (usually based on previous learning) which facilitate the evocation of a particular response by a particular stimulus. These mediators may vary on a number of dimensions. They may, for instance, be more or less complex, implicit (unconscious) or explicit (conscious), systematic or casual. They may link the items through their form or through their meaning. It would seem reasonable to suggest that a mnemonic technique is a particular kind of mediation technique, and the object of this chapter is to examine mediation with this suggestion in mind.

Implicit Mediation

Interest in mediational processes was greatly stimulated by the research of Bugelski and Scharlock (1952), using the AB-BC-AC learning paradigm, first used by Shipley (1933). They found that AC learning was greatly facilitated by the interpolation of the BC list. It appeared that here one was dealing directly with mediational processes, without having to ask the subject what was going on in his head. Indeed, the mediation seemed not to depend on conscious awareness.

(Bugelski and Scharlock call their paper "An experimental demonstration of unconscious mediation"). Mandler and Earhard (1964) have proposed an explanation for the facilitation of AC learning which does not depend on mediation at all, but on the whole the evidence favours the interpretation given by Bugelski and Scharlock.

It seemed reasonable to assume that longer mediational chains of the form BC-CD-------AX could be formed, and that pre-existing associations could take part in such chains, and a number of theorists, (e.g. Harlow, 1949; Miller, 1951), made these assumptions. An influential experiment by Russell and Storms (1955) confirmed both assumptions. Their subjects learned two paired-associate lists each composed of both experimental and control item-pairs. The stimuli were nonsense syllables. The responses were words taken from the Kent-Rosanoff lists. In the second list the experimental pairs consisted of the first list stimuli (A items) paired with the most common associates of the most common associates of the first list responses (B items). Thus the responses of the second list experimental pairs (the D items) could be seen as linked to the A items by the implicit chain of mediators B-C-D.

The control pairs in the second list consisted of A items paired with words (X items) irrelevant to the B items. The experiment was thus a mixed list design in which experimental terms took the form AB, followed by B-C-D (implicit), and AD, while control terms took the form AB, followed by B-C-D (implicit), and AX. In second list learning AD pairs were learned significantly faster than AX pairs.

McGehee and Schulz (1961) confirmed the finding of Russell and Storms. They also found that the paradigm can be extended to include situations in which implicit responses interfere with AB learning. There was some suggestion from their data that the effect operated on the associative aspect of learning, and not on the response-learning aspect.

Explicit Mediation

In 1918 H.B. Reed wrote a series of articles in which he explored the process of memorisation in terms of the techniques which the subject described himself as using when learning a constant order paired-associate list consisting of English words. He described a number of types of "associative aids", and tried to study their effects on the learning process. The subjects seemed to be actively searching for some way of associating each pair of items, making use of similarities of form or meaning, associations which the two items had in common, visual images of the juxtaposition of the items, the conscious construction of images and ideas in which the juxtaposition of the items became bizarre, and so on. There seemed to be little overt evidence of learning until such aids had been constructed, and they also seemed to have strong effects upon retention.

Reed noted that as learning proceeded such aids seemed to drop out, until, after a number of repetitions, the subject appeared to be proceeding from stimulus term to response term directly. However, when the subject was asked to recall the stimulus term upon presentation of the response term, the associative aid reasserted itself, or, if it did not, there was no recall. Indeed, after looking at a number of other "transfer" paradigms, Reed concluded that transfer did not occur without associative aids.

Although many other researchers subsequently noted the subjects' use of these associative aids they have not always agreed (Peters, 1935) as to their importance.

There is an obvious analogy between the concept of mediation as implied by Bugelski and Scharlock, Russell and Storms, and McGehee and Schulz, and that dealt with by such authors as Reed, and Peters. The former authors appear to be denying the analogy at some level, by stressing that the mediation they are concerned with is unconscious. The problem of awareness has, however, proved to be an intractable one in other fields, such as those of verbal conditioning and of incidental learning, so it is not surprising that a number of authors should have avoided it by treating conscious mediation as a quite separate concept. The suggestion that the two may nevertheless be closely related is still worth considering. One study (Martin and Dean, 1964), supports this suggestion by finding no support for Russell and Storms, except where the mediation is explicit.

The increasing potency of attacks against the strict S-R approach together with the existence of fairly good evidence for the occurrence of mediation in verbal learning, has led to a return to the topic of conscious mediation on the part of many experimenters. A number of approaches have been used, corresponding to the general research strategies discussed in Chapter 3 of this thesis.

Some researchers have asked subjects for descriptions of the mediation processes consciously carried out during the learning process. Thus Martin, Boersma and Cox (1965) asked subjects to describe the associative strategies they had used in learning a paired-associate paralog list. They then tried to classify subjects' descriptions of the mental devices which they had used, christening these devices "natural language mediators". Some researchers (e.g. Bugelski, 1962; Mattocks (reported in Underwood and Schulz, 1960)) have not set out specifically to investigate mediation, but have asked their subjects questions about mediation as part of the information gleaned after an experimental session.

A more direct approach has been to ask the subject to mediate, and to observe the effects of this instruction upon learning performance and upon reported mediation. This procedure was used by Rundquist (1965), McNulty (1966), and Martin and Dean (1966).

A third approach has been to describe to the subject a particular mediation technique, in more or less detail, to instruct him to use it, and to note the effects of such instruction on performance. Some of the research on mnemonics described above in Chapter 2 (e.g. Wallace, Turner, and Perkins, 1957; Smith and Noble, 1965) falls into this class, but not all of the research in this class is on mnemonics. Thus Yarmey and Thomas (1966) varied the type of material between abstract and concrete and instructed their subjects to use either verbal mediators or images. However, even if this sort of research is not <u>about</u> mnemonics it is clearly highly relevant to mnemonics.

In addition to the researchers who have made explicit reference to natural language mediation, or to mnemonics, it could be considered that Paivio and his associates (see pp. 82, 83, above) were dealing indirectly with these topics. Paivio's approach, which is closely related to that of Mowrer (1960), is based on the hypothesis that the ease with which words are learned in verbal learning tasks is related to their capacity to evoke images to mediate associative learning. This leads to research in which variables such as abstractness-concreteness and imageability are controlled. It would seem reasonable to suggest that one way in which these variables exercise their effect on the learning process is by controlling the extent to which particular kinds of natural language mediator, and particular kinds of mnemonic, can be used by the subjects.

Explicit Mediation and Learning

There has been a tendency in the literature for the question of whether explicit mediation is or is not facilitatory to be asked in a global way. In asking the question researchers seem to have assumed that type of material, presentation time, and conditions of testing are not relevant to the issue. The results have in consequence been inconsistent.

Thus Reed (1918), working with long presentation times and meaningful material, concludes that mediational strategies play a vital, if not the vital, role in verbal learning. Peters (1935), using nonsense material, and shorter presentation times, has mainly negative findings. Wallace, Turner and Perkins (1957) had their subjects learn 500 paired-associates in one presentation, using a visual imagery mnemonic. The subjects were self-paced up to a maximum time of 20 seconds, and made no errors in recall. It is difficult to believe that a control group of subjects who did not use the mnemonic technique would have matched this level of performance. Smith and Noble (1965), on the other hand, using a similar mnemonic system on nonsense syllable material, with relatively untrained subjects, found no significant difference in learning between their mnemonic and their control subjects, although there were small differences in recall and relearning.

In studies in which subjects have been asked after learning the material to report on mediators of which they were aware it has emerged fairly consistently that when subjects are able to report a mediator for a particular pair of items the connection between that pair of items tends to have been learned more quickly, and to be retained better, than is the case for item-pairs for which no mediator can be reported. Thus in the study of Mattocks reported in Underwood and Schulz (1960) the mean number of correct anticipations for items for which the subjects reported mediators was 14.88, whereas for items for which no mediator was reported it was 11.16, a highly significant difference. However, as Underwood and Schulz point out (p.298), it cannot be assumed that failure to report mediators always implies an absence of mediators. Nor does presence of mediators entail that they facilitate learning. A number of mechanisms could be envisaged which would lead to the reported findings without there being any implication that mediation facilitates learning. It could even be the case that success in learning reinforces recall of the mediator. On the other hand, the facts that some experimenters have found that instructions to mediate improve performance (e.g. Garskof, Sandak, and Malinowski, 1965), that subjects evidently expect mediators to assist them in the task, and that some researchers (e.g. Olton, 1966), have demonstrated situations in which presenting the subjects with a pre-constructed mediator enhances learning performance, all suggest that the naive interpretation is the correct one.

A number of studies in which instructions to mediate have been used have failed to show significant facilitation of learning. Thus McNulty (1966) found only an insignificant tendency for mediating subjects to learn faster than the controls when the material consisted of higher approximations to English. Martin and Dean (1966) found that instructions to mediate increased the number of reported mediators, but did not increase the number of correct responses. Both Wood (1966) and Smith and Noble (1965) interpret their results as negative in this regard, although their interpretations are arguable (see Chapter 2 above).

Some studies of reported mediation (e.g. Adams and McIntyre, 1967), have also failed to find evidence of facilitation of learning, but these are in the minority, in contrast to the situation with regard to instructions to mediate. This fact might give rise to a suspicion that mediation behaviour is not easily subject to control through instructions. Unfortunately this suspicion cannot yet be checked, as few studies have been reported in which mediation instructions have been given and descriptions of mediators collected.

There are, however, other hypotheses which might be put forward to account for many of the negative results which have been reported. There now seems to be much evidence to support the assertion that the occurrence of mediators is dependent on the presentation time used in the experiment. Thus Rundquist (1965) found more mediators were reported with a 4 sec. rate of presentation than with either a 1 or 2 sec. rate. Bugelski (1962) and Montague, Adams and Kiess (1966) had similar findings. Where instructions to mediate were used both Rundquist (1965) and Wood (1966) found, and here their conclusions conflict with those of Bugelski (1962), that performance was dependent on presentation time rather than on total learning time. This finding could be accounted for by suggesting that the longer the time spent in constructing a mediator (up to a certain limit), the more effective it is likely to be, and that the construction of a mediator is essentially a one-trial operation. In addition, it is likely that there is a lower limit on the length of the presentation time that a subject needs in which to construct an effective mediator. The subjects of Wallace, Turner and Perkins (1957), even after extensive practice, seldom took less than 5 seconds to construct a mediator. Rundquist and Farley (1964), using latency as a measure of mediation time, noted that subjects seldom started writing down their mediators in less than five seconds. Some investigators, (e.g. Martin, Boersma and Cox, 1965; Montague and Wearing, 1967), have also found that the effectiveness of mediators is related to their complexity.

In view of these considerations, it would be reasonable to suggest that subjects need at least 5 seconds to construct mediators, and, if they are naive, may need much more. If, as the results of Martin, Cox and Boersma (1965) suggest, meaningfulness is related to the ease with which natural language mediators are constructed, subjects learning material with low meaningfulness may need yet more time. The only study which has used instructions to mediate, meaningful material, and presentation times as long as 5 seconds, and which also has negative results, is that of Wood (1966), and it is debateable whether his results really are negative. His control group had free recall instructions biassed towards serial recall. His experimental groups had series of peg-words, printed on their answer sheets, and their instructions were to use various kinds of mediator to connect the peg-words to the words of the experimental lists. The groups with peg-words made many less errors on recall than did the control group, and Wood suggests that it was the presence of the peg-words rather than any mental manipulations that the subjects carried out, that produced the difference in performance. However, he did not have a control group which would enable this suggestion to be tested; that is, a group given normal constant-order paired-associates instructions; and he did have another experimental group whose superiority to the controls could not be accounted for by the same suggestion. This group did not have peg-words but was instructed to construct mediators to connect the list words in a sequential fashion, and Wood himself does not present any alternative to the hypothesis that their superiority to the control subjects reflected their carrying out of these instructions.

It was noted above (p.93) that Reed (1918) observed that associative aids appear to drop out as learning proceeds. A number of other researchers (e.g. Zinchenko, 1959; Dean and Martin, 1966; Adams and McIntyre, 1967), have also shown that when the material is repeatedly presented the number of natural language mediators reported first rises and then declines. In the Adams and McIntyre study subjects actually reported themselves as recalling in a rote fashion items for which they had previously reported natural language mediators. If, in analysis, responses which have been 'rote' from the first time they were correctly elicited are not distinguished from responses which were initially reported as mediated, the role of the mediator in the learning process may become obscured.

Explicit Mediation and Retention

Most studies which have been concerned both with acquisition and with retention have found the effect of explicit mediation to be greater on retention than on acquisition. Even when it has not seemed that learning was facilitated, some facilitation of retention has usually been reported (see Smith and Noble, 1965). Groninger (1966) concludes that natural language mediators aid retention by deterring proactive interference. Kiess and Montague (1965) tested retention for CVC pairs after 24 hours, and found that where the subjects had initially formed no natural language mediator there was negligible retention.

There could, however, have been some exaggeration of the effect of explicit mediation upon retention in the literature so far. In the first place, some researchers, including Kiess and Montague (1965) have required their subjects to write down their mediators after learning, and before the retention study. There seems little doubt that reporting mediators will itself aid retention independently of the effect of the mediators themselves. Indeed, Boersma, Conklin and Carlson (1966) found this to be the case, and failed to detect any effect upon either acquisition or retention of the mediator per se. Olton (1966) tried to equate the degree of learning of his mnemonic and control subjects by giving them different numbers of presentation of the material. When he had done this, and the results of a successive probability analysis suggested he had succeeded, retention tests after one week showed no difference between control and experimental groups. In view of these findings it would seem premature to conclude that explicit mediation has more effect upon retention than it does upon acquisition.

One consistent finding which has strong relevance to the view implied by Adams and McIntyre (1967) that there is a clear distinction to be made between rote and mediated learning is reported by Montague, Adams and Kiess (1966). They found that items for which natural language mediators were constructed but subsequently forgotten were recalled less well than items for which no mediator had been constructed. The results of Adams and McIntyre clearly confirm this finding. This result might be explained as follows. It could be the case that in rote learning the subject learns to produce each item as the response to a stimulus situation, which may be the previous item, or a cluster of items, or a particular serial position, or just the total situation. In mediated learning, on the other hand, he may learn to respond to an internal stimulus, usually a response to the previous item. If this internal stimulus is not present, through forgetting, he may be less likely to produce the appropriate response than he would be if he had even a weak S-R bond of the rote variety, since in this case there is some guarantee that the stimulus will be presented as it is presented by the experimenter. If one considers this view together with the finding discussed above that mediators tend to drop out when the material is repeatedly presented some very interesting possibilities arise.

One such possibility is that there may be circumstances in which a subject learning paired-associates by deliberately constructing mediating word chains will arrive at a rote mode of responding more rapidly than a subject who carried out the task in a rote fashion right from the start. Let us assume that one subject learns the pair of items 'mouse-tree' by interpolating the mediator 'hole-hollow', while another subject simply tries to connect 'mouse' to 'tree' by repeating the pair to himself again and again. If the paired-associate list is a long one, a large number of presentations may be required before the second subject has reliably connected this particular pair. The subject using the mediator may meanwhile have correctly responded to the stimulus 'mouse' enough times to have caused the mediator to drop out completely, and may thus have arrived at a rote mode of responding to this stimulus.

It is not difficult to envisage an operant conditioning model of the way in which mediators drop out with repeated recall. For instance, since there is a reward involved (i.e. being correct), the process might be seen as one in which the stimulus term is being established as a discriminative stimulus for the occurrence of the response, but in which a pre-existing chain of implicit responses is being evoked, which initially ensures the occurrence of the response. These implicit responses might be seen as dropping out later because the reinforcement is not in fact contingent on their occurrence.

Different kinds of Mediation

Martin, Boersma and Cox (1965), using low meaningfulness paralogs (pseudo-words) classified associative strategies into 7 categories along an ordered dimension of cue complexity. Their data suggested that complexity of strategy was positively related to performance. They found a correlation of .62 (n = 39) between the total strategy score of subjects and their total numbers of correct responses. Martin, Cox and Boersma (1965) followed up this experiment with one in which they varied stimulus and response meaningfulness. They confirmed the previous result, and found further that while stimulus meaningfulness was less potent than response meaningfulness in its effect upon learning, it was more potent than response meaningfulness in its effect upon the complexity of the strategies used by the subjects. The relation between meaningfulness and complexity of strategies was positive. Montague and Wearing (1967) using CVC lists, found that complex natural language mediators led to less error in recall than simple ones.

The compilers of memory courses, (e.g. Lorayne, 1963), often assert that particular kinds of mediating device are more effective than others. Most think that mediators using imagery, and particularly those using bizarre imagery, are more effective than verbal mediators. There is as yet little evidence in the psychological literature to support these claims.

It could be the case, however, that mnemonists tend to work with material for which this kind of device is more suitable. The work of Yarmey and Thomas (1966) suggests that imagery devices are more effective when the material consists of concrete nouns, while verbal devices work best for abstract nouns.

Moreover, much of the other research that has arisen as a result of Paivio's (1965) 'conceptual peg' hypothesis indicates that different types of material may lead to different modes of learning.

On the other hand Wood (1966) fails to find any differences in the performance of subjects given instructions designed to elicit bizarre imagery, common imagery, and verbal mediation. He also finds little support for the view that differences between abstract-concrete and concrete-abstract learning are due to differences in the capacities of the stimulus terms to evoke imagery. He cites work by Perrino (1966) which suggests that induced syntactical connotations in nonsense materials can lead to differences in the ease of learning pairs of the forms noun-adjective and adjective-noun.

It is not clear, however, that Wood's subjects fully carried out his instructions, nor that the work that has been carried out on the concrete-abstract dimension can be criticised through an attack on work on the noun-adjective dimension (see Lambert and Paivio, 1956), which led to it. Furthermore, the group of studies Wood is attacking show a degree of internal consistency which is sufficient to suggest that they should not be ignored on the basis of one negative finding.

CHAPTER 8.

Pilot studies and informal work on mnemonics.

Following examination of a number of commercial memory courses (e.g. Furst, 1954; Roth, 1961), the writer entered into informal discussion with nine professional and semi-professional conjurers who had used mnemonics for entertainment purposes. These all displayed great enthusiasm about mnemonics, and were very willing to demonstrate their prowess.

In general, their experiences with mnemonics were consistent with the statements about mnemonics made in the memory courses, and they were in very clear agreement with each other. Thus all agreed that bizarre associations worked better than common ones, that rehearsal of associations was unnecessary, and that the length of the list to be learned did not affect performance in any obvious way. All but one agreed that imagery worked better than verbal associations. With regard to peg-words, all agreed that it was possible to learn list after list without noticeable interference. One performer had learned more than 100 lists during a six month period, using the same peg-word series each time. He had not been aware of any tendency for intralist intrusions to occur. It is worth noting that none of them had ever learned more than one list in the same performance, or had occasion to recall a list after the end of the performance in which it had been learned.

Wood (1966), carried out an experiment in which he presented three lists to his subjects in one experimental session, requiring them to use the same peg-words each time. He found negative transfer, and concluded that the claims of mnemonists that interference did not occur were not justified. However, the claim made by these mnemonists does not entail the prediction of lack of interference under Wood's conditions.

There was some discussion of the difference in 'feel' between mnemonic and non-mnemonic learning, and all of the performers said that this difference was quite striking. One said, "Using mnemonics does not feel like learning at all".

All believed that they could not possibly perform without mnemonics the feats they performed with them, but that anyone could in fact learn to use them. Seven of them believed that if they "failed to make a good image", for a particular pair of items, there was no hope of making a correct response when required to recall the second item. Thus their beliefs were consistent with the implied view of Adams and McIntyre (1967) that learning is either rote or mediated, but not both. This view is also consistent with the finding of Adams and McIntyre that subjects who forgot their natural language mediators performed worse than subjects who had not constructed them in the first place.

The writer himself learned Furst's memory system (Furst, 1954), and has used it under a variety of conditions. His observations are essentially in agreement with those of the professional performers discussed above. On the other hand, he finds that his performance using non-bizarre imagery, and verbal mediation, is not very different from his performance using bizarre imagery. However, the subjective experience is certainly very different, and he does show some tendency to learn better with bizarre imagery as the length of the list is increased. Clearly, self-experimentation in this area can at best be no more than a source of hypotheses, but it is at least interesting that the writer's experience should be consistent with that of the professional users of mnemonics, who work under different conditions, and approach the matter with a very different attitude.

Pilot work on imagery

It has long been recognised (see Galton, 1880; McKellar, 1957), that people vary widely in their assessment of their own ability to produce imagery. If imagery were as important in mnemonics as some experts suggest, one might expect individual differences in ability to produce it to be a critical variable in the area. Furthermore, although there have been many studies in which information was sought as to whether the subjects experienced images (e.g. Short, 1953; Oswald, 1957), little work has been done in which the subject was required to use the imagery as a part of the technique by which the task was carried out, as is suggested in mnemonic instructions.

Accordingly, it seemed worthwhile to devise a task to see if subjects could manipulate imagery in such a way as to solve a problem through the agency of this manipulation. The task selected involved the use of a slide-rule on which one of the logarithmic scales had been replaced by a linear one. Pairs of two-figure numbers were sampled from a random number table, and the subjects were required to carry out in their heads a manipulation which, on a normal slide-rule, would have corresponded to multiplying these two-figure numbers by each other. This operation thus consisted in adding a number on the linear scale to a number on the logarithmic scale, and reading the sum on the logarithmic scale. They then carried out the manipulation on the actual slide-rule, and thus obtained information as to how accurate the mental manipulation had been. Four subjects attended 15 half-hour sessions during each of which they carried out about 30 of these operations. All of them improved steadily in the accuracy with which they could solve the problems. By the 15th session they were all performing with about 99% accuracy. These results suggest that imagery is amenable to being used as opposed to being merely experienced.

Two out of the four slide-rule subjects had previously described themselves as having poor ability to produce visual imagery, but stated that this did not seem to interfere in the task which they had been asked to carry out. Similar remarks were made by subjects in another pilot study, this time in relation to mnemonics rather than manipulation of imagery. 23 subjects had been taught the Link system, and had been learning lists of 16 items under a variety of (mostly fairly informal) conditions, with a viewto obtaining information about some of the parameters relevant to this kind of learning, and generating hypotheses about the nature of such learning. These subjects were asked how often they experienced visual, auditory, and 'other' imagery. 9 of them said that they did not normally experience visual imagery, but that this did not seem to affect them in the context of the task they had been set. 2 subjects remarked, with some surprise, that they did seem to be able to produce images in this context. Another said that he just interpreted the instructions as referring to 'imaginings', rather than images. There was no obvious difference between the performance of these 9 subjects and that of the remainder of the group. It appeared, then, that individual differences in mode of dominant imagery were not a strikingly powerful source of variability in either mnemonic or imagery manipulating performance, although it is clear that a much more formal and sensitive study would be needed before one could say that this variable was not relevant at all.

Retention, backwards recall, and prescribed images.

The 23 subjects mentioned in the preceding section each learned a number of word lists under self-paced conditions, and these were recalled and re-learned under various conditions, and after intervals ranging from 24 hours to 15 weeks. 28 control subjects were also used, who had not had any mnemonic training. The material used was unrelated object-names.

It soon became clear that, using this type of material, retention intervals, for both mnemonic and control subjects, needed to be much longer than those usually used with nonsense materials. Retention after 24 hours was close to 100% for both types of subject. It was still very high after one week, and it appeared that at least three weeks needed to elapse before comparisons between the groups became very meaningful. At this point the mnemonic subjects had higher savings scores than the controls, but the differences on recall were much smaller, though still in the same direction. Mnemonic subjects reported that on the initial re-learning (recall) presentation of the list they often could not remember their mnemonic image, and that when this was the case they could not even guess at the items. However, when they had seen both items of a consecutive pair they tended to remember what their image had been, and then they could use the mnemonic when the list was re-presented. This report was certainly consistent with the results, especially since the proportion of the recall errors that were omissions was much higher for the mnemonic than for the control subjects, 79% of errors by mnemonic subjects were omissions, and 51% of errors by control subjects.

It had been expected that mnemonic subjects would find backwards recall of the lists much easier than would the controls. This expectation had arisen because the mnemonic experts consulted had all found backward recall very easy, and because some experimenters (e.g. Young, Patterson and Benson, 1963), had failed to find transfer from forward to backwards serial learning. In fact neither our experimental nor our control subjects appeared to have any difficulty with this task, both achieving near perfect recall in the reverse direction. Subjective reports from the control subjects in this and the other pilot studies suggested that even without mnemonic instruction they made considerable use of mnemonic devices when given sufficient time. There was also some suggestion that they were aware of a distinction between learning by rote and learning in a way which made reference to the meanings of the items. Those few subjects who described themselves as learning in such a way that, "It would not have mattered what the items were", made more errors than the others on the backward task. It might be the case therefore that the use of highly meaningful material, together with long presentation times, can lead to a different type of learning from that which occurs in more traditional kinds of serial learning tasks.

It had been suggested in discussion (see Chapter 3) that the effectiveness of bizarre imagery mnemonics might depend on reward, practice, or attention factors associated with the creation of the image by the subjects. It therefore seemed worthwhile to check whether the subjects could make use of mnemonics provided by other subjects. In some of the earlier pilot studies tape-recordings had been made of the subjects' descriptions of their mnemonic devices. Fifteen of these descriptions, selected for coherence and apparent effectiveness, were re-recorded in appropriate sequence onto one tape, and played back to 11 naive subjects. These subjects were then given one anticipation recall test on the list. Their performance was intermediate between that of the mnemonic and control subjects on the same list. The mean error rates for a single anticipation trial were, for the 23 mnemonic subjects, 3.7; for the 11 prescribed-images subjects, 5.8; and, for the 28 control subjects, 9.0. It is worth noting, however, that neither the mnemonic nor the control subjects were naive. All differences were significant at the .05 level (using t tests). The most conservative interpretation of this result is that factors other than those associated with the creation of a mnemonic are related to its success. If it is borne in mind that descriptions of mnemonics are likely to be incomplete, and that mnemonics may be differentially effective for different subjects, the difference between the created mnemonic group and the prescribed mnemonic group seems less critical than the difference between the latter and the control group.

General findings and comment

All subjects in these pilot studies who were given mnemonic instructions learned how to apply them, and performed better than average control subjects, with a single exception. The one who did not learn to apply them actually performed worse than average, and thought that his attempts to apply the instructions were impeding his performance. This subject had described himself as frequently experiencing visual imagery. In later experiments three other subjects showed the same response to the instructions.

Average times taken by the mnemonic subjects to create images started at about 20 seconds and diminished with practice to about 5 seconds. This is essentially consistent with the data of Wallace, Turner and Perkins (1957).

