R842-05 ERIC REPORT RESUME

ED 010 263 2-28-67 24 (REV)

HOW LEARNING IS AFFECTED BY CHANGE IN SUBJECT MATTER--SOURCES OF INTERFERENCE IN VERBAL LEARNING.


CALFEE, ROBERT C.
XY081677 UNIVERSITY OF WISCONSIN, MADISON



CRP-S-321
BR-5-8134
-NOV-66 OEC-5-10-374
EDRS PRICE MF-$0.18 HC-$3.16 79P.



*STIMULUS BEHAVIOR, *LEARNING PROCESSES, *COGNITIVE PROCESSES, RETENTION, RETENTION STUDIES, LEARNING, LEARNING THEORIES, *MODELS, *VERBAL LEARNING, MADISON, WISCONSIN



THE EFFECTS OF INTERPOLATED ITEMS ON THE LEARNING AND RETENTION OF INDIVIDUAL STIMULUS-RESPONSE (S-R) UNITS IN PAIRED-ASSOCIATE LEARNING WERE INVESTIGATED. EXPERIMENTS WERE DESIGNED TO DETERMINE THE EFFECTS OF INTERFERENCE PRODUCED BY OTHER ITEMS WHICH OCCUR NATURALLY WITHIN THE CONTEXT OF A STANDARD PAIRED-ASSOCIATE TASK. IN ADDITION, SEVERAL INVESTIGATIONS WERE CARRIED OUT IN THE GENERAL AREA OF COGNITIVE PROCESSING OF VERBAL INFORMATION. THE EFFECTS OF VARIATIONS IN LIST LENGTH (NUMBER OF ITEMS TO BE LEARNED SIMULTANEOUSLY) WERE STUDIED WITHIN THE CONTEXT OF A PAIR OF LEARNING MODELS. A RELATED PROBLEM WAS THE RELATIVE EFFICIENCY OF VARIOUS PART-LIST PROCEDURES. A FINAL SERIES OF EXPERIMENTS WAS CONCERNED WITH SHORT-TERM MEMORY PHENOMENA IN THE PROCESSING OF VERBAL INFORMATION. DATA FROM ALL THE STUDIES SUPPORTED THE GENERAL CONCLUSION THAT RETENTION OF SINGLE S-R UNITS IS DETERMINED BY MORE THAN ONE FACTOR. IN ADDITION TO THE NUMBER OF TIMES A UNIT HAS BEEN PRESENTED, THE NUMBER AND SPACING OF OTHER ITEMS TO BE LEARNED, THE STATE OF LEARNING, COMPATIBILITY OF INPUT AND OUTPUT MODES, AND PRIOR INFORMATION ABOUT MEMORY LOAD WERE CONSIDERED IMPORTANT. THEORETICAL MODELS WERE PROPOSED AND DISCUSSED IN RELATION TO THE DATA OBTAINED IN THE STUDIES. (JC)

FINAL REPORT
Project No. S-321 (S-8134)
Contract No. OE-5-10-374

# HOW LEARNING IS AFFECTED BY CHANGE IN SUBJECT MATTER: SOURCES OF INTERFERENCE IN VERBAL LEARNING

November, 1966

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

ED010263

# How learning is affected by change in subject matter: Sources of interference in verbal learning

Project No. S-321
Contract No. OE-5-10-374

Robert C. Calfee

October 17, 1966

The research reported herein was performed pursuant to a Contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

University of Wisconsin

Madison, Wisconsin

## Table of Contents

## Acknowledgement

The research reported herein represents the collaborative efforts of several individuals. Donald Homa and Julianne Mapes assisted in the design of the experiments in Section II; Mr. Homa also ran the subjects and carried out most of the data analysis for these studies. Esther Macalady prepared the stimulus lists and ran subjects in the German-English study of Section III. Joel Goren prepared stimulus materials and carried out the research reported in Section IV on visual and acoustic encoding. Phyllis Walzer, Pat Gottlieb and Rogina Polesta performed the experiments in short-term memory in children. Special thanks are due to Helen Dawes of the University of Wisconsin Preschool Laboratory, and Mary Griggs of the Neighborhood House Playchool, Madison, for their helpful cooperation. The developmental work was done in collaboration with Mavis Hetherington, whose experience in this area was a godsend to one not used to work with children. Special appreciation goes to William Chase. The experiments reported in Section IV are largely his work. He has played an important role as sounding board, critic, and stimulating source of ideas in the area of human learning.

Portions of the computer time required for the work on optimization in Section III, and for analysis of the visual confusion data in Section IV, were made available by the Research Committee, University of Wisconsin, Madison, from the Wisconsin Alumni Research Committee.

## Section I. Introduction

The primary goal of the studies carried out under this contract was to investigate the effects of interpolated items on the learning and retention of individual stimulus-response (S-R) units in paired-associate learning. Specific experiments which are reported in Section II were designed to determine the effects of interference produced by other items which occur naturally within the context of a standard paired-associate task. The empirical results were useful in formulation and evaluation of alternative mathematical models for acquisition and retention of verbal information.

In addition to this primary goal, several investigations were carried out with the support provided by the contract in the general area of cognitive processing of verbal information. The thread which ties together the entire research program is the search for optimal strategies for the presentation of verbal material. In a series of investigations reported in Section III, the effects of variations in list length (the number of items to be learned simultaneously) are studied within the context of a pair of learning models, the first a discrete-state Markov process, and the second involving a convolution of incremental acquisition and forgetting processes. A related problem is the relative efficiency of various part-list procedures. A study is reported in which German-English pairs are learned under various part-list procedures, and the data are predicted by a two-operator linear model.

A final series of experiments, reported in Section IV, are concerned with short-term memory phenomena in the processing of verbal information. In the first study, a factorial comparison is made of visual and auditory input and output operations on reaction time in a simple memory task. Both visually and acoustically confusing sets of materials were used, as well as a neutral set. An auxilliary study was performed to obtain an empirical confusion matrix based on visually similar letters.

A final set of studies on short-term memory was carried out using pre-school children as subjects. The main variables of interest were memory load (number of items to be recalled) and load information (information prior to presentation about the number of items to be presented in a serial list).

## Section II: Interpresentation effects in paired-associate learning

Background of problem: In a paired-associate task, the sequence of presentations of a particular S-R item is imbedded in a complex structure comprised of the presentations of other items in the list. Between successive presentations of a given item from a list of N pairs are interpolated from zero to 2(N-1) other items, for each of which the state of learning may range from complete ignorance to well-acquired. If the anticipation technique is used, an additional factor is introduced, viz., the amount of time and/or effort which the subject expends in trying to come up with the correct response for each item.

The effects of variations in interpresentation events on the course of learning a given item are of interest in part because of recent theoretical developments in statistical learning models. For example, the trial-dependent forgetting (TDF) model (Calfee & Atkinson, 1965) incorporates explicit assumptions about the effects of interpolated items on the acquisition of a given pair. After a study presentation, a pair is assumed to go into a short-term or long-term memory store. An item in the latter state is considered to be learned. If the item is in short-term memory, then succeeding presentations may result in forgetting, represented as transition to a guessing state. Estimates of the transition probabilities from several independent sets of data, each of which the model handled rather well (Calfee & Atkinson, 1965; Calfee, 1966a), all indicated that learning was much more likely to occur if an item remained in short-term memory since its last presentation than if it had been forgotten. One implication is that massing of presentations should lead to faster learning, a prediction which is not in accord with numerous other findings. Greeno (1964), using a design in which some items occurred twice in succession on each trial, found that the second presentation produced little learning; the same number of trials to criterion (twice the number of presentations) was required for the experimental pairs as for those presented once per trial. Rothkopf and Coke (1963, 1966) found in a free-recall task that better retention took place if successive repetitions of items were spaced rather than massed.

Greeno (1966) has argued for a theoretical model in which an item must be forgotten to be learned; i.e., if an item is in short-term memory just prior to a study presentation, the study presentation will have no effect. An identifiable theory is stated involving transition parameters for transfer to long-term memory from either short-term or guessing states. The theory is applied to data from Peterson's laboratory (Peterson, Wampler, Kirkpatrick & Saltzman, 1963; Peterson, Hillner & Saltzman, 1962) as well as Greeno's. In general, transition from the guessing state is three to four times more likely than from the short-term state according to Greeno's analysis.

In the first experiment to be reported in this paper, Greeno's twice-per-trial technique is used to evaluate parametrically the effects of massing on learning rate. In a second experiment, a different type of mixed-list design is used to determine effects of massing presentations, with the proviso that at least one other item intervene between successive presentations of a given pair.

The third experiment tests another assumption incorporated in the TDF model concerning the effects of interpolated items, viz., that the amount of interference produced by unlearned items is substantially greater than interference from learned items.

Method

## Experiment I

Subjects. Forty undergraduates enrolled in introductory psychology at the University of Wisconsin, volunteered to serve as Ss, and were assigned in an alternating manner to the High and Low Meaningfulness conditions.

Materials. Two stimulus sets of 15 CVC trigrams were selected using Archer's (1960) norms, a high association (HA) set ($\overline{X}$=85, range 80-90) and a low association (LA) set ($\overline{X}$=14, range 10-17). To obtain sets with minimal intralist similarity, no consonant was used in the first or last positions more than once, and the vowels (including Y) were evenly distributed within each set. Response terms consisted of two digit numbers; the digits 1-9 were used with approximately equal frequency, repeating numbers were not allowed, and the choice was otherwise random. Pairings of CVC's with two-digit numbers was randomized separately for each S.

Each 15-item list was divided into 5 sublists of approximately equal association value. Sublists were then assigned to one of the 5 spacing conditions described below according to a Latin square design, which was randomly selected for each block of 5 subjects. In condition N, the 3 items were presented once per trial. In the other conditions, each item was presented twice per trial with 0, 1, 2 or 3 other pairs interpolated between the two presentations for the respective conditions. Hence, if the pair HYP-27 was in a sublist assigned to Condition 0, then on each trial that pair would be presented twice in succession with no other pairs interpolated. Between- and within-trial spacing are necessarily confounded -- the more interpolated pairs between the two presentations, on a given trial, the fewer pairs falling between the second presentation on trial n and the first presentation on trial n+1. A trial consisted of presentation of 27 pairs, however, and the between-trial difference in interpolated pairs is proportionately much smaller than within-trial variation.

An anticipation procedure was used. The stimulus term was presented by a Carousel projector, and the S was given as much time as needed to make a response. Response terms were always available for reference on a card in front of the S. Following S's response, the correct response was presented together with the stimulus for a 2-sec. feedback period.

Procedure. Following instructions about the paired-associate task, S was given a 3-item practice list (A, B and C paired with 3, 5 and 7) to a criterion of two errorless trials. Any questions were answered by appropriate rephrasing of the instructions. The experimental list was then presented with no trial breaks except to change slide trays. Within the constraints of the spacing conditions, order of presentation of items was random, except that an item at the end of one trial was never permitted to be the first item on the next trial. A response was required on each trial, and S was encouraged to guess if he didn't know the answer. Training continued to a criterion of three consecutive errorless trials.

Results. In Table 1A is presented the average trial of last error to a criterion of 3 consecutive correct responses based only on the first presentation per trial. An analysis of variance showed that spacing had a significant effect, $F(4,152) = 8.6$, $p < .01$, association level approached significance, $F(1,38) = 3.2$, $.10 < p < .05$, and there was no interaction between the two variables. The more spaced are the presentations, the more efficient is acquisition, a finding which is even more obvious in Table 1B where presentations to criterion is presented, rather than trials.

In Table 2 are presented (A) total errors to criterion based on first presentation only, (B) based on second presentation only, and (C) summed over both presentations. Analyses of variance on these statistics indicated that effects of spacing conditions were significant in all cases, (A) $F(4,152)=8.2$, $p < .01$, (B) $F(3,114)= 48.4$, $p < .001$, (C) $F(4,152)=8.3$, $p < .01$. Association level approached significance for statistic A, $F(1,38)=3.0$, $.10 < p < .05$. The F-ratio for association level by spacing was less than 1 for all statistics. Errors based on the second presentation reflect short-term retention. There are virtually no errors on the second presentation for zero interpolated pairs; as the number of interpolated pairs increases, errors on the second presentation increase in a negatively accelerated fashion. By the time three pairs intervene before the second presentation there is a difference of only one error, on the average, between performance on first and second presentations.

Errors to criterion for conditions 0 and N are virtually identical, while considerably more errors were required for the intermediate spacing conditions. The possibility was considered that short-term memory errors, i.e. those errors occurring on the second presentation, might be relatively ineffective in the learning process. However, the probability that a criterion run started on trial n+1 was equal to .093 in group HA if conditionalized on an error on the first presentation on trial n, and equal to .086 for an error on the second presentation; the corresponding values for group LA are .072 and .082. A criterion run is just about as likely to follow a second-presentation as a first-presentation error.

Table 1A

Mean Trial of Last Error to Criterion of Three Successive Correct Responses on First Presentation

| Group<br>Interpolated Item Condition | High Association | Low Association |
|---|---|---|
| 0 | 4.8 | 6.0 |
| 1 | 4.0 | 5.2 |
| 2 | 4.1 | 5.2 |
| 3 | 3.9 | 4.9 |
| N | 5.4 | 6.4 |

Table 1B

Total Presentations of Last Error to Criterion of Three Successive Correct Responses Starting with First Presentation on a Trial

| Group<br>Interpolated Item Condition | High Association | Low Association |
|---|---|---|
| 0 | 8.6 | 11.0 |
| 1 | 7.0 | 9.5 |
| 2 | 7.1 | 9.5 |
| 3 | 6.8 | 9.1 |
| N | 5.4 | 6.4 |

Table 2

Mean Total Errors on First, Second or Combined Presentations per Trial to Criterion of three Successive Correct Responses, with Predicted Values from Stimulus Fluctuation (SF) and Random-Acquisition-Forgetting model

| Group | | High Association | | | Low Association | | |
|---|---|---|---|---|---|---|---|
| Presentation | | 1st | 2nd | 1st&2nd | 1st | 2nd | 1st&2nd |
| Interpolated Item Condition | | | | | | | |
| 0 | Obs. | 4.3 | .4 | 4.7 | 5.3 | .5 | 5.8 |
| Pred. | SF | 4.7 | 0 | 4.7 | 5.6 | 0 | 5.6 |
| | RAF | 4.0 | 0 | 4.0 | 4.9 | 0 | 4.9 |
| 1 | Obs. | 3.6 | 1.7 | 5.4 | 4.6 | 2.2 | 6.8 |
| Pred. | SF | 3.4 | 1.3 | 4.7 | 4.0 | 1.6 | 5.6 |
| | RAF | 4.0 | 1.5 | 5.5 | 4.9 | 1.8 | 6.7 |
| 2 | Obs. | 3.7 | 2.4 | 6.1 | 4.7 | 2.8 | 7.5 |
| Pred. | SF | 3.0 | 1.7 | 4.7 | 3.6 | 2.1 | 5.7 |
| | RAF | 4.0 | 2.3 | 6.3 | 4.9 | 2.9 | 7.8 |
| 3 | Obs. | 3.5 | 2.5 | 6.0 | 4.3 | 3.1 | 7.4 |
| Pred. | SF | 2.9 | 1.8 | 4.7 | 3.3 | 2.4 | 5.7 |
| | RAF | 4.0 | 2.8 | 6.8 | 4.9 | 3.5 | 8.4 |
| N | Obs. | 4.7 | --- | 4.7 | 5.6 | --- | 5.6 |
| Pred. | SF | 4.7 | --- | 4.7 | 5.7 | --- | 5.7 |
| | RAF | 7.2 | --- | 7.2 | 9.3 | --- | 9.3 |

Note: Parameter values used for predictions for SF model were HA: $a^t$=.52, J=.20, LA: $a^t$=.52, J=.16; for RAF model, HA: c(1-$\alpha$)=.13, f=.44, LA: c(1-$\alpha$)=.10, f=.41.

## Experiment II

The general conclusion to be drawn from the first experiment would seem to be that spacing of items substantially enhances the effect of each feedback presentation. The second experiment was designed to evaluate distribution effects in paired-associate learning using a between-Ss design. It is occasionally the case in psychological research that significant effects of a given variable depend on within-S variation, and are not observed when varied between Ss. Results based on the first group of Ss run in this experiment in fact showed no effects of spacing, and the experiment was therefore replicated, with substantially the same results, viz., no effect.

### Method

Subjects. In the first replication, 45 undergraduates at the University of Wisconsin, Madison, served as Ss. In the second replication, 48 students volunteered from the same source. All Ss were enrolled in an introductory psychology course, and received experimental credits in return for their services. Within each replication, Ss were assigned at random to one of three experimental groups, with the restriction that there be an equal number of Ss in each group.

Procedure. Each S learned a 15-item paired-associate list by the anticipation technique. Stimulus members were the low association CVC's from Exp. I; response terms were the set of 2-digit numbers. Instructions and pretraining were identical to Exp. I, as was the general presentation procedure. There was no limit on response time, and there was no intertrial interval.

The training procedure consisted of 6 anticipation trials followed by a test trial (stimulus only), and then 6 more anticipation trials followed by a test trial. Subjects were then given a poem to read for about 2 min., and a final recall trial was administered. The three experimental groups differed in the relative massing of subsets of experimental items during the two 6-trial anticipation series. Group I served as a control group; the 15 items were presented in random order on each trial. In Group II, the list was divided into 3 sets of 5 items each. If these sets are labeled A, B, and C, then A comprised a control set, in that each of the 5 items was presented once on each trial. Items in sets B and C were presented in massed fashion; items in set B were presented twice on all odd-numbered anticipation trials, while items in set C were presented twice on even-numbered anticipation trials. In Group III, 6 items were selected for set A, and the remaining 9 items were divided into 3 sets (B, C, and D) of 3 items each. As before, items in A served as control items and were presented once per trial. Items in B, C, and D were presented three times per trial, but each set was used on only one-third of the trials. Set B items appeared on trials 1, 4, 10 and 13, set C items on 2, 5, 9, and 12, and D items on trials 3, 6, 8 and 11. (This assignment of trial numbers and sets, which provides partial counterbalancing of sets and order of presentation, was used only in the second replication. In the first replication, sets were run in the same order in both blocks of anticipation trials.) These assignment conditions achieve the following results: (1) each trial consists of 15 presentations; (2) each item is presented for 6 anticipation trials in each of two blocks; (3) a spaced control baseline is available both between and within conditions. From Group I through Group III, there is a marked increase in massing of items within the list. Thus, in Group I (and for control items within each other group), the mean number of interpolated pairs is 14; in Group II, the corresponding value for experimental items is 6.3, and in Group III it is about 4.2. In none of the massed conditions was an item repeated with 0 interpolated pairs.

## Results

The primary results can be stated succinctly. Performance on the two test trials during training and the final recall test showed no effect on massing either between- or within-subjects. Mean number of correct responses on each test is presented in Table 3. In the first replication, the differences between conditions, which show a tendency toward better performance in Group II, are not significant, $F(1,38)=1.2$, $p<.10$, and in the second replication, the means are almost identical at all test points. In Table 3 is presented proportion of correct responses for spaced and distributed items within Groups II and III. The difference between Groups II and III in the first replication approaches significance, $F(1,28)=3.9$, $.05<p<.10$, but this effect is apparently fortuitous, since it is not apparent in the second replication. The interaction between groups and spaced vs. massed items within groups in the second replication is significant $F(1,150)=8.0$, $p<.05$. Spaced items show somewhat better performance than distributed items in Group II, but this relation is reversed in Group III. The same trend is found in the first replication to a much lesser degree. However, there is no other difference, within groups, between spaced and massed items.

## Experiment III

This investigation is concerned with the amount of interference produced by learned versus unlearned items. In the anticipation technique, both input (study) and output (test) operations must be carried out by the subject, each of which has been shown to produce substantial interference (Tulving & Arbuckle, 1966).

Table 3

Proportion of correct responses on retention test trials for Experiment II

Replication 1 (N = 15 per group)

| Group | | Test Trial 7 | 14 | 2 Minutes Post-Acquisition |
|---|---|---|---|---|
| I | | .43 | .74 | .73 |
| II | Spaced | .52 | .84 | .83 |
| | Massed | .45 | .84 | .84 |
| III | Spaced | .35 | .68 | .68 |
| | Massed | .42 | .73 | .68 |

Replication 2 (N = 16 per group)

| Group | | Test Trial 7 | 14 | 2 Minutes Post-Acquisition |
|---|---|---|---|---|
| I | | .42 | .84 | .85 |
| II | Spaced | .54 | .86 | .86 |
| | Massed | .39 | .81 | .79 |
| III | Spaced | .40 | .88 | .88 |
| | Massed | .44 | .90 | .89 |

In the TDF model, it had been assumed that the largest part of interference-produced forgetting could be attributed to input of unlearned items. While this assumption is undoubtedly an oversimplification, it is of interest to determine the relative effects of learned versus unlearned items on acquisition of paired associates.

## Method

Subjects. A total of 24 students at the University of Wisconsin, Madison, served as Ss, and were assigned at random to one of two experimental conditions. Each S was paid $2.00 for his participation, which required from 35 to 80 min.

Procedure. A second set of 15 low association CVC's were selected from Archer's (1960) list, by the same criteria used to pick the original set of 15 items in the first set. The list of two-digit response terms was also expanded to 30 items. The expanded list contained no double digits (e.g. 33), multiples of 10, or digit reversals (e.g. 23 and 32). For groups of 4 Ss, two in each of the experimental groups, a list of 16 stimulus-response terms was selected at random from the pool of 30. Stimulus-response pairings were randomly determined, as was presentation order, for 6 blocks of 4 Ss each. Instructions, pretraining, and general procedure were the same as the preceding studies.

In Condition L (Learned), Ss were given a list consisting of 12 pairs from the basic 16-item list for 8 anticipation trials. There followed, without interruption, a series of trials on which 4 E (experimental) items, together with the 12 "learned" items, were presented until a criterion of 3 consecutive errorless trials was achieved for the 4 E items. Each E pair was presented twice per trial. Two of the pairs were repeated on each trial with 2 "learned" interpolated items, the other 2 pairs with 4 "learned" interpolated items.

In Condition U (Unlearned), Ss were given stimulus familiarization training on 12 stimulus terms by means of 8 free recall trials. Following this training, each S learned a 16-item list consisting of the 12 familiar items plus 4 E items, the latter repeated on each trial with either 2 or 4 "unlearned" interpolated pairs.

## Results

Mean trials to criterion for E pairs in Condition L was 3.5, while for Condition U this statistic was 5.5. This difference is statistically significant, $F(1,22)=5.4$, $p<.05$. Spacing had no effect, nor did any of the variables interact. Not all Ss had acquired the 12-item list by the end of the 8 training trials, and there is evidence that degree of associative performance on the prelearned list is directly reflected in rate of learning for the 4 experimental items. For example, in Condition L the rank-order correlation between number of errors on the last two training trials of the 12-item list and number of errors for E items on the second and third experimental trials was equal to .61 which differs significantly from 0 at the .05 level. That this correlation does not simply represent subject-selection factors (the better Ss might learn both lists more rapidly) is suggested by looking at the same rank-order correlation for Condition U, using number of items omitted on the last two free recall trials as the covariate, which yields a correlation of .11. Thus, it seems reasonable to state that the more completely learned were the items in the 12-item list, the less interference was produced and the more rapidly the performance criterion was achieved.

## Discussion

The results of the first experiment extend the finding of others (Greeno, 1964; Izawa, 1966), that acquisition efficiency, measured by trials to criterion, is improved by increasing the spacing between successive reinforcements of a paired-associate item. A number of different theoretical ideas can be put forward to account for this result.

The stimulus fluctuation model of Estes (Estes, 1955a, 1955b; Izawa, 1966) which was developed to handle distributional phenomena in verbal learning, makes the general prediction that under spaced conditions, fewer presentations should be required to reach criterion. The principle assumptions of the model are:

(1) The stimulus member of each item comprises a set of $N^*$ cues or stimulus elements, of which a subset of size $N$ is available for sampling on each trial.
(2) Individual elements move from the available to the unavailable subsets (and vice versa) over time, the probability of a transition occurring over any small time period $\Delta t$, being a constant value.
(3) On each trial, the entire set of available elements is sampled, and with probability $\underline{c}$, those elements in the sample not already associated to the correct response become conditioned.
(4) Each element is either conditioned to the correct response, or is unconditioned. Items remain in the conditioned state once there. The probability of a correct response during a test is equal to the proportion of conditioned elements in the sample.

A number of theoretical predictions can be derived from these assumptions which are applicable to the data from the first study in particular. In deriving these predictions, we will follow the notation introduced by Estes:

$q_n$ = probability of an incorrect response before the nth reinforced trial

$F_n$ = probability that a given cue is conditioned after the nth reinforced trial

J = $\bar{N}/N^*$ or the proportion of elements in the available sampling set

$a^t$ = probability of an interchange of cues between the available and unavailable subsets during a single presentation interval.

Two theorems, proven by Estes, are also useful. First, the probability that a cue in the unavailable set during the presentation of an item is available k presentations later is $J(1-a^{kt})$. Second, the probability that a cue available during a presentation, is also available k presentations later, is $J+(1-J)a^{kt}$. Finally, it will be assumed for simplicity that the conditioning parameter, c, is equal to 1.

From these assumptions, it can be shown that the underlying conditioning process takes the following form for Experiment I:

$$
\begin{aligned}
F_0 &= g \\
F_1 &= g+(1-g)J=(1-J)g+J \\
F_2 &= F_1+(1-F_1)J(1-a^{kt}) \\
&= 1-(1-g)[1-J(1-a^{kt})] \\
F_3 &= F_2+(1-F_2)J(1-a^{(25-k)t}) \\
&= 1-(1-g)(1-J)[1-J(1-a^{kt})][1-J(1-a^{(25-k)t})]
\end{aligned}
\qquad [1]
$$

where k is the number of interpolated items between the two presentations on a trial. The change from one reinforcement to the next follows this pattern: $F_n$, the number of conditioned cues following the nth reinforcement, is equal to $F_{n-1}$, plus that proportion of the unconditioned cues on n-1 which were unavailable on n-1, but became available on n, viz, a proportion equal to $J(1-a^{mt})$, where m is the number of other interpolated items between n-1 and n. The general form of $F_n$ on even presentations, which is the value of $F_n$ following the second presentation on a trial for repeated items, is

$$
\begin{aligned}
F_0 &= g \\
F_n &= 1-(1-g)(1-J)[1-J(1-a^{kt})]^{\frac{n-2}{2}-1}[1-J(1-a^{(25-k)t})]^{\frac{n-2}{2}}, \\
&\qquad n = 2, 4, 6\ldots
\end{aligned}
\qquad [2]
$$

The form of the equation for odd presentations, i.e., following the first presentation on a trial, is

$$F_n = 1-(1-g)(1-J)\{[1-J(1-a^{kt})][1-J(1-a^{(25-k)t})]\}^{\frac{n-1}{2}}, \quad n = 1, 3, 5... \qquad [3]$$

The probability of an error on the test trial immediately preceding the nth reinforcement is equal to the probability that each individual element is in the unconditioned state, times the probability that the each element which was in the available sample on the preceding feedback event (and hence was in the conditioned state) has moved to the unavailable state during the intervening time period (cf. Izawa, 1966, p. 912 for a more detailed derivation). Hence

$$q_n = (1-F_n)(1-a^{mt}) \qquad [4]$$

where m is the number of interpolated items since the last study item. For even values of n, i.e., for the first test presentation on each trial, m is equal to 0, 1, 2 and 3 for the various groups of items. For odd values of n, m is equal to 25, 24, 23 and 22 for the same respective groups. By summing the error probabilities over all trials, a theoretical expression for total errors to criterion on the first and second presentations on each trial for repeated items can be obtained:

$$T_1 = (1-g)+\frac{(1-g)(1-J)(1-a^{(25-k)t})[1-J(1-a^{kt})]}{1-[1-J(1-a^{kt})][1-J(1-a^{(25-k)t})]}$$

$$T_2 = \frac{(1-g)(1-J)(1-a^{kt})}{1-[1-J(1-a^{kt})][1-J(1-a^{(25-k)t})]} \qquad [5]$$

Finally by summing $T_1$ and $T_2$, it can be shown that overall total errors to criterion, T, should be constant over all spacing conditions,

$$T = T_1+T_2 = \frac{(1-g)}{J} \qquad [6]$$

The effect of increasing the number of interpolated items is to increase the errors on the second presentation on a trial, but there should be a corresponding decrease in the number of errors on the first presentation. In fact, total errors in the N and 0 item groups were equal, as predicted. However, the number of total errors in the 1, 2 and 3 item groups were significantly greater than the other two groups. Estimates were obtained by a least-squares procedure for the Low and High association lists. It was found that the parameter, $a^t$, was constant over the two association levels at a value of .52. The parameter J was equal to .16 in the Low group, and to .20 in the High group. The fact that material of high association value has a higher proportion of cues available for sampling could arise either because more cues are available on each trial from a base set which is invariant in number over association levels, or because with meaningful material, a smaller base set of cues is necessary for adequate performance, while the number of cues in the available sample is constant over association levels. Without a much more detailed analysis, which is not appropriate in view of the failure of the basic prediction of constant total errors, it is not possible to choose between these two alternative interpretations.

These parameter values were used to predict the eighteen observed statistics in Table 2. As may be seen, the fit is not very good for either first or second presentations in item groups 1, 2 and 3, reflecting the fact that the prediction of constant total errors is not met by the data. The error in excess of the predictions are shared about equally by the first and second presentations.

A second theoretical analysis, which rests on a two-component memory (i.e., postulation of separate long-term and short-term memory processes), assumes that for a transition into long-term memory storage to occur, an item must have been forgotten since the previous reinforcement. Greeno (1966) presents a Markov model for verbal learning in which on each presentation, an item can be in one of three states, L (long-term memory), S (short-term memory) and F (forgotten). The general form of the model is described by the transition matrix and column vector of response probabilities below,

$$\begin{array}{ll} & \text{State on trial } \underline{n}+1 \\ \text{State on trial } \underline{n} & \begin{array}{c|ccc|} & L & S & F \\ L & 1 & 0 & 0 \\ S & c & (1-c)h & (1-c)(1-h) \\ F & d & (1-d)h & (1-d)(1-h) \end{array} \end{array} \qquad \begin{array}{c} \Pr(\text{Correct} \mid \text{row state}) \\ \begin{bmatrix} 1 \\ 1 \\ g \end{bmatrix} \end{array} \qquad [7]$$

As mentioned earlier, Greeno presents evidence from several experiments in support of the statement that transition to L occurs only from F, i.e., $c = 0$. In order to apply the model with this restriction ($c = 0$) to the present study, let $h = 1-(1-x^m)$, where m is the number of interpolated items since the last reinforcement. (That is, h is a decreasing exponential function of m, such that following zero interpolated items, a pair remains in S with probability 1, while after a large number of items, a pair is very likely to have been forgotten.) On the first presentation of each trial, $h_1 = 1-(1-x^{25-k})$, which will differ negligibly from 0 in the present study (e.g., consider Peterson & Peterson, 1959). Following the second presentation, $h_2 = 1-(1-x^k)$, where $k$ is the number of interpolated items. If the state probabilities are represented by lower case letters subscripted with the presentation numbers, then $p_n$, the probability of a correct response on the $n$th presentation, and the state probabilities are described by the following equations,

$$
\begin{aligned}
p_n &= \ell_n + s_n + gf_n \\
\ell_n &= \ell_{n-1} + df_{n-1} \\
s_n &= h_i s_{n-1} + h_i(1-d)f_{n-1} \\
f_n &= (1-h_i)s_{n-1} + (1-d)(1-h_i)f_{n-1}
\end{aligned}
\qquad [8]
$$

where i is 1 on even presentations and 2 on odd presentations. It can be shown by induction that for even values of n, (on the first test presentation of a trial)

$$
\begin{aligned}
f_n &= \left\{(1-d)\ [1-d(1-h_2)]\right\}^{n-1} \\
s_n &= 0
\end{aligned}
\qquad [9]
$$

For odd values of n, the state probabilities are

$$
\begin{aligned}
f_n &= (1-h_2)(1-d)\left\{(1-d)\ [1-d(1-h_2)]\right\}^{n-1} \\
s_n &= h_2(1-d)\left\{(1-d)[1-d(1-h_2)]\right\}^{n-1}
\end{aligned}
\qquad [10]
$$

These equations can be used to derive the expected total errors on the first and second presentations. Except for labeling of the parameters, the expressions are identical to the stimulus fluctuation model, and by inspection of the transition matrix, it can be seen that prior to entering L, the process will have to be in F for $1/d$ trials, on each of which an error is 1-g, so that total number of errors over the first and second presentations is again predicted to be constant over all spacing conditions at

$$T = \frac{1-g}{d}\ . \qquad [11]$$

Another model, which involves incremental acquisition and forgetting processes, is described in Section 3. This latter model, which was developed to handle list-length variation in paired-associate learning, does not predict constancy of total errors, but instead comes close to predicting the observed pattern of total errors. The model is not without its own problems, however, and these will be considered in some detail in Section III.

To anticipate one aspect of that discussion, a problem which arises in studies of the effects of spaced presentations on paired-associate learning is that, when test trials are given, with increased spacing there is a concomitant increase in the rate of errors on tests. It may well be that the deleterious effects of massing arise because the subject is not receiving appropriate feedback about the adequacy of his storage strategies with massed presentations. I.e., errors may provide information to the subject that he hasn't learned particular S-R pairs and needs to do additional encoding. Evidence on the role of active responding (and, in particular, of errors) on acquisition of paired associates is sparse. It has been shown that an uninterrupted sequence of study trials, during which the subject makes no responses, and hence receives no information about the effectiveness of his storage operations, is a relatively inefficient technique (Izawa & Estes, 1966). Data from a study now in progress (Watters, 1966) in which the subject times his own feedback interval under an anticipation procedure show a big difference in intervals following precriterion successes and errors, the latter intervals being much longer. These findings suggest the importance of further investigation of the role of the response in learning verbal associates. For the present, theoretical models based on spacing effects, short-term memory effects, or the idea that effective feedback is more likely to be associated with errors than correct responses, cannot be differentiated from one another on the basis of available data.