A high proportion of the control subjects reported some use of mnemonic devices, though this was seldom done on a systematic basis. One of them had done a commercial memory course, and his performance was indistinguishable from that of the mnemonic subjects. The spontaneously applied techniques of the other control subjects seldom involved imagery, much less bizarre imagery. Many of these applied to particular adjacent pairs of items, other pairs being described as 'learned by rote'. Where overall strategies were spontaneously applied these were often highly idiosyncratic, and also, to judge by performance, relatively ineffective. Thus one subject said that she had noticed that the items were alternately round and angular. The list she had learned began, 'woman, pencil, car, cupboard, horse, envelope, cake, spectacles.......'. Her performance was not better than average.

Since there had seemed to be no clear basis for choosing the items for the lists from any of the published analyses a procedure was adopted which had as its object the obtaining of items which were fairly familiar, visualisable, and independently drawn. This consisted in the item-collector positioning himself in a public place, stopping passers-by, and requesting them to name the first object which came into their mind. In this way a pool of object-names were collected, from which, after repeated items and complex expressions had been eliminated, the lists were made up by random selection. This technique was also used in several of the later experiments.

Although most of the pilot work was carried out with subjects who were not naive, at least in the later stages of their work, it was realised that the bulk of the experimentation could not be done in this way. Very few of the student-subjects available could be obtained for long periods. Furthermore, if any experiment using a large number of subjects was allowed to run for too long, there was a considerable risk of inter-group contamination. In some of the later experiments there was in fact reason to think that some contamination had taken place as a result of the interest which the mnemonic techniques had aroused in the subjects who had used them. The use of naive subjects undoubtedly resulted in some degree of variability in the effectiveness of the instructions in the later work, and it is unfortunate that it could not be avoided.

CHAPTER 9.

Experiment I. Mnemonic versus non-mnemonic learning under ideal mnemonic conditions.

A serial learning task involves the storage of two different kinds of information. Firstly there is information about the nature of the items (e.g. the word "task" occurs in the previous sentence). Secondly there is information about the position of each item in the list (e.g. the word "task" is the 4th word in the sentence in question). It would therefore be likely to be informative to use an additional control group in which the mode of presentation did not stress order. Another way of obtaining information relevant to this distinction would be to give some subjects the opportunity to exercise choice in regard to the order in which they recalled the material. If the mnemonic subjects tended to conserve order in recall more than did the control subjects, this would suggest that the mnemonic operated on order-learning as well as or instead of on item-learning.

If the mnemonic system really did work in the way that a naive interpretation of the instructions would suggest, the serial position effect would not be as evident in the performance of mnemonic as in that of control subjects. An implication of the instructions usually given to learners of the link system is that the major determinant of correct responding is the making of effective images. The ease with which this can be done is likely to be determined by the nature of particular pairs of consecutive items rather than by any considerations of where they are in the list.

Experiment 1, then, represents an attempt to obtain information relevant to three issues. These are, firstly, do mnemonic subjects learn better than control subjects under ideal mnemonic conditions? Secondly, is there a difference between mnemonic subjects and control subjects in their preference for ordered recall? Thirdly, is there a difference in error distributions between mnemonic and control subjects.

Material

This consisted of 25 names of objects. Since frequency was neither necessarily relevant nor easily applied as a criterion to object names the method used to compile the list consisted in asking 25 people each to supply one item, and randomising the resulting series. A further list of 13 items was constructed in the same way, and subsequently used in demonstrating the mnemonic system to those subjects who were to use it. Both of these lists appear in Appendix 1.

Design and terminology

The design was factorial, with three different modes of administration of the lists, and two different modes of recall. These will be referred to as the 'input' and 'output' conditions. Since all combinations of input and output condition were used there were six different groups of subjects. The input conditions will be referred to as 'Mnemonic', 'Non-mnemonic', and 'Free'. The output conditions will be referred to as 'Ordered', and 'Free'. Thus a group described as 'Non-mnemonic, free', used non-mnemonic input and free output. The meanings of these terms will be clarified by consideration of the instructions given below.

Instructions

Group 1. (Mnemonic, Ordered). These subjects were instructed in the use of the mnemonic system as follows:

"In order to learn any given item in the list you must connect it to the previous item by thinking up as fantastic a connecting image as you can. The second item will thus be connected to the first, the third to the second, and so on, until the end of the list. The connecting images must be as unusual, complicated, and active as possible. Thus if I were connecting 'pig' with 'piano' I might picture myself sitting at an enormous pig, trying to play a piano keyboard let into its side. As I tried to play the pig would be bucking about and generally making things difficult. The noise it made would be a sort of plunking squeal. If the next item was 'apple' I would proceed to make an image connecting 'apple' to 'pig'. I would then know the list, 'piano, pig, apple'.

As soon as your image is completed stop thinking about it and go on to the next item.

Do not try to review the list as you go along. You will not be able to do this and it may alarm you unnecessarily. You will find that in spite of this you will have little difficulty in recalling the list as a whole".

The technique was then demonstrated using a recording memory drum which presented the 13 word demonstration list at the rate of 20 seconds per presentation, with an interval of 0.5 seconds between presentations.

The subject was then required to apply the system to the experimental list of 25 words presented at the same rate as in the demonstration. As soon as he had finished the memory drum was turned back and he was required to say each word within the period during which the machine was presenting the previous word. Thus while the machine was presenting the word 'start', he was to produce the first word of the list. The machine would then present the first word and the subject would attempt to produce the second. This would continue until the end of the list was reached.

<u>Group 2. (Mnemonic, Free)</u>. The input conditions were the same as for Group 1., (Mnemonic, Ordered), but instead of the machine being used to organise the output the subject was provided with an answer sheet with spaces numbered from 1 to 25. He was asked to write down against these numbers all the words from the list that he could remember, in the

order in which they occurred to him. He was given a total time in which to do this equal to that used by Group 1., i.e. 512 seconds. Thus a record was obtained of the order in which he recalled the items.

<u>Group 3. (Non-mnemonic, Ordered).</u> This group had the list presented in the same way as did Group 1., (Mnemonic, Ordered), except that the mnemonic system was not mentioned, and no demonstration was given. Instead the subjects were simply told that each word would be presented for 20 seconds, and that they were to try to learn the list. Output was the same as for Group 1.

<u>Group 4. (Non-mnemonic, Free).</u> Input was as for Group 3., (Non-mnemonic, Ordered), and output as for Group 2., (Mnemonic, Free).

<u>Group 5. (Free, Ordered).</u> These subjects were presented with the entire list printed in order, and were given 512 seconds to learn it, no special instructions being given. Output was as for Group 1., (Mnemonic, Ordered).

<u>Group 6. (Free, Free).</u> Input was as for Group 5., (Free, Ordered), and output as for Group 2., (Mnemonic, Free).

Subjects

The subjects were 90 first-year psychology students ranging in age from 18-24 years. They were allocated randomly to each of the experimental groups, so that each group consisted of 15 subjects. No subject admitted to any previous conscious experience with mnemonic systems or devices.

Incentive

Subjects in all groups were offered a reward of three pence for each word correctly recalled. This was regarded as necessary because for at least two of the groups the task was such as to seem discouragingly difficult. It was hoped that the provision of a financial reward would overcome this discouragement.

Results

The main results appear in Figure 1, which shows the number of items recalled by each subject, and means and standard deviations for each group. Individual histograms represent groups, and the histograms are based on a common scale representing for each column the score shown on the base-line. Vertical scales for each histogram represent the number of subjects making a given score.

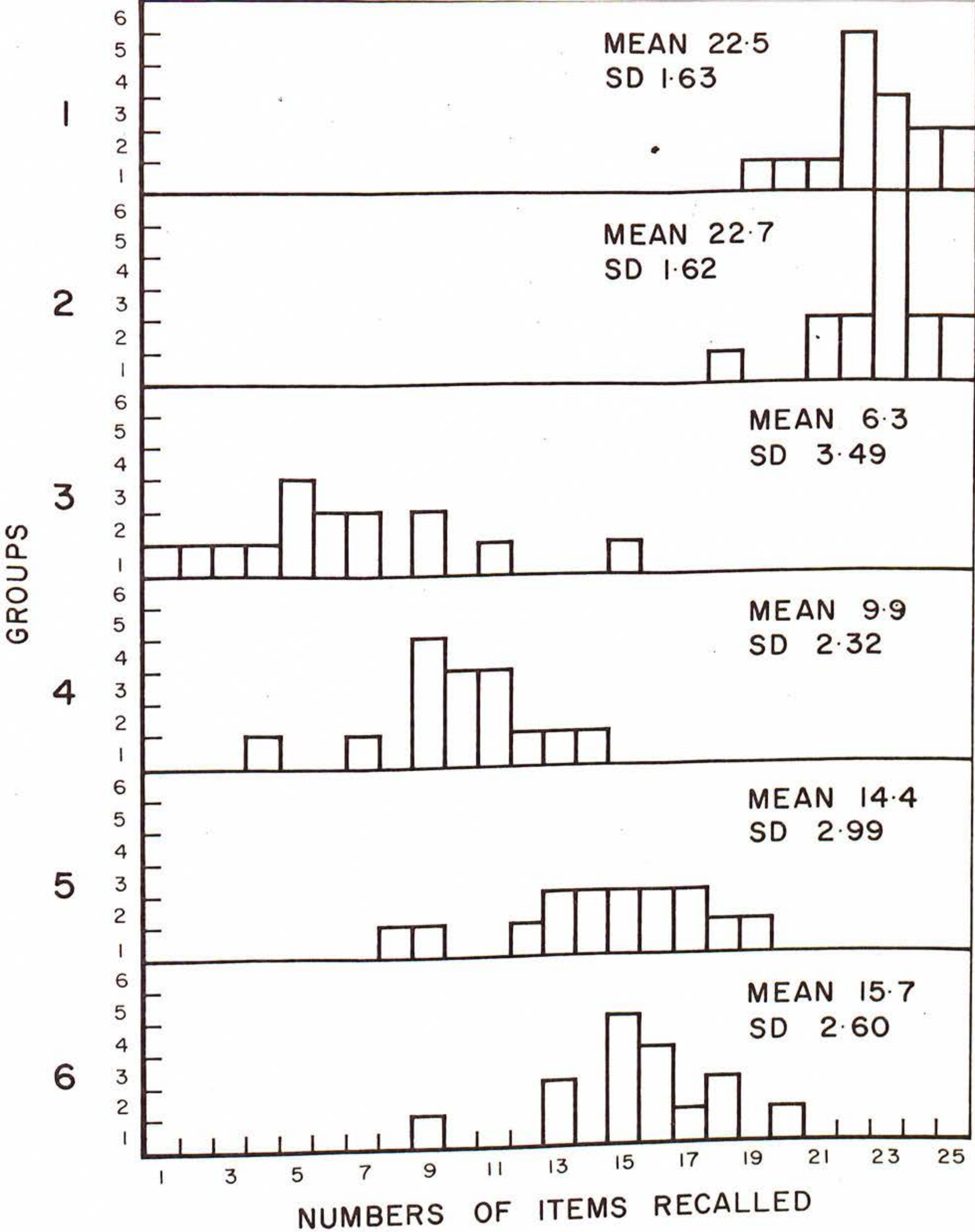

Fig. 1 Data for all groups. Numbers of subjects in each group making each number of correct responses.

An analysis of variance was carried out. The results are shown in Table 1. The marginally significant interaction appears to result from the large increment in performance produced by the free output condition under the non-mnemonic input conditions. Under the other input conditions this instruction made relatively little difference (see Figure 1.). However, a Bartlett test for homogeneity of variance gave a X^2 significant at the .05 level, a finding which might be considered to cast some doubt upon the reliability of this interaction, although the main effects are clear enough not to be seriously affected by such doubts.

We can therefore say both that the different input conditions produced different levels of performance and that the output conditions did so, the former set of difference being clearer than the latter.

Since all paired-comparisons of means were of interest the data were further analysed using the method of Scheffé (see Hays, 1963) at the 0.5 level. All paired-comparisons proved significant except those between Groups 5 and 6, and between Groups 1 and 2.

TABLE 1.

Experiment 1. Analysis of Variance

(Fixed Constants Model)

Source of Variation	Sums of Squares	D.F.	Variance Estimate	F.
Output conditions	64.2	1	64.2	F = 9.3 p .01
Input conditions	3126.8	2	1563.4	F = 226.6 p .001
Interaction	45.3	2	22.7	F = 3.2 p .05
Within Cells	582.3	84	6.9	
Total	3818.6	89		

Figure 2. represents for each group the number of correct reproductions of items 1-5, 6-10, etc. It is clear from the figure that the two mnemonic groups showed little or no serial position effect, whereas the other four groups showed quite pronounced effects.

In order to represent quantitatively the tendencies in Groups 2, 4 and 6 (the free output groups) towards reproducing the list in its original order, some index was needed. This was computed in the following way.

Each subject in a group had either recalled each item at some point in his response series or had failed to recall it at all. If it was the second item he had remembered it was given the number 2 as its ordering number. If it was the eighth it was given the ordering number 8, and so on. In the same way the items that he had recalled were given ordering numbers representing the order in which they had occurred in the list as it was presented. For each subject a coefficient of correlation was computed, using Kendall's tau (see Kendall, 1948), which represented the extent of the tendency of that subject

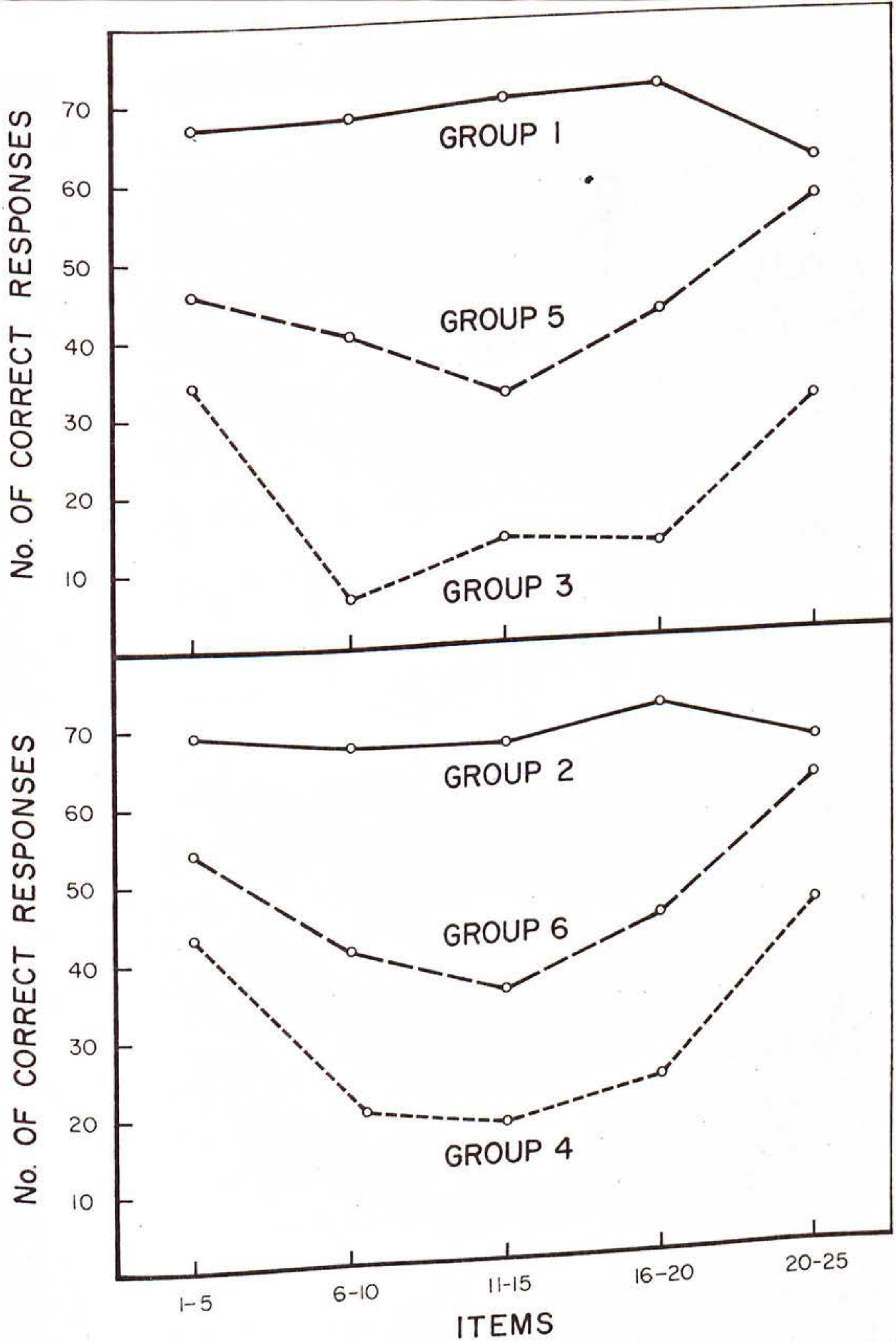

Fig. 2 Serial Position Data for all groups. Correct responses summed for successive series of 5 items.

to recall the items in the order in which they had been presented. For each of the three groups a weighted average of the individual tau value was computed, to give an index of the mean tendency within that group to recall the items in the order in which they had been presented.

The results were:-

Group 2 (Mnemonic, Free)	tau = .93 $p < .01$
Group 4 (Non-mnemonic, Free)	tau = .23 $p < .05$
Group 6 (Free, Free)	tau = .51 $p < .01$

These three values for tau were compared by setting limits on each (see Kendall, 1948, Chapter 4), and all proved significantly different from each other at the .05 level.

Conclusions

The results show that mnemonics are superior to either of the other two methods of presentation, and that the free recall groups do not do better than the relevant ordered recall groups except in the case of Groups 3 and 4, where there is a significant difference (see Figure 1.).

The mnemonics groups do not show a serial position effect whereas the other groups do so.

The mnemonics subjects show a greater tendency than the others to adhere in the free recall situation to the original order of the items.

Discussion

The differences between the mnemonics groups and the analogous groups without mnemonics (Groups 3 and 4) show that the success of the mnemonic system cannot be ascribed either to the length of the presentation time or to the use of a single presentation of the material. Indeed the superiority of the free input groups (Groups 5 and 6) to Groups 3 and 4 suggests that the manner in which the material is presented to the mnemonic groups is one which would normally be most unfavourable. The manner in which the material is presented to Groups 5 and 6 is one which would normally be expected, especially in the case of Group 6, to be most favourable, but even this group does not approach the level of performance of the mnemonic groups.

The lack of difference between the mnemonic groups (3 and 4) was expected, since there would seem to be little reason to expect that mnemonic subjects would abandon their mnemonics after they had formed them, but the lack of difference between the two free input groups (5 and 6) was not expected, as it seemed reasonable to expect that, in the absence of mnemonics, a recall task calling for the retrieval of less information would be easier than one calling for the retrieval of more information. The difference in means between these two groups (see Figure 1.) was, however, in the expected direction, even if not significant, and it is possible that the present result was anomalous.

The absence of a serial position effect in the mnemonic groups was expected, but the regularity of the error distribution (see Figure 2.) for these groups was not. If the effect of the mnemonic system is to turn the serial task into a constant order paired-associates one, as it seems reasonable to suggest it is, one would expect the presence of the bizarre images to reduce even those serial properties that such a list normally possesses. For if success in recall becomes dependent on these associations it should also become dependent on the ease with which such associations are formed for particular pairs within the list. There is no particular reason to expect this factor to be in any way related to serial position within the list. However, one would expect that particular pairs would be easy to connect for most subjects, while others would be hard for most subjects, and that this would lead to considerable irregularity in the serial distribution of errors. There was no evidence that such irregularity occurred. On the other hand if, as the recall scores suggest, the task was a particularly easy one for the mnemonic subjects, this effect might become obscured.

The tendency for the mnemonic subjects to adhere to the serial order of the list in the free recall situation is again in accordance with expectations, and could well be a function of the way in which ordering information is built into a series of mnemonic images. It may be that, for the user of this type of mnemonic system, to recall the items is to recall the order in which they were presented.

The differences in performance between the mnemonic groups and the others may seem quite extreme, but there is reason to think that the present study underestimates these differences. The subjects in all groups were given 20 seconds to learn each item in the list, because this was the time allowed in studies of a similar mnemonic system by Wallace, Turner and Perkins (1957). However, these researchers noted that with practice this time could be reduced to 3 or 4 seconds. With trained subjects, then, the total input time in the present experiment could have been reduced from 512 seconds to 100 seconds, in which case one might expect that the differences between the mnemonic subjects and the others would have been even more extreme.

Wallace, Turner and Perkins also found that, for their mnemonic subjects, input time was relatively invariant with length of list. This suggests that if the lists had been longer in the present experiment the differences between mnemonic and non-mnemonic subjects might have been more extreme.

It might be suggested that one effect of the mnemonic instructions is to force the subjects to spend more time looking at the items than is normally the case. In the present study, however, presentation times were the same for all groups, and observation of the subjects suggested that the mnemonic subjects spent less time actually looking at the display than did the other groups. Many mnemonic subjects appeared to just note what the item was, and then to shut their eyes while they constructed their mnemonics. The free input subjects, on the other hand, seemed to spend most of their time actually looking at the material.

Many of the mnemonic subjects expressed some surprise about the effectiveness of the system, and about the subjective differences from non-mnemonic memorisation, which seemed to be similar to those described by the subjects in the pilot studies (see p. 114 above). This fact, coupled with the performance differences observed in the experiment, might well prompt the suggestion that in the mnemonic/non-mnemonic dichotomy one is dealing with, to use a computer analogy, two different modes of storage addressing. Although it is clearly too early to put forward models with any degree of confidence both the types of errors made and the subjective reports of the subjects suggest that the link system behaves like a pigeon-hole filing system in which each item is stored in a location provided with a bright label which contains information about what is to be found on the label of the location containing the next item. Non-mnemonic retrieval, on the other hand, gives the impression of a filing system in which item locations are determined by their relevance to other items filed in juxtaposition to them. It would not be surprising if, as suggested by the work of Adams and McIntyre (1967) multiple outputting of the material led to one mode of storage being replaced by the other.

CHAPTER 10.

Experiment 2. Mnemonic and Non-mnemonic Learning with Multiple Presentations.

Introduction

Part of the object of Experiment 1 was to enable mnemonic subjects to operate under conditions as close as possible to those under which the professional mnemonists described themselves as operating. One disadvantage of this procedure is that the results obtained cannot be compared with other results reported in the psychological literature, since most experiments on serial learning employ some variant of the anticipation method, and involve multiple presentations and learning to criterion. Indeed, it could be maintained that it did not necessarily follow from the results of Experiment 1 that subjects with mnemonic instructions would learn faster than controls if multiple presentations were used. Thus it might be the case that on trials subsequent to the first, control subjects would improve fast enough relative to the mnemonic subjects to cancel out the first-trial advantage of the latter.

Since, as previously noted (p.58) there is some tendency for subjects using mnemonics to see the task as being essentially one-trial, it was decided to reduce presentation time below that required by most naive subjects under self-timing conditions, and to add an instruction to the effect that subjects should attempt to improve their images on trials subsequent to the first. It was hoped also that reduction of presentation time would reduce the tendency for control subjects to use mnemonic techniques spontaneously.

Pilot runs suggested that with lists of 25 items (as used in Experiment 1) some of the control subjects would require longer periods than were available in which to learn the list, so list length was reduced to 16 words. The items used were randomly selected from those used in the previous experiment.

Since this could be done without interference with the main object of the experiment it seemed worthwhile to attempt simultaneously to collect some data relevant to three other issues. The first of these was retention. Although some of the previous studies on natural language mediation (e.g. Groninger, 1966) and on mnemonics (e.g. Smith and Noble, 1965) make some reference to retention little data exists on the long-term retention of material of the kind used in these experiments, and nothing appears to be known about such long-term retention when mnemonic instructions are used. Accordingly it was decided that some subjects would be asked to relearn the list in this study after delay-periods ranging from one week to ten weeks.

Secondly, Wallace, Turner and Perkins (1957) reported that there was a considerable learning-to-learn effect with mnemonic subjects, resulting, they thought, from increasing efficiency in constructing images. However, since they did not use control subjects, these too might have shown such an effect. It was therefore decided to present some subjects with a second list made up from the residual items from Experiment 1.

Thirdly, as reported in Chapter 8 (p.114), mnemonists claim that the mnemonic techniques they use enable them to recall the list backwards with the same facility with which they recall it forwards. Although a number of authors (e.g. Young, Patterson and Benson, 1963) have investigated backward serial learning they have not used the materials or presentation times involved in the present experiment. Thus it was decided that some subjects in the present experiment would be required, after reaching criterion on the experimental list, to relearn it in the reverse direction.

Method

The subjects, 64 first-year psychology students, were randomly assigned to Mnemonic (Mn.) and Non-mnemonic (N-Mn.) instruction groups. Mn. subjects received the same instructions as were used in Experiment 1, with the addition of the suggestion that if they made an incorrect anticipation they should spend the remainder of the presentation time 'strengthening' the image they had made previously. The N-Mn. subjects were given standard serial anticipation instructions.

The material was presented on a memory drum which was set for 7 seconds presentation time and a 4 second inter-item interval, during which subjects were expected, in all trials after the first, to attempt to anticipate the next item. Learning was carried to a criterion of one error-free trial. Inter-trial interval was 30 seconds. The object of the inter-item interval was to enable the experimenter to record response latency in such a way that this would be distinct from recognition time. A pen recorder was attached to the memory drum, and recorded the point at which each item was presented, and the start of the anticipation interval. The experimenter could place another mark on the tape by pressing a button when the subject made a response.

12 subjects in each group learned a second list (L2) under the same conditions and instructions as the first (L1). Both lists are reproduced in Appendix 2. 10 subjects in each group relearned the experimental list in the reverse direction during the experimental session. In each group the 10 remaining subjects and 6 of those who had learned L2 were recalled at various intervals after the initial learning and required to relearn the L1. The procedure in relation to retention and backward learning was modified as the experiment proceeded.

Results (Mn./N-Mn.)

As can be seen from Table 1. Mn. subjects performed better than N-Mn. subjects in terms of both trials and errors to reach criterion. The differences between means, tested with a Mann-Whitney U test (two-tailed), were both highly significant. For errors

TABLE 1.

Means, Standard Deviations and Medians for trials and errors to criterion.

	Mn.		N-Mn.	
	Trials	Errors	Trials	Errors
Mean	3.27	12.00	5.08	26.46
S.D.	1.64	5.31	1.43	8.63
Median	3.00	9.00	5.00	28.00

p was smaller than .002 and for trials it was smaller than .005.