The failure to find any significant difference in Experiment II as a function of massing of items either within or between subjects seems surprising in view of the results of the initial study. However, the greatest decrement in efficiency due to massing is seen with zero interpolated pairs. Even in Group III of Experiment I, where massed items are presented three times per trial, the random ordering was constrained so that an item was never repeated twice in immediate succession. The number of presentations per trial was also shorter in Experiment II (15 as compared to 27 in Experiment I). Even taking these factors into account, the result is unexpected. Based on the low association data (presentations to criterion) from Experiment I together with the average number of interpolated items separating successive presentations of massed and spaced items in the second study, one would predict that the massed items would require about 8.8 presentations to reach criterion, compared with about 7.4 for spaced items. At none of the test points for either replication is there any indication of superior performance for the spaced items. At the same time, the short-term retention effects of massing can be observed by a decline during <u>training trials</u> in the probability of an error over successive repetitions within a trial. Averaged over all trials, the probabilities are .65 and .55 for Group II, .62, .44 and .29 for Group III. These values may be compared with the corresponding presentations of spaced items, .57 and .51 for Group II, .69, .59 and .54 for Group III: Comparison of these two sets of values shows a much greater decline over pairs (or triplets) of presentations for the massed items versus the spaced items. Apparently, as long as variation in the degree of spacing of items does not exceed certain bounds, the subject is able to allocate his associative efforts in such a fashion as to learn all items with about equal efficiency. A complete specification of these boundaries is not possible from these data, but one significant limit is immediate retest and reinforcement of a previously presented item. Such an operation is simply a waste of time.

The results of the final study show that a major source of interference in verbal learning is produced by unlearned items in a list, and, by implication, results from the active processing and storage of information by the subject. Using a short-term retention design, Tulving and Arbuckle have shown that significantly greater interference results from input or study presentations as compared with output or test presentations. Posner and Rossman (1965) have also found that as the amount of information processing required during an interpolated task is increased, higher rates of retention loss occur in short-term memory. The design used in Experiment III was such that not all the items in the prelearned list were well-acquired prior to training on the experimental list. In other words, some of the "learned" items had not in fact reached criterion. Four subjects in Group L had reached a criterion of one perfect recitation of the first list. Trials to criterion for the experimental items for these subjects were 2, 3, 4 and 4.5, which don't seem unreasonable values for learning a 4-item list. The four best subjects on the preliminary recall task in Group U required 4, 7, 9 and 14 trials on the experimental items.

An alternative interpretation, to be contrasted with interference-produced forgetting, is that unlearned items are identified by the subject and "selected" for learning. Thus, on any trial the subject's attention is limited to a relatively few items. In Group L, the selected items are E items, whereas in Group U, any of the items may be chosen. Battig (1966) has suggested that subjects can identify and respond in a differential fashion to learned items in a paired-associate list. In the present experiment it might be reasonable to suppose that subjects isolate unlearned items.

## Section III: Incremental acquisition and forgetting processes in paired associate learning

In studies of verbal learning, acquisition rate typically decreases disproportionately with total amount of material (McGeoch & Irion, 1952, pp. 487-496). For example, mean trials to criterion is greater as the number of S-R items in a paired-associate list is increased (Carroll & Burke, 1965; Runquist, 1965, 1966). This decrement in learning rate has been attributed (Atkinson & Crothers 1964; Calfee & Atkinson, 1965) to the additional interference which naturally occurs with increased list length. As represented in the trial-dependent-forgetting (TDF) model, each reinforced presentation of an S-R pair has the effect of moving that pair either to the long-term (L) or short-term (S) memory states of a Markov process. For items in either of these states, the correct response occurs with probability 1. Every time some other pair is presented, as a consequence of the interference produced, some items in short-term memory may be forgotten, which is represented in the model as a return to an unlearned (U) state. The overall interference increases with longer lists, leading to the poorer performance which is usually found.

In a study in which paired-associate lists of three different lengths were used, the TDF model in a slightly revised form gave an excellent account of the data, compared with several alternative formulations. Nonetheless, in two respects, this model is less than satisfactory. First, the estimated parameter values imply that following a reinforced presentation learning (i.e., transition to state L) is far more probable if an item is in state S than state U prior to the feedback presentation. An optimum schedule would therefore involve massed presentations of each item in the list, contrary to results obtained by Greeno (1964) and Calfee (1966b). Secondly, state L is an absorbing state in which probability of a correct response is unity. The model thus predicts that if one list is learned to a strict criterion, and then a second list is learned, subsequent test performance on the first list should show no retention loss. In fact, considerable forgetting is usually observed under these conditions. One might argue, of course, that such a procedure exceeds the boundary conditions under which an absorbing-state model can be expected to hold up.

The TDF model is a special case of a class of three-and four-state Markov models for verbal learning in which the assumption is made that associations are formed by transition through two or three discrete stages, in each of which probability of a correct response takes on a constant value (Atkinson & Crothers, 1964; Bernbach, 1965, Greeno, 1964; Restle, 1964). An alternative representation of the learning process is the random-trial-incremental (RTI) model of Norman (1964), in which each pairing of stimulus and response is effective with probability $\underline{c}$, and is otherwise ineffective. Effective feedback events produce a linear increment in the probability of a correct response, while ineffective events leave the probability unchanged. The probability of a correct response following the nth reinforcement, $p_n$, is therefore a linear function of $p_{n-1}$,

$$p_n = A(p_{n-1}) = c(\alpha p_{n-1}+1-\alpha) + (1-c)p_{n-1}$$
$$= [1-c(1-\alpha)]p_{n-1}+c(1-\alpha) \qquad [12]$$

where A is the acquisition operator. This model was compared to various alternatives, including both Markovian and linear-operator processes (Atkinson & Crothers, 1964), and proved more adequate than other linear models, but generally yielded poorer predictions than the long-short Markov model which was presented.

The RTI model may also be arrived at in terms of a dual-process memory. Suppose that a stimulus-response pair can be represented in memory by a set of distinctive features. The probability of a correct response depends on the number of features which have been stored in long-term memory and are available for reference when an item is presented. On some feedback presentations, no effort is made by the subject to transfer features to permanent storage, perhaps because of attentional factors. On other presentations, the subject affects a transfer of some number of features; the number of features transferred is assumed to be proportional to the number remaining to be transferred, which implies a linear operator if the total number of features is very large.

In its original form, the RTI model cannot handle list length effects without additional parameters for each list. By suitable modification to take into account the effects of interference produced forgetting, the model can be extended to account for variations in list length, and its performance relative to Markov models is also substantially improved. Further, the revised model can make predictions about overlearning and distribution effects, as well as retention losses due to learning of interpolated lists, which are generally consonant with experimental findings.

The basic modification proposed is to assume that when a particular stimulus-response pairing takes place, the entire feature list for that pair is stored in short-term memory. With probability c, the S also engages in an encoding operation, in which some of the features are transferred to long-term memory. Following the feedback interval, there is a gradual loss of information stored in short-term memory about this particular pair, because of the limited capacity of short-term memory and the need to process the other items being presented. This loss is also assumed to take the form of a linear operator, which is a function of the number of interpolated items, where the limit of the forgetting operator is determined by the acquisition operator. Specifically, the probability of an error for a particular item on trial n following m interpolated items, and

$\underline{k}$ effective feedback events, where $\underline{f}$ is the forgetting parameter, is

$$q_{k,m,n} = F(q_{k,m-1,n}) = q_{k,m-1,n} + f(\alpha k_{q,} - q_{k,m-1,n})$$
$$= \alpha k_{q,} \; [1-(1-f)^{m-1}] \qquad [13]$$

The process is presented graphically in Fig. 1.

Although the process may appear somewhat involved, derivation of various statistics is remarkably straightforward. This presentation will be somewhat abbreviated; for more detail on the RTI model, see either Norman (1964), or the treatment in Atkinson, Bower, and Crothers (1965). Consider a particular item for which there has been $\underline{k}$ effective reinforcements through trial $\underline{n}$-1, and let $\underline{S}_{n-1}$ be the indicator random variable for the number of effective reinforcements on trial $\underline{n}$-1. Then it can be shown that $\underline{S}_{n-1}$ has the binomial distribution with parameter $\underline{c}$,

$$P(S_{n-1} = k) = \binom{n-1}{k} c^k (1-c)^{n-k-1} \qquad [14]$$

The $\underline{r}$th raw moment of the distribution of response probabilities on trial $\underline{n}$, $\underline{V}_{r,n}$, is therefore

$$V_{r,n} = \left(\frac{1}{X}\right)^2 \sum_{i=0}^{X-1} \sum_{j=0}^{X-1} \sum_{k=0}^{n-1} (q_{k,(i+j),n})^r \; P(S_{n-1} = k) \qquad [15]$$

The term $\left(\frac{1}{X}\right)^2$ and the sums on $\underline{i}$ and $\underline{j}$ take into account the possible distribution of interpolated pairs between the presentation of an item on trials $\underline{n}$-1 and $\underline{n}$, respectively (cf. Calfee & Atkinson, 1965, for additional detail). There can be between 0 and X-1 other pairs following the presentation on trial $\underline{n}$-1, and each of these events has probability $\frac{1}{X}$. A similar analysis holds for the number of interpolated pairs preceding the presentation on trial $\underline{n}$. The sum on $\underline{k}$ takes into account the number of effective reinforcements.

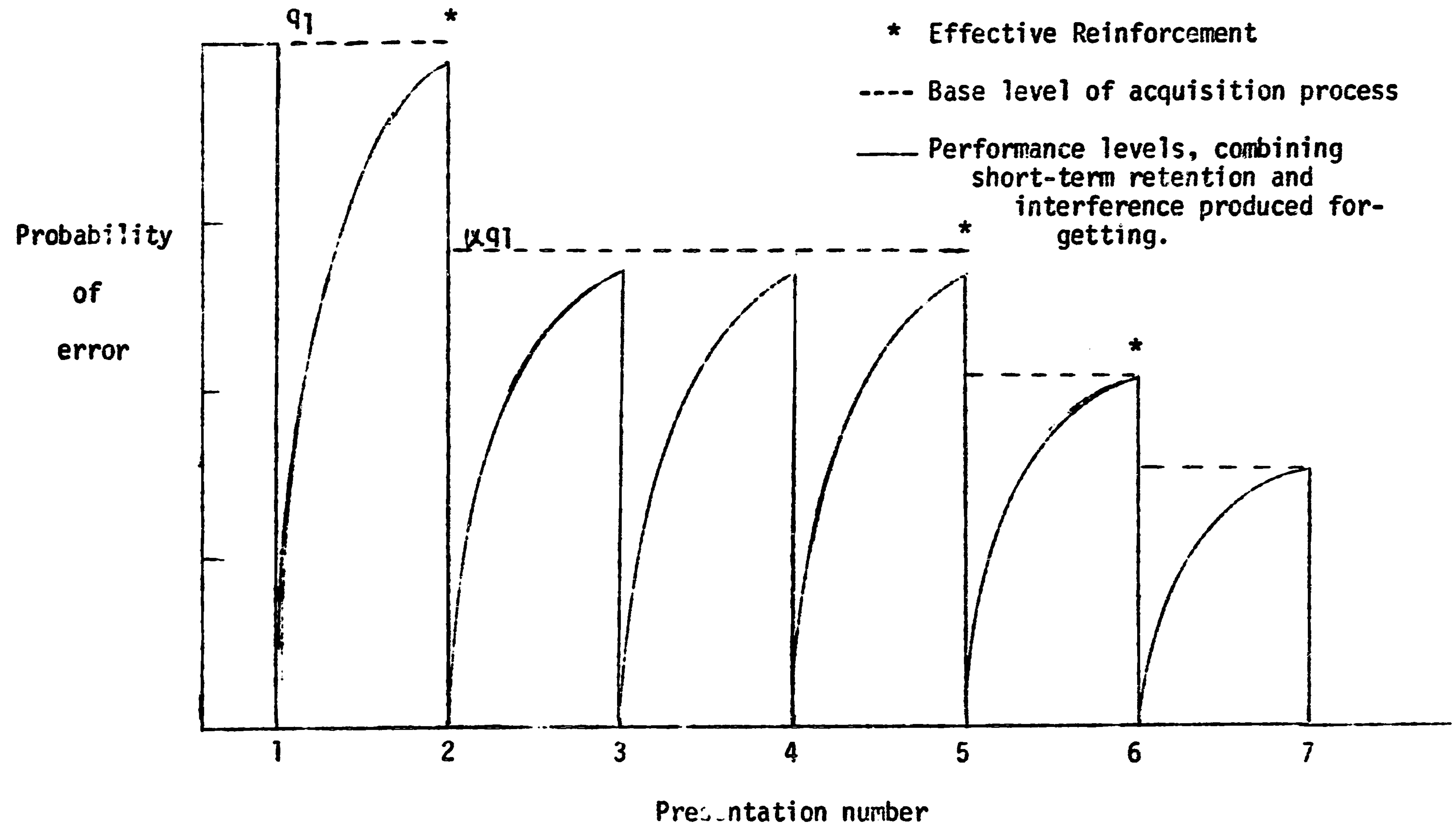
Making the appropriate substitutions in Eq. 15.

$$V_{r,n} = \left(\frac{1}{X}\right)^2 \sum_i \sum_j \sum_k \left\{\alpha k_{q,} \; [1-(1-f)^{i+j}]\right\}^r \binom{n-1}{k} c^k (1-c)^{n-k-1}$$
$$= \left(\frac{1}{X}\right)^2 q_1^{\,r} \; [1-c(1-x^r)] \sum_i \sum_j \left\{1-(1-f)^{i+j}\right\}^r \qquad [16]$$

In the case of the first raw moment or mean learning curve, th expression reduces to

$$V_{1,n} = q_1 \; [1-c(1-\alpha)] \left\{1-\left[\frac{1-(1-f)^{X-1}}{Xf}\right]^2\right\} \qquad [17]$$

Fig. 1. Graphical representation of random-trials-acquisition-and-forgetting model.



For very long lists, the function approaches the RTI model in the limit. I.e., the short-term memory is of no help in improving performance, because the probability that any information remains in short-term memory between one presentation and the next is vanishingly small.

This model, which will be referred to as the RAF model (random-trials-acquisition and forgetting), has several properties in common with the TDF model. Acquisition due to feedback and forgetting associated with interpolated items are both explicitly accounted for. The model is semi-Markovian, in that if the proportion of features stored during effective feedback events is reasonably large, then there will be a small number of states which can be differentiated on the basis of response probability. Thus, if $\alpha$ is fairly small, e.g. less than .25, then after two or three effective reinforcements, the probability of a correct response is so close to 1 that subsequent feedback events may produce no measurable change in performance, and the subject may appear for all intents to have reached an absorbing state. Performance may exceed learning to a substantial degree with short lists, where there is little interference-produced forgetting between successive presentations. However, the effects of feedback presentations are not washed out by subsequent interference; instead, performance declines to the level of the basic acquisition process. For example, suppose a subject is given five trials on a 4-item list, by which point performance should be almost perfect. Then five trials are given on a second 4-item list, followed by a test trial on the first list items. According to the RAF model, during the second-list learning, probability of a correct response on the first list items will decline to $1-q_1\,[1-c(1-\alpha)]^5$. In the TDF model, on the other hand, first-list items in state L will remain there during the learning of the second list, while any item not in state L will return to state U, or a chance level of performance. Finally, the RAF model makes the rather interesting prediction that the effects of variation in list length should appear only during acquisition performance. If a list of length X is presented for k reinforced trials, and a suitable delay involving some kind of rehearsal-preventing activity is interpolated, then on a test following the delay, the proportion of correct responses is predicted to be a constant, invariant with changes in X, and equal to $1-q_1\,[1-c(1-\alpha)]^k$.

The relative adequacy of the TDF and RAF models was determined initially by comparing their performance on two sets of paired-associate data from experiments in which list length was varied. Experiment SH is reported in detail in Calfee and Atkinson (1965). Briefly, each subject learned a list of 9, 15 or 21 pairs, in which the stimulus members consisted of two-digit numbers, and the response terms were one of the three nonsense syllables, RIX, FUB, or GED. The list was presented in a standard anticipation mode for a total of ten trials with no breaks between trials.

Experiment RH involved more difficult stimulus materials, four rather than three response terms, and list lengths of 8, 16, 24 or 32 items. Otherwise, the basic procedure is quite similar between the two studies. The stimulus items in the second study consisted of 32 four-letter Russian nonsense words. Each of the 31 letters in the Russian alphabet was used with approximately equal frequency, and appeared no more than once in a given word. The basic list of 32 items was divided into four sublists of 8 items each so as to yield minimal intralist similarity in the judgment of the experimenters. Within each sublist, the first and last letters are different for each item in the list. The responses consisted of the digits 1 through 4. The subject responded by marking the appropriate column on an IBM mark-sense form. Two random assignments of each response term to two stimulus items within a sublist were prepared. A total of 16 subjects were run under each list length condition, either 8, 16, 24 or 32 pairs. Each sublist was used equally often for every list length, as were the two random stimulus-response pairings. The number of trials was fixed within each condition, an attempt being made to chose a number which would reduce error rate on the last trial to about .25. The number of trials for 8, 16, 24 and 32 pairs was 9, 12, 14 and 15, respectively. Subjects were given instructions about the task and a brief pretraining list to familiarize them with the procedure. After brief rest, the experimental list was presented. Stimuli were printed on 3x5 cards. The first four items were dummy pairs, which did not appear thereafter, and were included to eliminate any primacy effects. Responses were self-paced, in that subjects were given as long as necessary to record their answer. After the response, the correct answer was then pronounced by the experimenter, and three seconds later, the next stimulus card was presented. The stack of cards was prepared beforehand, and kept behind a shield, so that the subject could not see how much longer he had to go.

The mean learning curves for Experiment II are presented in Table 4. There is a substantial effect apparent as list length increases from 8 to 24 pairs; the further increment to 32 pairs produces little additional change in the performance curve. The hypothesis of precriterion stationarity based on Vincent quartile (Suppes & Ginsberg, 1963) can be rejected at better than the .001 level for all lists. As may be seen from the quartile data which are presented in Table 5, the functions show a tendency to flatten out at the shorter list lengths.

The data used in parameter estimation and primary evaluation of the models were 4-tuple sequences of successes and errors for individual stimulus-response pairs on trials 2 through 5 and 6 through 9. The 16 possible combinations of successes (s) and errors (e) for each of these four-trial blocks are listed in Table 6 for the three groups in Experiment SH, and Table 7 for the four groups in Experiment RH.

Table 4

Probability of an incorrect response, trial by trial, Experiment RH

| | List Length | | | |
|---|---|---|---|---|
| Trial | 8 | 16 | 24 | 32 |
| 1 | .66 | .70 | .76 | .78 |
| 2 | .61 | .72 | .73 | .72 |
| 3 | .56 | .65 | .73 | .72 |
| 4 | .48 | .65 | .66 | .67 |
| 5 | .44 | .54 | .67 | .66 |
| 6 | .35 | .54 | .64 | .64 |
| 7 | .30 | .48 | .58 | .56 |
| 8 | .22 | .46 | .56 | .56 |
| 9 | .16 | .40 | .53 | .49 |
| 10 | | .35 | .47 | .46 |
| 11 | | .35 | .39 | .41 |
| 12 | | .31 | .37 | .36 |
| 13 | | | .36 | .37 |
| 14 | | | .33 | .35 |
| 15 | | | | .31 |

Table 5

Probability of an incorrect response precriterion in Vincent quartiles, Experiment RH

| Vincent Quartile | List Length 8 | 16 | 24 | 32 |
|---|---|---|---|---|
| 1 | .67 | .74 | .76 | .77 |
| 2 | .57 | .71 | .73 | .73 |
| 3 | .64 | .69 | .72 | .71 |
| 4 | .62 | .66 | .66 | .65 |

Table 6A

Observed and predicted frequencies for 4-trial response sequences, trials 2-5, Experiment SH

| | 9 Items | | | | 15 Items | | | | 21 Items | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trials 2 - 5 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 |
| cccc | 83 | 77.2 | 66.1 | 67.7 | 98 | 90.7 | 92.8 | 88.2 | 97 | 107.5 | 123.8 | 110.0 |
| ccce | 3 | 4.2 | 7.1 | 4.5 | 10 | 6.7 | 8.3 | 7.4 | 11 | 9.0 | 10.4 | 11.6 |
| ccec | 10 | 8.0 | 9.9 | 8.4 | 13 | 11.1 | 12.1 | 11.8 | 14 | 13.7 | 15.2 | 16.8 |
| ccee | 4 | 3.7 | 6.0 | 3.1 | 10 | 9.2 | 8.9 | 7.4 | 12 | 14.5 | 11.6 | 11.4 |
| cecc | 18 | 17.2 | 16.6 | 17.1 | 25 | 22.7 | 22.7 | 22.7 | 35 | 27.3 | 29.5 | 30.7 |
| cece | 2 | 4.4 | 6.2 | 4.3 | 4 | 9.9 | 9.4 | 8.6 | 14 | 15.1 | 12.3 | 12.8 |
| ceec | 10 | 8.5 | 9.1 | 8.2 | 7 | 16.2 | 14.6 | 14.8 | 17 | 23.3 | 19.7 | 20.5 |
| ceee | 3 | 3.9 | 6.6 | 3.4 | 12 | 13.6 | 14.5 | 12.5 | 20 | 24.5 | 21.4 | 21.1 |
| eccc | 40 | 39.5 | 33.0 | 41.1 | 58 | 54.6 | 54.4 | 56.6 | 78 | 67.6 | 75.9 | 73.7 |
| ecce | 3 | 4.9 | 6.8 | 5.8 | 6 | 10.5 | 10.6 | 10.5 | 15 | 15.6 | 14.2 | 15.6 |
| ecec | 12 | 9.4 | 9.7 | 11.1 | 16 | 17.4 | 16.2 | 17.6 | 22 | 24.0 | 22.1 | 24.2 |
| ecee | 2 | 4.4 | 6.6 | 4.4 | 12 | 14.3 | 14.5 | 13.5 | 30 | 25.3 | 21.6 | 21.8 |
| eecc | 14 | 20.2 | 16.9 | 23.1 | 31 | 35.4 | 32.7 | 36.7 | 47 | 47.6 | 47.4 | 49.2 |
| eece | 2 | 5.1 | 6.9 | 6.1 | 11 | 15.5 | 15.3 | 15.9 | 16 | 26.5 | 22.9 | 24.5 |
| eeec | 13 | 9.9 | 10.1 | 11.8 | 32 | 25.7 | 24.0 | 27.4 | 42 | 40.6 | 36.8 | 39.7 |
| eeee | 6 | 4.6 | 7.4 | 4.9 | 30 | 21.2 | 24.0 | 23.3 | 55 | 42.8 | 40.4 | 41.6 |

Table 6B

Observed and predicted frequencies for 4-trial response sequences, trials 6-9, Experiment SH

| | 9 Items | | | | 15 Items | | | | 21 Items | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trials 6 - 9 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 |
| cccc | 205 | 197.2 | 165.0 | 176.9 | 271 | 260.3 | 257.7 | 255.5 | 319 | 317.1 | 351.0 | 329.1 |
| ccce | 0 | 1.1 | 4.9 | 1.8 | 6 | 3.3 | 7.2 | 5.4 | 8 | 5.2 | 9.3 | 10.9 |
| ccec | 0 | 2.6 | 6.6 | 4.1 | 8 | 6.6 | 9.8 | 9.1 | 13 | 9.2 | 12.7 | 15.8 |
| ccee | 0 | .3 | 2.0 | .0 | 2 | 2.6 | 3.7 | 2.1 | 4 | 6.1 | 4.9 | 4.0 |
| cecc | 12 | 6.4 | 9.6 | 8.9 | 13 | 14.4 | 15.2 | 16.1 | 27 | 19.2 | 20.2 | 24.7 |
| cece | 0 | .5 | 2.1 | .7 | 1 | 3.1 | 3.8 | 2.8 | 6 | 6.8 | 5.1 | 4.9 |
| ceec | 1 | 1.2 | 2.9 | 1.6 | 2 | 6.2 | 5.7 | 5.4 | 11 | 12.1 | 8.0 | 8.3 |
| ceee | 0 | .2 | 1.3 | .0 | 5 | 2.4 | 4.0 | 2.4 | 10 | 8.0 | 6.8 | 5.7 |
| eccc | 13 | 15.4 | 15.3 | 19.4 | 24 | 33.7 | 27.5 | 32.7 | 55 | 45.8 | 39.5 | 46.3 |
| ecce | 0 | .6 | 2.2 | 1.0 | 2 | 3.6 | 4.2 | 3.7 | 10 | 7.5 | 5.8 | 6.2 |
| ecec | 0 | 1.5 | 3.1 | 2.5 | 11 | 7.2 | 6.2 | 7.0 | 5 | 13.2 | 8.8 | 10.2 |
| ecee | 0 | .2 | 1.3 | .0 | 1 | 2.8 | 4.0 | 2.9 | 3 | 8.8 | 6.8 | 6.3 |
| eecc | 1 | 3.7 | 4.7 | 5.6 | 15 | 15.8 | 11.0 | 14.6 | 17 | 27.4 | 17.4 | 20.6 |
| eece | 0 | .3 | 1.4 | .5 | 5 | 3.4 | 4.2 | 4.0 | 7 | 9.8 | 7.2 | 7.7 |
| eeec | 0 | .7 | 1.9 | 1.2 | 5 | 6.8 | 6.3 | 7.8 | 11 | 17.3 | 11.4 | 13.7 |
| eeee | 0 | 0.1 | .9 | .0 | 4 | 2.7 | 4.5 | 3.7 | 19 | 11.5 | 10.2 | 10.6 |

Table 7A

Observed and predicted frequencies for 4-trial response sequences, trials 2-5, Experiment RH

| | 8 Items | | | | 16 Items | | | | 24 Items | | | | 32 Items | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trials 2 - 5 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 |
| cccc | 17 | 18.2 | 10.0 | 11.6 | 19 | 16.5 | 13.7 | 14.6 | 14 | 18.8 | 19.7 | 17.7 | 20 | 22.3 | 26.2 | 19.7 |
| ccce | 3 | 2.0 | 4.1 | 3.1 | 4 | 3.1 | 4.1 | 3.9 | 6 | 4.3 | 5.5 | 6.2 | 9 | 5.6 | 7.3 | 8.7 |
| ccec | 3 | 2.9 | 4.6 | 3.9 | 3 | 3.7 | 4.7 | 4.6 | 6 | 4.9 | 6.3 | 7.1 | 8 | 6.3 | 8.3 | 9.8 |
| ccee | 6 | 3.8 | 5.8 | 4.4 | 8 | 8.2 | 8.1 | 7.5 | 10 | 12.3 | 11.3 | 11.2 | 16 | 16.3 | 14.9 | 15.3 |
| cecc | 6 | 5.5 | 5.8 | 5.6 | 5 | 5.9 | 6.6 | 6.7 | 19 | 7.3 | 9.1 | 9.8 | 12 | 9.1 | 12.0 | 13.0 |
| cece | 4 | 4.0 | 5.9 | 4.9 | 3 | 8.3 | 8.3 | 7.8 | 5 | 12.3 | 11.6 | 11.7 | 19 | 16.4 | 15.3 | 16.1 |
| ceec | 5 | 5.8 | 6.6 | 6.2 | 9 | 9.9 | 9.7 | 9.4 | 15 | 14.1 | 13.7 | 13.9 | 23 | 18.4 | 18.1 | 18.7 |
| ceee | 6 | 7.6 | 9.4 | 7.5 | 20 | 22.0 | 21.8 | 21.0 | 27 | 35.0 | 32.8 | 32.2 | 39 | 47.6 | 43.8 | 43.4 |
| eccc | 13 | 13.3 | 8.7 | 10.6 | 17 | 14.0 | 13.4 | 14.3 | 13 | 16.7 | 19.5 | 19.2 | 19 | 20.3 | 25.8 | 23.4 |
| ecce | 6 | 4.1 | 6.2 | 5.9 | 6 | 8.3 | 9.0 | 8.6 | 10 | 12.4 | 12.6 | 13.2 | 20 | 16.4 | 16.7 | 18.4 |
| ecec | 4 | 6.0 | 6.9 | 7.4 | 13 | 10.0 | 10.5 | 10.4 | 10 | 14.1 | 14.8 | 15.5 | 15 | 18.4 | 19.6 | 21.2 |
| ecee | 5 | 7.8 | 9.4 | 8.8 | 19 | 22.1 | 21.9 | 21.3 | 34 | 35.1 | 33.0 | 32.5 | 36 | 47.7 | 44.0 | 44.0 |
| eecc | 10 | 11.3 | 8.8 | 10.7 | 20 | 15.9 | 15.6 | 16.3 | 22 | 21.0 | 22.9 | 23.3 | 28 | 26.6 | 30.4 | 30.3 |
| eece | 7 | 8.2 | 9.6 | 9.9 | 17 | 22.3 | 22.4 | 22.2 | 41 | 35.2 | 33.8 | 34.0 | 40 | 47.8 | 45.1 | 46.2 |
| eeec | 14 | 12.0 | 10.8 | 12.4 | 32 | 26.7 | 26.4 | 27.0 | 27 | 40.3 | 40.1 | 40.4 | 50 | 53.8 | 53.6 | 54.2 |
| eeee | 19 | 15.5 | 15.4 | 15.2 | 61 | 59.2 | 59.7 | 60.7 | 125 | 100.1 | 97.3 | 96.2 | 158 | 139.0 | 131.1 | 129.7 |

## Table 7B

Observed and predicted frequencies for 4-trial response sequences, trials 6-9, Experiment RH

| | 8 Items | | | | 16 Items | | | | 24 Items | | | | 32 Items | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trials 6 - 9 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 | Obs. | TDF | RAF | RAF2 |
| cccc | 64 | 58.9 | 34.8 | 43.2 | 69 | 64.6 | 56.7 | 60.2 | 60 | 76.8 | 82.1 | 76.9 | 90 | 92.6 | 108.6 | 89.1 |
| ccce | 0 | 3.0 | 5.7 | 4.7 | 9 | 3.7 | 6.9 | 5.8 | 12 | 4.4 | 9.2 | 10.1 | 10 | 5.3 | 12.0 | 15.2 |
| ccec | 5 | 4.2 | 6.4 | 6.0 | 6 | 4.5 | 7.7 | 6.7 | 10 | 5.1 | 10.3 | 11.2 | 10 | 6.1 | 13.3 | 16.7 |
| ccee | 1 | 2.8 | 4.9 | 3.5 | 6 | 7.5 | 7.1 | 6.1 | 7 | 11.1 | 9.2 | 9.1 | 11 | 14.5 | 11.9 | 12.8 |
| cecc | 6 | 6.6 | 7.6 | 8.1 | 12 | 6.9 | 9.8 | 8.9 | 15 | 7.7 | 13.2 | 14.1 | 15 | 9.1 | 17.2 | 20.3 |
| cece | 1 | 3.1 | 5.0 | 4.1 | 1 | 7.7 | 7.3 | 6.5 | 7 | 11.2 | 9.5 | 9.5 | 8 | 14.6 | 12.2 | 13.4 |
| ceec | 2 | 4.3 | 5.6 | 5.3 | 5 | 9.5 | 8.6 | 8.0 | 11 | 13.1 | 11.4 | 11.4 | 19 | 16.8 | 14.8 | 15.8 |
| ceee | 4 | 2.9 | 5.7 | 3.5 | 11 | 15.9 | 16.0 | 15.1 | 17 | 28.3 | 24.5 | 24.0 | 23 | 39.7 | 32.6 | 32.6 |
| eccc | 16 | 11.8 | 10.1 | 12.0 | 15 | 14.6 | 16.0 | 15.9 | 23 | 17.2 | 22.9 | 23.5 | 36 | 20.7 | 30.3 | 31.4 |
| ecce | 1 | 3.4 | 5.1 | 4.7 | 4 | 8.0 | 7.8 | 7.1 | 14 | 11.3 | 10.4 | 10.6 | 14 | 14.6 | 13.4 | 15.1 |
| ecec | 2 | 4.8 | 5.7 | 6.1 | 10 | 9.8 | 9.2 | 8.8 | 15 | 13.2 | 12.4 | 12.6 | 24 | 16.9 | 16.0 | 17.7 |
| ecee | 1 | 3.2 | 5.7 | 4.0 | 14 | 16.3 | 16.0 | 15.6 | 21 | 28.6 | 24.5 | 24.3 | 33 | 39.9 | 32.8 | 33.0 |
| eecc | 7 | 7.5 | 7.1 | 8.3 | 16 | 15.0 | 13.5 | 13.9 | 16 | 20.0 | 19.5 | 19.7 | 23 | 25.2 | 25.8 | 26.3 |
| eece | 5 | 3.5 | 5.8 | 4.6 | 12 | 16.8 | 16.4 | 16.5 | 21 | 28.9 | 25.3 | 25.3 | 30 | 40.2 | 33.8 | 34.5 |
| eeec | 6 | 4.9 | 6.5 | 6.0 | 22 | 20.6 | 19.5 | 20.6 | 31 | 33.9 | 30.7 | 31.0 | 43 | 46.3 | 41.2 | 41.6 |
| eeee | 7 | 3.3 | 6.7 | 4.0 | 44 | 34.6 | 37.5 | 40.3 | 104 | 73.2 | 69.1 | 70.6 | 123 | 109.7 | 96.[illegible] | 96.7 |

For each of the models, the theoretical expression for the probability of each four-trial sequence was determined. Following Atkinson and Crothers (1964), let $O_{i,j,n}$ be the ith four-tuple for Group $j$ where the sequence begins on trial $n$. Let $\hat{N}(O_{i,j,n})$ be the observed frequency of this sequence, and $N(O_{i,j,n};p)$ be the predicted frequency for a particular choice of the parameters, $p$, of the model. Then define the function

$$X^2_{i,j,n} = \frac{[N(O_{i,j,n};p) - \hat{N}(O_{i,j,n})]^2}{N(O_{i,j,n})} \qquad [18]$$

A measure of the discrepancy between a model and the data from Group $j$ is

$$X^2_j = \sum_{i=1}^{16} X^2_{i,j,2} + \sum_{i=1}^{16} X^2_{i,j,6} \qquad [19]$$

A discussion of the properties of this statistic can be found in Atkinson, Bower and Crothers (1965), and in Holland (1966). Briefly, the quantity $X^2_j$ is distributed approximately as chi square with 30 degrees of freedom, disregarding loss of df due to parameter estimates for the moment.