The latency measure that was taken, depending as it did on the reaction time of the experimenter, was considered fairly rough, and was only scored to the nearest $\frac{1}{4}$ second, thus producing a 17 point scale. For the first anticipation trial, counting only those items on which a response was made, the weighted mean for Mn. subjects was 11.4 and that for N-Mn. subjects was 8.5. This difference was significant on a t test ($t = 2.94$, $p < .01$). There was some tendency for these latency scores to converge. Thus on the second anticipation trial the mean scores were 9.1 (Mn.) and 7.6 (N-Mn.), while on the third they were 7.5 (Mn.) and 6.7 (N-Mn.).

There was some difference between the Mn. and N-Mn. subjects in the relative proportionality of different types of error. For both groups extra-list intrusions were very rare (only 7 were made altogether). The remaining errors could therefore be classified as errors of omission (O), and of commission (C). Thus for Mn. subjects the means were 8.23 (O) and 3.75 (C), while for N-Mn. subjects they were 15.63 (O) and 10.79 (C).

In order to assess the significance of the difference the O-scores of the Mn. subjects were multiplied by 26.46/12.00, the ratio of the error means for the two groups. This gave a mean for the transformed scores of 18.15. A Mann-Whitney U test between the transformed O-scores for Mn. and the raw O-scores for N-Mn. was significant at the .05 level.

The results on the second list were similar to those on the first list in that the Mn. subjects made less errors and took less trials to reach criterion than did the N-Mn. subjects. However, as Table 2. indicates both groups performed better on the second list.

Although in absolute terms the N-Mn. groups improved more than did the Mn. group the proportional improvement of the Mn. group was greater. Although it is difficult to evaluate this difference, the fact that the difference between error-scores on L2 was more significant than on L1 (Mann-Whitney $U = 15$, $p < .001$) in spite of the fact that the samples were so much smaller, suggests that transfer was greater for the Mn. group. The difference in trials to criterion, though significant ($p < .02$) was not as impressive, but the small range (1-5) did not leave much scope for differences to manifest themselves.

TABLE 2.

Means on trials and errors to criterion on lists L1 and L2.

	Mn.		N-Mn.	
	Trials	Errors	Trials	Errors
L1 (32 Ss)	3.27	12.00	5.08	26.46
L2 (12 Ss)	2.13	3.50	3.33	10.08

The distribution of errors of omission and commission followed the same pattern as in the L1 results. The error means were, for Mn., 3.17 (O) and 0.33 (C), while for N-Mn. they were 5.58 (O) and 4.50 (C). Multiplying the O scores for Mn subjects by 10.08/3.50, the ratio of the error means for the two groups gave a transformed mean of 9.13 which was again significantly different from the mean for N-Mn. (Mann-Whitney $U = 30$, $p < .01$).

Retention

The main object in collecting data on retention was to obtain information of an exploratory kind, so that later retention studies would use suitable time intervals. As a result, there was no time interval which had more than 4 subjects assigned to it, and no formal analysis was possible. The time intervals used were 1, 2, 5 and 10 weeks for the single list subjects, and 1 week and 5 weeks for the two list subjects. (Where the retention interval was 5 weeks or more 4 days were allowed for obtaining access to defaulters). None of the 1 or 2 week subjects showed much forgetting as measured by the savings method (errors made on relearning expressed as a percentage of errors made on initial learning). 5 week subjects showed average savings of 58% (Mn.) and 47 % (N-Mn.). 10 week subjects showed average savings of 12% (Mn.) and 15% (N-Mn.). For those subjects who relearned L1 after learning L2, mean savings were rather less, 60% (Mn.) and 71% (N-Mn.) after 1 week, and 33% (Mn.) and 17% (N-Mn.) after 5 weeks. However, the variability of these subjects was such that little reliability can be assumed for these means. The only conclusion that can be drawn from this data is that with this type of material it would be wise to use delay-periods of 5 weeks or more in studies of retention.

Backward learning

Again the intentions of the experimenter were exploratory, and conditions were altered as the data collection proceeded, but this time the alterations depended on the performance of subjects under the previous conditions. After 3 subjects from each group had been run under conditions of immediate backward relearning it was found that only one, a N-Mn. subject, had made any errors (he had made 2), and a further 3 subjects from each group were required to relearn the list backwards after a 20 minute delay period. There was again little sign that the subjects found this difficult, since only two subjects made any errors. A Mn. subject made 1 error and a N-Mn. subject made 2. Finally, 4 subjects from each group were given delay periods of 30 minutes during which they worked on mathematical puzzles. This time 2 Mn. subjects and 1 N-Mn. subject made errors, but only 1 error was made by each. There is then no indication from this data that either Mn. or N-Mn. subjects find backwards recall with this material more difficult than forward recall.

Discussion

The main results from this study make it clear that the advantages conferred by mnemonic instructions persist when the material is presented more than once. This result becomes all the more impressive in view of the fact that the Mn. subjects were not this time given a prior demonstration of the system. A number of subjects were asked after the session to give descriptions of the images they had constructed, and all subjects, in both groups, were invited to give any comments about the experiment which occurred to them. It became evident that not all Mn. subjects had succeeded in carrying out the instructions fully, either in the sense that they had made images of the type suggested, or in the sense that they had made images for all pairs of items. In particular few of the images described seemed to be very bizarre. Furthermore, many of the N-Mn. subjects described themselves as carrying out mental processes, in relation to some pairs of items, which did not seem to be different in kind from those which the Mn. subjects had been instructed to carry out. These observations suggest that the observed differences in performance underestimate those in principle attainable.

On the other hand it was also clear from the comments made that most of the Mn. subjects had felt that what they were doing was very different from just trying to remember a list of words "by heart". Many of the N-Mn. subjects described themselves as having "just tried to recite the words off", and made statements like "I didn't take any notice what the words meant". From these kinds of remarks it appeared that the subjects, rightly or wrongly, distinguished at least two kinds of learning, which might be called loosely "rote" and "non-rote".

The findings on latency and the differences between error distributions are not inconsistent with these subjective observations. Thus it might be suggested that mnemonic subjects initially took longer to respond because they were carrying out a mental process equivalent to recapturing and decoding their mnemonic, whereas rote subjects had merely to call for the item that came next. The convergence of the latencies in later trials is consistent with the findings of Adams and McIntyre (1967) that natural language mediators tend to drop out with successive reproductions of the list. Similarly, the greater tendency of the mnemonic subjects to make their errors in the form of omissions is consistent with the claim of professional mnemonists that if they cannot remember their mnemonic they do not remember anything at all. It is also consistent with the notion that rote learners, when they call for an item out of the memory store, are not expecting it to come complete with a mnemonic certificate of identity, and are therefore more likely to make an error of commission by naming the first item they think of.

The findings on second list learning are consistent with the view that mnemonic subjects improve rapidly in their use of mnemonics, especially if the change in distribution of O and C errors is taken into account. The difference in improvement between Mn.

and N-Mn. subjects is not, however, very great, and it is surprising that the N-Mn. subjects improved as much as they did. One possibility is that the second list was much easier than the first, but there is nothing in the structure of the lists which would lead one to expect this. It may be worth noting in this connection that the subjective reports of the subjects suggested that the Mn. subjects had applied the mnemonic instructions more successfully than they did on their first list, while the N-Mn. subjects had made more use of natural language mediators, mainly ones making use of word structure rather than word meaning.

The findings on retention are too sketchy to require or deserve much comment. The backward learning results, on the other hand, while similarly informal, are somewhat surprising, and appear to reduce in impressiveness the claims of mnemonists on behalf of their systems. It was clear from the remarks of the N-Mn. subjects in this part of the experiment that they too were surprised at their ability to carry out this task.

Having established that mnemonic instructions improve performance both under ideal mnemonic conditions and under conditions more nearly approaching standard serial learning ones there would appear to be two different ways of continuing the investigation. One is to continue to check the claims made on behalf of mnemonics, and to attempt to discover, to the extent that the research justifies these claims, what the mnemonic subject is doing that produces the differences in performance. The other is to attempt to fit the observations about mnemonics into the field of other observations about verbal learning which bear some formal resemblance to them, and, by making use of unifying theoretical concepts, to attempt to arrive at propositions about what verbal learners will do under specified conditions, not necessarily involving mnemonic instructions.

CHAPTER 11.

Experiment 3. Rated inter-item association value and incidental retention through the creation of bizarre images.

Introduction

The link mnemonic system is usually used with material consisting of object names, and, given the nature of the instructions, this does not seem surprising. It would be much harder to form images of abstract concepts, and, if the items were are adjectives, or worse still, adverbs, although one might still make images, it would be difficult to make these distinctive. It is, in fact, very easy to find material on which the mnemonic subject cannot use his techniques, and a corollary of this observation is the implication that subjects attempting to use mnemonic techniques will not always perform better than subjects lacking such a technique. There are, almost certainly, some circumstances in which they will perform worse.

These reflections lead to the conclusion that mnemonic performance is not theoretically interesting through being superior to non-mnemonic performance, but through being different from it. On this view an experiment which finds mnemonic performance to be superior to non-mnemonic performance merely shows that whatever it was that the mnemonic subjects were doing which was different from what the non-mnemonic subjects were doing was peculiarly suited to the materials and mode of presentation used. An implication of this view is that differences of a qualitative kind between mnemonic and non-mnemonic subjects, such as differences in response latency, error type, and error distribution, should be of at least as much interest as differences in the level of performance. In addition, this view implies that there would be much value in presenting mnemonic subjects with different materials to learn, and observing their performance on these different materials.

Mnemonic subjects frequently remark of particular pairs of words that it is difficult to make an image connecting them, or that it is difficult to make a bizarre image connecting them. This difficulty appears to be a function both of the imageability of the individual items, and of the degree of relationship between them. If in fact the mental operations that the mnemonic subject is carrying out are responsible for the observed differences between his performance and that of the non-mnemonic subject it would be of some interest to see what effect differences in the material have on what he does. In fact, in Experiments 1 and 2, all of the items were highly imageable, so the most relevant variable for us to try to vary would be the degree of relationship between pairs of items.

Wood (1966) found no evidence that instructing the subjects to make their images bizarre made their performance any different from that of subjects simply instructed to make images. On the other hand, there is no evidence that the difference in instructions produced any difference in the behaviour of the subjects. It could be that under the conditions used by Wood it was not possible for most subjects to produce bizarre images. Also, although this would not further the claims of the mnemonists, the difference between bizarre and mundane imagery might be expressed in differences between variables other than those reflecting level of performance.

In view of the above considerations it was decided that a sample of images would be collected and analysed to see whether differences in interitem relations affected the subject's implementation of bizarre imagery instructions. To avoid the employment by the subjects of other memory techniques than those implicated in the instructions it was decided that they would be asked to make images of a bizarre, active nature, linking pairs of items, without being told that this was a memory task. If at a later date they were able, without having expected to be required to do this, to recall the material on the basis of their images, this would be an interesting finding, as it would militate against an explanation of mnemonic performance in motivational terms.

Material

Although it was initially intended to use only three pairs of items in the present study, the possibility was envisaged of carrying out other studies later using inter-item association as an independent variable. For this reason data was collected relating to the degree of association between all pairs of items in lists L1 and L2 from Experiment 2. 20 judges were presented with two matrices, one for each list, in which each cell represented a pair of words. They were asked to place in each cell a number from 1 to 5 representing the degree of relatedness of the pair of words in question. They were asked to fill out the matrices at random rather than in a systematic way.

For each item-pair the ratings of the 20 judges were averaged to produce a mean inter-item association (a.) rating. For each list a split-half reliability coefficient was computed by dividing the judges randomly into two groups of 10, taking the average ratings for each item separately for each group of judges, and correlating these two sets of average ratings. Rho was .797 for list L1 and .806 for list L2. Both distributions of mean a. ratings were heavily skewed towards the low end, with modes of 1.5 (L1) and 1.6 (L2).

Initially three item-pairs were selected, having low (L), medium (M), and high (H) inter-item association values. These were EGG-PIANO (a.=1.05), TORCH-HANDBAG (a.=2.25), and WINDOW-LADDER (a.=3.70). Later it was decided to replicate the study, and a further three item-pairs were selected, having the same a. values. These were PENCIL-HORSE (L), ARMCHAIR-WINDOW (M), and SPECTACLES-BOOK (H).

Method

At a time when 107 second-year psychology students were gathered in a lecture theatre expecting a lecture the experimenter entered and requested their cooperation in a study he was conducting. He informed them that he was collecting a sample of bizarre associative images for analysis, and gave descriptions of two sample bizarre connective images. Paper was handed out, and the students were asked to make active, bizarre, multi-modality connective images, and to write descriptions of these images, for each of the three experimental pairs of items. They were presented in the order L-H-M in order to avoid an obvious sequence. Three minutes were allowed for each pair. They were asked to write their names on the paper, the reason given being that the experimenter might wish to contact some of them for the purpose of obtaining a further sample of images. The image descriptions were rated by the experimenter and one other rater on a 5 point scale for the degree to which they conformed to the instructions.

The ratings of the two judges were intercorrelated, giving a rho of .793. Since this represented a reasonably high level of reliability each image description was assigned a "goodness" rating equal to the mean of the ratings given by the two judges.

15 weeks after the initial session the experimenter returned to obtain recall data. The locale and situation were the same as before. The students were reminded of the events of the previous session and were asked to write down the second word of each pair as a response to the first. The pairs were presented in the same order as before, and 20 seconds was allowed for each pair. The subjects wrote their names on their answer-papers, and these were collected.

As already mentioned it was later decided that, because there had been so few items, and because the order in which they were presented might have been relevant to the results obtained, the study should be replicated. This was done, using a different class of 121 second-year psychology students. The only differences from the first version of the study were that three different item-pairs were used, that these were presented in the reverse order from the point of view of their a. values (i.e. the order M-H-L), and that the ratings of the images were carried out by three judges not including the experimenter. The only other difference was that for administrative reasons it was necessary to hold the sessions after the lectures rather than before.

Because not all of the students were present at both experimental sessions it was not possible to include in the analysis the data from all students who had constructed images. In the first version of the experiment the results of 88 subjects were analysed, and in the replication, the results of 94.

<u>Results</u>

To facilitate comparison the results of the original study (A), and the replication (B), will be presented in parallel.

The results are summarised in Tables 1. and 2. The differences between the mean ratings for those subjects who did and did not recall the items suggest that the ratings are positively related to recall performance. This suggestion was tested by means of Mann-Whitney U tests corrected for ties and converted to normal deviates. Table 3. summarises the results of these tests.

These results, then, strongly suggest that the ratings relate to recall performance.

Friedman two way analyses of variance were carried out in order to test the suggestion that subjects made images of different degrees of "goodness" on the different items. These analyses were both significant, at the .001 level for (A) and at the .01 level for (B). Wilcoxon signed-ranks matched-pairs tests were used to compare the L pairs with the H pairs, and the H pairs with the M pairs. The T values were converted to normal deviates. For the L/H comparisons the results were $z(A) = 3.28$ ($p < .001$), and $z(B) = 2.84$ ($p. < .01$). For the H/M comparisons the results were $z(A) = 2.73$ ($p < .01$) and $z(B) = 3.11$ ($p < .001$). Thus each item-pair produced a mean rating significantly different from each of the others.

In order to compare the levels of recall performance on the different item-pairs McNemar tests were carried out for comparisons L/H, L/M and H/M. For the (A) data the numbers involved in the L/H comparisons were small, so this was evaluated by means of a Sign test. This was significant at

TABLE 1.

Mean "goodness" ratings and numbers of subjects succeeding and not succeeding in recalling each item. (A).

	L		H		M	
	Number	Rating	Number	Rating	Number	Rating
Correct Recall	67	3.69	76	3.30	45	3.22
Not Correct Recall	21	3.33	12	2.75	43	2.56
Total	88	3.61	88	3.23	88	2.90

TABLE 2.

Mean "goodness" ratings and numbers of subjects succeeding and not succeeding in recalling each item. (B).

	L		H		M	
	Number	Rating	Number	Rating	Number	Rating
Correct Recall	63	3.48	54	3.21	34	3.04
Not Correct Recall	31	3.00	40	2.62	60	2.13
Total	94	3.22	94	2.96	94	2.46

TABLE 3.

Results of U-tests on differences in "goodness" ratings of images from subjects who did and who did not recall the items.

	L		H		M	
	z value	p	z value	p	z value	p
(A)	2.11	$<.05$	1.77	.08	2.82	$<.01$
(B)	2.38	$<.05$	2.07	$<.05$	3.02	$<.01$

the .05 level (two-tailed), in the direction opposite to that of the L/H difference in "goodness" ratings. All of the other differences were evaluated by Chi-square, and were consistent in direction with the differences in mean ratings. The other differences for the (A) data were significant at the .001 level, giving $\chi^2 = 13.78$ (L/M), and $\chi^2 = 22.40$ (H/M). For the (B) data L/H gave $\chi^2 = 6.81$ ($p < .01$), L/M gave $\chi^2 = 12.43$ ($p < .001$), and H/M gave $\chi^2 = 8.34$ ($p < .005$). Apart from the reversal involved in the L/H comparison for (A) all of these comparisons are consistent with the "goodness" ratings comparisons.

Discussion

The results are highly consistent with the view that imagery mnemonics have a direct effect upon performance; an effect which it would be difficult to account for in terms of such factors as motivation, since the subjects did not know when they made their images that they were to be asked to recall the material.

Since one of the rating criteria was the level of bizarreness of the images the results are in contrast to those of Wood (1966), who was unable to show that either imagery or bizarreness had any effect. However, there was no independent evidence that Wood's subjects produced bizarre imagery, or indeed any imagery at all. Laboratory experience would have led the present writer to predict that under the conditions used by Wood little bizarre imagery would be produced.

The differences between the ratings for the images produced for the different items are not quite as convincing as those between successful and unsuccessful images. It is possible that content differences in the images could have led the judges to rate inconsistently in such a way as to produce these rating differences. On the other hand the factors which would lead to this inconsistency are the same ones which might be expected to lead the subjects to make images of different degrees of "goodness" for the different item-pairs. Although the rated differences in the inter-item association value have high face validity it could be argued that it was not these differences which led to the observed differences in the images described. If this was correct our second prediction could not be regarded as clearly confirmed, but it would still be possible to maintain a less specific but nevertheless important position, namely, that the ability of the subjects to carry out the instructions is affected by the nature of the items. And, by extension, this would lead to the reasonable hypothesis that the nature of the items will determine to some extent what types of natural language mediators will be used by uninstructed subjects.

The relation that appears in our data between inter-item association and "goodness" of image is not monotonic. It is not difficult to find hypotheses to account for this. With low inter-item association practically any image would be bizarre. With high inter-item association the form of the association is usually fairly specific (consider for example the association between EGG and CUP) and hence relatively easy to ignore deliberately. But with moderate inter-item association there is usually a large number of ways in which the items could be related, and it might be difficult to put them all aside in favour of a bizarre one. Irrespective of such ad hoc theorizing there is no particular reason to predict a monotonic relation.

The reversal of direction in the differences between the ratings and levels of recall performance for pairs L and H in (A) is of some interest. It suggests that the relationship between the ratings and successful recall is not a simple one. In both runs of the experiment the ratings for H showed higher variability (though not significantly) than did the ratings for L. It may be that some subjects in (A) did not succeed in setting aside the pre-existing association between WINDOW and LADDER in favour of a bizarre one, but that for some of these subjects that pre-existing association was sufficient to mediate a recall. There may well not be a straightforward relationship between ratings of inter-item association and the capacity of the pre-existing associations to mediate recall, and this could account for the reversal not occurring in (B). In other words, if the natural language association between SPECTACLES and BOOK forms a less satisfactory mediator than that between WINDOW and LADDER there will be less tendency for subjects who have failed to carry out the mnemonic instructions to recall correctly in spite of this failure.

Clearly the technique used in this study is capable of considerable development and extension. It would be valuable to attempt to devise more efficient means of analysing and evaluating image-descriptions. It would also be helpful to discover what is the relation between inter-item association ratings and recall in the absence of mnemonic instructions.

CHAPTER 12.

Experiment 4. Learning and retention with successive approximations to the complete mnemonic instructions.

Introduction

Although the mnemonic system studied in Experiments 1, 2 and 3 is one of the simplest general-purpose systems in use, the instructions given to the subjects are fairly complex. The subject is asked to consider the items in pairs, and to connect these pairs by images. He is asked to construct his images in such a way that they are vivid, active, and bizarre. It is suggested that he use as many sensory modalities as possible in his images.

Although it is clear that the instructions as a whole have a pronounced effect on the performance of the subjects it could be that it is not the whole set of instructions that is having this effect. Thus subjects simply instructed to think of the items in pairs might exhibit performance indistinguishable from that of subjects with the complete mnemonic instructions.

There are some problems attached to any attempt to assess the differential effectiveness of the elements in a set of instructions such as the ones under consideration. The first problem is that of identifying the elements. What may, from the experimenter's point of view, seem like a distinct element may not be so from the point of view of the subject. However, it would seem reasonable in the absence of any obvious alternative to assume that what the experimenter sees as a distinct instruction can be implemented as a distinct instruction by the subjects.

The natural approach of using a factorial design is not easily applied to this situation for two reasons. Firstly the number of elements is such that a design of this sort would be extremely unwieldy, and secondly the elements do not exhibit the degree of independence which is needed to make such a design meaningful. Thus one cannot instruct the subject to make his images bizarre without also instructing him to make images.

An alternative to a factorial design is to use a series of instructions of increasing degrees of mnemonic completeness. One problem with this technique is again that some of the instructions may not be independent from the others, and that this lack of independence may not be symmetrical. Thus it could be that vivid images are not usually bizarre, whereas bizarre images are usually vivid. Thus it is important that the order in which the elements are added to the instructions be such as to take into account this sort of overlap.

The order in which it was decided that the elements should be added, bearing the above considerations in mind, was as follows:-

(a) Pair each item with the previous one.
+(b) Make images of the paired items.
+(c) Make the images vivid.
+(d) Make the images active.
+(e) Use as many sensory modalities as possible.
+(f) Make the images fantastic (bizarre).

It was, however, realised that this was not the only order which would have been acceptable. Thus there was no very strong argument against reversing the order of elements (e) and (f).

Since some subjects had indicated that they had been unsure how to make use of their images in recall it was decided that a recall instruction should be tried with each of the above sets of mnemonic instructions. There would thus be 12 experimental groups plus a control group with standard serial anticipation instructions.

Since it was thought that some aspects of the mnemonic instructions which did not affect learning might affect retention it was decided that each subject should be asked to relearn the experimental list after one of three retention intervals. One consequence of this decision was the further decision that the learning phase of this experiment should run for a fixed number of presentations of the list, rather than to a criterion of one error-free anticipation, as in Experiment 2.

Since a variety of instructions were to be used, some of which would take less time to carry out than would others, it was felt that a relatively long presentation time should be used. However, considerations of subject availability required that a short list be used. The combination of a short list and long presentation times might have led to an unfortunately high level of performance on the part of the control subjects, which would vitiate the analysis. It was therefore decided that the experimental list should be one having low inter-item association.

Material

Two 10 item lists were used, one for practice (List P), and one experimental list (List E). Both lists were low inter-item association orderings of items out of the lists which had been rated for inter-item association as described in Chapter 11 (p.157). The average inter-item association ratings for both lists were 1.12.

Subjects

The design of the experiment called for 273 subjects. First-year psychology students were used. However, 17 of the students who attended the initial session of the experiment failed to return for the retention test. These subjects were replaced. It was considered that the selection of 273 subjects out of a total of 290 could not be regarded as a serious source of bias in this particular experimental situation.

Method

Subjects were run individually. All subjects learned List P. This was presented by memory-drum at a 5 second rate, followed by 5 anticipation runs at the same rate. Inter-trial interval was approximately 10 seconds. The subjects' error scores on this task were used as an index of learning ability, and on this basis they were placed in one of 7 performance categories. They were then assigned randomly to one of the experimental groups, with the restriction that the number of subjects in each learning-ability category was to be kept as near equal as possible, at all stages of the experiment. They were also assigned randomly but with the same restriction to one of three retention intervals. These were 5 weeks, 10 weeks, and 15 weeks (plus or minus 4 days). Each subject was informed that a large number of different sets of instructions were being used in the experiment, and that he should not necessarily assume that the instructions he received were expected to facilitate his learning of the experimental list.

The subject was then given the learning instructions, and recall instructions, if any, appropriate to the group to which he had been assigned, and the experimental list was presented by memory-drum at a 7 second presentation rate, followed by 3 anticipation trials at the same rate. The inter-trial interval was again approximately 10 seconds. The subject was then told when he was to return to relearn the experimental list, and that he would be reminded nearer the date. He was told in general terms the object of the experiment. It was made clear to him that it was the particular set of instructions he had been given that was under test, and not his own memory, and that any rehearsal would remove all point from the experiment. All subjects seemed to understand this, and appeared willing to cooperate.

At the second session all subjects were given four anticipation trials with the experimental list, under the same conditions as before. They were then asked whether they had recalled their mediators on the first re-presentation of the list, or later. Those who did recall mediators from the first session were asked to describe some of them.

Instructions

The subjects were divided into 13 groups of 21. The 13 were made up of a control group (C), which had standard serial anticipation instructions, and 12 experimental groups. These latter consisted of 6 groups having one of the 6 sets of learning instructions, and 6 groups with the learning instructions and a recall instruction. Since the instructions were the experimental variable they are given below.

The learning instructions were as follows:-

(a) In learning the list I want you to think of each item as being paired with the previous item. So when, say, the fourth item appears, think of the name of the third item in relation to it. Do you understand?

(b) In learning the list I want you to think of each item as being paired with the previous item, and to make some kind of a mental image of the pair of items. So when, say, the fourth item appears, think of the third item, and make an image of the two of them together. Do you understand?

(c) In learning the list I want you to think of each item as being paired with the previous item, and to make a really vivid mental image of the pair of items. So when, say, the fourth item appears, think of the third item, and make a really vivid image of the two of them together.

<u>(d) In</u> learning the list I want you to think of each item as being paired with the previous item, and to make a really vivid, active, image of the pair of items. Make the items <u>do</u> something in the image. So when, say, the fourth item appears, think of the third item, and make a really vivid image of the two of them doing something together. Do you understand?

<u>(e) In</u> learning the list I want you tothink of each item as being paired with the previous item, and to make a really vivid, active, image of the pair of items. Make the items <u>do</u> something in the image, and use as many sensory modalities as you can; see them, feel them, hear them, even smell them if you can. So when, say, the fourth item appears, think of the third item, and make a really vivid image, using as many sensory modalities as you can, of the two of them doing something together. Do you understand?