Parameter estimates were found for each experiment which yielded minimum values of the sum of $X^2_j$ over all groups within an experiment. (The parameter search was carried out to three decimal places on a computer using a search program for parameters for non-linear equations developed by the Mathematics Research Center, University of Wisconsin. (University of Wisconsin Computing Center, 1966). This program, based on a method due to Marquardt (1963), uses the method of steepest descent for initial closing-in to the minimum in a space of several dimensions, and then gradually switches to the method of Gauss for final determination.) In Table 8 are presented the parameter estimates and total chi square values from both experiments for the two models. Since both the RAF and TDF models are three-parameter models, there are 87 df for Experiment SH and 117 df for Experiment RH.

Both models provide a reasonably good fit to the data. The TDF model is somewhat more accurate in Experiment SH, a little less accurate in Experiment RH. Closer examination of the discrepancies between the RAF model and the data in Tables 6 and 7 shows certain systematic trends; specifically, too many errors are predicted for the shorter lists, too few errors for longer lists. This pattern becomes more evident if one considers total errors from trials 2 to 9, presented in Table 9. One modification of the RAF model which seems reasonable considering relative interference produced by learned and unlearned items is to assume that the forgetting function should reflect the error rate; specifically, that the amount

Table 8

Parameter estimates and $X^2$ values for Experiments SH and RH

| | Model | Parameter | | $X^2$ Value Trials 2 - 5 | $X^2$ Value Trials 6 - 9 | Total |
|---|---|---|---|---|---|---|
| Exp. SH | TDF | a | .42 | 49.6 | 65.9 | 115.5 |
| | | b | .11 | | | |
| | | f | .19 | | | |
| | RAF | a | .09 | 60.7 | 90.2 | 150.9 |
| | | c | .21 | | | |
| | | f | .15 | | | |
| | RAF2 | a | .008 | 47.5 | 66.5 | 114.0 |
| | | c | .20 | | | |
| | | f | .56 | | | |
| Exp. RH | TDF | a | .33 | 73.8 | 147.2 | 221.0 |
| | | b | .07 | | | |
| | | f | .25 | | | |
| | RAF | a | .13 | 81.2 | 126.7 | 207.9 |
| | | c | .05 | | | |
| | | f | .20 | | | |
| | RAF2 | a | .009 | 68.5 | 97.8 | 166.3 |
| | | c | .09 | | | |
| | | f | .65 | | | |

Table 9

Observed and predicted mean total errors on trials 2 to 9, Experiments SH and RH

| | | Obs. | TDF | RAF | RAF2 |
|---|---|---|---|---|---|
| Exp. SH | 9 Items | 1.21 | 1.29 | 1.65 | 1.53 |
| | 15 Items | 2.03 | 2.07 | 2.02 | 2.10 |
| | 21 Items | 2.50 | 2.58 | 2.16 | 2.36 |
| Exp. RH | 8 Items | 3.11 | 3.10 | 3.72 | 3.47 |
| | 16 Items | 4.41 | 4.51 | 4.57 | 4.57 |
| | 24 Items | 5.10 | 4.92 | 4.72 | 4.76 |
| | 32 Items | 5.00 | 5.07 | 4.77 | 4.84 |

of forgetting produced by each interpolated item should be proportional to the probability of an error for that item. In the original version of the model, the probability of an error on trial n after m interpolated items and k effective reinforcements was

$$q_{k,m,n} = \alpha^k q_1 \, [1-(1-f)^{m-1}] \qquad [20]$$

In the revised model, the equation for $q_{k,m,n}$ is the same, except that the exponent of the forgetting process, $m-1$, becomes $(m-1)V_{1,n}$, which alters the forgetting function in the desired fashion.

A second modification arose from the following considerations. As the number of items in a list becomes greater, it should be necessary to store in memory more features to allow successful retrieval of the correct response. For larger lists, there is almost always a corresponding increase in formal intralist similarity because of limits on the number of distinctive units, in the present case, alphabetic characters. If the number of features stored on each effective feedback is more or less constant, then the requirement of a larger total number of features for longer lists implies that $\alpha$, the rate parameter for the acquisition process, should be larger for longer lists. (The parameter $\alpha$ is inversely related to speed of learning; the larger $\alpha$ is, the slower acquisition occurs.) The specific assumption is that $\alpha$ is an exponential function of list length, X,

$$\alpha_X = 1 - \alpha^{X-1}$$

These two modifications only slightly affect the mathematical tractability of the model. Although the equations for the learning process take on a more formidable appearance, only minor alterations were necessary in the theoretical expressions used in the parameter-search computer program. The predicted frequencies for the 4-triple response sequences from the revised model, RAF2, are presented in Tables 6 and 7; parameter estimates and $\chi^2$ values are in Table 8. As may be seen, the fit of the model is substantially improved with these modifications. (Analyses of the effects of each of the modifications in turn showed that each modification contributed about equally to the reduction in $\chi^2$.) However, the model continues to deviate systematically from the data in that too many errors are predicted for short lists, too few for long lists.

Two other extensions of the RAF model have been considered, but no final results are available at the present time because of problems in carrying out the mathematical analysis. First, if the idea about acquisition of verbal material by storage of discrete features is to be taken seriously, then a more appropriate formulation would be to replace the linear operator process with an N-element pattern process (Estes, 1959; Atkinson & Estes, 1963). (The feature-list process has been suggested in one form or another by numberous investigators, e.g., Atkinson & Shiffrin, 1965; Bower, 1966; Feigenbaum, 1963). As N, the number of features or patterns, becomes small, there are significant changes in the nature of the theoretical acquisition process. For example, the limiting form of the proposed pattern-RAF model is the one-element all-or-none model, whose properties are well known (Bower, 1961) and differ considerably from the RAF model. (Actually, a more suitable one-element analog to the RAF model is the TDF model, since the one-element all-or-none model does not take into account short-term memory and interference-produced forgetting.) The feature-list process has been worked out for the two limiting cases, $N=1$ and $N=\infty$. The derivations for intermediate values of N are presently under investigation, but the derivations have not been completed.

The second extension concerns the assumption that learning or storage of information is more likely to occur following an error than a success. One approach would be to assume that effective reinforcements occur only on error trials. The resulting model is a two-operator linear process. The forgetting operator is experimenter-controlled (Bush & Mostellor, 1955) in the sense that the subject's responses do not determine the number of interpolated items. The acquisition operator is subject-controlled, according to the relation,

$$q_{n+1} = \alpha q_n \quad \text{if an error on } n$$

$$q_{n+1} = q_n \quad \text{if a success on } n$$

This particular model is not too difficult to work with, and some analyses are presently underway. It appears that this modification of the RAF model will take care of a problem first noted by Atkinson and Crothers (1964) and also found in the two list length studies, viz., the random-trials-incremental process predicts too many errors during the first few trials. The modification has the effect of increasing acquisition rate during the early trials. However, this form of the model also predicts constancy of total errors over the various spacing conditions of

Experiment I in the preceding section. Hence, a more interesting extension would be the case where learning occurs following both successes and errors, but with different probabilities of occurrence. This process is currently under study.

Next, the application of the basic RAF model to the error data of Experiment I in Section I will be considered briefly. From Eq. 17, it can be seen that on the first presentation on each trial, the probability of an error is

$$q_i = q_1 [1-c(1- )]^{i-1} , i = 0, 2, 4, \ldots \qquad [21]$$

since with more than 20 interpolated items from the last presentation of the item, it is likely that short-term retention of the pair will be negligible. Hence, total errors to criterion on the first presentation for repeated items will be

$$T_1 = q_1 \sum_{i=1}^{\infty} [1-c(1-\alpha)]^{2(i-1)} = \frac{q_1}{1 - [1-c(1-\alpha)]^2} \qquad [22]$$

The model makes the strong prediction of constant errors to criterion on the first presentation for repeated items. Moreover, total errors for the once-per-trial items should be $q_1 / \{1 - [1-c(1-\alpha)]\}$, implying that there should be noticeably more errors to criterion for the once-per-trial items than for the first presentation on repeated items. Total errors to criterion from the second presentation for repeated items can be found in a similar fashion to be

$$T_2 = q_1 \left\{1 - \left[\frac{1-(1-f)^k}{kf}\right]^2\right\} [1-c(1-\alpha)] \sum_{i=1}^{\infty} [1-c(1-\alpha)]^{i-1}$$
$$= q_1 \left\{1 - \left[\frac{1-(1-f)^k}{kf}\right]^2\right\} \frac{[1-c(1-\alpha)]}{\{1-[1-c(1-\alpha)]^2\}} \qquad [23]$$

where $\underline{k}$ is the number of interpolated items between the first and second presentations. Parameter estimates of $c(1-\alpha)$ and f were obtained by a least-squares procedure, and the predicted total errors are to be found in Table 2. It is obvious that total errors from the sum of first and second presentations should increase with increased spacing, according to the model.

In this process, the number of presentations required to reach a reasonably stringent criterion is constant, independent of spacing. The effect of massing is to reduce total errors to criterion; this increase in correct responses prior to reaching criterion reflects short-term retention. The failure of the basic model to account for increases in presentations to criterion with greater massing results from the assumption that effective feedback is equally likely every time a stimulus-response pair is presented for study. Comments made previously about modification of this characteristic of the model (e.g. differential probabilities of conditioning after successes and errors) are relevant here also.

The final study in this section concerns the use of learning models to specify optimal strategies for presentation of verbal material. Specifically, a two-operator linear model is examined with reference to the optimal allotment of presentation time to the two halves of a "split" paired associate list. This work is an extension of Suppes' (1964) investigation of optimal block size in paired-associate learning.

Suppose a list consists of $m$ stimulus-response pairs divided into blocks of $k$ items, each block to be presented $n$ times. The problem was to find that value of $k$ which yielded the highest number of correct responses on a test following the training series. Suppes' finding based on the two operator model described below, was that if learning took place at a more rapid rate than forgetting, then the block size should be as large as possible; i.e., choose $k = m$. If forgetting occurred more rapidly than learning, then the block size should be as small as possible, $k = 1$. The first procedure has been termed the whole-list method, and much available data suggests that this method is more efficient.

The present paper represents an extension of Suppes' original paper, in which the requirement that each item be presented a fixed number of times is relaxed. Suppose the material to be learned consists of a list of 2m items, in which the first block of m items is presented for $t$ trials. On each trial, the m items are all presented in random order. The second block of m items is then presented for T-t trials, following which a retention test of the entire set of 2m items is administered. The problem is to find a value of $t$ which yields a mean error probability during the test which is as small as possible. Since interference produced by the learning of items in the second block may lead to considerable forgetting of the first item learned, it would seem intuitively that the items in the first block should receive somewhat more than half of the training trials. The analysis below substantiates this conjecture, and indicates how the choice of $\underline{t}$ depends on the learning and forgetting rate parameters.

It is helpful in reaching a solution for the more complex case to consider first the very simplest situation, viz., a list of two items. The first item is presented for $\underline{t}$ trials, and the second for T-t. Assume that acquisition and forgetting processes may be described by a two-operator linear model. On each reinforced training trial for item i, the probability of an error is reduced by applying an acquisition operator $Q_L$,

$$q_{i,n+1} = Q_L\,(q_{i,n}) = aq_{i,n} \qquad [24]$$

whereas each time the other item is presented, the probability of an error for item i is increased by applying a forgetting operator $Q_F$,

$$q_{i,n+1} = Q_F\,(q_{i,n}) = \underline{b}q_{i,n} + (1-\underline{b})q \qquad [25]$$

where q is the initial error rate. The operators $Q_L$ and $Q_F$ are first-order linear difference equations of the form

$$Q(q_n) = Rq_n + S$$

where R and S are constants. The well-known solution to this process is

$$q_{n+1} = R^n q_1 + S\,\frac{1-R^n}{1-R} \qquad [26]$$

The derivations in this paper are based on this result in large part, and intermediate steps are omitted for brevity. The derivations generally follow the pattern in Suppes' paper (1964) to which the reader is referred for further detail.

This two-operator model is similar in some respects to the RAF model. Two important differences are (1) c = 1, so that every reinforcement is assumed to be effective, and (2) the forgetting operator has the effect of reducing response probability from the acquisition level reached following a reinforcement to the base guessing level. Work is presently underway to investigate the properties of the RAF model with regard to block size and split list presentations. Since the forgetting function affects only training performance and not long-term retention, then the basic model must predict no effect of block size. For the RAF model revised so that learning is more likely after errors, then whole list learning should be somewhat more efficient in any circumstance. The relative inefficiency of various part-list procedures remains to be determined.

For the special case of a two-item list, following t trials on item 1, $q_{2,t}$ will still be equal to $q$, while $q_{1,t}$ will equal $\alpha^t q$. At the end of training, the error probabilities for the two items will be

$$E(q_1) = Q_L{}^t \, Q_F{}^{T-t}(q) = [1-(1-a^t)b^{T-t}]q$$

$$E(q_2) = Q_L{}^{T-t}(q) = a^{T-t}q \qquad [28]$$

Hence the expected number of errors on the posttest will be

$$E(\overline{q}) = (1-b^{T-t} + a^t\, b^{T-t} + a^{T-t})q \qquad [29]$$

The value of t which minimizes $E(\overline{q})$ can be found by standard methods; find dE/dt, set the derivative equal to zero, and solve for t. From Eq. 5 we have

$$\frac{dE}{dt} = \left\{ b^T \left[b^{-t} \log b + t\left(\frac{a}{b}\right) \log\left(\frac{a}{b}\right)\right] - a^{T-t} \log a \right\} q \qquad [30]$$

Setting this derivative equal to zero yields the equation

$$1 - a^t = \left[\left(\frac{a}{b}\right) T-t - a^t\right] \left(\frac{\log a}{\log b}\right) \qquad [31]$$

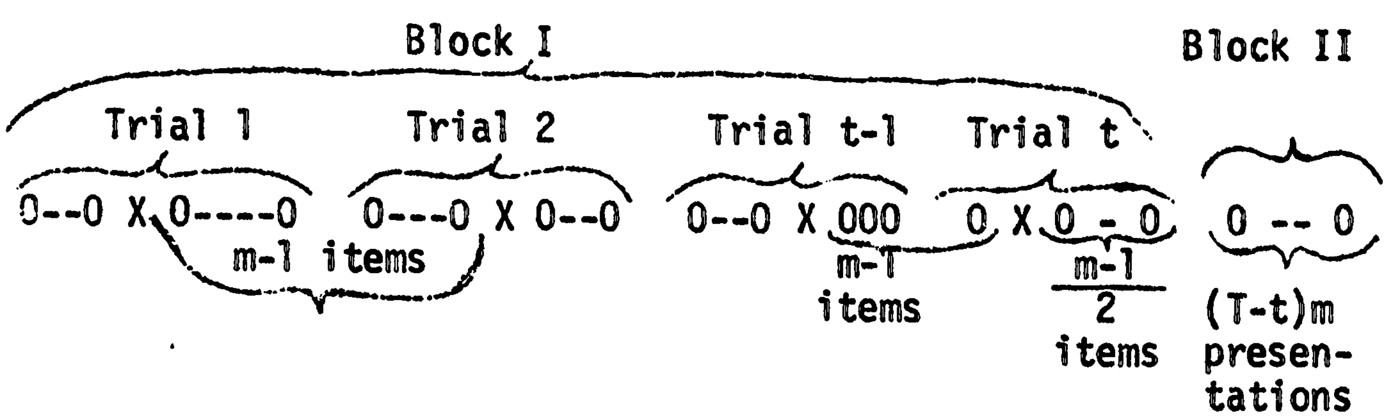
A concerted effort to find a solution in t for this equation was unsuccessful. However an approximate solution was arrived at by noting that if a is not too close to 1 or t is relatively large, the term, $a^t$, may be small enough to be disregarded. Assuming that $a^t$ is negligible, then the optimal value of t is

$$t_{op} = T - \frac{\log\left(\frac{\log a}{\log b}\right)}{\log b - \log a} \qquad [32]$$

To the extent that the approximation holds, then for any value of a and b, the number of trials which should be allotted to the first item is a linear function of T. Moreover, the number of trials remaining to be allocated to the second item is a function of a and b, and is independent of T.

As a tends toward 1 (slower learning), then item 2 should receive a larger share of the trials, whereas if this parameter is close to 0 (rapid learning) a single presentation of item 2 may be optimal. On the other hand, if b is near 0, then it may be the case that item 2 should not be presented at all, since any presentations of this item will lead to virtually complete loss of the newly formed association for item 1. (There is no requirement that the two items be learned equally well.) If b is near 1, then more trials may be allotted to the second item, because interference will be minimal.

Turning to the more general problem, assume that the list consists of 2$\underline{m}$ items, where $\underline{m} > 1$. The items in a block are presented in random order on each trial, so that a particular item may be the first, second, ...mth item presented on the trial. Hence, between successive presentations of a given item, 0 to 2($\underline{m}$-1) other items may be interpolated; on the average $\underline{m}$-1 items are interpolated between successive presentations of a given item. A precise derivation of the expected error rate following training would involve computing the error rate for each random sequence which might occur, and finding the expectation over sequences. This derivation is very cumbersome, and so as an approximation we will find E(q) based on the average location of an item in the sequence. That is, consider an item which is presented in the middle position on each trial as indicated in the diagram below for block I where X represents a reinforced presentation of the "average" item, and 0 represents presentations of other items:




From this description of the presentation sequence, it can be seen that the probability of an error for an item in block I at the end of training will be in terms of the two operators $Q_L$ and $Q_F$

$$E(q_1) = (Q_L Q_F^{m-1})^{t-1} Q_L Q_F^{\frac{m-1}{2}} Q_F^{(T-t)m} q \qquad [33]$$

There are t-1 complete cycles of a reinforced presentation followed by m-1 other items, and then on trial t, there is a partial cycle consisting of a reinforced presentation followed by (m-1)/2 other items.

The items in block II then constitute a total of (T-t)m presentations on each of which the forgetting operator is applied. Substituting the first-order difference equation for the operators yields

$$E(q_1) = [1 + X(1-Z) - Y(1-aZ)]q$$

where $X = \left(\frac{a}{b}\right)^t b^{tm - \frac{m-1}{2}}$ [34]

$$Y = b^{(T-t)m + \frac{m-1}{2}}$$

$$Z = \frac{1-b^{m-1}}{1-ab^{m-1}}$$

From similar considerations it can be seen that

$$E(q_2) = (Q_L Q_F{}^{m-1})^{T-t-1} Q_L Q_F{}^{\frac{m-1}{2}}$$

$$= [1+X' (1-Z) - Y' (1-aZ)] \quad [35]$$

where

$$X' = (ab^{m-1})^{T-t-1} b^{\frac{m-1}{2}}$$

$$Y' = b^{\frac{m-1}{2}}$$

The average error rate, averaged over items in both blocks, is simply

$$E(q) = \frac{1}{2} q \; [2 + (1-Z)(X + X') - (1-aZ)(Y + Y')] \quad [36]$$

Since only the quantities X, X′ and Y involve the parameter t, with respect to which a minimum is being sought, the optimization problem is reduced to finding a minimum in t for the function

$$F = (1 - Z)(X + X') - (1-aZ)(Y+Y'). \quad [37]$$

The derivative is

$$\frac{dF}{dt} = (1-Z) \frac{dX}{dt} + \frac{dX'}{dt} - (1-aZ) \frac{dY}{dt} \quad [38]$$

where

$$\frac{dX}{dt} = b^{Tm - \frac{m-1}{2}} \left(\frac{a}{b}\right)^t \log\left(\frac{a}{b}\right)$$

$$\frac{dX'}{dt} = -\left(ab^{m-1}\right)^{T-t-1} b^{\frac{m-1}{2}} \log\left(ab^{m-1}\right) \qquad [39]$$

$$\frac{dY}{dt} = -\, b^{(T-t)m + \frac{m-1}{2}}\, m \log b.$$

When $\frac{dF}{dt}$ is set to zero, there is again no apparent solution in $t$ for the resulting equation. Following the strategy used for $m = 1$, assume that if (1) $a$ is not close to 1, (2) $b$ is reasonably large, and (3) T is not too small, then $\frac{dX}{dt}$ will be negligible, so that an approximate value of $t_{op}$ may be found by solving the following equation for t:

$$0 = -(1-Z)\frac{dX'}{dt} - (1 - aZ)\frac{dY}{dt}. \qquad [40]$$

Making the appropriate substitutions for the derivatives, eliminating the common factor $b^{m(T-t)}$, expressing $\log(ab^{m-1})$ as $(\log a + (m-1)\log b)$, and dividing through by log b yields the result,

$$\left(\frac{a}{b}\right)^{t-(T-1)} = \frac{(1-Z)}{mb^m(1-aZ)}\left[\frac{\log a}{\log b} + m-1\right]. \qquad [41]$$

Letting the quantity on the right side of the equation be represented by C and taking the logarithm, we have

$$t - (T-1)\log\left(\frac{a}{b}\right) = \log C \qquad [42]$$

or

$$t_{op} = T-1 + \frac{\log C}{\log\left(\frac{a}{b}\right)} \qquad [43]$$

If $a < b$, which will be the case with data to be considered later, and if log C is positive, then the final term in the equation above will be negative. Hence, the optimal number of trials will be approximately a linear function of T, the total number of trials available for both blocks, such that one or more trials should be given to the second block, but the exact number is independent of the total available trials.

The adequacy of the approximation was determined by programming a computer to search for optimal values of t, for which the derivative of Eq. 37 is equal to 0. The parameter space looked at included values of a from .3 to .7 in increments of .2, b from .90 to .999 in a roughly exponential series (more closely spaced as the boundary of 1.0 was approached), list lengths of 20 and 50, and total trials, T, from 2 to 25. The approximation is quite good (the difference between actual and approximate $t_{op}$ is .1 or less) except for a and b close to 1 (.995 or greater), and for small values of m and T (lists of less than 10 items, or fewer than 10 trials).

Two properties of the parameters a and b are worth noting briefly. First, suppose $a > b$. In this case, the proportionate reduction in error probability which results from a study presentation is less than the reduction in success probability which occurs when another item is presented. (Recall that the operator $Q_F$ is applied once for each interpolated item in a list.) From Suppes' (1964) results, when a list is divided into blocks of size k, and the items in each block receive an equal number of trials, optimal block size is 1. That is, the first item should be presented T/m times, then the second item, etc. Because forgetting takes place at a faster rate than acquisition, perfect performance is impossible. The expected error rate, E(q), following a total of n trials per item for each of m items, with a block size of k, was shown by Suppes (1964) to be

$$E(q) = q\left\{1 - \frac{(1-b^{nm})(1-a)}{(1-b)m}\left(\frac{1-Y^nX^n}{1-YX}\right)\left(\frac{1-X}{1-X^n}\right)\right\} \qquad [44]$$

where $Y = a/b$ and $X = b^k$. If $a > b$, and the optimal block size of 1 is used, then Eq. 44 becomes

$$E(q) = q\left\{1 - \frac{(1-b^{mn})(1-a^n)}{m\,(1-b^n)}\right\} \qquad [45]$$

As n becomes large, E(q) approaches a limit of $q(1-1/m)$. The situation is perhaps more obvious if the guessing rate is close to zero. Then the prediction of the model is that the probability of a correct response after a large number of trials is 1/m; i.e., that only one item from the list will be learned. Performance generally reaches a higher level than acquisition of a single item in most situations, and hence it seems reasonable to suppose that generally $a < b$, at least within the context of this particular model. Crothers (personal communication) has found better retention under part-list than whole-list presentation with very long lists of Russian vocabulary items, which implies $a > b$. However, it still seems doubtful that a block size of 1 would be optimal. Rather, one might question the adequacy of the model, which cannot predict that intermediate block sizes will be optimal.

Second, taking into account the magnitude of retention loss typically observed in paired-associate learning, it is likely that under most conditions $\underline{b}$ will be very close to 1. Suppose, for example, that a list of 10 items is learned to a rigorous criterion, so that it is reasonable to assume that the probability of a correct response is close to 1. Next the subject is given five trials on a second, 10-item list, for a total of 50 presentations, on each of which the probability of a correct recall of items in the first list is reduced exponentially with parameter $\underline{b}$. If $\underline{b}$ is .999, the retention loss is .05, while if $\underline{b}$ is as low as .99, the loss is about .60; complete loss of the material in the first list is expected if $\underline{b}$ is less than .95.

More generally the asymptotic probability of a correct response after many training trials using the whole-list method, given $\underline{a} < \underline{b}$, will be (c.f. Eq. 44)

$$\lim_{n \to \infty} E(q) = 1 - \frac{(1-a)(1-b^m)}{m(1-b)(1-ab^{m-1})} \qquad [46]$$

If $\underline{b}$ is not fairly close to unity, then even for relatively short lists (e.g. 10 to 20 items), the asymptotic error rate is much higher than is typically observed in paired-associate learning, in which "perfect" performance (i.e. a very low error rate) is readily reached by most subjects. Given that $\underline{b}$ is approximately 1 and m not too large, then $(1-b^m) \cong (1-b)m$, and $(1-ab^{m-1}) \cong (1-a)$, so that the error rate will be close to zero. To suggest that $\underline{b}$ be very near unity means that the interference produced by a single interpolated S-R presentation is slight; the retention loss over a series of interpolated pairs of even moderate length rapidly reaches sizable proportions, however.

The final theoretical question concerns the efficiency of split-list learning, given optimal allocation of trials to the first and second blocks, compared with whole-list presentation. While it has not been possible to prove this result analytically, it appears that the whole-list procedure is always more efficient if $\underline{a} < \underline{b}$. The expected error rate, computed numerically over the same range of parameter values used in evaluating the approximation to $t_{op}$, was always lower for the whole-list method than optimal split-list. Plots of the efficiency function, $E(q)_{split} - E(q)_{whole}$, show that the exact value of the difference bears a complex relation to the parameters, $\underline{a}$ and $\underline{b}$, and the number of trials. However, for values of $\underline{a}$ between .3 and .7, the difference in error probability between the whole-list and split-list procedures is slight (generally less than .05). Empirical support for the theoretical prediction of a small loss in efficiency when using the split-list procedure, would be significant because often there are practical reasons to present material in blocks rather than using the whole list. For example, the error probability during training typically remains quite high for a considerable number of trials with the whole-list method, which may cause the student to feel that he is making very little progress, with a resultant lowering of motivation (Hovland, 1951). Under the special circumstances of the experimental laboratory, this motivational decrement may often prove inconsequential, especially since experimental lists are typically short (less than 50 items). Outside the laboratory, and with fairly long lists, the problem is probably of more concern, and it is of importance to be able to specify the conditions under which the loss in efficiency is slight.

The experiment described below was carried out to obtain some preliminary data on the effects of variations in blocking procedures on efficiency of learning foreign language vocabulary associates.

## METHOD

Subjects. -- Ten graduate students in psychology at the University of Wisconsin, Madison, served as Ss. None of the students had any previous experience with the German language. Each S received ten dollars (about $1.00 per hour) for his services.

Materials. -- The vocabulary items consisted of five lists of 50 German words. Each list was composed of approximately the same number of nouns, verbs, adjectives, and other parts of speech. Obvious english cognates were avoided. Within-list similarity was kept low in the stimulus members by not permitting two or more words from the same root to occur in a single list, and in response terms by eliminating synonomous response terms within a list. Each German stimulus word was printed on a 1.5 x 3.5 in. card, and the appropriate english equivalent was printed on the reverse of the card.

Procedure. -- Five different presentation conditions were used in the study:

W Whole-list; a list of 50 words was presented in random order for five trials.

S-5/5 Split-list with a 5-5 division; the list of 50 words was divided into two 25 word blocks. The items in the first block were presented in random order for 5 trials, and then the items in the second block were presented for 5 trials.

S-7/3 Split list with a 7-3 division; the items in the first block were presented for 7 trials, and then the items in the second block were presented for 3 trials.

S-9/1 Split list with a 9-1 division.

P A part-list presentation; the 50 words were divided into 5 blocks of 10 words, and each block was presented for 5 trials.

Subjects were run under a different condition-list combination for five consecutive days using a Graeco-Latin square design so that all lists and conditions were represented in a sequentially counterbalanced order over a set of five Ss. Two different squares were made up for the two sets of 5 Ss. At the conclusion of each training session, a recall test (T-1) on the material just learned was administered. A second test (T-2) was given 24 hours later, just prior to the training trials on the next list. (The second test for the fifth day's material was also administered 24 hours after initial training.) A final test (T-F) on all 250 words from the five lists, randomly ordered, was performed from 3 to 5 days following the final training session.

At the beginning of the first session, the subject was given the following instructions:

"In this experiment you are going to learn some german words. On each trial, I will show you a card which has a german word printed on it, and I will pronounce the word for you. You must either try to give the english equivalent, or say 'No answer'. After you have made your response, I will tell you the correct english answer. You should try to take no longer than 10 seconds to make your response. You will be tested on the words you learn several times, so you should try to retain them, but no discussions with other subjects, persons speaking German, or dictionary study are allowed."

After answering any question by rephrasing the instructions, the first training session began. A standard anticipation procedure was used. On each presentation, the experimenter held up the card with the german word before the subject, and pronounced the word. (The experimenter had two years of undergraduate german.) After the subject's response, the english equivalent was spoken by the experimenter. About 5 seconds later, the next stimulus was presented. The card decks were shuffled between trials. A Masonite divider separated the subject and the experimenter, so that neither the stimulus decks nor the data sheets were visible to the subject.

## Results

In Table 10 is presented the mean error probability as a function of condition, list and training day for each of the three tests. Analysis of variance of these data indicated that the presentation technique was a significant variable during all three tests:

Table 10

Observed error probabilities

at each test interval and for each condition

and predicted error probability

during initial test, T-1, in parentheses

| Test Interval | Condition W | S-5/5 | S-7/3 | S-9/1 | P |
|---|---|---|---|---|---|
| T-1 (10 min.) | .190 | .178 | .192 | .380 | .212 |
| | (.176)* | (.193)* | (.218) | (.367) | (.202)* |
| T-2 (24 hr.) | .228 | .242 | .294 | .459 | .314 |
| T-F (5 da.) | .490 | .478 | .552 | .653 | .571 |

* Used for parameter estimation.

T-1, $F(4,28) = 17.9$; T-2, $F(4,28) = 16.1$; T-F, $F(4,28) = 6.2$; $p < .01$ for each ratio. The day variable was significant during the final test only $F(4,28) = 7.4$, $p < .01$. No other sources of variance were significant. Thus, the lists were about equal in difficulty, and there was no evidence of "learning-to-learn" effects over days. The 9/1 split-list procedure was the least effective, followed by the part-list procedure. The whole-list and 5/5 split-list procedures lead to similar performance on all tests, with somewhat more than half of the items being correctly recalled on the final test under these conditions. Error probability on the final test was a decreasing function of day number; i.e., the later in the training series a list was learned, the better the list was recalled on the final test. This result could be attributed to either greater delay between training and test, or to the fact that more lists were learned in the interval, or both.

In order to determine whether the data from the experiment may be described by the model, it is necessary to estimate the parameters $\underline{a}$ and $\underline{b}$. This estimation was carried out by finding numerically those parameter values which minimized the squared deviations of predicted and observed error probability on the first test using conditions W, P and S-5/5. The least squares estimates were $\underline{a} = .64$ and $\underline{b} = .999$. The predicted values for each of the conditions on the first test are presented in Table 1. According to model, the optimal division of the ten trials under the split-list condition should be 5.4 trials for the first block. Thus the S-5/5 condition should be optimal, but a 7/3 division should be only slightly worse, according to the model.

In addition to first test performance, it is possible to predict performance during the training sessions. For example, let $T_n(m)$ be the expected total errors during the first n training trials of a list containing m items. It can be shown that

$$T_n(m) = \frac{1-(ay)^n}{1-ay} + \left[n - \frac{1-(ay)^n}{1-ay}\right]\left[\frac{1-y}{1-ay}\right]$$

where $y = b^{m-1}$.

Observed total errors from the first five trials of the training series for condition P, the average of the three S conditions, and condition W provide information about learning under list lengths of 10, 25 and 50 items, respectively. (The first block data only were used from conditions P and S.) The observed total errors in five training trials were 2.41, 2.94 and 3.11, and the predicted errors were 2.52, 2.58, and 2.67. The predictions are not bad, but leave something to be desired. (With $\underline{b}$ so close to 1, slight changes in this parameter cause big differences in the predicted values of the total errors. For example, if $\underline{b}$ is set equal to .996, the predictions are much closer to the observed values -- 2.62, 2.84, and 3.14. Test performance predictions become much worse with this estimate of $\underline{b}$, however).

Without additional assumptions, the model cannot generate predictions about performance on the second and final tests. It is designed to account primarily for interference-produced forgetting which may occur while a list is being learned, and it is well-known that considerable extra-experimental forgetting occurs. It is important to note, however, that the relative ordering of performance under the different experimental conditions is the same for the long-term tests as well as the immediate test. The effects of the presentation procedures did not wash out over a period of about a week, and in this sense can be called long-term effects.