<u>(f) In</u> learning the list I want you to think of each item as being paired with the previous item, and to make a really vivid, active, image of the pairs of items. Make the items <u>do</u> something in the image, and

make the relationship between the items really bizarre, fantastic, and absurd. Even if a pair of items fit together in some quite natural everyday way do not make use of this relationship in your image, but rather put them together in some quite unnatural way, for instance having a cart hollowed out of a horse rather than a horse pulling a cart. Use as many sensory modalities as you can in your image, bringing in vision, feeling, hearing, even smell if you can. So when, say, the fourth item appears, think of the third item, and make a really vivid, lively, thoroughly fantastic image, using as many sensory modalities as you can, of the two of them. Do you understand?

Ideally, the same recall instructions should have been used for all 6 groups having such an instruction. The full recall instruction was:-

(B) When trying to recall an item, try to remember what you did with the previous item, and then try to recognise the item you are looking for in the resulting image.

Clearly this instruction was not applicable to the group which had instruction (a), so it was necessary to construct for this group a reduced version which was:-

(A) When trying to recall an item bear in mind the previous item and try to think which item you paired it off with.

The 13 groups, 6 with recall instruction, either A or B, and 7 without, labelled with the symbols for the instructions which they received, were C, a, b, c, d, e, f, aA, bB, cB, dB, eB and fB.

Results

The mean error scores for each group in the learning phase of the experiment are shown in Figure 1. For all groups the error distributions were strongly skewed, as can be inferred from a comparison of the means and the standard deviations (also given in Figure 1.). Significance tests were therefore carried out on the error scores transformed by adding 1 and taking logs.

An analysis of variance was carried out on the error scores of the 12 instruction groups. This was a two-way analysis with replications, there being 6 learning instructions, 2 recall instructions and 21 individuals in each group. The results are summarised

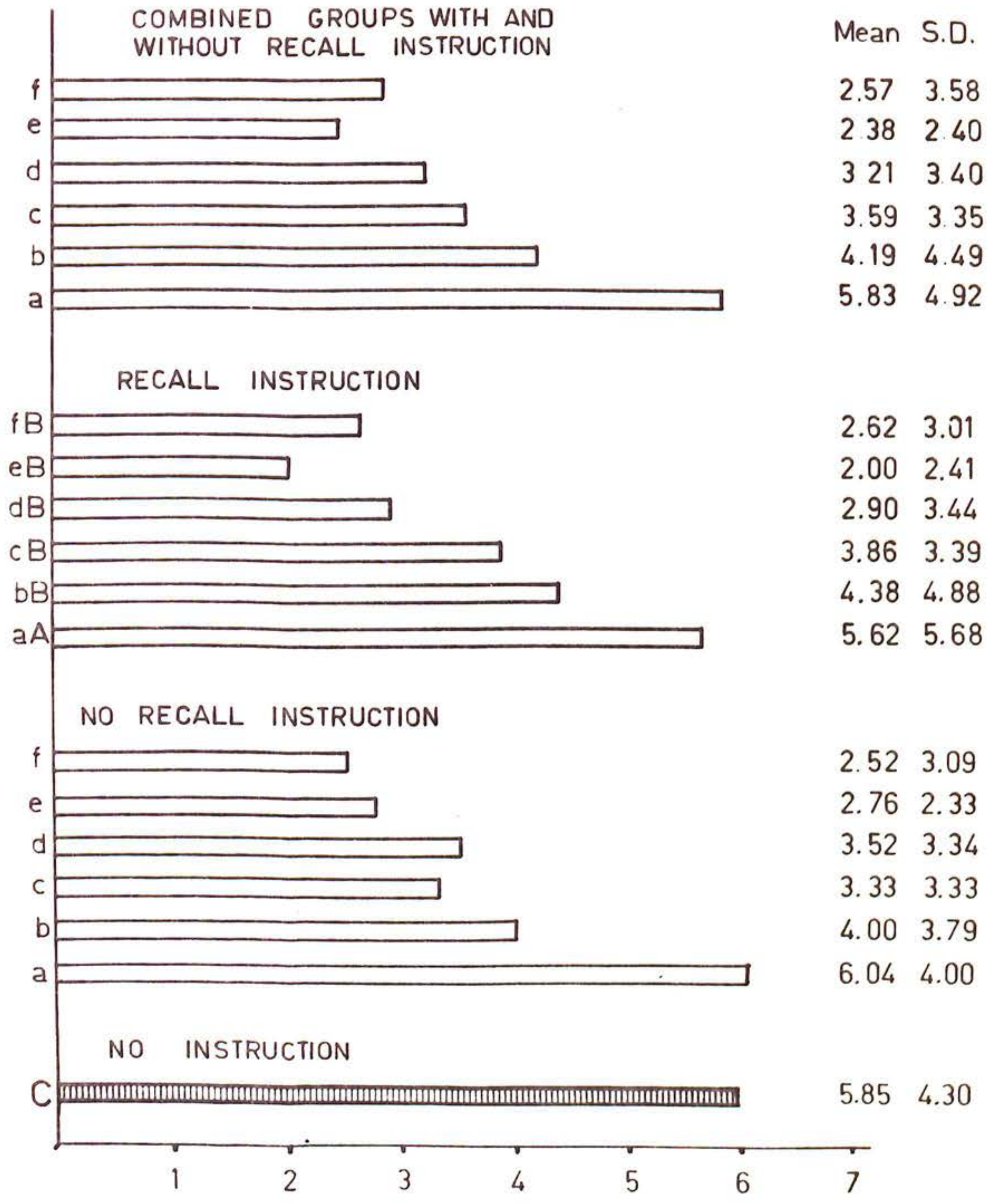

FIGURE 1. Experimental List Error-Score Means and Standard Deviations.

in Table 1. The only significant effect was that of learning instructions ($p < .001$).

The variance due to recall instructions being lower than that due to error, the groups with and without recall instructions were pooled, and a one-way analysis of variance was carried out, this time using seven different levels of learning instruction, since it was now possible to add the control group into the analysis. For this analysis F was 6.01 (df 6/266), which was significant at the $p = .001$ level.

Post-hoc comparisons were carried out on the means of the 7 groups of the one-way analysis of variance, using the method of Scheffé (see Hays (1963) pp.483-489). This method was used in spite of its relative lack of power, because although the instructions were ordered on the dimension of mnemonic completeness no clear expectations had been stated as to the direction or size of differences between particular groups. In fact the only pair-comparisons which were significant at the .05 level were those between the control group and groups (e) and (f), and between group (a) and groups (e) and (f). Another series of ordered

TABLE 1.

Two-way analysis of variance on error scores of instruction groups in the learning session.

	SS	Df	Variance estimate	F
Recall instructions	.055	1	.055	
Learning instructions	3.050	5	.610	5.17 ($p < .001$)
Error (Within groups)	28.455	240	.118	
Interaction	.268	5	.054	
Total	31.828	251		

comparisons were carried out between the means of all groups above and all groups below each inter-group point on the instructions scale. Thus the mean of the control group was compared with the combined mean of groups (a) to (f). The combined mean of groups (C) and (a) was compared with that of groups (b) to (f), and so on. Only two of these comparisons were not significant at the $p = .05$ level, those between (C) and the rest, and between (f) and the rest. To summarize these analyses, they suggest that the recall instructions did not affect performance, and that each addition to the learning instructions, with the exception of the bizarreness instruction and the possible exception of the pairing instruction, produced an improvement in learning performance.

One of the findings of Experiment 2 was that subjects with mnemonic instructions made a higher proportion of their errors in the form of omissions than did control subjects. It seemed worthwhile to see whether this tendency increased with the completeness of the mnemonic instructions, as appeared to be the case with learning performance. This hypothesis was in terms of relative proportions of errors per subject falling into two categories, but since many subjects had made either no errors at all or no errors in one of the categories it was not possible to carry out a trend test on the proportional error scores. An alternative was to ask whether the proportion of subjects making more errors of commission than of omission decreased as one moved up the series of instructions. This alternative involved throwing away a lot of information, since subjects who had made the same number of errors of each kind had to be excluded, and since the original hypothesis was consistent with the possibility that most subjects made more errors of one of the two kinds, but it did represent a conservative test of the hypothesis. The statistical test selected was the rank t-test (see Bross, 1954). This gave $t^2 = 9.40$ ($p < .005$), a convincing indication of a trend in the expected direction.

Although the trend was as predicted it might have been explicable without assuming that the use of mnemonic techniques gives rise to errors of omission rather than errors of commission. Thus, combining all groups, the proportion of all 1st-trial errors which were errors of commission was .338, whereas the equivalent proportion for trials after the first was .541. This being so the observed trend might be a reflection of the fact that the subjects with higher level mnemonic instructions made less errors after the first trial. To check this possibility further rank t-tests were carried out separately on errors made in the first anticipation trial and on errors made in later anticipation trials. The first of these gave $t^2 = 5.41$ ($p < .02$), and the second gave $t^2 = 7.22$ ($p < .01$). These results suggest that the overall trend is not explicable in terms of a decreasing tendency to make errors in later trials.

The mean error scores for the 39 retention groups are set out in Table 2., together with their standard deviations. These error scores, particularly in the 5-week retention groups, tended to be very low, and their distributions were again strongly skewed. These scores were again transformed by adding 1 and taking logarithms, and a three-way analysis of variance was carried out. Table 3. gives the results of this analysis, which used 3 retention intervals, 2 recall instruction conditions, and 6 levels of mnemonic instruction (the control group being excluded). The results show highly significant effects for instruction level and retention intervals, and the interaction between these just fails to reach significance at the $p = .05$ level. Recall instructions were again not significant, although they gave an F value greater than 1, so that although the groups with and without recall instructions were again combined for further analysis, this was done less confidently than it was in relation to the acquisition data.

The mean error scores for the control subjects and its combined recall-instruction and no-recall-instruction groups are shown in Table 4. and in Figure 2. Trend tests were carried out for each retention interval (see Jonckheere, 1953). For the 10 and 15-week retention intervals these were significant at the $p = .05$ level. For the 5-week interval S_o was 1.95, which is very slightly below the 1.96 value required for significance at the $p = .05$ level.

The reversal of the trend which the (f) group subjects exhibit at all three retention intervals is difficult to evaluate statistically, since the groups involved are too small to enable comparisons of the uncombined group by pairs means to reach significance,

TABLE 2.

Means and SDs of retention groups.

(n for each group was 7).

Instruction condition	Retention Interval					
	5 weeks		10 weeks		15 weeks	
	Mean	SD	Mean	SD	Mean	SD
C	3.7	2.63	9.4	5.21	13.2	6.18
a	4.2	5.24	11.6	4.54	12.4	4.31
b	3.8	3.18	12.7	5.61	14.7	6.05
c	4.3	5.13	9.5	4.31	8.0	3.76
d	2.5	2.72	8.0	3.55	9.4	4.13
e	1.2	3.43	6.7	2.98	7.2	3.93
f	3.1	2.47	8.2	4.35	8.1	3.71
aA	3.8	1.96	10.3	5.19	13.1	5.21
bB	3.0	4.11	7.9	3.44	11.4	4.43
cB	2.2	2.10	11.6	6.71	9.8	5.10
dB	1.9	1.04	7.3	2.93	7.4	2.87
eB	1.4	1.22	5.1	3.67	6.8	3.33
fB	2.9	1.28	7.2	4.77	9.3	3.56

TABLE 3.

Analysis of variance on transformed retention error-scores.

	SS	Df.	Variance estimate	F
Learning Instructions (A)	3.300	5	.660	4.923 ($p < .001$)
Recall Instructions (B)	.164	1	.164	1.227
Retention Intervals (C)	1.466	2	.733	5.473 ($p < .001$)
A X B	.275	5	.055	
A X C	2.480	10	.248	1.851 (F for $p = .05$ is 1.88)
B X C	.122	2	.061	
A X B X C	1.771	10	.177	1.322
Within Groups	28.944	216	.134	
TOTAL	38.522	251		

TABLE 4.

Error scores for Control groups (each n is 7) and for combined groups with and without recall instructions (each n is 14) on the retention trials.

Instructions	Retention Interval 5 weeks	10 weeks	15 weeks
C	3.7	9.4	13.2
a	4.0	10.9	12.8
b	3.4	10.3	13.0
c	3.2	9.6	8.9
d	2.2	7.7	8.4
e	1.3	6.8	7.0
f	3.0	8.8	8.7

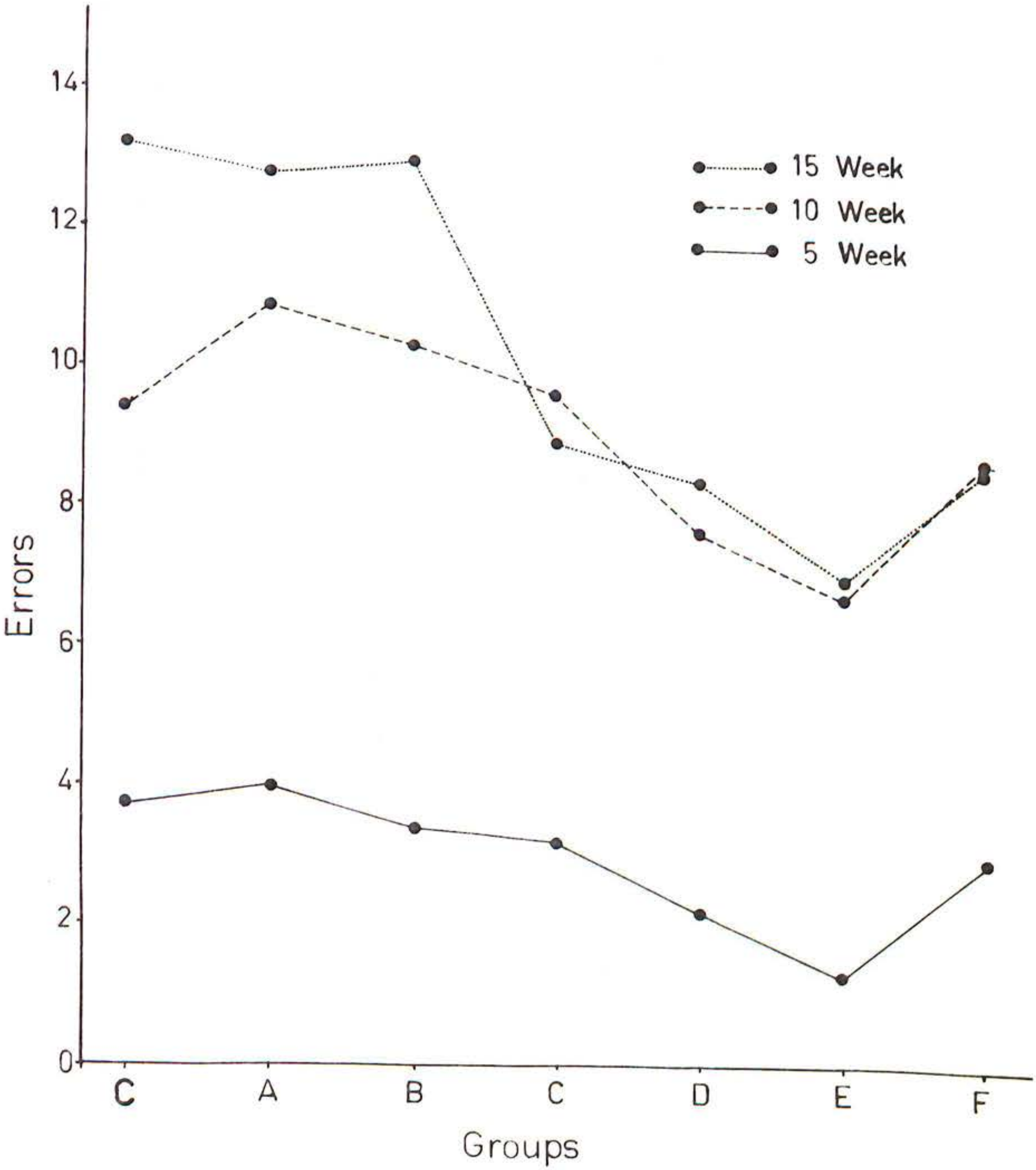

FIGURE 2. Mean Error Scores of Retention Groups at 3 Intervals with Recall- and No-Recall-Instruction Groups combined.

while post hoc comparisons of the Scheffé type cannot easily be used after a three-way analysis of variance. On the other hand since this reversal occurred in all six relevant pairs of groups it is unlikely to be a statistical artifact.

An analysis was carried out on the relative frequency of errors of omission and commission in the retention data. Overall relative proportions of errors of commission were .291 for the first retention trial, and .592 for trials after the first. There was, however, no evidence in the data of systematic between-group differences.

Some introspective data was collected from the subjects after the retention trials. They were asked whether they had used any learning techniques other than those suggested by the experimenter. Where appropriate they were asked to give descriptions of the type of image they had used. They were also asked to indicate which of the items anticipated correctly on the first retention trial had been recalled through the image they had made, and whether they had recalled the rest of their images after seeing the items again.

The answers to these questions were highly informative. Only 7 of the 21 control subjects said that they had learned entirely by rote, while 6 of them described themselves as having used some systematic mnemonic technique, for instance as having positioned the objects mentally around the experimental room. In contrast to this, only 8 of the remaining 252 subjects described themselves as having used mnemonic techniques other than those suggested by the experimenter. On the other hand many of the instructed subjects described themselves as having had difficulty in carrying out their instructions, and the images they described bore witness to the accuracy of this information, since many of them did not seem to be in any way vivid. The (f) and (fB) groups were by far the worst in these respects. In these groups 7 out of 42 subjects said they had been unable to make images at all, and only 10 of the remainder produced any sample images which the experimenter would have classified as being in any way bizarre.

Although it would be hard to quantify the extent to which subjects recaptured their original images during the retention trials their reports gave the impression that there was an increasing tendency to do this as the mnemonic instructions became more complete, at least up to the level of the (e) instruction, although, rather oddly, many of the subjects who said they did recall their images after the first retention trial said that they did not think the images had in any way facilitated recall. Nevertheless there was an increasing trend in the numbers of subjects ascribing first-trial recall to the effect of images.

11 subjects in groups (C) to (c) said that they did not even recognise the items in the list during the retention trials. No subject above group (c) said this, although 3 of them were convinced that the items were presented in a different order during the retention trials.

Discussion

Much of the detailed analysis which it had been hoped would be possible with this experimental design was made impossible or worthless by the relative paucity of errors. It appears that the task was not difficult enough in spite of the use of low inter-item association lists. On the other hand it also appears that there was a sense in which the task was too difficult, as the (f) and (fB) subjects seem, both from their error scores and from their subjective reports, to have had great difficulty in carrying out their instructions.

The results on the initial learning confirm those of the earlier experiments, and suggest that most of the elements in the mnemonic instructions make some difference to performance, although the subjective data suggest that naive subjects are unable to respond to mnemonic instructions as easily as some authors, (e.g. Wood, 1966) assume. Although the results of Experiment 3 suggest that bizarreness is a factor relevant to performance the present results do not confirm this. In the retention trials the groups having this instruction actually did worse than some of the groups without it. However, this fact can be used in partial refutation of one criticism which might be made of the design of this experiment. It happens that the instructions increase in length as they increase in mnemonic completeness, and it might have been argued, if there had been no reversal in the trend, that it was the length and not the content of the instructions which mediated increments in performance.

If the apparent ineffectiveness of the bizarreness instruction is, as has been suggested, due to the subjects not being able to carry out this instruction under the conditions of this experiment it is likely that there are as yet few, if any, experiments in the literature in which this type of imagery was used by the subjects. The most obvious explanations for the failure of the subjects in this matter are that the presentation time per item (7 seconds) was too short, or that training is required before subjects can produce this type of imagery. The latter explanation is supported by the relative paucity of bizarre imagery in Experiment 3, when the subjects had 60 seconds in which to construct images. Only Wallace, Turner and Perkins (1957) have used trained subjects, and their instructions did not mention bizarreness. Since they did not sample their subjects' images we do not know if any proportion of them were bizarre. Their findings suggest a presentation rate of about 5 seconds as being about the fastest rate at which trained subjects can use even non-bizarre imagery mnemonics effectively, which leads to the suspicion that presentation time may have been an important limiting factor in the present experiment even for some subjects without the bizarreness instruction. This suspicion is borne out by the subjective reports of the subjects.

The trend found towards making errors in the form of omissions as the instructions became more complete is consistent with the view that the fully mnemonic subject is learning in a way different in kind rather than in degree from that of the rote subject. The result confirms that obtained in Experiment 2, and, given that even the most successful mnemonic subjects in these experiments are not very successful, it would not be very surprising to find that professional mnemonists were correct in thinking that a fully mnemonic learner would make no errors of commission.

Although no questions were formally asked about this aspect of their experience, many subjects spontaneously made remarks implying an awareness of some qualitative difference between mnemonic and non-mnemonic learning. Remarks like, "It is like magic", and, "It felt quite different from ordinary learning", were frequent. It may be relevant that although 76 subjects made remarks of this sort, no subject in groups (C), (a), or (aA) made one. This might be taken to suggest that imagery is the important distinguishing feature of mnemonic learning, if it was not the case that our series of instructions was defective in that it missed out a possibly important instruction at this point. It is in fact arguable that there should have been an instruction between (a) and (b) according to which the subjects would have been asked to think of the items as being in pairs, and to connect these pairs in some way. The absence of this instruction reduced the force of the above observation. Remarks made about the practice task by many subjects confirm both the suggestion that the subjects were aware of two kinds of learning, and the reservation about assuming that imagery was the critical variable. Many of them described some attempt to connect up the items sequentially, not necessarily through the use of imagery, and then went on to say something like, "But the items went past too quickly, so I gave up and learned the list by rote".

The retention data are not by any means as clear as the data obtained during acquisition. Although there was a trend at each retention interval it was not as straightforward a trend as that which appeared in acquisition. Furthermore the absolute differences between the mean error scores of groups adjacent on the series were smaller than was the case on the acquisition phase of the experiment. This was an unexpected result, since other studies on mnemonics (e.g. Smith and Noble, 1965) have suggested that the effect of a mnemonic upon retention may be greater than its effect upon acquisition. However, there is not sufficient evidence about the matter for this result to call for special explanation.

Although the interaction between retention interval did not quite reach significance at the .05 level it may be worthwhile to consider what it might mean if it is nevertheless veridical. It would suggest that once the subjects are using vivid imagery the forgetting curve levels off after about 10 weeks and there is little if any more forgetting. For subjects not using vivid imagery, however, Figure 2. suggests that if longer retention intervals had been used error scores quite a lot larger would have been obtained. Another possibility is that for all groups only items not mediated by vivid imagery are forgotten, and most of these are forgotten by the 10th week. If the introduction of the vivid imagery instruction produces a sudden increase in the proportion of items mediated by such imagery results of the kind obtained in this experiment will result. These explanations are clearly over-generous to the data we have. However, it is worth noting in this connection that the memory prodigy Shereshevskii, studied by Luria (1960) over a number of years, showed little, if any, forgetting of the material he had memorised, and the most striking aspect of the way in which Shereshevskii remembered material is that he always made use of vivid imagery mediators.

The failure to find in the retention data the same sorts of trend in the proportions of errors of omission and commission as were found in the acquisition data is not altogether surprising. In the first place the groups were small and so were the numbers of errors. Secondly, there is no particular reason to suppose that subjects used in relearning material they failed to recall on the first retention trial the same techniques they had used in learning it originally. Thirdly, we do not know enough about the kinds of change that take place in a memory trace over time to justify any surprise at this particular kind of change.

The introspective reports of the subjects were, on the whole, consistent with the formal results.

However, in addition to acting as some kind of a check on the formal analysis they suggest some general comments outside that analysis. They suggest, for instance, that the control subjects cannot be regarded as non-mnemonic, nor the people receiving the fullest mnemonic instructions as fully mnemonic. Thus the intergroup differences observed in this study probably underestimate those which might have been observed with fuller control over the behaviour of the subjects. The subjective reports also suggest that part of the effect of the instructions was to inhibit whatever tendencies to explicit mediation of other kinds the subjects may have had.

CHAPTER 13.

Experiment 5. Learning and retention with and without mnemonic instruction, for two different lengths of list.

Introduction

One of the most striking findings of Wallace, Turner and Perkins (1957) was that once their subjects had learned a number of lists by using mnemonics they were able to learn lists of widely disparate lengths at very similar rates of presentation. Thus they present a table (p.6) showing, among other things, average item-input-time for one of their subjects, on lists ranging in length from 25 to 700 items. For 9 lists, the rank order correlation between list length and average item-input-time was -.37. It is, however, worth noting that the subject was self paced, and that his average input-time for the 9 lists was 25.2 seconds per item, which may be compared with input rates of from 3 to 5 seconds that were achieved when the subjects were asked to learn the items as fast as they could. In fact, in another table, Wallace, Turner and Perkins present (p.8), for another subject, item-input-rates on a series of 10 lists of 25 items, and 3 lists of 50 items, when the subject had been given instructions to learn as fast as possible. The subject took 20.4 seconds per item on the first list, but rapidly increased his speed to reach asymptotes of 3 seconds on the 25 words lists, and 5 seconds on the 50 words lists. These results could be seen as consistent with the rule suggested by Thurstone (1930) that input time per item increases as the square root of the length of the list.

In view of the equivocality of the results of Wallace et al, and of the wide currency among mnemonists (see Chapter 8, p.107) of the view that length of list makes little difference to performance, it was decided to compare the performance of mnemonic and control subjects on lists of two different lengths. This experiment was not intended to provide a direct test of the thesis that item-input-time is invariant with respect to list length, but rather to explore further some of the differences between mnemonic and non-mnemonic performance found in Experiments 1, 2 and 4, with particular reference to any interactions which might appear between these differences and the variable of list length.

Since it was intended also that error distributions, retention, and the relations between learning and retention should be examined more closely than previously, it was felt that the very considerable within groups variance shown in the previous studies should be reduced as much as possible. It was therefore decided to use male subjects only, and to make use of the normative practice list performance data obtained in Experiment 4 in order to match subjects into the different experimental groups as accurately as possible.

Subjects

The subjects were 72 male first-year psychology students aged between 18 and 22 years. All of these attended the initial session, but, because the length of the retention period was specified to the day in this study, whereas in the previous studies four days had been allowed for tracing and running defaulters, 9 of these were not included in the retention phase of the experiment, as they were not available on the required day.

Method

Four groups of subjects were used, each group being assigned to one of two instruction conditions and one of two lengths of experimental list. The four groups were mnemonic long-list (fB-L), mnemonic short-list (fB-S), control long-list (C-L), and control short-list (C-S).

The material and the basic plan of the sessions were very similar to those used in Experiment 4. Thus the same practice list was used, and the shorter of the experimental lists was the same low inter-item association list of 10 items as was used in that experiment. The longer experimental list was list L2 of 16 items used in Experiment 2, rearranged to minimise inter-item association (see Chapter 11, p157), and its first 10 items were the same as the shorter experimental list used in the present experiment. All of these word-lists are reproduced in Appendix 1.

In the first session all subjects (who were run individually) learned the practice list, which was presented by memory drum at a 5 second rate. The list was presented once, followed by five anticipation runs at the same rate. Inter-trial interval was approximately 10 seconds. On the basis of their error performance on this task, the subjects were assigned to one of six categories of learning ability, defined in terms of the ranges within which successive "sextiles" of the 290 subjects of Experiment 4 had fallen. They were then assigned randomly to one of the four experimental groups of the present experiment, with the restriction that the six learning ability categories had to be equally represented in each group.

All subjects then learned the experimental list appropriate to their particular group, the mnemonic and control subjects receiving instructions identical to those received by subjects in groups fB (the fully mnemonic group) and C (the control group) in Experiment 4. The list was presented once, followed by 3 anticipation trials, at a presentation rate of 6 seconds per item, with an inter-trial interval of approximately 10 seconds. The presentation rate was made faster than in Experiment 4 with a view to increasing the numbers of errors made by the mnemonic subjects to the point where an error analysis would become practicable. After the learning of the experimental list each subject was given an appointment to return after 6 weeks, and was cautioned against rehearsal.

In the second session subjects were given 4 anticipation trials on the experimental list under the same conditions as before. Mnemonic subjects were then asked to give samples where possible of the mediators they had used, and whether they had recalled their original mediators. All subjects were asked whether they had used any learning techniques other than those they had been instructed to use. As was previously indicated, not all subjects returned for the retention sessions, and the numbers in the groups were reduced from 18 to 14 (fB-L), 16 (C-L), and 15 (C-S). All 18 fB-S subjects returned.

Results

The mean error scores for both learning and retention phases of the experiment are displayed in Table 1., together with the appropriate standard deviations. The short list error scores were weighted to make it possible to compare them with those obtained on the long lists.

It can be observed from the table that whereas there were, in the learning phase of the task, gross differences in mean error score between the mnemonic and control subjects, and between the long and short list subjects, the latter differences are much less striking in the retention data. For both mnemonic and control subjects the short list groups show much less variability than the long list groups in the learning phase of the experiment, but this is reversed in the retention phase.

Tables 2. and 3. show the results of two-way analyses of variance on the learning and retention data. In the learning phase both list length and instructions show significant effects, but only instructions show a significant effect in the retention phase. In the retention data variance due to list length was in fact less than that due to error.

The suggestion that variation in list length affects mnemonic performance less than non-mnemonic performance implies an interaction between list

TABLE 1.

Error scores on learning and retention: Means and SDs.

	LEARNING			RETENTION		
	n	Mean	S.D.	n	Mean	S.D.
fB-L	18	9.8	7.28	14	15.3	6.82
fB-S	18	5.1	4.27	18	14.2	12.07
C-L	18	19.3	8.04	16	28.4	8.79
C-S	18	11.4	6.77	15	22.7	9.88

Note: (a) Learning means do not take into account the first presentation, when the subject was not required to anticipate.

(b) Short list error scores have been multiplied by 1.6 to adjust for the difference in opportunity to make errors.

TABLE 2.

Two-way analysis of variance of error scores on learning.

	SS	df	Variance estimate	F	p
Instructions	1062.6	1	1062.6	21.4	$<.01$
List length	759.2	1	759.2	15.2	$<.01$
Interaction	59.6	1	59.6	1.2	n.s.
Error	3385.3	68	49.8		
Total	5266.5	71			

TABLE 3.

Two-way analysis of variance of error scores on retention.

	SS	df	Variance estimate	F	p
Instructions	1691.5	1	1691.5	16.7	$<.01$
List length	80.2	1	80.2	.8	n.s.
Interaction	387.1	1	387.1	3.8	c.f. critical F of 4.0 for p =.05
Error	5973.5	59	101.2		
Total	8132.4	62			

length and instructions. There was no sign of such an interaction in the learning data, but the retention data do show some degree of interaction in the appropriate direction. It does not, however, quite reach significance at the .05 level.

The observed difference in retention between mnemonic and control subjects may not indicate that mnemonics improve retention per se, as it could be a direct consequence of improved learning. In an attempt to separate these two factors each subject's total of correct responses in the learning phase of the task was subtracted from his total of correct responses on the retention phase (short-list scores having been multiplied by 1.6 to make them comparable with the long-list scores). The resulting difference scores should represent performance on retention with the effects of individual differences in learning held constant. The means of the four groups on these derived scores were 10.5 (fB-L), 6.9 (fB-S), 6.9 (C-L), and 5.0 (C-S). The differences between these means were consistent in direction with those obtained on learning and retention, but were nowhere near large enough to be significant. In a two-way analysis of variance it was found that neither the interaction nor either of the main effects showed variance greater than that due to error. This data, then, is consistent with the view that mnemonic instructions do not affect the number of errors produced during relearning independently from the effect they have on learning.

It seemed worthwhile to look at the errors obtained on the first anticipation run for both learning and retention. In the case of the learning errors it could be argued that since mnemonics may be essentially one-trial (see Rock, 1957), anticipation trials subsequent to the first are likely to obscure differences due to mnemonics. In the case of retention the first anticipation trial can be seen as essentially different from the later ones, as correct responses on this trial represent recall, whereas correct responses on the later trials may represent relearning, or even new learning. It would be possible for relearning to be faster for mnemonic subjects than for controls even if recall was no better.

The results of analyses of variance on these two sets of first trial errors are shown in Tables 4. and 5.

TABLE 4.

Two-way analysis of variance of errors made on the first anticipation trial on learning.

	SS	df	Variance estimate	F	p
Instructions	223.3	1	223.3	24.70	$<.01$
List length	113.5	1	113.5	12.50	$<.01$
Interaction	6.2	1	6.2	.69	n.s.
Error	615.0	68	9.0		
Total	958.1	71			

TABLE 5.

Two-way analysis of variance of errors made on the first anticipation trial on retention.

	SS	df	Variance estimate	F	p
Instructions	123.9	1	123.9	7.04	$\simeq$.01
List length	0.0	1	0.0		
Interaction	30.6	1	30.6	1.7	n.s.
Error	1039.0	59	17.6		
Total	1193.5	62			

There are no differences in general results between these two analyses and those shown in Tables 2. and 3. However, the differences in relative size of the F ratio between Tables 2. and 4. are consistent with the view that mnemonic instruction has more effect on the first anticipation trial than on later trials. Similarly, in the two retention analyses the difference between the two F ratios for instructions is consistent with the view that mnemonic instructions have more effect on re-learning than on recall. Clearly this method of analysis can be at best no more than suggestive, but it is probably safer not to enter into too many extensive analyses in the absence of clear predictions as to the outcomes.

One of the findings of Experiment 2 was that mnemonic subjects make a higher proportion of their errors in the form of omissions than do control subjects. The results of Experiment 4 suggested that this difference is not explicable as an interaction between the faster learning rate on the part of mnemonic subjects and a tendency for all subjects to make a higher proportion of commission errors in earlier anticipation trials. An attempt was made to check this suggestion in relation to the present data. Mean error scores were computed for all groups, separately for total errors, errors made on the first anticipation trial, and errors made on later anticipation trials. In addition means were obtained separately for errors of omission and commission under each of three restrictions. For each group the mean commission errors under each restriction were expressed as a decimal fraction of the relevant mean total error score. The results are exhibited in Table 6. Table 7. represents the equivalent analysis for the retention data, with the addition of the percentage savings scores for the four groups. It can be observed that, for both tables, the sets of ratios of commission to total error scores are entirely consistent with the expectations that (a) the proportion of errors of commission would rise as the task proceeded and, (b) under all comparisons the proportion of errors of commission produced by the mnemonic subjects would be lower than that produced by the control subjects.

These proportional differences are not themselves suitable for further analysis because

TABLE 6.

Error means on learning for total, first, and later trials, and for O and C errors separately.

Note that short-list means have been multiplied by 1.6 to make them comparable with long-list means.

Error-Means		fB-L	fB-S	C-L	C-S
Total	(A)	9.8	5.2	19.3	11.0
1st trial	(B)	4.9	3.0	9.0	6.0
Later trial	(C)	4.9	2.1	10.3	5.0
Total O	(D)	6.8	2.8	9.9	3.6
Total C	(E)	3.0	2.3	9.4	5.3
1st trial O	(F)	4.1	1.7	5.9	3.1
1st trial C	(G)	.8	1.3	3.1	2.8
Later trial O	(H)	2.8	1.2	4.0	2.6
Later trial C	(I)	2.2	1.0	6.3	2.5
Proportional Differences (see p. 214)					
E/A		.30	.45	.49	.48
G/B		.17	.44	.34	.48
I/C		.44	.46	.62	.49

TABLE 7.

Error means on retention for total, first, and later trials, and for O and C separately.

Note that short-list means have been corrected to make them comparable with long-list means.

Error-Means		fB-L	fB-S	C-L	C-S
Total	(A)	15.3	14.2	28.4	22.7
1st trial	(B)	9.8	9.5	13.4	12.9
Later trial	(C)	5.5	4.7	14.9	9.8
Total O	(D)	8.8	8.9	12.1	10.6
Total C	(E)	5.7	5.2	16.2	12.5
1st trial O	(F)	6.8	7.1	8.1	7.6
1st trial C	(G)	3.1	2.4	5.3	5.3
Later trial O	(H)	2.0	1.8	3.9	2.9
Later trial C	(I)	2.6	2.8	10.9	7.2
Proportional Differences (see p. 214)					
E/A		.37	.36	.57	.55
G/B		.31	.25	.40	.41
I/C		.48	.58	.73	.73
Savings %		40.7	32.8	19.4	15.3

individual subjects make differential contributions to them depending on the total number of their errors.

The total consistency of the results is, however, moderately impressive. For the learning data an additional analysis was attempted, using difference scores generated by subtracting commission errors from omission errors for each subject, for total trials, first trial, and later trials. The difference scores of the mnemonic subjects were multiplied by the ratios of the control subjects' mean error scores to the mnemonic subjects' mean error scores. This procedure made it possible to compare separately the O and C error-scores for mnemonic and non-mnemonic subjects in spite of the differences in mean total error-score. Mann Whitney U tests were then carried out comparing mnemonic with control groups, separately for long and short list groups, for total trials, 1st trial, and later trials. All of the differences were in the predicted direction, but the only one which was significant at the .05 level was the difference for long list, total trials. For the long list the U values were 92 (total trials), 101 (1st trial), and 111 (later trials). For the short list the values were 129, 141, and 139 respectively. For samples of 18 $\mu_u = 162$ and the critical value of U for $p = .05$ (two-tailed) is 99.

The results of Experiment 1 suggested that mnemonic instructions affect the form of the serial position curve. The conditions of Experiment 1 were not as close to those of a "regular" serial learning experiment as were those of the present study, in that there was only one presentation of the list, and the rate of presentation was 20 seconds. It might be held that the flattened serial position curve produced by the mnemonic subjects in that study resulted from an interaction between the instructions and the unusual experimental conditions, rather than from the instructions per se. Figures 1(a) and 1(b) show the serial position error curves for the present data, for learning and retention phases of the experiment. In order to achieve strict comparability of the curves the error scores were converted using Jensen's (1962b) index of relative difficulty. Jensen's index involves the correction

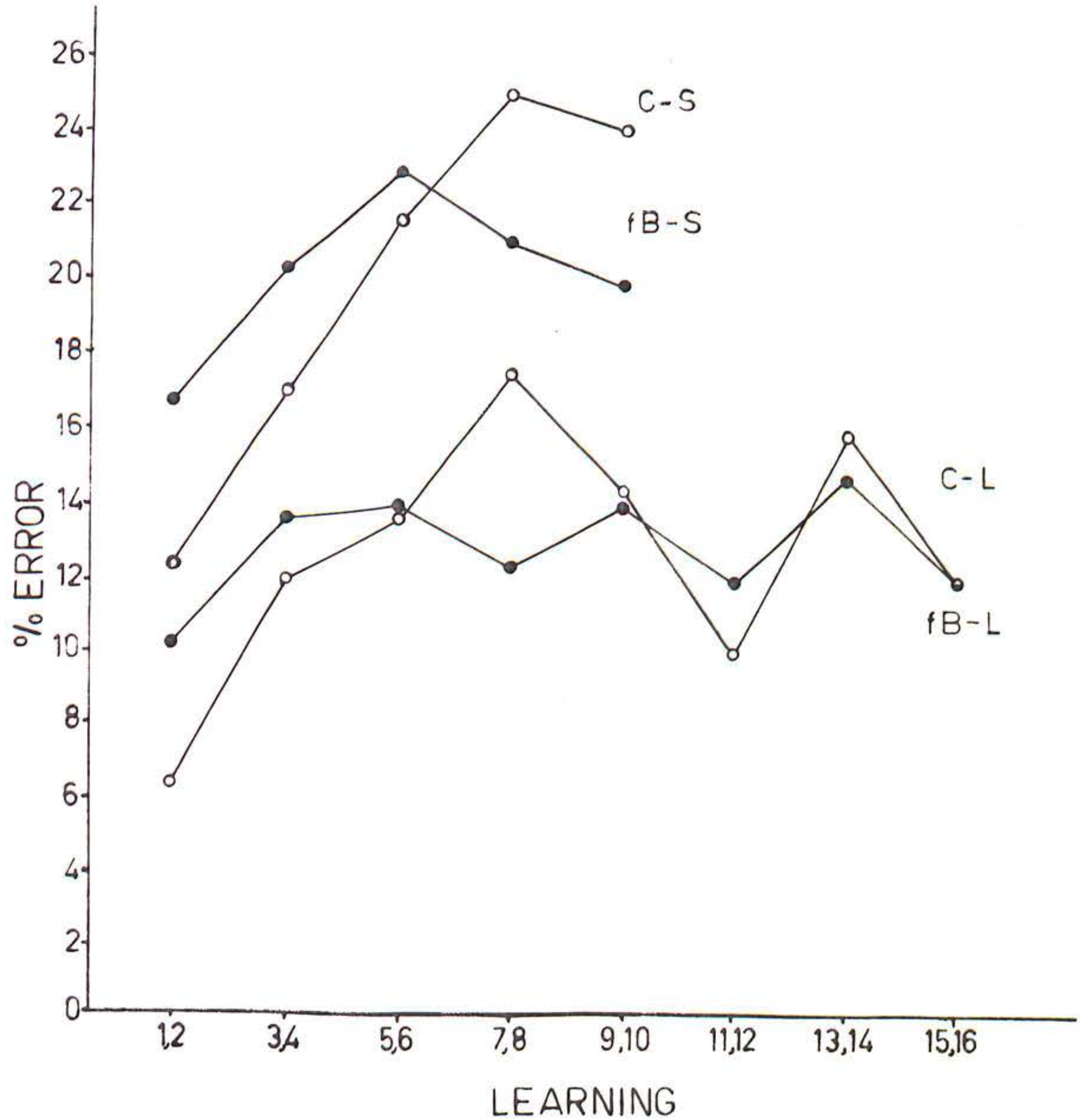

FIGURE 1(a). Jensen Position Error Index
(points represent indices for consecutive pairs of items)

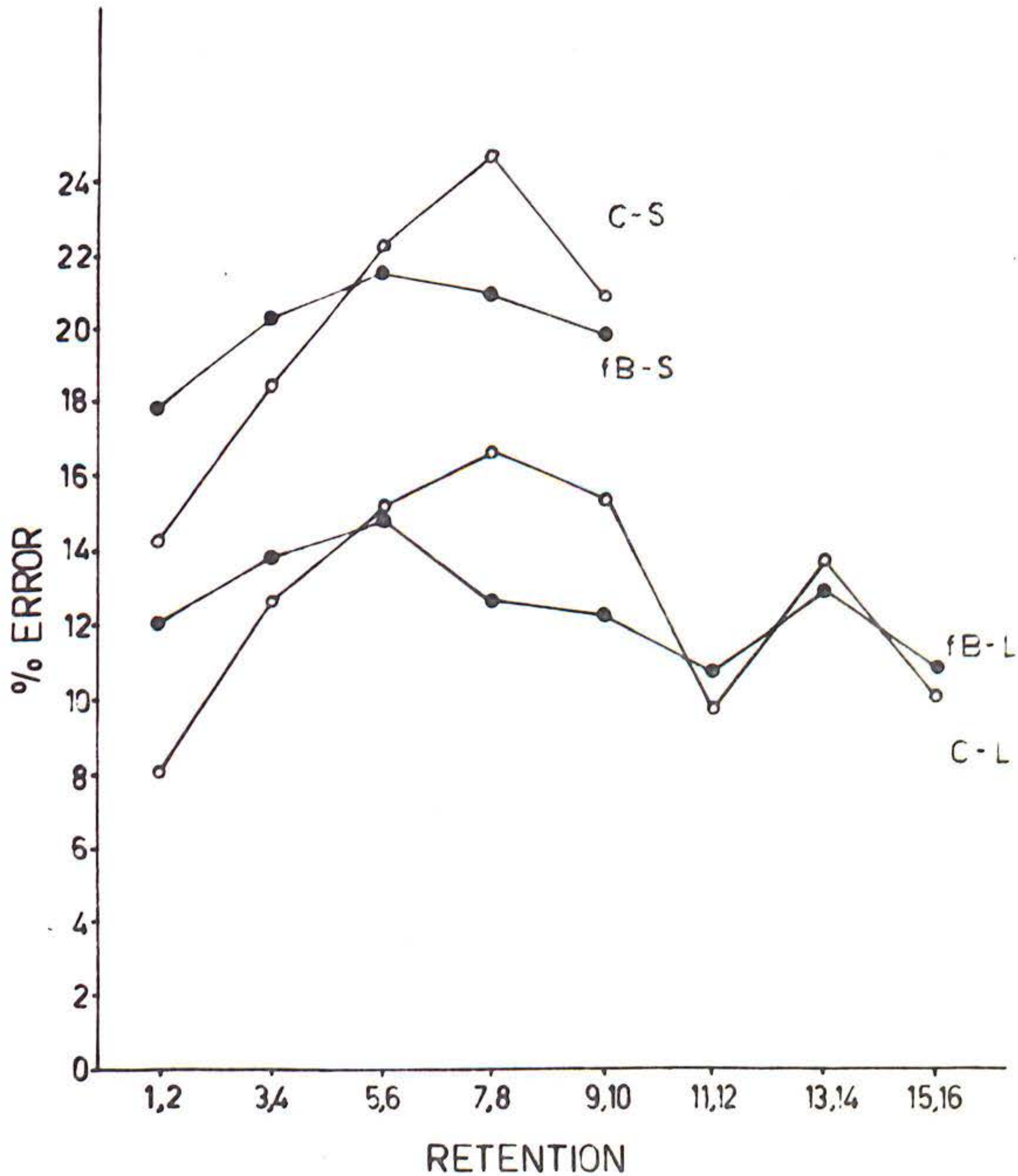

FIGURE 1 (b) Jensen Position Error Index
(points represent indices for consecutive pairs of items)

of the error-scores at each position in the list to adjust for between-list differences in total error rate, followed by conversion of the corrected error-scores to percentages of total errors. It has the advantages of rendering serial position curves comparable even when the data were collected under very different conditions, and of resulting in identical curves whether errors or correct responses are being considered. The figures show that for each appropriate pair comparison, the curve produced by the mnemonic subjects is flatter than that produced by the control subjects.

Discussion

The results of this experiment confirmed several of the findings of the previous experiments, in direction, if not always in degree. Thus both mnemonic groups showed lower error scores in learning and relearning than the equivalent control groups. Similarly, mnemonic subjects made a higher proportion of their errors in the form of omissions than did control subjects, although these differences did not reach levels of significance which would, in the absence of previous evidence on the matter, have been regarded as more than suggestive. The flattening of the serial position error curve, observed in Experiment 1, was also found in this experiment, in spite of considerable differences in the conditions of the two experiments.

The main findings of this study which had not been found in the previous studies were, firstly, that there was no sign of an interaction between list length and the effect of mnemonic instruction, secondly, that the effect of mnemonic instruction on retention is primarily a consequence of its effect on learning, and thirdly, that list length affects performance, for both mnemonic and control subjects, much less in retention than in learning.

The absence of an interaction between list length and instructions is not necessarily inconsistent with the hypothesis of relative invariance of mnemonic performance with list length, as put forward by professional mnemonists. This hypothesis was initially framed in terms of speed of learning, whereas our results are in terms of errors made over a fixed number of trials. The lists in this experiment were in fact short enough that a trained mnemonist would not have been expected to make any errors at all. On the other hand if we are to claim that the advantage in performances of the mnemonic subjects in these experiments is mediated by their implementation of the instructions, we should expect that qualitative aspects of their performance would have more in common with what would be expected of trained mnemonists than would equivalent aspects of the performance of control subjects. The absence of an effect of list length on retention is not too surprising given that most subjects had learned the list by the end of the learning session. The chief effect of list length that has been commented on in the literature (see McGeoch & Irion, 1952, p.487), has been in relation to speed of learning. There is no particular reason to expect that once the material has been learned this factor will have any great effect. In the absence of a significant effect of list length on retention the near-significant interaction between list length and instructions would be difficult to interpret and it is probably safer to assume that it was artifactual.

Stage performers and those anxious to promulgate the use of mnemonics tend to emphasize the fact that only one presentation is required for perfect recall performance when mnemonics are used. In view of this claim it is interesting to see that in the analysis of variance table for errors on the first trial of learning (Table 4.) the F ratio for Instructions is greater than in the equivalent table for total errors, which is what one would expect if the mnemonics had their greatest effect on the first trial. The equivalent ratios for the retention data show the opposite trend, suggesting that mnemonics do not improve recall as much as re-learning. This result should, however, be treated cautiously. It could easily be a function of the retention interval used. Many of the subjects reported themselves as having recalled their images once they had seen the items again (i.e. as a result of the first anticipation trial on retention). In other words, for these subjects relearning may have consisted simply in being reminded of their images. A comparatively small lessening of the retention interval might, if this description is accurate, have resulted in no reminder being needed. By the same token, if retention had been tested by recognition, the mnemonic subjects might have recalled their images as soon as they saw the items they had received previously.

The analysis of errors of omission and commission gave results broadly consistent with the previous finding that mnemonically instructed subjects make proportionately fewer errors of commission. The differences were more evident in the long-list data. This could be seen as suggesting that long-list subjects were more successful in applying the instructions. Several short-list subjects remarked that they had not gone on trying to use images after the first anticipation trial during the learning phase, as they had at that point only a few more items to learn. Whether this is a correct interpretation or not it is clear that length of list may affect both learning and retention in rather complex ways. Thus, in spite of the fact that the performance of short-list subjects was less variable than that of long-list subjects on learning, it was more variable, for both mnemonic and control subjects, on retention. It is interesting, in this connection, to note that savings scores (see Table 7.) were higher, for both mnemonic and control subjects, on long-list than on the short-list.

In relation to the serial position data, while the mnemonic subjects produced flatter curves than did the control subjects, it is worth noting that the curves produced by the control subjects were by no means regular. While they did show primacy and recency effects they also showed obvious and convincing dips at points other than the beginning and end. It could be argued that the short-list curves were not different from the equivalent section of the long-list curves and that the data were consistent with an interpretation in terms of specific item difficulty rather than serial position difficulty. Such an interpretation would be consistent with the suggestion that with the type of material used in these experiments most subjects make considerable efforts to construct mediators even if not instructed to do so. In the present experiment more than half the control subjects reported that they had made use of mediators. Four of them described themselves as having used techniques similar to the one the mnemonic subjects were instructed to use.

As with the previous experiments, then, it is probably safer to regard the experimental and control groups as having been samples richer and poorer in mnemonic performance, rather than to see them as having been composed of mnemonic and non-mnemonic subjects respectively.

CHAPTER 14.

Experiment 6. The effects of high and low meaningfulness and inter-item association upon the learning performance of subjects without mnemonic instruction.

Introduction

It was suggested in Chapter 10 (p.152) that one might use the insights about learning obtained from research on mnemonics to make predictions as to what would happen in experiments not necessarily involving mnemonic instruction. In Experiment 3 it was found that the ability of subjects to carry out mnemonic instructions depended at least in part upon the nature of the material, and that inter-item association as rated appeared to be a variable relevant to this dependency. It was suggested (p.166) that if there is any similarity between situations in which a subject carries out mnemonic instructions and those in which he uses mediating techniques he selects for himself, the nature of the material to be learned should to some degree determine which techniques will be applied, and how well they will be applied.

In Experiment 3 pairs of items were used for which ratings for inter-item association strength had been obtained. Deese (1959) showed that, in free recall, a measure of inter-item association in terms of the degree of associative overlap of the list items was strongly related to list difficulty. This variable has been further explored by Deese and others (e.g. Deese, 1962; Garskof and Houston, 1963), but has usually been measured in terms of the extent to which pairs of items share associative hierarchies (Deese, 1959, 1962), or of the frequency with which they elicit each other as free associations (Amster, 1967; Garskof et al, 1967). However, the use by subjects of natural language mediators (Montague and Wearing, 1967; Martin and Dean, 1966) makes it likely that factors other than associative overlap may be relevant in determining the "degree of relatedness" of a pair of items. Thus pairs like "whimper-chaos" may be much more highly related (in terms of the ease with which they can be associated with one another) than their degree of associative overlap would suggest.

Prior to Experiment 3 a number of judges rated pairs of items from 1-5 for degree of inter-item association. Their subjective reports on the rating process were consistent with the view that rated inter-item association value is almost directly a measure of the ease with which natural language mediators can be constructed. Wearing and Montague (1967) found that two paired-associates lists which had been equated on most dimensions of meaningfulness were still not equally easy to learn, and that the difference appeared to be related to a difference in inter-item association, as estimated by the numbers of natural language mediators produced by the subjects. Their concept of inter-item association appears to be similar to that used by the raters of our word lists.

The use of ratings in this context confers the advantage that the researcher can make use of a wider range of materials in his experiments. He is not restricted to the materials for which published norms are available, since he can always have material rated as he needs it. This in turn means that factors such as meaningfulness can be held constant by using the same items in more than one list, but arranging them in different orders so as to achieve different average levels of inter-item association.

Experiment 6 was intended to demonstrate, at two different levels of meaningfulness, that rated inter-item association affects list difficulty. Subjective reports were also collected, with a view to deciding whether any of the performance variation could be ascribed to differences in the learning techniques used by the subjects. In the analysis an attempt was made to examine types of error in such a way as to relate the results of this experiment to those obtained in the experiments using mnemonic instructions.