In summary, the two-operator model handled several important aspects of the data very creditably. Given an estimate of the learning and forgetting rate parameters for a subject matter and for a particular student population, it is possible to predict performance with longer lists and the additional training which may be required to reach any criterion level. It appears that when it is convenient, long lists may be divided into two sublists, and with optimal assignment by presentation time to each of the sublists, this split-list procedure will be very nearly as efficient as the whole-list technique. The model is inadequate in several respects, as mentioned previously. Its chief virtue is simplicity, and even so, problems of mathematical analysis arise. The theoretical and empirical work described above nevertheless provide a stepping stone to further progress in the study of efficiency in learning.

## Section IV. Acoustic and visual confusions in immediate memory

Recent investigations of human memory suggest that the stimulus input may undergo several transformations over a period of perhaps less than a minute or so (Sperling, 1966; Wicklegren, 1966; Dale & Gregory, 1966). Restricting attention to visually-presented verbal stimuli, the sequence of information processing may take the following form: a brief visual storage period of no more than a few seconds, an auditory or acoustically-encoded period of around 20 to 30 seconds (as if the subject had spoken or repeated the material to himself), and finally an associatively-encoded representation is achieved, which may last over in indefinite time interval. This representation of the functioning of human memory rests largely on the patterns of confusion errors in memory with variation in presentation and retention intervals. The primary study reported in this section was designed to shed further light on the effects on short-term recall of the relation between the mode of input--visual or acoustic--and the mode of output or testing, which used analogs of the visual and acoustic input procedures. (This study was conceived independently by William Chase, and he was largely responsible for its design and execution.) A second study was performed to obtain a visual confusion matrix based on perceptual errors.

### Method

Subjects.--The Ss were 11 female introductory psychology students who received class credit for participation in the experiment, and one female graduate student who was paid for her participation.

Design.--The experiment consisted of visual or acoustic presentation of the list for memorization, visual or acoustic presentation of the test stimulus, and 3 types of material (visually confusing, acoustically confusing, or neutral letters), comprising a 3x2x2 factorial within-Ss design. Each S participated in five 45-minute sessions. On days 1 and 5 the Ss performed under all 4 combinations of visual and acoustic presentation of the list with visual and acoustic presentation of the test, using as stimulus materials the digits 1-8. The first session provided the subject with some experience in the task; the last session permitted an evaluation of the magnitude of improvement over sessions. On days 2, 3, and 4 the Ss received 4 tasks a day, and over the 3 days, received all combinations of 4 tasks and 3 types of material. The order of the 12 conditions was counterbalanced by means of a 12x12 latin square with one S per row of 12 conditions.

Materials and procedure.--The stimulus materials consisted of lists of visually confusing letters (BCDGOQRU, cf., Chase, 1965), acoustically confusing letters (BCDEPTVZ, cf., Conrad, 1964), or neutral letters (ADHIMQYZ). With each type of material, lists of 1, 2, or 4 letters were composed. Sixteen lists were made up at each length, with each letter being used an equal number of times. There were 8 lists with the test letter present and 8 lists with the test letter absent. Each letter within a type of material was used once as a test letter, and each serial position contained an equal number of positive test letters. The set of 48 lists, 16 at each of 3 list lengths, was then randomly permuted, so that list length and type of material varied randomly within each task condition. The same random sequence was used for all 4 conditions within a day. Three practice lists, one of each list length, preceded each set of 48 experimental lists. Prior to each condition within a day, S was informed of the new task conditions.

Visual stimuli were typed on plain white index cards in capital elite letters. Memory lists were presented on 3x5 cards for a 5 sec. interval. Test letters were on 5x7 cards centered in the exposure field of a Polymetric two-channel tachistoscope. The letters subtended 33' of visual angle when viewed in the tachistoscope.

Acoustic stimuli were recorded in a female voice on a Wollensak stereophonic tape recorder. The memory list was presented twice at a rate of .5 sec. per letter, with the second presentation starting 2.5 sec. after the first.

A warning click sounded 5 sec. after the start of the presentation of the memory list, and then 2 sec. later, the test letter was presented, either visually or acoustically. The S's task was to press one of two buttons marked "YES" or "NO", depending upon whether or not the test letter was a member of the previously memorized list of 1, 2 or 4 items. Half the Ss were assigned at random to press "YES" with their dominant hand and half with their non-dominant hand. The Ss were instructed to respond as quickly as possible without making any errors. They were also informed that there were an equal number of "YES" and "NO" responses arranged in a random sequence. The latency of each response was measured on a Standard Electric timer to the nearest .01 sec.

## Results

The median reaction time for correct responses was determined for each task by type-of-material by list-length combination for a S, based on the pooled reaction times from 8 "YES" and 8 "NO" responses. A least squares estimate of the slope of the reaction time function over list lengths was then computed for each subject and for each of the 12 basic conditions. This slope measure, which represents rate of search through the list in memory, is the main dependent variable. An analysis of variance showed that the only significant main effect was due to type of material, $F(2,22) = 4.68$, $p < .05$. Rate of search through acoustically confusing lists was much slower (64.0 msec./item) than through the visually confusing and neutral lists, which did not differ from one another (46.1 and 48.0 msec./item, respectively).

There was a significant interaction between mode of list presentation and mode of test presentation, $F(1,11) = 13.05$, $p < .005$. As can be seen in Table 11, search rates are much higher if the presentation and test is in the same mode (46.7 ms/item) than if they are in different modes (58.8 ms/item).

The intercepts of the reaction time functions were also analyzed. The only significant effect was mode of test presentation; reaction time was faster to an auditory test stimulus than a visual test stimulus by 117 ms., $F(1,11) = 40.8$, $p < .001$.

## Discussion

The most significant results of this study were (1) the large effect of acoustically confusing material, (2) the absence of any effect due to visually confusing material and (3) the faster processing rates when identical input and output modes are used. The first result is not surprising in view of the studies of Conrad and Wicklegren. For the time intervals used, there is good reason to believe that the primary representation of verbal material in short-term memory is of an acoustic or phonetic form. The second finding is somewhat surprising in light of an earlier study by Chase (1965) using identical materials and a very similar memory search task, in which search rates through the visually confusing material were slower than through either acoustically confusing or neutral material. We have not come up with a very satisfactory explanation for these contradictory results, but some preliminary ideas are currently being investigated.

The last finding mentioned above has important bearing or theories of memory. The ubiquitous finding of acoustic confusions in memory would seem to imply a common format and storage location for verbal material after the first few seconds of processing have taken place. Since information is more readily available for test when similar sensory modes are used for input and output, then either the "acoustic" memory tr ce might retain some type of sensory input tag which facilitates retrieval differentially depending on the output mode, or there might be different temporary storage locations for material input in visual or auditory modes, even though the form of the stored material in both locations might be acoustic. In any event, the mode of sensory input continues to be an important factor in retrieval, even after processing to a common format has been achieved.

Table 11

Search rates in msec. per item as a function of modality of presentation of list and test

| | | Test Modality | | |
|---|---|---|---|---|
| | | Visual | Acoustic | |
| List modality | Visual | 44.1 | 57.4 | 50.8 |
| | Acoustic | 60.1 | 49.2 | 54.6 |
| | | 52.1 | 53.3 | 52.7 |

While selecting materials for the preceding study, a review of the literature revealed that there existed no recent evaluation of visual confusions among letters of the alphabet. Consequently, the following study was carried out to obtain a complete visual confusion matrix. This matrix was to be compared with a visual overlap measure devised by Chase (1965) as a basis for selection of a visually confusable set for his study.

## Method

Subjects.--The Ss were 77 introductory psychology students run in 9 groups of from 7 to 10 Ss per group. The records of 5 additional Ss were rejected because they did not have data for every trial.

Stimuli.--Block capital letters were photographed to produce a clear black image on a white background. The masking stimulus was a random black and white checkerboard pattern with an equal number of black and white squares.

Apparatus.--The stimuli were projected from behind onto an opaque glass screen 8" high by 10" ~~wide~~ situated in the center of a 5' by 3' masonite panel. The Ss were seated 7-10 feet from the screen.

Two Anscomatic II slide projectors were used to present the letters and the masking stimulus. Two Wollensak alphax camera shutters were used to time the stimulus presentations of the letters and blanking stimulus. The intertrial interval was timed with a cam-timer controlling, a Cramer model 940A timer which controlled onset of the letter stimuli and the masking stimulus, respectively.

Procedure.--The sequencing of events during a trial was as follows. The trials were spaced 5 seconds apart. A brief apparatus click served as a warning signal, followed in 1/2 sec. by the letter and masking stimulus. The letter was exposed for about 20 msec., followed by about 30 msec. of darkness, and then followed by a masking stimulus for 1 1/2 seconds. The masking sitmulus was used to control precisely the duration of the visual afterimage.

A preliminary study was conducted in which the time between letter onset and masking stimulus onset was varied in order to determine an exposure duration which would yield about 50% error. The duration chosen was approximately 50 msec.

Two hundred and sixty slides, 10 copies of each letter, were arranged in random order, 40 slides in each of 6 trays, and a final tray with 20 slides. The order of presentation of the 7 trays of letters was counterbalanced across the 9 groups so that each tray appeared at least once and not more than twice in any serial position.

The answer sheets consisted of 14 columns of 20 boxes on a single sheet. Every other column was separated by a double line so as to correspond to a block of 40 trials. The Ss were instructed to make a response in every box, even if they had to guess. They were first shown a few exposures to acquaint them with the apparatus. The duration was gradually decreased to the exposure duration used in the experiment so the Ss could first get an idea of the letters used and the exposure duration used. They were then run on 40 trials at a pace of 5" per exposure while the experimenter called off the number of the trial in advance. A short rest was given after each block of 40 trials while the experimenter changed slide trays.

## Results and Discussion

The stimulus-response confusion matrix is presented in Table 12. The error rate varied from 60 to 90 percent for various items. Substantial response bias was observed, some letters occurring as responses with twice the frequency of others (e.g. Z vs L). The pattern of confusions is generally along the lines to be expected from structural considerations, with some exceptions. For example, the following sets are mutually confusable: C-G, E-F, I-J-L, O-Q, K-X, and V-Y. However, the following sets are not confused, though it would have been predicted from a visual overlap measure (Chase, 1965): A-H, C-D, U-V, etc. The confusion matrix is being treated by the multi-dimension monmetric scaling procedure of Shephard (1964) in an effort to find a space of fewer than 26 dimensions to describe the matrix. Tentative results lead to the conclusions that (1) a space of about 6 dimensions is required to describe the data, and (2) the dimensions represent specific complex combinations of the sort represented by E-F vs I-J-L, and C-G versus O-Q, rather than more general characteristics such as vertical and horizontal lines, curved segments, etc. Additional work remains to be done in analyzing these data.

## Table 12A

Visual confusability matrix, response frequency to each stimulus

Stimuli

| Responses | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 386 | 21 | 20 | 26 | 23 | 33 | 16 | 13 | 28 | 31 | 25 | 16 | 19 |
| B | 18 | 377 | 15 | 18 | 18 | 18 | 18 | 28 | 23 | 33 | 22 | 24 | 26 |
| C | 15 | 17 | 295 | 17 | 28 | 15 | 48 | 25 | 30 | 28 | 24 | 29 | 17 |
| D | 19 | 24 | 18 | 370 | 28 | 15 | 15 | 14 | 21 | 33 | 13 | 22 | 24 |
| E | 19 | 23 | 13 | 18 | 279 | 37 | 13 | 27 | 17 | 18 | 24 | 31 | 12 |
| F | 21 | 19 | 17 | 19 | 87 | 299 | 23 | 32 | 37 | 28 | 30 | 34 | 38 |
| G | 5 | 20 | 130 | 24 | 20 | 17 | 395 | 27 | 22 | 19 | 22 | 17 | 22 |
| H | 9 | 16 | 8 | 9 | 18 | 12 | 7 | 300 | 31 | 21 | 30 | 26 | 49 |
| I | 7 | 1 | 8 | 1 | 8 | 6 | 7 | 11 | 103 | 25 | 9 | 23 | 9 |
| J | 11 | 7 | 6 | 5 | 3 | 7 | 7 | 5 | 9 | 135 | 10 | 9 | 10 |
| K | 24 | 13 | 8 | 15 | 19 | 19 | 9 | 19 | 32 | 24 | 269 | 17 | 18 |
| L | 31 | 13 | 19 | 20 | 34 | 25 | 10 | 37 | 92 | 73 | 37 | 244 | 25 |
| M | 13 | 20 | 16 | 18 | 16 | 20 | 10 | 17 | 26 | 25 | 30 | 15 | 248 |
| N | 23 | 10 | 7 | 13 | 14 | 16 | 10 | 43 | 25 | 22 | 29 | 18 | 36 |
| O | 25 | 22 | 36 | 29 | 25 | 31 | 33 | 34 | 34 | 41 | 30 | 30 | 27 |
| P | 28 | 40 | 26 | 34 | 38 | 58 | 32 | 45 | 44 | 37 | 17 | 36 | 26 |
| Q | 8 | 9 | 26 | 13 | 8 | 8 | 25 | 9 | 12 | 7 | 17 | 13 | 17 |
| R | 27 | 43 | 26 | 27 | 30 | 23 | 26 | 37 | 34 | 37 | 40 | 34 | 28 |
| S | 17 | 19 | 25 | 25 | 12 | 24 | 23 | 23 | 31 | 35 | 27 | 26 | 23 |
| T | 24 | 14 | 11 | 23 | 19 | 37 | 22 | 17 | 38 | 29 | 17 | 27 | 20 |
| U | 12 | 11 | 12 | 14 | 13 | 9 | 6 | 8 | 13 | 14 | 7 | 15 | 15 |
| V | 4 | 5 | 4 | 9 | 8 | 10 | 1 | 5 | 10 | 13 | 4 | 8 | 12 |
| W | 4 | 6 | 5 | 5 | 5 | 7 | 1 | 10 | 11 | 15 | 7 | 10 | 12 |
| X | 10 | 9 | 11 | 4 | 3 | 11 | 4 | 8 | 20 | 11 | 18 | 21 | 19 |
| Y | 1 | 6 | 6 | 5 | 4 | 6 | 2 | 3 | 12 | 4 | 3 | 8 | 5 |
| Z | 9 | 5 | 2 | 6 | 10 | 7 | 7 | 6 | 15 | 12 | 9 | 17 | 13 |

## Table 12B

Visual confusability matrix, response frequency to each stimulus

Stimuli

| | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 21 | 14 | 17 | 14 | 18 | 19 | 20 | 24 | 22 | 28 | 20 | 15 | 23 |
| B | 18 | 11 | 15 | 19 | 37 | 25 | 8 | 21 | 23 | 20 | 28 | 19 | 13 |
| C | 20 | 18 | 12 | 23 | 16 | 17 | 25 | 22 | 12 | 23 | 19 | 15 | 17 |
| D | 30 | 20 | 23 | 13 | 29 | 22 | 19 | 24 | 21 | 27 | 21 | 22 | 11 |
| E | 28 | 10 | 21 | 9 | 11 | 14 | 18 | 22 | 16 | 30 | 25 | 21 | 15 |
| F | 48 | 30 | 17 | 18 | 25 | 19 | 31 | 31 | 28 | 37 | 33 | 51 | 23 |
| G | 16 | 27 | 20 | 21 | 17 | 26 | 18 | 25 | 19 | 22 | 22 | 17 | 8 |
| H | 35 | 5 | 9 | 12 | 9 | 10 | 17 | 19 | 18 | 33 | 41 | 14 | 8 |
| I | 7 | 11 | 7 | 7 | 9 | 7 | 23 | 15 | 10 | 11 | 18 | 19 | 13 |
| J | 7 | 5 | 10 | 3 | 4 | 9 | 6 | 5 | 14 | 5 | 13 | 9 | 16 |
| K | 28 | 9 | 14 | 12 | 23 | 5 | 18 | 11 | 33 | 26 | 66 | 14 | 16 |
| L | 46 | 22 | 26 | 19 | 12 | 26 | 39 | 31 | 30 | 37 | 42 | 52 | 33 |
| M | 38 | 13 | 6 | 18 | 16 | 9 | 21 | 18 | 20 | 31 | 27 | 25 | 22 |
| N | 183 | 16 | 15 | 9 | 14 | 15 | 23 | 22 | 15 | 24 | 32 | 30 | 22 |
| O | 31 | 363 | 23 | 134 | 31 | 23 | 34 | 44 | 20 | 23 | 32 | 24 | 26 |
| P | 33 | 45 | 396 | 31 | 59 | 38 | 37 | 45 | 37 | 34 | 28 | 40 | 25 |
| Q | 14 | 59 | 9 | 312 | 11 | 14 | 11 | 23 | 14 | 13 | 13 | 11 | 8 |
| R | 42 | 21 | 36 | 33 | 358 | 35 | 41 | 38 | 39 | 36 | 33 | 37 | 20 |
| S | 20 | 19 | 22 | 16 | 17 | 359 | 41 | 30 | 24 | 24 | 24 | 26 | 23 |
| T | 35 | 15 | 20 | 20 | 22 | 21 | 261 | 19 | 28 | 29 | 33 | 45 | 40 |
| U | 13 | 6 | 13 | 5 | 7 | 11 | 9 | 233 | 27 | 22 | 17 | 12 | 13 |
| V | 5 | 6 | 8 | 4 | 3 | 2 | 7 | 11 | 227 | 20 | 7 | 49 | 10 |
| W | 17 | 5 | 7 | 3 | 4 | 4 | 9 | 8 | 23 | 185 | 12 | 6 | 7 |
| X | [illegible] | 10 | [illegible] | 7 | 8 | 15 | 12 | 15 | 13 | 17 | 138 | 18 | 13 |
| Y | [illegible] | 5 | 5 | 5 | 3 | 6 | 6 | 5 | 10 | 6 | 8 | 156 | 6 |
| Z | 15 | 5 | 11 | 3 | 7 | 19 | 16 | 9 | 18 | 7 | 18 | 23 | 339 |

Section V: Short-term memory in children

The studies reported in this section investigated short-term retention in children as a function of the amount to be retained, and information prior to presentation about the amount to be retained. The initial study will be described only briefly, since it is already available in a published report (Calfee, Hetherington & Walzer, 1965). The study was designed to investigate the effects of within-subject variation in list length on STM of pre-school children.