Materials

Two sets of 12 words were selected from the list published by Saltz (1967) in such a way that one set had high meaningfulness (mean m = 8.6) and the other had the lowest meaningfulness (mean m = 6.8) that could be achieved consistent with all words being nouns. In addition to having higher meaningfulness all words in the high m set were at least ten times more frequent than those in the low m set.

The words in each set were next arranged in two different orders with the intention first of maximising and then of minimising the degree of association (a.) between adjacent items. This was carried out by the experimenter on an intuitive basis. It was hoped that the next step in the preparation would validate this procedure.

The four word lists were broken down into their constituent word pairs (making 44 pairs), and these were presented in different orders to 20 judges. The judges were instructed to indicate by placing a mark on an un-numbered scale running from "Very weak association" to "Very strong association" how strongly each pair of words seemed to them to be associated. These rating marks were then converted by measurement to scores ranging from 0 to 10. The split half reliability of the ratings was computed as rho = .89. Upon reassembly of the two lists the two high-m lists were found to have mean a. indices of 2.64 and 7.06, and the low-m lists gave means of 1.65 and 5.96, there being practically no overlap in a. values between high and low a. lists. These findings were seen as justifying the subjectivity of the initial ordering procedure. The four lists are reproduced in Appendix 1, together with the Noble m values of the words, and the a. values whose computation is described above.

Method

Each of the four lists (High m, High a.(HH); High m, Low a. (HL); Low m, High a. (LH); Low m, Low a. (LL), was learned by 16 first-year psychology students using the anticipation method. Presentation was by memory drum at a 4 second rate, and learning was carried to a criterion of one error-free anticipation of the list. Inter-trial interval was 10 seconds. Each subject initially learned a practice list consisting of 7 two-digit numbers, and the error scores in this task were used to achieve a rough matching of the subjects assigned to learn each experimental list. Four levels of learning ability were used in this matching procedure.

When the learning tasks had been completed all subjects were asked whether they had noticed anything particular about the experimental list, and also whether they had been aware of employing any particular technique in learning the list.

TABLE 1.

Means, medians, and SDs for untransformed errors and trials data.

	HH		HL		LH		LL	
	Errors	Trials	Errors	Trials	Errors	Trials	Errors	Trials
Mean	4.1	2.8	24.3	6.2	8.5	4.6	25.5	7.0
S.D.	4.56	1.58	25.62	4.16	6.56	2.93	16.80	3.00
Median	2.5	2.5	13.5	5.0	7.0	4.0	17.0	6.5

Results

The main results are set out in Table 1. Because of the high degree of skew in the data analyses of variance were carried out on logarithmic transformations of both errors to criterion and trials to criterion.

In the analysis of variance of errors both main variables and the interaction between them proved significant. Inter-item association showed the most striking effect ($F = 42.20$ $df = 1/60$ $p < .001$), but meaningfulness was significant at the .025 level ($F = 5.68$ $df = 1/60$) and interaction at the .05 level ($F = 4.59$ $df = 1/60$). HH was the easiest list, followed by LH, HL, and finally LL. The differences HH/LH and LH/HL were significant at the .02 level (two-tailed), using t tests on the transformed scores ($t \neq 2.65$ and $t = 2.66$), but the HL/LL difference was insignificant ($t = .97$).

In a similar analysis of trials to criterion both main effects were significant (meaningfulness $F = 4.60$ $p < .05$, inter-item association $F = 27.83$ $p < .001$), but there was no interaction. The ordering of the lists was the same, but the only difference between consecutive groups significant according to the t test was that between LH and HL ($t = 2.16$ $p < .05$). The other differences would have been significant if one-tailed tests had been used.

The answers to the questions asked after the learning session showed some striking inter-group differences. Thus all but three HH subjects noticed that all items were interconnected, and most of them described themselves as having based their learning on this. LH subjects tended to notice particular pairs as connected, and to supplement these connections by the use of mixed strategies involving imagery and similarities in word shape, spelling and meaning. Neither HL nor LL subjects showed much awareness of the experimental arrangement of the lists. Rote learning (without much reference to meaning) was described by about half the members of both groups as the main technique used, but subjects in LL showed greater awareness of remote associations, and several appeared to have made some use of these, while several of the HL subjects tried to make use of images. For both of these groups the mean error rate was higher for subjects who described themselves as having learned by rote, and, in the case of group LL, this difference was significant at the .05 level (two-tailed) on a Mann-Whitney U test. Attempts to relate the meanings of adjacent words were more common in HL, whereas LL subjects used a variety of techniques related to form rather than to content. Thus subjects noted that 'Shilling' and 'Malice' both had 'l' in the middle, that the initial letters of the sequence 'Malice-Outfit-Poison' spelled MOP, and that 'Poison' and 'Bonus' had approximately the same shape in that the only tall letter they contained was the initial one.

There was some suggestion in the data that the differences in list structure led to differences in both type and distribution of errors. Subjects with highly meaningful lists made many more of their errors in the form of intralist intrusions than of omissions, whereas the reverse was true of subjects learning the low meaningfulness lists. Thus the HL subjects made in absolute terms more intrusion errors than did the LL subjects although the difference in mean total errors was in the opposite direction.

There was a marked difference between high and low inter-item association groups in the proportion of all item errors which occurred after one correct anticipation of the item in question had been made. Thus, expressing these errors as fractions of the total errors for each group HH made 17/66 errors of this type while LH made 12/136. The equivalent figures for the other two groups were 113/389 for HL and 119/408 for LL. This difference might be accounted for in terms of a difference between rote and non-rote learning such that once a non-rote connection has been made there is little likelihood of further error.

Given the differences in total performance between the groups, and the high within-groups variability in this study it would be difficult to assess the significance of these post hoc observations, but they certainly look as if they would warrant more direct examination in experiments designed for the purpose.

Discussion

The results of the analyses of variance show that both meaningfulness and inter-item association (in terms of judges ratings of the degree of relatedness of word pairs) affect the level of difficulty of serial lists of English nouns. Thus the degree of ease with which a list is learned may be as much affected by the orders of the items in a list as by considerations of what the items are.

The fact that the same list of words may be more or less difficult to learn if arranged in a different order has implications for theories about the nature of the functional stimulus. Thus it would be difficult to account for this phenomenon if it was assumed that associative links between items in the list played no part in serial learning.

The less formal findings of this study lend some support to the view that the structure of the list has some determining effect both on the strategy that a subject will adopt and on the type and distribution of the errors he makes. The high variability of the subjects, in both strategy and performance, make it impossible to discover much about the sort of relations which might exist between the sorts of strategy which subjects adopt of their own accord and those which were imposed on them by instruction in the earlier experiments in this series.

The finding that the manipulation of inter-item association leads to differences in performance does not entail an explanation in terms of the use by subjects of different mediation techniques. An associative frequency theory could cope with our data. Even if our observations of differences in error type were accepted as reliable it would be possible to insist that these were a by-product of the differences in performance, for although in the previous experiments of this series such differences have been associated with differences in explicit mediation they have also been associated with higher performance. On the other hand, the correlation of these observations with differences in the subjective reports of the subjects cannot help but suggest an explanation in terms of mediational differences, even though such reports must be treated as being individually unreliable.

It would seem perverse to deny all validity to such reports, given that in the past they have been associated with both qualitative and quantitative differences in performance, and that there is some reason (c.f. Experiment 3) to regard them as other than epiphenomenal.

Although neither the quantity nor the quality of the data really warrant such speculation, it may be worthwhile to set out a tentative explanation in mediational terms of the results of this experiment. The primary object of doing this will be to give examples of the concepts and terms which a theory consistent with the data might use.

The HH subjects learned a list in which the items were highly available. In addition each item had fairly strong pre-established connections with its neighbours. For most of these subjects, then, the learning task consisted in noting what the items were, and rehearsing the meaningful connections between them. Little in the way of supplementation of the pre-existing connections seemed necessary, since the recall phase of the task could now be carried out by calling for the most recently rehearsed of the readily available associations for each item as that appeared in the memory drum window.

The HL list consisted of the same items as the HH list, so item availability was no problem, but the absence of pre-established connections posed some problems. Some subjects tried to learn by rote, simply trying to make the right noise at the right point in the list, relying on the ready availability of the appropriate noises. Others tried to construct associations between adjacent items which would give some assistance, when re-evoked, in determining their order. Such devices as constructing sentences, noting structural similarities in adjacent words, and incorporating adjacent items in compound images were used.

List LH consisted of rather less available items, but there were again pre-existing associations present. Some subjects tried simply rehearsing these associations in the hope that the re-presentation of the items would re-evoke the associations, and, through them, the adjacent items. This was not always successful, because of the low availability of the items themselves as responses. Many of the errors made by this group took the form of intruding synonyms or near-synonyms. As a result several subjects "went back to learning by rote". Others supplemented their associations by manoeuvres designed to increase item availability, for instance noting spelling similarities, the shapes of the words, etc.

LL subjects were confronted with a fairly difficult task, as they had neither pre-existing connections nor readily available items. If they were going to use mediational devices these would have to be very effective ones. And the presentation time used in the experiment did not allow for the construction of very elaborate mediators. Seven of them did not try, they simply settled down to learn the list by rote. Of the remainder, some supplemented rote learning by noting formal similarities between words in relation to the relative positions of those words in the list. Others, although they had discounted the possibility of establishing meaningful connections between adjacent words, noted that, for instance, 'chaos' and 'turmoil', two easily associated words, were three words apart in the list, an observation which facilitated the learning of 'turmoil', as well as providing clear location-cues for the intervening two words.

Explanations of the kind illustrated are not necessarily inconsistent with the more conventional sorts of explanations of between groups differences in verbal learning tasks, but they do differ importantly in assigning a causal role to mental events. A preparedness to do this would undoubtedly have an initially deleterious effect on the precision of the language used by verbal learning theorists, but it is arguable that it would also increase the area of verbal learning available for study. The increasing work on natural language mediation, and recent interest in such topics as the tip-of-the-tongue phenomenon (Freedman and Landauer, 1966) suggest that it is beginning to be regarded as more acceptable to take seriously what the subject in an experiment says about what is going on in his mind.

CHAPTER 15.

General results and discussion.

In this chapter an attempt will be made to collate the results of the entire series of experiments in such a way as to make evident any general principles which emerged.

Mnemonic instruction and improvement in performance

Experiments 1, 2, 4 and 5 used experimental groups with mnemonic instructions and control groups without such instructions. In each case the mnemonically instructed subjects performed better than the uninstructed on all performance measures considered. These included errors on the first anticipation trial (Experiments 1, 4 and 5), trials to criterion (Experiment 2), and total errors in a fixed number of anticipation trials (Experiments 4 and 5). The difference also occurred when the subjects were required to learn a second list (Experiment 2).

Similar differences in retention in favour of the mnemonically instructed subjects occurred in experiments 4 and 5, although the results here were not quite as clear. In the former experiment the complexity of the design together with significant interactions between the main variables suggest some caution in interpretation. In Experiment 5 the difference between mnemonic and non-mnemonic subjects was, if anything, clearer on retention than on learning, but when the number of correct responses in the learning phase of the task was subtracted, for each subject, from the number of correct responses in the retention phase, the difference between mnemonic and non-mnemonic performance virtually disappeared. This last result is consistent with that of Olton (1966), who found that when he attempted to equalise learning for mnemonic and control subjects by giving the mnemonic subjects fewer presentations of the material, differences in performance after a retention interval disappeared.

The implementation of mnemonic instructions as an explanation for improved performance

In Chapter 3 (pp. 40-44) a number of hypotheses were briefly considered as alternatives to the suggestion that implementation of mnemonic instructions is responsible for improvement in learning. The hypotheses considered involved the production, by the giving of the instructions, of differences in time taken per item (when self-paced), in motivation, in expectation, or in attention. All of these factors are ones which might reasonably be expected to have some effect upon learning, and one course which might have been suggested for a research project on mnemonics would have been to proceed systematically to exclude them.

This course was not followed, for a number of reasons. The first reason was that, with the exception of time-per-item, these are all factors which it would be extremely difficult to control, or to measure. Secondly, the pursuit of this course would have set very narrow limits on the theoretical ground that the research could cover in the time available. Finally, the results of the pilot work suggested that implementation of the instructions was a causal factor in the improvement of performance, irrespective of any effect that these other variables might have. The data do, however, suggest some comments about the effect of these factors.

If one takes the simplest interpretation of the time-per-item hypothesis, treating average total presentation time as the measure of time-per-item, the mnemonically instructed subjects in Experiments 1, 4 and 5 used the same amount of time as the controls, while in Experiment 2 they used less time. If it is interpreted as average presentation time-per-item up to the first correct anticipation mnemonic subjects in Experiments 2, 4 and 5 took less time than the controls.

Observation of the subjects (see p.134 above) suggested that the time they spend looking at an item may not always be determined by the time for which it is displayed. This suggests that the time spent actually looking at the items could be taken as another interpretation of time-per-item. Our results do not permit comment on this interpretation. On the other hand previous researchers, such as Bugelski (1962), who have dealt with time-per-item, would be similarly unable to comment on this interpretation.

The most convincing evidence against motivation as the critical factor in mnemonic superiority is contained in the result of Experiment 3. In this experiment subjects who reported "good" images recalled the material better than those who reported less "good" images, in spite of having been unaware when they constructed the images that they were going to be required to recall anything. An attempt was made in Experiment 1 to equalise motivation by using a financial reward. This was sufficiently large to ensure that both mnemonic and control subjects would be reasonably highly motivated. On the other hand, throughout the series of experiments, most subjects showed symptoms of being very highly motivated, sweating, giving signs of acute distress when they made errors, and showing a strong desire to know how their performance compared with that of other subjects.

In Experiments 1 and 2 the instructions were presented in such a way as to suggest to the mnemonic subjects that their implementation would improve performance. In Experiments 4 and 5, however, subjects who received mnemonic instructions were told that these would not necessarily make it easier to learn the material; that different groups of subjects were receiving different sets of instructions, some of which might improve performance, and some of which might interfere with it. They were given to understand that the experimenter did not have any clear expectation as to what would be the effect of any particular set of instructions. It cannot be shown that this information was accepted by the subjects, but the questions subsequently asked by many of them suggested that it was. In Experiment 3 the subjects were not initially expecting a recall task, and it is therefore unlikely that they would have formed any expectancies as to the success of the instructions in mediating recall.

No attempt was made to control attention, but it would be expected that this factor would correlate highly with motivation. In any case aspects of the information obtained from the experiments are more consistent with the view that the carrying out of the instructions mediated improved performance.

Firstly, probability of correct recall in Experiment 3 was related to rateable characteristics of the images produced. Secondly, in Experiments 1, 2, 4 and 5, mnemonic subjects were less variable in performance than were control subjects. If mnemonic instructions only produced effects upon performance through their effects upon such factors as motivation, expectations, and attention, one would expect, if anything, greater variability among subjects receiving these instructions than among the controls. Thirdly, in Experiment 4, the degree of difference in performance between the control subjects and those receiving mnemonic instructions was related to the degree of completeness of the mnemonic instructions. Fourthly, mnemonic instructions are associated with qualitative changes in performance, flattening the serial position curve (Experiments 1 and 5), reducing the proportion of errors which are intrusions (Experiments 2, 4 and 5), and increasing response latency (Experiment 2). These changes are not easily explained in terms of such factors as motivation, expectation, and attention. Fifthly, in Experiment 6, where no instruction was given, performance within groups appeared to be related to awareness on the part of the subjects that they had been using some mnemonic technique. Finally, the subjective reports of the subjects throughout the series of experiments were consistent with the hypothesis that the carrying out of mnemonic instructions leads to improved performance.

While none of these arguments is, individually, conclusive, their total effect must be to suggest that at least some, if not most, of the effect of mnemonic instructions on performance should be attributed to the carrying out of those instructions.

Discussion of the ways in which the implementation of mnemonic instructions might facilitate learning

It might be suggested that mnemonics improve performance by forcing the subjects to do more work in learning the material than they would otherwise have done. It is, however, questionable whether mnemonic subjects do work harder than rote subjects, since the latter subjects may expend much effort in activities like rehearsal. Moreover, if the work involved in creating a mnemonic was the crucial factor in mediating the improved learning performance it should not be possible for a subject to make use of a mnemonic generated by someone else. Our pilot results (see pp. 115-116) were consisted with the work of Olton (1966) in suggesting that the subject can in fact make use of a pre-constructed mediator. Olton's experimental subjects, who were provided with pre-constructed verbal mediators, learned much faster than his controls. It could be that the mnemonic subjects are receiving more reward than the control subjects, in that the creation and contemplation of vivid or bizarre images may be highly rewarding. Many of the mnemonic subjects did make in the experimental situation remarks like, "This is fun", and expressions of amusement were quite common. On the other hand, many of the images described, even by apparently highly successful mnemonic subjects, were not obviously amusing, or even very good according to the standards implied by the instructions (although it is worth noting that such descriptions would hardly be expected to be complete). Thus one subject described her image connecting CAR and PENCIL as follows, "I imagined a pencil lying on the seat of a car". It would clearly have been worthwhile to have asked subjects to rate their images for entertainment value, so that this variable could have been correlated with performance. It might also have been worthwhile to have asked them to rate their degree of enjoyment of the experiment as a whole, since reinforcement which is not item-specific may nevertheless affect learning. The hypothesis of differential reinforcement of mnemonic and control subjects can certainly not be dismissed on the basis of our data, but it is difficult to see how it could account for the observed qualitative differences between mnemonic and non-mnemonic performance

The hypotheses we have been considering share the characteristic of assuming that the mechanisms involved in learning with mnemonics are the same as those involved in learning without mnemonics. There are two ways in which this assumption could be wrong. Firstly, it could be the case that learning involving mnemonics is different in kind from non-mnemonic learning. The subject implementing mnemonic instructions could be operating in a different mode from the subject not implementing such instructions. Secondly, it could be that there is more than one kind of verbal learning, and that one effect of the implementation of mnemonic instructions is to change the relative preponderance of one or more of these types of learning in the performance of the subjects. Without going to the trouble of inventing new categories of learning there are several dichotomies which might be considered as possibly representing dimensions along which an attempt to operate mnemonically might shift the performance of the subject.

A dichotomy which might well be considered is that between learning involving position cues and learning involving associative cues. It was suggested in Chapter 5 of this thesis that subjects may use the previous item as their functional stimulus in some circumstances, and not in others. Our data do in fact give some grounds for suggesting that when a subject carries out mnemonic instructions he does make more use of the previous item as a cue than do most uninstructed subjects.

Firstly, one difference between a paired-associate task and a serial learning one is that in the former type of task there is little doubt that the functional stimulus is the stimulus item, or some part of it. And there is little evidence of a serial position effect in a constant-order paired-associate task (Battig, Brown, and Nelson, 1963). If part of the effect of carrying out mnemonic instructions was to cause the subject to use the previous item as the functional stimulus this would have the effect of turning the serial list into one in many ways resembling a constant-order paired-associate one. The flattening of the serial position curve observed in Experiments 1 and 5 would be a corollary of this change. Secondly, the verbal reports given by the subjects in these experiments are consistent with this hypothesis. Remarks like, "TABLE should come about now", were made by many control subjects, but were not made by mnemonic subjects in any of the experiments. Similarly, when asked how they set about trying to learn the list, many of the control subjects said that they were just trying to say the right word at the right point in the list. Thirdly, the mnemonic instructions specifically suggested that the subjects should try to connect each item with the one which immediately preceded it.

Experiment 2, perhaps unfortunately, took place shortly after the first-year students, as part of their course-work, had received two lectures on verbal learning. During these lectures the issue of the nature of the functional stimulus in serial learning had been raised. Many of the control subjects in this experiment said, in discussion after the experimental session, and clearly with their lectures in mind, that they had been trying to connect each item with its position in the list. Even if the subjects' reports can be relied upon it is possible that the lectures influenced their learning behaviour, but again it is noteworthy that none of the mnemonically instructed subjects gave such reports.

A second dichotomy which might be considered is that between rote and mediated learning, although this is not such a clear distinction as the previous one. This distinction has been made by some researchers, but has seldom been examined closely in the context of theory construction. Thus Adams and McIntyre (1967) pointed out differences between the performance of their rote subjects and that of their mediating subjects, but, although they seemed to feel that there were important theoretical differences between the two types of learning, they did not attempt to analyse what might lie behind the relatively crude behavioral criteria they were using; namely, whether the subjects did or did not report natural language mediators.

One reason for the avoidance of the rote-mediated distinction in theoretical analyses may be that it represents the confounding of at least two dimensions, and possibly three. One of these dimensions is that of directness. If a subject connects the response with the functional stimulus through a long chain of associative links, we may describe his learning as mediated; whereas if he responds quite directly to the functional stimulus, as if the response had been conditioned to the stimulus, we are likely to describe him as having learned by rote. This distinction refers only to the length of the chain of associations which leads from stimulus to response, and does not entail a difference in the type of associative bond involved.

Secondly, the subject may or may not make use of the meaning of the items in his learning of them. If he appears simply to be trying to produce the right noise as a response, and he makes errors which take the form of more or less accurate approximations to the shape or sound of the correct response, he may be described, and may describe himself, as learning by rote. Some subjects who thought of themselves as having learned in this way went so far as to say that it would not have mattered what the items had been; that they had not taken the meanings of the items into account at all. It has long been recognised (see Pillsbury, 1897) that subjects in short-term memory experiments may respond on the basis of the sound of a word rather than its meaning, and that this may even be the case when the material is visually presented. Recent work by Baddeley (e.g. Baddeley, 1968) suggests that acoustic coding is much more general in short-term memory than is semantic coding. On the other hand Baddeley and Dale (1966) found that semantic coding was more general in long-term memory. The results in this area suggest that there is a tendency, as presentation time is lengthened, for visual coding of visually presented material to give way to acoustic coding, which in turn gives way to semantic coding. It would not, however, be surprising if such factors as the levels of inter-item association within the list, the familiarity of the material, and the set of the subjects also acted to determine what type of coding was used. Some of the subjects in the present series of experiments described themselves as using mediators which exploited similarities between adjacent words in spelling, visual shape, or sound. These subjects did not appear to see themselves as operating in a rote fashion, as did those who described themselves as just trying to make the right noises, and it may be that there is a distinction to be made between the use of visual or acoustic mediators and the use of visual or acoustic coding of the items. It seems clear that the subject who gives the response "spectacle" in place of "spectacles" has been operating in a different way from the one who gives the response "glasses", but one cannot tell whether the response "spectacle" arises from decay of an acoustic mediator, or from a simple confusion of the sounds or shapes of the words "spectacles" and "spectacle".

A third distinction which appears to arise in connection with the rote-mediated dichotomy is that between the subject who tries consciously to connect the response to the stimulus by making use of some mnemonic or mediating technique as the one who does not do so. The latter subject will often be described, or may describe himself, as learning by rote.

Although there may be a tendency for particular positions on these three dimensions to go together, they are not perfectly related. Thus "chair" may be connected to "table" through the association that already exists between these two words, without the subject deliberately making use of this association. This is one reason why the subject who says, "It would not have mattered what the items were", is almost certainly mistaken. That he can make this error is, however, interesting in itself. Again, a subject may deliberately associate "actor" with "enter", without taking their meanings into account at all. Tachistoscopic studies suggest that words can be connected in this sort of way without the subjects being aware that an association has been made, far less making it deliberately.

Let us consider our data in relation to each of these distinctions. There do appear to be grounds for suggesting that mnemonic subjects tended to use relatively indirect methods of associating adjacent pairs of items. In the first place the instructions, which many of them thought they had to some extent succeeded in carrying out, suggested that they should do this. Secondly, many of the images described by them contained quite complex linking sequences and structures. Thirdly, the relatively long response latencies of the mnemonic subjects, observed in Experiment 2, are consistent with the subjects having to explore a complex linking structure before being able to produce a response. The decrease in latency on the later anticipation trials is consistent with the results of Adams and McIntyre (1967), and with the suggestion that with multiple recall the mnemonic is dropped, and the connection between stimulus and response becomes more direct. This suggestion was put forward by Reid, as early as 1918.

The grounds for suggesting that mnemonic subjects made more use of the meaning of the items are not quite as strong, since although they made more intrusion errors in which the intrusion was meaningfully related to the response neither mnemonic nor control subjects produced very many extra-list intrusions. Also, some of the extra-list intrusions produced by the mnemonic subjects were unrelated to the correct response. These appeared, however, upon later questioning of the subjects, to have been extraneous objects which had been introduced into the images as they were created. The control subjects, on the other hand, often said quite specifically that they had made no reference to the meaning of the items, and, when they had made use of any mediation technique, had tended to make use of the form of the items rather than their content.

Finally, there can be little doubt that the mnemonically instructed subjects showed a greater tendency than the control subjects towards deliberately making use of a learning technique. In the first place they were instructed to do so, and most of them appeared to have tried. A few mnemonic subjects did, however, say things like, "I tried to carry out the instructions, but the words went by too fast, so I gave up and went back to learning them off by heart". In the second place many of the control subjects made it clear that they had made no attempt to apply any technique to learning the list, some of them even appearing not to understand what the question meant. Even where techniques were described the subjects often made it clear that they had only used them in order to connect particular pairs of items, the rest having been learned by "rote".

This phase of the discussion has clearly relied more heavily than is desirable upon the subjective reports of the subjects, and it could be that one or more of the distinctions that have been made will turn out upon further research to be associated with no clear performance criteria. On the other hand the search for these criteria would seem to constitute a reasonably hopeful programme for future research.

So far it has been suggested that the mnemonic subject takes the previous item as his functional stimulus, and deliberately makes a relatively indirect association between stimulus and response, with reference to the meanings of the two items. But the instructions which he has been given suggest more than this. He is instructed to use imagery, and to make his images vivid, active, bizarre, and involving as many sensory modalities as possible. Moreover, the results of Experiment 4 suggest that these additional instructions do affect his performance. No reason has yet been given for their doing so.

Montague and Wearing (1967) found some support for the suggestion made by Martin, Boersma and Cox (1965) that the effectiveness of natural language mediators was positively related to their complexity, and it is possible to see several of the elements in our instructions as being conducive to complexity in the resulting mediator. Such complexity might facilitate learning through providing a system of multiple links between the stimulus and response terms such that, even if part of the mediator was forgotten, either it could be re-created, or enough might remain to make it possible to find the response term. Experimentation to confirm this might make use of carefully designed pre-constructed mediators.

Paivio and his associates (e.g. Paivio, Yuille and Smythe, 1966) have consistently found that imageability, particularly in the stimulus term, is related to the ease with which the material is learned. Wood (1966) did not find that instructions which involved imagery were any more effective than those which did not, but, as remarked earlier (see pp. 27,165) there is some doubt that Wood's subjects carried out his instructions. In our Experiment 4 (see Fig.1., Chapter 12) the addition of an instruction to use imagery (element (b)) produced more improvement in performance than did the addition of any other element of the instructions.

None of our data are relevant to the question of why the use of imagery should improve performance. One possibility is that, since a picture can convey information very much more economically than can words, an image involving two objects can contain many more connecting links than could a phrase, or a sentence. Paivio sees images as collections of conditioned sensations which are evoked by the stimulus term, and provide the possibility of multiple associations with the response term. This sort of view suggests that imagery is just another way of achieving complexity, and thus multiplicity, in the linkage between stimulus and response.

Several of the elements of our instructions seem to suggest different ways of making the images more striking, or vivid. In Experiment 4, elements (c), (d) and (e) (vividness, activity, and the use of as many modalities as possible) had between them more effect on performance than did (b) (the use of images). Although our data do not contain any positive justification for assuming that these instructions affect performance by increasing the vividness of the images it is worth noting that vividness was one of the criteria used in the rating of images in Experiment 3, and that these ratings did relate to performance. It is difficult to formulate a hypothesis to account for the relevance of vividness to the effectiveness of mnemonics (if indeed it is so relevant) without introducing concepts apparently more relevant to theories of perception than of verbal learning.

What has been said so far about the visual imagery mnemonic suggests that the image that is created by the subject provides a highly efficient system of links between the stimulus and the response, such that if the subject can recapture the image upon external presentation of the stimulus item, he will have no difficulty in extracting the response item from the image. But it is worthy of note that it is not the real stimulus, or even the functional stimulus, that is incorporated in the image; nor indeed is the actual response a part of the image. Both stimulus and response items are represented in the image by some sub-image which the subject has constructed as symbols for them. There is thus nothing automatic about the link between the stimulus and the image, and it may be that the vividness of the image is relevant to the success of the subsequent presentation of the stimulus arousing it. Presumably, out of all the images which might come to mind when a word such as 'apple' is presented, a recent, highly vivid one is more likely to arise than an old, dull one.

Questions about whether one word is likely to be evoked as an association to some other word may quite reasonably be answered in terms of the relative frequency of association between the two words in the spoken or written language, but it does not seem to make sense to try to use the same technique in discussing whether a particular word will evoke a particular image as an associate. One reason why it does not make sense is that there are an infinite number of possible images. It is quite possible that for most people words do not usually evoke complex images in the way that they may evoke other words, and the difficulty that subjects appear to have when asked to produce images is consistent with this possibility. In other words, the subjects seem to be constructing images, and not just re-evoking previously learned responses. But this means that recalling a previously constructed image is very different from recalling a previously evoked word. It doesn't make sense to say, "Now, which image was it?", as it does to say, "Now, which word was it?". In fact, it would seem reasonable to expect that the rules which govern the recall of images would be similar to those which govern the recall of percepts, rather than of words. One might envisage the image as being constructed against the background of other images, forming an image field, and apply to it such concepts as that of the figure-ground relationship.

Given this sort of conceptualisation it would not be difficult to fit in such variables as vividness and bizarreness, which could be determinants of the degree of figure-ground contrast of the image, and hence the probability that it would be recalled. Or it could be that there is an imagery version of the Von Restorff effect (see Buxton and Newman, 1940), such that vividness or bizarreness lead to 'isolation' of the mnemonic image in the image field, and hence to a heightened probability of recall. Future research in this area might explore the applicability of the principles of perceptual interaction developed by Köhler (1929) and Koffka (1935) to the formation of the mnemonic image. It might be useful in such research to make a clear distinction between the recall of the image and the recall of the response. It would not, for instance, be too surprising to find that there was an imagery version of the Zeigarnik effect (Zeigarnik, 1927), such that images which the subject did not have time to complete were better remembered than ones which were completed, whether or not such incomplete images mediated recall of the response.

In the above discussion bizarreness has been treated as similar in its effects to vividness. It is worth pointing out that our data does not confirm that bizarre images are any more effective in recall than are mundane ones. In fact few of the images produced in any of the experiments, except Experiment 3 where the subjects were given much more time than in the other experiments, appeared to be very bizarre. In Experiment 4 it could easily have been the case that the only effect of the addition of the bizarreness instruction was to give the subjects more to do than they could cope with in the time available to them. The only reasons that can be adduced for suggesting that bizarreness is relevant to performance are, firstly, the fact that the ratings used in Experiment 3, which successfully predicted recall, used bizarreness as one of their criteria, and, secondly, the insistance of professional mnemonists that it is so. On the other hand, if, in a well controlled study, bizarreness was found not to be relevant to recall of the image, this finding would count quite heavily against the quasi-perceptual-contrast theory that we have advanced.

To sum up, it has been suggested that the subject who uses imagery mnemonics forges an extremely strong link between an image standing for the stimulus and one standing for the response. This link derives its strength from the multiplicity of the connections it contains, and possibly too from intrinsic rewards attached to its creation and contemplation. One of the penalties the subject pays for this gain is that he now has to recall the image, and whether he can do this is dependent upon how vivid and, possibly, on how bizarre, it is.

There are many other mnemonic techniques which may not all be subject to all of the same principles. They are likely to vary in their degree of indirectness, in the extent to which they use the previous item as the functional stimulus, and in the extent to which they take the meanings of the items into account, in addition to varying in their use of imagery. It is possible that there are also situations in which the subject does not consciously apply a learning technique, but in which his learning behaviour is most usefully described as conforming to one or other of these mnemonic techniques.

Simultaneous rote and mnemonic learning

It was noted in Chapter 8 (p.108) that the professional mnemonists consulted felt that if they did not "make a good image" for a particular pair of items there was no hope of recalling the second item when presented with the first. Montague, Adams and Kiess (1966), and Adams and McIntyre (1967), found that where a natural language mediator is constructed and subsequently forgotten, the subject performs less well on recall than do subjects who did not construct a mediator in the first place. Adams and McIntyre imply the view that rote learning and mediated learning cannot take place simultaneously.

In the present series of experiments it was found consistently (Experiments 2, 4 and 5) that the mnemonic subjects made a higher proportion of their errors in the form of omissions than did the control subjects. This finding could be interpreted as consistent with the view of Adams and McIntyre. If mnemonic subjects try with most pairs of items to construct an image most of the errors they make will be associated with the loss of mediators. If the attempt to construct image mediators prevents rote learning, errors will tend to take the form of omissions.

The view that rote learning is incompatible with mediated learning is not, however, unambiguous. It could be, for instance, the case, not that the two types of learning are incompatible, but that the two types of responding are so. A subject trying to recall a mediator may be unable to call simultaneously for the word that the mediator was designed to assist him to find. He may in this respect function as a single channel system. On some views of rote learning the words in the list become associated through being presented in sequence, the subject playing a passive role in the matter. On this sort of view it is difficult to see how the subject could prevent some rote learning from taking place by trying to construct a mediator. It would be interesting to attempt to discover how subjects who had constructed mediators would perform if they could be prevented from trying to use their mediators in recall. It would also be interesting to see how mnemonically instructed subjects compared with controls on a recognition test of learning. Some subjects in the present series of experiments misunderstood the instructions and only attempted to construct mediators to connect alternate pairs of items. In the anticipation trials they tended to make alternate correct responses and errors. A study using longer lists in which the subjects were instructed to learn in this way might well throw some light on the relationship of passive rote learning to mediated learning, through analysis of the types of errors which occurred on the two interlocking sets of items. Our data suggest that, whether rote learning takes place simultaneously with mnemonic learning or not, the mnemonic subject who 'loses' his mnemonic has difficulty in responding in a rote fashion.

Mediated learning in the absence of instructions

It was suggested in the discussion of Experiment 6 (Chapter 14) that the nature of the list and its mode of presentation exercise determining effects on the technique which a subject, left to himself, will use to assist his learning.

Given that techniques vary in effectiveness with different materials and modes of presentation, and that subjects are actively trying to find ways of learning the material, it would be surprising if they did not modify their behaviour when confronted with different kinds of material, or with different rates of presentation.

On the other hand it would be surprising too if the ways in which subjects modified their learning procedures were always entirely appropriate, or were consistent from subject to subject, or from occasion to occasion. Furthermore, even within the learning of a particular list, the subject may change his approach to different pairs of items, either because he has decided that the techniques he has been using will not be effective, or because he observes variations in the material which call for new strategies. The degree of heterogeneity in the subjective reports given by members of groups of control subjects in the experiments here reported was consistent with these suggestions. Some subjects reported attempts at several different techniques within the learning of the same list, and different subjects reported the use of widely different techniques in learning the same list. The mnemonically instructed subjects, by contrast, seldom reported any technique except the suggested one, even when they had failed to carry out the instructions.

If the above observations are correct, the giving of an instruction has several effects, some of which may be of a deleterious kind from the point of view of learning performance. In particular, the subject may be prevented from taking advantage of features of the material which he would, left to himself, make use of in learning it. One might make experimental use of this effect as a control on the use of idiosyncratic techniques. Thus subjects might be given an instruction expected to be ineffective in facilitating performance specifically with the intention of reducing the use of such techniques.

In Experiment 6, where no mediating instructions were given, there appeared to be a wide variation between groups in the heterogeneity of the mediating techniques used by the subjects. The group which had a highly meaningful, highly associated list behaved much more consistently than the group which had an experimental list which was low on both of these variables. This minor finding suggests a number of interesting research possibilities. Thus one might take homogeneity of reported techniques as a dependent variable, and, by using lists constructed on different principles, attempt to discover what variables in the construction of the list produce such homogeneity. Conversely, one might attempt to discover, by analysis of the verbal reports of subjects who learned under the uninstructed condition, which were the most common types of mediator used, and then examine the effects of instructing subjects to use these particular techniques upon their learning of lists constructed in particular ways.

To judge by the reports of subjects in the present series of studies, the particular mnemonic system we have been investigating is not a 'natural' one. Few uninstructed subjects reported using it, and most of these had heard about it from other subjects, or had seen a demonstration of it on television. One or two had seen commercial memory courses which contained this particular system or one similar to it. There would perhaps be advantages to using in future research techniques which are within the range of methods which uninstructed subjects are likely to use.

It would be unwise at this stage to lay too much stress on the effects of inter-item association on the use by the subjects of particular types of mediator. It is likely that this is just one of a wide range of variables which have such effects, and it may do so, not directly, but through its association with some other, and possibly simpler, variable (e.g. frequency of occurrence of the two words in the same utterance). Experiment 6 was intended to show, not that the variable of inter-item association was the crucial one, but that there are some variables which affect learning performance at least partly through the control they exert upon the use of explicit mediation devices by uninstructed subjects.

CHAPTER 16.

Summary of experiments and conclusions.

The main findings of each of the six experiments will now be summarised, and the main conclusions will be briefly presented. Some suggestions will be made for future research.

Experiment 1.

This experiment was designed to test the efficacy of mnemonic instructions under conditions suggested by pilot work to be relatively good ones from the point of view of the user of mnemonics. Subsidiary aims were to check whether mnemonic subjects showed the same serial error effects as did controls, and to see whether the mnemonic operated to facilitate order-learning as well as item learning.

The results showed: (a) that with a single presentation at a 20 second rate mnemonic subjects recall far better by the anticipation method than do control subjects, (b) that under the conditions of this experiment mnemonic subjects reveal little or no serial position effect in the distribution of their errors, (c) that under free recall output conditions mnemonic subjects show a greater tendency than control subjects to adhere to the order in which the items were presented.

Experiment 2.

The main object of Experiment 2 was to test the performance of mnemonically instructed subjects under conditions more similar to those normally used in experiments on serial learning. Multiple presentations were used, and presentation time per item was reduced to 7 seconds.

The main findings were: (a) that under these conditions the mnemonic subjects took less trials and made less errors, (b) that on the first anticipation trial mnemonic subjects showed greater response latency than did controls, (c) that mnemonic subjects made a significantly higher proportion of their errors in the form of omissions than did control subjects. Subsidiary findings were: (d) that mnemonic subjects appeared to show proportionately more improvement in learning a second list than did control subjects, (e) that the shortest delay period that shows any great loss with this type of material and presentation, for either mnemonic or control subjects, is about 5 weeks, (f) that neither mnemonic nor control subjects seemed to have any difficulty in recalling the lists backwards.

Experiment 3.

The main objects of this experiment were to discover whether differences in the associative structure of item pairs would affect the ability of subjects to carry out mnemonic instructions, and whether ratings of the images produced by the subjects would be related to the probability of later successful recall of the response items.

The findings were: (a) that three pairs of items having different levels of rated inter-item association were treated differently by the subjects, the images produced for the medium association pair being not nearly as good as those produced for the high and low association pairs, and those produced for the low association pair being slightly but significantly better than those produced for the high association pair, (b) that the probability of later recall of the response items was related to the "goodness" of the images produced, (c) that of the three pairs of items, the high inter-item association pair was best recalled, and the medium association pair was worst, the difference between low and high association pairs being small but significant.

Because the order of the mean "goodness" ratings for the three pairs of items happened to be the same as the order in which they were presented, and because there were only three pairs, it was decided that the study should be replicated. Another three pairs of items were selected, having the same inter-item association ratings as the previous set, and these were arranged in the reverse order, and presented to a new group of subjects.

The results were substantially the same as those obtained in the original study. The only significant finding that was inconsistent with the previous results was that this time the pair with high inter-item association was less well recalled than the pair with low inter-item association.

Experiment 4.

Experiment 4 was designed as an endeavour to identify any particular aspect of the complex mnemonic instruction which might be responsible for more of the difference between mnemonic and control subject performance than were other parts of the instructions. The design involved the receipt by different groups of subjects of mnemonic instructions differing in their degree of completeness. In addition to learning the experimental list for a fixed number of anticipation trials the subjects were required to return after one of three retention intervals to undergo a fixed number of relearning trials.

The main results were: (a) that learning performance was positively related to the degree of completeness of the instructions, with the exception that the group lacking only the bizarreness instruction performed slightly better than the group having the complete instruction, (b) that performance on retention was related in the same way to completeness of the instructions, (c) that the proportion of subjects showing more errors of omission than of commission increased with the degree of completeness of the instructions, and that this trend was not explicable in terms of an interaction between the tendency for all subjects to make omissions during the earlier anticipation trials and the tendency for mnemonic subjects to complete their learning earlier than the control subjects, (d) that there were no systematic differences between groups in the retention data as regards the proportions of errors of omission and commission, (e) that the data gave some reason for the suspicion that the instruction to make imagery bizarre was one that the subjects had great difficulty in carrying out under the conditions of the experiment, (f) that failure to carry out the instructions adequately was relatively common throughout the groups, (g) that there was, over the data as a whole, a tendency for the proportion of the subjects ascribing correct recall during the retention trials to recall of their images to increase with the completeness of the instructions.

Experiment 5.

The objects of Experiment 5 were: (a) to compare the performance of mnemonic subjects and controls learning lists of two different lengths, (b) to investigate the differences between mnemonic and non-mnemonic performance in both learning and retention under conditions designed to reduce variation between groups in learning ability,

(c) to examine more closely the differences in type and distribution of errors found in the earlier experiments.

The results were: (a) that there was no interaction, either in learning or in retention, between the effects of list length and those of mnemonic instruction, (b) that the effect of mnemonic instruction upon retention was almost entirely a consequence of its effect upon learning, (c) that, during learning, mnemonic instructions affected performance in the first anticipation trial proportionately more than in later trials, whereas, after a retention interval, mnemonic instructions affected relearning trials more than the recall trial, (d) that, for both learning and retention, the proportion of errors which were omissions decreased as the trials continued, but was higher for mnemonically instructed subjects, (e) that for both long and short lists, and for both learning and retention, the serial position error curve among mnemonically instructed subjects was flatter than that among control subjects.

Experiment 6.

The objects of Experiment 6 were to confirm the expectation, generated by Experiment 3, that interlist differences in mean rated inter-item association would affect list difficulty, to obtain subjective data relevant to the question of whether such differences could be explained in terms of differences in the type of explicit mediator used by the subjects, and to check whether the meaningfulness of the items has similar effects.

The results were: (a) that inter-item association and meaningfulness were both positively related to learning, the former much the more strongly, (b) that there was a significant interaction between these two variables when the measure used was errors to criterion, but not when it was trials to criterion, (c) that subjects learning the different lists gave quite different descriptions of the techniques which they had used to assist their learning, (d) that there were differences between the groups learning the different lists as regards both type and distribution of errors.

Summary of main conslusions.

1. Under a wide variety of presentation conditions and measures of performance the learning of serial lists of concrete nouns is facilitated by the giving of imagery mnemonic instructions.

2. This facilitation is at least partly, and seems largely, contingent upon the subjects carrying out the instructions.

3. Facilitation is at least partially independent of the subject's awareness that he will be required to recall the material.

4. Mnemonic instructions produce qualitative as well as quantitative differences in performance, flattening the serial position curve, and producing differences in type and distribution of errors.

5. The extent of both the qualitative and quantitative differences in performance of subjects using mnemonics from that of control subjects is related to the completeness of the mnemonic instructions.

6. Mnemonic instructions have their primary effect upon learning rather than upon retention.

7. Differences in such factors as the pre-existing associations between items can affect performance through their effect upon the ability of the subjects to carry out the mnemonic instructions.

8. In the absence of mnemonic instructions variables such as meaningfulness and inter-item association affect the subject's choice of mediating technique, and in this way to some degree determine both the level of his performance and some of its qualitative characteristics, such as the distribution and predominant types of the errors made.

9. The particular mnemonic technique primarily investigated in this series of experiments is not one which subjects are likely to use in the absence of specific instruction in its use.

Suggestions for future research on mnemonics.

In view of the apparent frequency of failure on the part of the subjects to carry out the instructions fully it is still not possible to make any firm statement about the effects of bizarreness of the mnemonic image upon learning. Possible ways of remedying this deficiency would be to make use of trained subjects, to collect full descriptions of all images from the subjects, and to analyse the data in terms of the rated bizarreness of these images. Alternatively the subjects might be presented with preconstructed images (in the form of descriptions) designed to conform with the instructions.

There is nothing in the data obtained in this series of experiments which could help in deciding which characteristics of mnemonic performance stem from the use of this particular system of mnemonics, and which would attach to the use of any kind of mnemonic. It would, for instance, be desirable in future research to compare the performance of control subjects with that of subjects given types of mnemonic instruction not involving imagery. In view of the results of Experiment 6 it might be worthwhile to attack this problem from the opposite end, not giving instructions, but trying to devise experimental materials which would cause the subjects to make use of particular kinds of mediator.

The general consistency of the verbal reports of the subjects in this series of experiments with measures of their performance suggests that some refinement of the methods used to collect and analyse such verbal reports would repay the effort involved.

It might be valuable, for example, to take sub-samples from groups of subjects given particular materials to learn, or particular instructions to carry out in their learning, on the basis of their answers to questions about the directness of the techniques which they used, or the extent to which they made use of the physical form of the materials in constructing mediators.

With regard to imagery mnemonics per se, the two-phase hypothesis put forward in Chapter 15 (p.269), according to which some aspects of the mnemonic procedure are concerned with building a link between two items, while others are concerned with increasing the availability in recall of the link so created, is certainly worthy of further elaboration and investigation. It might be possible to design instructions which would manipulate separately these two aspects of the mnemonic created by the subjects. One could then compare quantitative and qualitative aspects of the performance of subjects who had received either of these sub-instructions with those of subjects who had received neither or both of them.

Finally, it is worth noting that although we have been concerned in this thesis with the recall of words which it could reasonably be assumed the subjects knew before they were encountered during the experiments, little recognition has been given to the theoretical importance of this fact. It must be the case that we have, in some sense, a dictionary or dictionaries in our heads (see Treisman, 1960), and the nature and organisation of these dictionaries can hardly be irrelevant to the question of what happens when we memorise and recall words which are already entered therein. It may then be the case that radical progress in this area must wait upon a fuller understanding of the nature of the cerebral dictionary.

APPENDIX 1.

Word lists used in the experiments.

Experiment 1. Items obtained as described in Chapter 9.

Experimental list (25 items)

Piano	Pen
Spade	Hammer
Horse	Bus
Table	Pencil
Aeroplane	Pyjamas
Book	Nail
Matchbox	Cupboard
Egg	Fish
Spectacles	Woman
Car	Explosion
Spider	Cross
Clock	Tree
Cake	

Demonstration list (13 items)

Armchair	Mouse
Envelope	Fan
Python	Handbag
Ladder	Torch
Toothbrush	Boot
Ruler	Cup
Window	

Experiment 2.

List L.1. (16 items)	List L.2. (16 items)
Handbag	Cupboard
Aeroplane	Pencil
Matchbox	Spade
Torch	Envelope
Fan	Nail
Pen	Spectacles
Mouse	Car
Armchair	Woman
Table	Cake
Piano	Horse
Clock	Cross
Egg	Explosion
Ladder	Book
Cup	Python
Window	Hammer
Spider	Toothbrush

Experiments 4 and 5.

List P is the practice list used in Experiments 4 and 5. List L-L is the longer of the two experimental lists used in Experiment 5. Its first 10 items were used as the experimental list in Experiment 4, and as the shorter of the experimental lists in Experiment 5. The numbers placed against consecutive item pairs are ratings of inter-item association computed as described in Chapter 11.

List P

Window	
	1.10
Matchbox	
	1.15
Clock	
	1.15
Pen	
	1.00
Spider	
	1.20
Table	
	1.25
Aeroplane	
	1.10
Mouse	
	1.00
Fan	
	1.10
Cup	

List L-L

Spade	
	1.05
Envelope	
	1.25
Car	
	1.05
Toothbrush	
	1.10
Book	
	1.20
Nail	
	1.25
Spectacles	
	1.05
Python	
	1.05
Cake	
	1.10
Hammer	
	1.50
Woman	
	1.70
Explosion	
	1.05
Pencil	
	1.00
Horse	
	1.10
Cupboard	
	1.20
Cross	

Experiment 6.

The 4 lists used in this experiment were high inter-item association and low inter-item association orderings of two 12-item lists of nouns selected from the lists published by Saltz (1967). The m values in the present table are those given by Saltz. The inter-item association (a.) values were obtained as described in Chapter 14 above. All words in the high m lists (HH and HL) had Thorndike-Lorge frequencies above 1000 per $4\frac{1}{2}$ million. All words in the low m lists (LH and LL) had Thorndike-Lorge frequencies in the range 1-200 per $4\frac{1}{2}$ million.

List HH

Word	m	a.
Dinner	8.87	
		9.20
Table	8.18	
		6.60
Family	8.68	
		8.88
Husband	8.14	
		7.10
Office	7.98	
		8.42
Building	9.02	
		7.95
Street	8.90	
		5.46
Window	9.15	
		6.62
Garden	8.89	
		7.58
Country	9.16	
		3.72
Letter	8.39	
		6.10
Story	7.82	

List HL

Word	m	a.
Letter	8.39	
		1.81
Garden	8.89	
		2.97
Office	7.98	
		2.65
Family	8.68	
		4.56
Country	9.16	
		2.73
Table	8.18	
		2.07
Street	8.90	
		2.00
Story	7.82	
		2.24
Window	9.15	
		3.13
Dinner	8.87	
		2.14
Building	9.02	
		2.73
Husband	8.14	

List LH

Word	m	a.
Outfit	6.83	
		6.63
Armour	7.49	
		6.47
Padlock	7.36	
		4.19
Poison	7.48	
		7.72
Malice	6.04	
		4.50
Turmoil	6.68	
		9.16
Chaos	6.08	
		3.40
Whimper	6.87	
		4.10
Shelter	7.30	
		7.07
Kindness	6.49	
		5.33
Bonus	6.68	
		6.99
Shilling	6.30	

List LL

Word	m	a.
Chaos	6.08	
		3.99
Armour	7.49	
		1.59
Kindness	6.49	
		1.03
Turmoil	6.68	
		1.78
Padlock	7.36	
		2.30
Whimper	6.87	
		1.32
Shilling	6.30	
		1.03
Malice	6.04	
		0.64
Outfit	6.83	
		0.98
Poison	7.48	
		1.21
Bonus	6.68	
		2.26
Shelter	7.30	

APPENDIX 2.

Examples of image descriptions given by subjects in Experiment 3.

These examples are selected from those obtained in the replication of Experiment 3. For each of the three pairs of words used 5 image descriptions are given, arranged in order of 'goodness' according to the ratings obtained in the experiment (see Chapter 11). The numbers given after each description represent the mean ratings obtained.

ARMCHAIR-WINDOW

1. Children pushing the chair around a room and smashing a French window. The chair eventually floats down the river and there is a devastating mess in the lounge room. 1.3

2. An armchair set into a wall, with its back serving as a window, which is able to be opened to let people through. 2.0

3. Trying to look through a window which is in an armchair - all that is visible is stuffing. If this breaks - broken glass as stuffing is not the best when you sit down. 3.3

4. An armchair is jumping in and out of a window which is closed but does not break. Finally the armchair comes to rest on the window ledge, raises its front legs, and smiles. 4.0

5. Chair hanging by its arms from the ledge of a window on the top of the Empire State Building. It had been trying to clean the window with its cushion. Now, losing its grip slowly, it watches the traffic below and creaks with horror. 4.7

SPECTACLES-BOOK

1. A large library of books in an old home. 1.0

2. The man donned his spectacles in order that he might read the book. 1.7

3. Book in the shape of a pair of spectacles, having pages where the lenses should be. It is opened and shut by folding side arms of spectacles. 3.0

4. The spectacles were skimming through the book in a rather haphazard fashion, turning over pages with a dexterous flip of one glass. 3.7

5. I'm sitting there trying to read a pair of spectacles through the pair of rimless books I am wearing. The lenses of the spectacles appear very blurred because the books are the wrong prescription. 4.3

PENCIL-HORSE

1. A horse treading on a twig. 1.0
2. The boy drew a picture of a horse, using a pencil for the purpose. 1.7
3. A horse writing a letter with a large pencil. 2.7
4. A horse sitting in an office, on the boss's knee, taking dictation with a pencil stuck in her hoof. 4.0
5. Horse-pencil with its head as the lead. Trying to write with it but this is hard as you have to hold it with your hand round its middle, and in any case the horse is alive and keeps kicking you on the elbow. 5.0

APPENDIX 3.

Discussion of the reasons for using two-tailed tests.

In the course of the research embodied in this thesis a large number of significance tests have been carried out. In a number of instances the situation was such that most experimental psychologists would have used a one-tailed test. In spite of the apparent loss of power involved, two-tailed tests have been used in each case. The decision to eschew one-tailed tests was not taken casually, and it would seem desirable to give a brief account of the reasons for it.

Jones (1952) suggests (p.46) a very simple rule for deciding when to use one-tailed tests when he says, "....... whenever an alternative to the null hypothesis is stated in terms of the direction of expected results, the one-tailed test is applicable". In view of the facts that no objections to this rule have been raised by mathematical statisticians, and that its application leads to an apparent gain in power to reject the null hypothesis, it is not surprising that many psychologists should have been ready to use it. Objections to Jones's rule have, however, been raised by psychologists, on a number of grounds.

Burke (1953) argues that the user of a one-tailed test should be prepared to argue that a difference in the non-predicted direction, however large, is consistent with the null hypothesis. Jones (1954) attempts to cope with this difficulty by redefining the null hypothesis to include the non-predicted side, so that, in the case of a mean difference, H_0 now states that the difference between population means is equal to or less than zero. It remains the case, however, that, whatever null hypotheses Jones might wish to test, the variety of statistical techniques he is discussing (e.g. the t test) can only test exact null hypotheses.

This being so, another difficulty arises in relation to the one-tailed test. The null hypothesis and the experimental hypothesis are no longer exhaustive. It seems reasonable, if the probability in a two-tailed test is within one's pre-stated region of rejection, to reject the hypothesis of no difference in favour of its contradictory, namely that there is a difference. Any decision about direction in this case is made on the basis of a further (and usually implicit and unrecognised) statistical process. Thus, if it is unlikely that the sample was drawn from a population with the mean specified in the null hypothesis, it is even less likely that it was drawn from a population with a mean differing from the null hypothesis mean in the opposite direction from the observed mean. In the case of a one-tailed test the rejection of the null hypothesis does not in the same way entail acceptance of the experimental hypothesis, since it is logically possible, even if statistically unlikely, that a difference in one direction significant at a given level could be obtained in samples from populations with parameters differing in the opposite direction.

Further difficulties arise from the fact that by using a one-tailed test one binds oneself to ignore data differing in the non-predicted direction. As Burke (1953) says (pp. 385-386), "It is to be doubted whether experimental psychology, in its present state, can afford such lofty indifference towards experimental surprises". Even if we could afford it, it is to be doubted if such indifference is possible in practice. Let us suppose I predict that a coin is biassed towards Heads, and prepare to test the result of 100 tosses with a one-tailed test. If the coin turns up Tails 90 times I lose my right to use a one-tailed test. Indeed, my original hypothesis has been falsified. It remains the case, however, that I would be foolish to act in relation to that coin as if I had no reason to think it was biassed.

If it be argued that what I have done here was to replace my one-tailed test by an implicit two-tailed test, and that this procedure is acceptable, I would point out that another absurdity arises from this point of view. If it is acceptable to use a two-tailed test upon the falsification of a directional hypothesis it becomes worthwhile to make a prediction in every experiment, even in the absence of any grounds for predicting direction, since one would then be able to make use of the extra power of the one-tailed test half of the time.

Irrespective of the question of what happens if the data falsify one's directional prediction it would seem strange that the mere making of a prediction should give one the right to use the extra power of the one-tailed test, and it quickly became accepted that the experimenter needed to be able to give some reason for his prediction before he gained this right. Three sorts of reasons have been put forward as being relevant. Firstly, he might have grounds arising out of previous research for the prediction he makes. McNemar (1955) presents the right to use the one-tailed test as a sort of reward for sagacity. Secondly, he might only be interested in results in a particular direction. The test might for instance be one in which a proposed improvement in industrial process is being compared with the process previously used. Thirdly, a population difference in the non-predicted direction might be impossible. There are very few exemplars of this situation to be found in the field of experimental psychology, although there are many prediction situations where one's confidence that there will not be population difference in a particular direction is such that one might wish to claim that a situation of this sort obtains. For example, if the scores of subjects on a German vocabulary test before and after a course in this language are compared it might reasonably be claimed that a population difference in means in a negative direction is impossible.

With regard to the first of these three reasons for using one-tailed tests it seems not unreasonable that a researcher should project his expectations into his choice of a region of rejection. We will, after all, accept shorter odds against an event which we think to be likely than against one which we think to be unlikely. But there seems something logically odd about projecting our expectations onto the computation of the probability of the event on the null hypothesis since it is the relative objectivity of this probability that makes significance testing worthwhile. In any case, if there is to be a reward for the experimenter's sagacity and knowledge of his field, it seems odd that this should always take the form of a one-for-one probability bonus. Sagacity and knowledge do, after all, vary on a continuous scale. Furthermore, an experimenter may have good reasons for expecting a difference, without having good reasons for predicting a direction, and it should be possible for him, too, to project his confidence into his choice of a region of rejection. A.H. Winefield, of this department, recently suggested that there is nothing to stop a researcher projecting the differential strengths of his expectations of positive and negative differences into his choice of the regions of rejection on both sides of the null hypothesis. Thus he might decide to set the region of rejection at the .04 level on the positive side and at the .01 level on the negative side giving a total region of rejection of 5%.

The second of the three reasons for choosing a one-tailed test seems, in the light of the above discussion, neither to need nor to deserve much comment. It seems to involve the assimilation of the statement that one has no interest in rejecting the null hypothesis in a particular direction to the statement that there is no reason to expect the null hypothesis to be falsified in that direction; an assimilation for which there appears to be no grounds. It becomes, in application, a straightforward use of Jones's rule, and thus opens the door to completely indiscriminate use of one-tailed tests.

The third type of reason for using one-tailed tests seems to involve an extreme case of directional expectation, with the added implication that the experimenter is able to prove in advance that his expectation is correct. Looked at from the point of view of Winefield's suggestion it is as if the experimenter was able to give good grounds for setting the region of rejection on one side at zero probability. This would be an event which occurred very seldom indeed in psychology, and certainly not in any of the experiments reported in this thesis. It is worth pointing out, however, that even in this sort of case the size of the region of rejection on the side on which one <u>is</u> set may be expected to vary with the strength of the experimenter's expectation that a difference will be found.

In view of the above arguments it was felt that the status of one-tailed tests was not clear, and that, in the absence of a clear willingness to accept the null hypothesis if a difference occurred in the unexpected direction, they should be avoided. It was felt that a Ph.D. thesis was not the place to introduce a radical change in the procedure of significance testing such as that proposed by Winefield.

BIBLIOGRAPHY

Abramczyk, R.R. & Bousfield, W.A. Sequential ordering in repeated free recall as a function of inter-item associative strength. Psychol. Rec. (1967), 17, 183-192.

Adams, J.A. & McIntyre, J.S. Natural language mediation and all-or-none learning. Canad. J. Psychol. (1967), 21, 436-449.

Amster, H. Convergent association norms for ten-year-old children and college age adults. Psychon. Monog. Suppl. (1967), Vol. 2, No. 1, (Whole No.17.)

Archer, E.J. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychol. Monogr. (1960), 74, 10, (Whole No. 497.)

Baddeley, A.D. How does acoustic similarity influence short-term memory? J. exp. Psychol. (1968), 20, 249-264.

Baddeley, A.D. & Dale, H.C.A. The effect of semantic similarity on retro-active interference in long- and short-term memory. J. verb. Learn. verb. Behav. (1966), 5, 217-220.

Bartlett, F.C. Remembering: A study in experimental and social psychology. Cambridge: Cambridge Univ. Press, 1932.

Battig, W.F., Brown, S.C. & Nelson, D. Constant versus varied serial order in paired-associate learning. Psychol. Rep. (1963), 12, 695-721.

Battig, W.F., Brown, S.C. & Schild, M.E. Serial position and sequential associations in serial learning. J. exp. Psychol. (1964), 67, 449-457.

Berlyne, D.E. Conflict, arousal, and curiosity. New York: McGraw-Hill, 1960.

Boersma, F.J., Conklin, R.C. & Carlson, J.E. Effects of reporting associative strategies on the retention of paired associates. Psychon. Sci. (1966), 5, 463-464.

Bousfield, W.A. The occurrence of clustering in the recall of randomly arranged associates. J. gen. Psychol. (1953), 49, 229-240.

Braun, H.W. & Heymann, S.P. Meaningfulness of material, distribution of practice, and serial-position curves. J. exp. Psychol. (1958), 56, 146-150.

Bross, I.D.J. Is there an increased risk? Federation Proceedings. Vol.13, 1954, pp.815-819.

Bugelski, B.R. Presentation time, total time, and mediation in paired-associates learning. _J. exp. Psychol_. (1962), 63, 409-412.

Bugelski, B.R., Kidd, E. & Segmen, J. Image as a mediator in one-trial paired-associate learning. _J. exp. Psychol_. (1968), 76, 69-73.

Bugelski, B.R. & Scharlock, D.P. An experimental demonstration of unconscious mediated association. _J. exp. Psychol_. (1952), 44, 334-338.

Burke, C.J. A brief note on one-tailed tests. _Psychol. Bull_. (1953), 50, 384-386.

Buxton, C.E. & Newman, E.B. The forgetting of 'crowded' and 'isolated' materials. _J. exp. Psychol_. (1940), 26, 180-198.

Chomsky, N. Review of Skinner (1957) _Language_. (1959), 35, 26-58.

Clark, L.L., Lansford, T.G. & Dallenbach, K.M. Repetition and associative learning. _Amer. J. Psychol_. (1960), 73, 22-40.

Cofer, C.N. _Verbal learning and verbal behavior_. New York: McGraw-Hill, 1961.

Dean, S.J. & Martin, R.B. Reported mediation as a function of degree of learning. Psychon. Sci. (1966), 4, 231-232.

Deese, J. The psychology of learning. New York: McGraw-Hill, 1952.

Deese, J. Influence of inter-item associative strength upon immediate free recall. Psychol. Rep. (1959), 5, 305-312.

Deese, J. From the isolated verbal unit to connected discourse. In C.N. Cofer (Ed.), Verbal learning and verbal behavior. New York: McGraw-Hill, 1961.

Deese, J. On the structure of associative meaning. Psychol. Rev. (1962), 69, 161-175.

Dowling, R.M. & Braun, H.W. Retention and meaningfulness of material. J. exp. Psychol. (1957), 54, 213-217.

Dukes, W.F. & Bastian, J. Recall of abstract and concrete words equated for meaningfulness. J. verb. Learn. verb. Behav. (1966), 5, 455-458.

Ebbinghaus, H. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Leipzig: Duncker und Humbolt, 1885. Trans. as: Memory: a contribution to experimental Psychology by H.A. Ruger and C.E. Bussenius. New York: Teachers College. Columbia Univ., 1913.

Ebenholtz, S.M. Serial learning: position learning and sequential associations. J. exp. Psychol. (1963), 66, 353-362.

Ebenholtz, S.M. Serial-position effect of ordered stimulus dimensions in paired associate learning. J. exp. Psychol. (1966), 71, 132-137.

Estes, W.K. Learning theory and the new "mental chemistry". Psychol. Rev. (1960), 67, 207-223.

Estes, W.K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. Psychometrica. (1961), 26, 73-84.

Estes, W.K., Hopkins, B.L. & Crothers, E.J. All-or-none and conservation effects in the learning and retention of paired associates. J. exp. Psychol. (1960), 60, 329-339.

Farber, I.E. The things people say to themselves. Amer. Psychologist. (1963), 18, 185-197.

Freedman, J.L. & Landauer, T.K. Retrieval of long-term memory: "Tip-of-the-tongue" phenomenon. Psychon. Sci. (1966), 4, 309-310.

Furst, B. Stop forgetting. Garden City Press, 1954.

Gagné, R.M. _The conditions of learning_. New York: Holt, Rinehart and Winston, 1965.

Galton, F. Statistics of mental imagery. _Mind_. (1880), 5, 300-318.

Gannon, D.R. & Noble, C.E. Familiarization (n) as a stimulus factor in paired-associate verbal learning. _J. exp. Psychol_. (1961), 62, 14-23.

Garskof, B.E. & Houston, J.P. The measurement of verbal relatedness: an idiographic approach. _Psychol. Rev_. (1963), 70, 277-288.

Garskof, B.E., Houston, J.P. & Mednick, S.A. Overlap and direct associative strength norms for 480 word pairs. _Psychon. Monogr. Suppl_. (1967), Vol.2, No.2. (Whole No.18).

Garskof, B.E., Sandak, J.M. & Malinowski, E.W.A. Controlling the "fate" of first list associates. _Psychon. Sci_. (1965), 2, 315-316.

Glaze, J.A. The association value of non-sense syllables. _J. genet. Psychol_. (1928), 35, 255-267.

Groninger, L.D. Natural language mediation and covert rehearsal in short term memory. _Psychon. Sci_. (1966), 5, 135-136.

Guthrie, E.R. The psychology of learning. New York: Harpers, 1935.

Harlow, H.F. The formation of learning sets. Psychol. Rev. (1949), 56, 51-65.

Hays, W.L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Horowitz, L.M. & Izawa, C. Comparison of serial and paired-associate learning. J. exp. Psychol. (1963), 65, 352-361.

Hull, C.L. The meaningfulness of 320 selected nonsense syllables. Amer. J. Psychol. (1933), 45, 730-734.

Hunter, I.M.L. Mnemonic systems and devices. Science News. (1956), No. 39, 75-97.

Jensen, A.R. Transfer between paired-associate and serial learning. J. verb. Learn. verb. Behav. (1962a), 1, 269-280.

Jensen, A.R. An empirical theory of the serial position effect. J. Psychol. (1962b), 53, 127-142.

Jensen, A.R. & Rohwer, W.D. What is learned in serial learning? J. verb. Learn. verb. Behav. (1965), 4, 62-72.

Jonckheere, A.R. A distribution-free k-sample test against ordered alternatives. _Biometrica_. (1953), 40, 133-145.

Jones, L.V. Tests of hypotheses: One-sided vs. two-sided alternatives. _Psychol. Bull_. (1952), 49, 43-46.

Jones, L.V. A rejoinder on one-tailed tests. _Psychol. Bull_. (1954), 51, 585-586.

Kausler, D.H. _Readings in verbal learning_. New York: Wiley, 1966.

Kendall, M.G. _Rank correlation methods_. London: Griffin, 1948.

Kiess, H.O. & Montague, W.E. Natural language mediators in paired associate learning. _Psychon. Sci_. (1965), 3, 549-550.

Koffka, K. _Principles of Gestalt Psychology_. New York: Harcourt, Brace, 1935.

Köhler, W. _Gestalt Psychology_. Liveright, 1929.

Krueger, W.C.F. The relative difficulty of nonsense syllables. _J. exp. Psych_. (1934), 17, 145-153.

Lambert, W.E. & Paivio, A. The influence of noun-adjective order on learning. Canad. J. Psychol. (1956), 10, 9-12.

Lockhead, G R. A re-evaluation of evidence for one-trial associative learning. Amer. J. Psychol. (1961), 74, 590-595.

Lorayne, H. Develop a super-power memory. London: The New English Library, 1963.

Luria, A.R. Memory and the structure of mental processes. Problems of psychology, No.1. U.K.: Pergamon Press, 1960.

Mandler, G. Associative frequency and associative prepotency as measures of response to nonsense syllables. Amer. J. Psychol. (1955), 68, 662-665.

Mandler, G. Verbal learning. In New Directions in Psychology, T.M. Newcomb (Ed.). New York: Holt, Rinehart & Winston, 1967.

Mandler, G. & Campbell, E.H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. J. exp. Psychol. (1957), 54, 269-273.

Mandler, G. & Earhard, B. Pseudomediation: Is chaining an artifact. Psychon. Sci. (1964), 1, 247-248.

Martin, C.J., Boersma, F.J. & Cox, D. A classification of associative strategies in paired-associate learning. Psychon. Sci. (1965), 3, 455-456.

Martin, C.J., Cox, D.L. & Boersma, F.J. The role of associative strategies in the acquisition of paired-associate materials: an alternative approach to meaningfulness. Psychon. Sci. (1965), 3, 463-464.

Martin, C.J. & Saltz, E. Serial versus random presentation of paired-associates. J. exp. Psychol. (1963), 65, 609-615.

Martin, R.B. & Dean, S.J. Implicit and explicit mediation in paired-associate learning. J. exp. Psychol. (1964), 68, 21-27.

Martin, R.B. & Dean, S.J. Reported mediation in paired-associate learning. J. verb. Learn. verb. Behav. (1966), 5, 23-27.

McGehee, N.E. & Schulz, R.W. Mediation in paired-associate learning. J. exp. Psychol. (1961), 62, 565-570.

McGeoch, J.A. & Irion, A.L. The psychology of human learning. New York: KcKay, 1952.

McKellar, P. _Imagination and thinking_. New York: Basic Books, Inc., 1957.

McNemar, Q. _Psychological Statistics_. New York: John Wiley & Sons, 1955.

McNulty, J.A. The effect of "instructions to mediate" upon paired-associates learning. _Psychon. Sci_. (1966), 4, 61-62.

Millar, G.A. _Language and communication_. New York: McGraw-Hill, 1951.

Miller, G.A. Comments on paper, "One trial learning?" by L. Postman. In C.N. Cofer and B.S. Musgrave (Eds.), _Verbal learning and behavior_. New York: McGraw-Hill, 1963.

Miller, G.A., Gallanter, E. & Pribram, K.H. _Plans and the structure of behavior_. New York: Holt, Rinehart & Winston, 1960.

Miller, G.A. & Selfridge, J.A. Verbal context and the recall of meaningful material. _Amer. J. Psychol_. (1950), 63, 176-185.

Montague, W.E., Adams, J.A. & Kiess, H.O. Forgetting and natural language mediation. _J. exp. Psychol_. (1966), 72, 829-833.

Montague, W.E. & Wearing, A.J. The complexity of natural language mediators and its relation to paired-associate learning. Psychon. Sci. (1967), 7, 135-136.

Mowrer, O.H. Learning theory and the symbolic behavior. New York: Wiley, 1960.

Noble, C.E. An analysis of meaning. Psychol. Rev. (1952a), 59, 421-430.

Noble, C.E. The role of stimulus meaning (m) in serial verbal learning. J. exp. Psychol. (1952b), 43, 437-446.

Noble, C.E. Meaningfulness and familiarity. In C.N. Cofer and B.S. Musgrave (Eds.), Verbal learning and behavior. New York: McGraw-Hill, 1963.

Noble, C.E. & McNeeley, D.A. The role of meaningfulness (m) in paired-associate verbal learning. J. exp. Psychol. (1957), 53, 16-22.

Noble, C.E., Stockwell, F.E. & Pryor, M.W. Meaningfulness (m') and association value (a) in paired-associate syllable learning. Psychol. Rep. (1957), 3, 441-452.

Olton, R.M. The effect of a mnemonic upon the retention of paired-associate verbal material. Ph.D. Dissertation, University of California, Berkeley, 1966. (Abstract only).

Osgood, C.E. The nature and measurement of meaning. Psychol. Bull. (1952), 49, 197-237.

Osgood, C.E., Suci, G.J. & Tannenbaum, P.H. The measurement of meaning. Urbana Ill: University of Illinois Press, 1957.

Oswald, I. The use of the E.E.G. in the study of mental imagery. Quart. J. exp. Psychol. (1957), 9, 113-118.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canad. J. Psychol. (1963), 17, 370-379.

Paivio, A. Abstractness, imagery, and meaningfulness in paired-associate learning. J. verb. Learn. verb. Behav. (1965), 4, 32-38.

Paivio, A. & Yarmey, A.D. Pictures versus words as stimuli in paired-associate learning. Psychon. Sci. (1966), 5, 235-236.

Paivio, A. & Yuille, J.C. Word abstractness and meaningfulness, and paired-associate learning in children. J. exp. Ch. Psychol. (1966), 4, 81-89.

Paivio, A., Yuille, J.C. & Smythe, P.C. Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. Canad. J. Psychol. (1966), 20, 362-377.

Perrino, C. An evaluation of the effect of grammatical structure on learning. Unpublished Masters thesis, Northwestern University, 1965. Cited by Wood (1966).

Peters, H.N. Mediate association. J. exp. Psychol. (1935), 18, 20-48.

Pillsbury, W.B. A study in apperception. Amer. J. Psychol. (1897), 8, 315-393.

Postman, L. One-trial learning? In C.N. Cofer and B.S. Musgrave (Eds.), Verbal behavior and learning. New York: McGraw-Hill, 1963.

Postman, L. Rewards and punishments in human learning. In L. Postman (Ed.), Psychology in the making. New York: Alfred A. Knopf, 1964.

Postman, L. & Schwartz, M. Studies of learning to learn: I. Transfer as a function of method of practise and class of verbal materials. J. verb. Learn. verb. Behav. (1964), 3, 37-49.

Primoff, E. Backward and forward association as an organising act in serial and paired-associate learning. J. Psychol. (1938), 5, 375-395.

Reed, H.B. Associative aids. I. Their relation to learning, retention, and other associations. II. Their relation to practise and the transfer of training. Psychol. Rev. (1918), 25, 128-155, 257-285.

Restle, F. Significance of all-or-none learning. Psychol. Bull. (1965), 64, 313-325.

Rock, I. The role of repetition in associative learning. Amer. J. Psychol. (1957), 70, 186-193.

Rock, I. & Heimer, W. Further evidence of one-trial associative learning. Amer. J. Psychol. (1959), 72, 1-16.

Rock, I. & Steinfeld, G. Methodological questions in the study of one-trial learning. Science. (1963), 140, 822-824.

Roethlisberger, F.J. & Dixon, W.J. Management and the worker. Harvard University Press, 1940.

Roth, D.M. Roth memory course. Santa Monica: Motivation, 1961.

Rundquist, W.N. Mediation speed, reported mediators and recall, as related to presentation rate. Psychon. Sci. (1965), 3, 51-52.

Rundquist, W.N. & Farley, F.H. The use of mediators in the learning of verbal paired associates. J. verb. Learn. verb. Behav. (1964), 3, 280-285.

Russell, W.A. & Jenkins, J.J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. Tech. Rep. No. 11, Contract N8, ONR-66216, 1954.

Russell, W.A. & Storms, L.H. Implicit verbal chaining in paired-associate learning. J. exp. Psychol. (1955), 49, 287-293.

Saltz, E. Thorndike-Lorge Frequency and m of stimuli as separate factors in paired-associates learning. J. exp. Psychol. (1967), 73, 473-478.

Seibel, R., Lockhart, R. & Taschman, C.S. The learning of 100 paired-associates in a single, self-paced trial. Psychon. Bull. (1967), 1(2), 11. (Abstract only).

Shipley, W.C. An apparent transfer of conditioning. J. gen. Psychol. (1933), 8, 382-391.

Short, P.L. The objective study of mental images. Brit. J. Psychol. (Gen.), (1953), 44, 35-51.

Skinner, B.F. Verbal Behavior. (The Century Psychology Series). New York: Appleton-Century-Crofts, 1957.

Smith, K. & Noble, C.E. Effects of a mnemonic technique applied to verbal learning and memory. Percept. Mot. Skills. (1965), 21, 123-134.

Thorndike, E.L. & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University Press, 1944.

Thurstone, L.L. The relation between learning time and length of task. Psychol. Rev. (1930), 37, 44-53.

Treisman, A.M. Contextual cues in selective listening. Quart. J. exp. Psychol. (1960), 12, 242-248.

Tresselt, M.E. & Spragg, S.D.S. Changes occurring in the serial reproduction of verbally perceived materials. J. genet. Psychol. (1941), 58, 255-264.

Underwood, B.J. Stimulus selection in verbal learning. In C.N. Cofer and B.S. Musgrave (Eds.), Verbal behavior and learning. New York: McGraw-Hill, 1963.

Underwood, B.J. & Erlbacher, H.H. Studies of coding in verbal learning. Psychol. Monog. (Gen. and Applied), (1965), Vol.79, No.13.

Underwood, B.J. & Keppel, G. One-trial learning? J. verb. Learn. verb. Behav. (1962), 1, 1-13.

Underwood, B.J., Rehula, R. & Keppel, G. Item selection in paired-associate learning. Amer. J. Psychol. (1962), 75, 353-371.

Underwood, B.J. & Schulz, R.W. Meaningfulness and verbal learning. New York: J.B. Lippincott, 1960.

Wallace, W.H., Turner, S.H. & Perkins, C.C. Preliminary studies of human information storage. D.A. Project No. 3-99-12-023. S.C. Project No. 1320. University of Pennsylvania, 1957.

Wearing, A.J. & Montague, W.E. Associability of CVC-word pairs and its relation to list difficulty. _Psychon. Sci_. (1967), 7, 133-134.

Wees, W.R. & Line, W. The influence of the form of a presentation upon reproduction: the principle of determination. _Brit. J. Psychol_. (1937), 28, 167-189.

Weinland, J.D. _How to improve your memory_. New York: Barnes and Noble, 1957.

Williams, J.P. Supplementary report: A selection artifact in Rock's study of the role of repetition. _J. exp. Psychol_. (1961), 62, 627-628.

Winzen, K. Die Abhängigkeit der paarweisen Assoziation von der Stellung des besser haftenden Gliedes. _Z. Psychol_. (1921), 86, 236-252.

Witmer, L.R. The association value of three-place consonant syllables. _J. genet. Psychol_. (1935), 47, 337-360.

Wood, G. Mnemonic systems in recall. Unpublished Ph.D. dissertation. Northwestern University, 1966.

Woodworth, R.S. & Poffenberger, A.T. _Textbook of experimental psychology_. Mimeographed ed. New York: Columbia University, 1920.

Woodworth, R.S. & Schlosberg, H. Experimental psychology. New York: Henry Holt, 1954.

Yarmey, A.D. & Thomas, K.A. Set and word abstractness-concreteness shift in paired-associate learning. Psychon. Sci. (1966), 5, 387-388.

Young, R.K. A comparison of two methods of learning serial associations. Amer. J. Psychol. (1959), 72, 554-559.

Young, R.K. The stimulus in serial verbal learning. Amer. J. Psychol. (1961), 74, 517-528.

Young, R.K. Tests of three hypotheses about the stimulus in serial learning. J. exp. Psychol. (1962), 63, 307-313.

Young, R.K. & Clark, J. Compound stimulus hypothesis in serial learning. J. exp. Psychol. (1964), 67, 301-302.

Young, R.K., Patterson, J. & Benson, W.N. Backward serial learning. J. verb. Learn. verb. Behav. (1963), 1, 335-338.

Zinchenko, P.I. Ob osobennostiakh, mnemicheskoĭ poznavatl'noĭ orientviovki v protsessakh pamiati. Vop. Psikhol, (1959), 5(1), 76-92. (Abstract only).

Zeigarnik. B. Das Behalten eledigter und uneledigter Handlungen. Psychol. Forsch. (1927), 9, 1-85.