## Method

The Ss were 38 preschool shildren between 3.5 and 5 years of age ($\bar{X}$=4.1, S.D.=.51) from the Unitarian Society Nursery School, Madison, Wisconsin. Each child was asked if he would like to play a game; if the child was willing, he was brought to the experimental room. There he was shown a collection of small toys, and told that he might choose one toy as a prize for playing the game well. All Ss received the toy of their choice at the end of the session.

Three Ss who asked to be run again were given a second session a week or more after the first session. There were no noticeable differences between sessions, and these data are included in the analyses to follow.

The stimulus materials consisted of a set of eleven brightly colored animal cards. Each card was shown to S, who was asked to name the animal. If there was no response, S was told the answer, and the process was repeated until S was familiar with each card in the deck. The experiment consisted of 24 trials, each trial requiring about 1 minute, for a total session time of about half-an-hour.

On each trial a subset of 3, 4 or 5 cards was randomly selected for presentation. The randomization for a child was arranged so that each display size was used for 8 trials, and each serial position was tested at least once for every display size. The selected cards were shown one at a time to S for a 1 sec. interval, S called out the name of the animal, and then the card was placed face-down in a horizontal array in front of S. After the last card was presented, a cue card which was identical to one member of the presentation set was held up, and S was asked to turn up the matching card in the array. If the response was incorrect, S continued to turn up additional cards until a match was obtained. The intertrial interval was about 40 sec., during which time E arranged cards for the next trial and chatted with S.

## Results and Discussion

In the bottom marginals of Table 13 is presented the mean proportion of correct responses at each position for the three display sizes. Position 1 corresponds to the last card displayed prior to the retention test. The proportion of correct responses is a decreasing function of the number of cards intervening between the test position and recall. The probability that the last card in the display is correctly identified when it is the test position varies between .85 and .90, and appears to be unrelated to display size. It may be that this performance represents the best obtainable with preschool children in this task. The proportion of correct responses in Position 2, the next-to-last card presented, varies from .34 to .67, and is monotonically related to display size. As the display size is decreased (i.e., as fewer cards are presented), S is more likely to recall the card presented in Position 2, and the amount of this improvement is greater than can be accounted for by changes in the guessing rate.

The number of times each position was the first choice at each test position is presented in the cells of Table 13 for the three display sizes. Investigation of these data indicates no generalization around the correct position was observed. (Such generalization was observed in a study by Atkinson, Hansen, and Bernbach, 1964.) Rather, there was a tendency at all display sizes to choose one of the middle cards in the array when an error occurred. This tendency is apparent in the distribution of error probabilities in the right-hand margin of Table 13.

Since there appeared to be no generalization with the relatively small display sizes used in this study, the hypothesis was entertained that, when the cue card was presented, either the child was able to retrieve the position of the matching card from a short-term store, or else the child simply guessed at random according to the non-uniform distribution in Table 13. The following analysis was carried out with this hypothesis in mind. Each child was allowed to turn up cards until the matching card appeared. In Table 14 is presented the mean number of cards turned up at each test position, based on those trials when the first response was incorrect. The predicted values are obtained from the all-or-none retrieval hypothesis mentioned above, and the empirical error distributions in Table 13. (Independence of successive choices from irrelevant alternatives was also assumed, so that following the choice of a card, the error probabilities for the next choice were obtained by eliminating that choice and renormalizing the empirical frequencies. For example, if the initial choice of Position 1 was incorrect on a trial, the theoretical probability of choosing Position 2 next would be .74 for display size 3.) The predicted values in Table 14, in fact, give a very good account of the observed data. If there is any noticeable trend

Table 13

Frequency of first response in each position for each test position and error frequency distributions for three display sizes.

| First Response | Test Position 1 | 2 | 3 | 4 | 5 | Total Errors | P (E) |
|---|---|---|---|---|---|---|---|
| | | 3 - Item List | | | | | |
| 1 | 93 | 12 | 10 | - | - | 22 | .18 |
| 2 | 12 | 72 | 61 | - | - | 73 | .60 |
| 3 | 4 | 22 | 42 | - | - | 26 | .22 |
| P(E) | .85 | .68 | .38 | | | | |
| | | 4 - Item List | | | | | |
| 1 | 71 | 14 | 4 | 4 | - | 22 | .15 |
| 2 | 7 | 40 | 32 | 16 | - | 45 | .31 |
| 3 | 2 | 16 | 35 | 33 | - | 51 | .36 |
| 4 | 3 | 11 | 11 | 29 | - | 25 | .18 |
| P(E) | .85 | .49 | .43 | .35 | | | |
| | | 5 - Item List | | | | | |
| 1 | 62 | 9 | 8 | 5 | 1 | 23 | .14 |
| 2 | 3 | 23 | 15 | 14 | 6 | 38 | .20 |
| 3 | 1 | 25 | 25 | 23 | 26 | 101 | .40 |
| 4 | 1 | 8 | 12 | 13 | 8 | 42 | .16 |
| 5 | 0 | 2 | 6 | 10 | 22 | 18 | .10 |
| P(E) | .92 | .34 | .38 | .20 | .35 | | |

Table 14

Mean number of cards turned up before a correct match, given that the first response was an error observed and predicted (in parentheses).

| List Length | Test Position 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 3 | 2.69 (2.65) | 2.12 (2.25) | 2.30 (2.53) | - | - |
| 4 | 3.00 (3.14) | 2.56 (2.58) | 2.45 (2.70) | 3.12 (3.18) | - |
| 5 | 3.60 (3.62) | 3.07 (3.31) | 2.71 (2.80) | 2.81 (3.55) | 3.68 (3.98) |

to the discrepancies, the predicted values tend to be slightly higher than the observed, which might indicate that the subject may occasionally have some idea about the correct position, even though the first response is wrong. This interpretation would accord with the finding of both Atkinson, et al, (1964) and Hansen (1965) that the second choice, following an initial error, tends to be correct more frequently than chance would predict.

The second study in this series was originally designed as an extension of the first study, using a wider range of list lengths, varied again within subjects. In the first study, the variation in size of the list was quite small in absolute terms (3 to 5 items), while in the present study, each child was shown lists of 4, 8 or 11 items. A substantial decrement in performance occurred even with short list lengths for the wider-range condition, and the behavior of the children tended to follow inefficient stereotypes, such as looking for the test card by starting at one end of the list and working toward the other end. The thought occurred to us that, with the wider range of list length variation, information about the number of cards to be shown prior to each list presentation might serve an important role both in re-establishing a stronger recency effect, and also by permitting other, "memory allocation", mechanisms to come into play.

Method

The subjects in Group NK (No Knowledge) were 16 children (mean age, 4.6 years, range, 4.2-4.9) from the Preschool Laboratory, University of Wisconsin, Madison who were tested during the fall of 1965 by Pat Gottleib. Each child participated in two testing sessions, with an interval of from 6 to 8 weeks between sessions. In the first session, a presentation interval of 2 sec. per card was used; in the second session, the presentation interval was 4 sec. per card. There was no significant difference between session , so the data were collapsed over sessions for analysis.

The subjects in Group K (Knowledge) were 14 children from the Neighborhood House School, Madison (mean age, 4.6 years, range 3.5-5.0), who were run in the spring of 1966 by Rogina Polesta. Each child participated in a single session, in which a 2 sec. presentation interval was used.

Each child was asked if he would like to play a game; if the child was willing, he was brought to the experimental room. There he was shown a selection of toys in a box and told that he could choose any toy as a prize for trying hard and playing well. All Ss received the toy they desired at the end of the session.

The stimulus materials consisted of the same set of eleven animal cards used in the first study. Each card was shown to the S, who was asked to name the animal. If the S did not know the name, he was told the answer, and the process was repeated until S knew the names of all the cards in the deck. The experiment consisted of two sessions of 23 trials each. Each trial with a presentation interval of two sec. lasted about a minute, for a total session time of about 30 minutes; each trial with a presentation interval of four sec. lasted about a minute and a half, for a total session time of about 40 minutes.

On each trial a subset of 4, 8, or 11 cards was randomly selected for presentation. The randomization for a S was arranged so that for each subset, every serial position was tested once per session. The selected cards were shown one at a time to S for a two sec. interval in the first session and a four sec. interval in the second session. S called out the name of the animal, and the card was placed facedown in a horizontal row in front of S. After the last card in the subset was presented, a cue card which was identical to one card of the presentation set was held up, and S was asked to turn up the card which he thought matched the cue card. If his response was incorrect, S was allowed to continue turning up cards until the correct card was obtained. The intertrial interval was about 15 sec. during which time E picked up the completed trial and took out the cards previously arranged for the next trial.

In Group K, the following modification to the basic procedure was introduced to give the child information about the number of cards to be presented prior to each list. The cards were laid out during the presetnation on a 30x8 in. board, which was marked off into 11 sections by red stripes. Prior to each list, the experimenter covered all positions not to be used in the next list with a masonite cover, so that prior to a 4-item list, only 4 sections remained uncovered, etc. The purpose of the board was explained to the child following the familiarization training.

## Results and Discussion

In Table 15, the probability of a correct response at each serial test position is presented for each group and the three list lengths. Of the 23 entries in Table 15, there are only two for which the performance of Group NK is superior to Group K. The overall mean probability of a correct response shows the performance of the children with prior knowledge of the number of cards to be clearly superior to the no-knowledge group (.37 versus .24). The largest differences occur when the most recently presented cards are tested, but substantial performance gains also are observed for the first card presented as well as intermediate positions.

Table 15

Probability of correct response at each serial position as a function of list length with knowledge (K) and no knowledge (NK) of number of items to be stored away

| List Length | | Position 1 | 2 | 3 | 4 | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | K | 1.00 | .64 | .21 | .29 | | | | | | | |
| | NK | .72 | .32 | .19 | .26 | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | | |
| 8 | K | .93 | .50 | .29 | .21 | .21 | .07 | .29 | .36 | | | |
| | NK | .71 | .46 | .35 | .13 | .09 | .00 | .10 | .20 | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 11 | K | .93 | .57 | .43 | .36 | .29 | .29 | .07 | .07 | .07 | .07 | .36 |
| | NK | .85 | .26 | .10 | .10 | .23 | .23 | .10 | .07 | .00 | .00 | .19 |

It is true, of course, that other variables are confounded with the knowledge-no knowledge dimension -- viz., subject population, time of year, and experimenter. It is our opinion that these other variables do not contribute significantly to the observed differences in behavior. Not only is there a large performance difference; the pattern of responding is quite different in Group K. For example, very little choice stereotypy is observed in Group K. Children in this group tend to select cards in the vicinity of the correct test card; i.e., the response generalization observed by Atkinson, Hansen and Bernbach, 1965, is also observed in Group K. The markedly smaller amount of generalization in Group NK is apparent by inspection of the test-response matrices in Table 16. A more concise comparison of the relative amounts of generalization is obtained by comparing average absolute deviations of first position chosen from correct position, shown in Table 17. When the children in Group NK make an error at any position, they tend to make a much wilder guess than children in Group K.

The improved performance of the children in Group K might be attributed in part to a Von Restorf effect. With prior knowledge about list length, the last card or two might have special saliency. From a limited-capacity hypothesis of short-term memory, the reduced retention at the most recent positions should be accompanied by better recall at some earlier positions. In fact, Group K subjects exhibited better retention at all positions, including the first card presented.

While additional research needs to be performed to rule out alternative explanations, an interesting possibility raised by these results concerns the utilization of short-term memory capacity by a subject. Most theories of short-term memory have assumed, either implicitly or explicitly that the primary memory system consists of a fixed capacity buffer in which incoming verbal material is stored in a more or less serial fashion (Atkinson & Shiffrin, 1965; Bower, 1966; Waugh & Norman, 1965). The data from the present study suggest that the subject may allocate space in the short-term memory system in a more dynamic fashion. Prior information about memory requirements may be useful in setting up the memory system so that more efficient storage becomes possible. Hypotheses about the mechanisms responsible for the increase in efficiency must necessarily be speculative at present. One possibility is that the subject attempts to organize incoming material on the basis of first item (or two), last item or two, and middle items, where organization entails storage in separate "memory bins". When a test card is presented, the subject proceeds in a hierarchical fashion, asking first whether the test card is first, last or middle, then tests within general position. Free recall studies of subjective organization (e.g., Tulving, 1966) indicate that encoding strategies similar to that just described play an important role in retention of verbal material. In the absence of information about list length, the subject would not be able to carry out this operation.

Table 16A

Frequency of first response at each test position

for Group K

| First Response | Test Position 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 4-Item List | | | | | | |
| 1 | 22 | 11 | 9 | 9 | | | | | | | |
| 2 | 8 | 10 | 9 | 9 | | | | | | | |
| 3 | 0 | 3 | 6 | 5 | | | | | | | |
| 4 | 1 | 7 | 7 | 8 | | | | | | | |
| | | | | | 8-Item List | | | | | | |
| 1 | 22 | 5 | 6 | 5 | 8 | 12 | 8 | 9 | | | |
| 2 | 0 | 14 | 6 | 8 | 6 | 4 | 3 | 1 | | | |
| 3 | 4 | 5 | 11 | 7 | 4 | 4 | 6 | 3 | | | |
| 4 | 1 | 3 | 2 | 4 | 5 | 4 | 6 | 4 | | | |
| 5 | 0 | 0 | 3 | 1 | 3 | 4 | 4 | 6 | | | |
| 6 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | | | |
| 7 | 0 | 0 | 0 | 2 | 1 | 1 | 3 | 2 | | | |
| 8 | 4 | 3 | 3 | 3 | 2 | 2 | 1 | 6 | | | |
| | | | | | 11-Item List | | | | | | |
| 1 | 26 | 1 | 7 | 3 | 5 | 3 | 7 | 3 | 7 | 4 | 6 |
| 2 | 2 | 4 | 3 | 10 | 3 | 2 | 2 | 3 | 4 | 3 | 2 |
| 3 | 1 | 4 | 3 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 |
| 4 | 0 | 2 | 3 | 3 | 3 | 7 | 2 | 4 | 3 | 2 | 3 |
| 5 | 1 | 4 | 9 | 3 | 7 | 1 | 7 | 6 | 2 | 5 | 4 |
| 6 | 0 | 6 | 1 | 4 | 3 | 7 | 4 | 5 | 3 | 1 | 4 |
| 7 | 0 | 1 | 3 | 1 | 3 | 3 | 4 | 3 | 3 | 6 | 2 |
| 8 | 0 | 2 | 2 | 1 | 2 | 2 | 0 | 2 | 2 | 2 | 1 |
| 9 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 1 | 0 | 2 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 11 | 1 | 3 | 2 | 4 | 4 | 2 | 3 | 2 | 6 | 5 | 6 |

Table 16B

Frequency of first response at each test position

for Group NK

| First Response | Test Position 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | **4-Item List** | | | | | | |
| 1 | 14 | 0 | 1 | 0 | | | | | | | |
| 2 | 0 | 9 | 8 | 1 | | | | | | | |
| 3 | 0 | 4 | 3 | 9 | | | | | | | |
| 4 | 0 | 1 | 9 | 4 | | | | | | | |
| | | | | | **8-Item List** | | | | | | |
| 1 | 13 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| 2 | 1 | 7 | 4 | 5 | 2 | 3 | 0 | 3 | | | |
| 3 | 0 | 1 | 4 | 6 | 1 | 3 | 2 | 1 | | | |
| 4 | 0 | 2 | 5 | 3 | 6 | 4 | 2 | 0 | | | |
| 5 | 0 | 1 | 1 | 0 | 3 | 2 | 2 | 1 | | | |
| 6 | 0 | 0 | 0 | 0 | 1 | 1 | 3 | 2 | | | |
| 7 | 0 | 0 | 0 | 0 | 0 | 1 | 4 | 2 | | | |
| 8 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 5 | | | |
| | | | | | **11-Item List** | | | | | | |
| 1 | 13 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 8 | 4 | 4 | 2 | 1 | 1 | 0 | 2 | 1 | 0 |
| 3 | 0 | 1 | 6 | 2 | 2 | 1 | 0 | 1 | 0 | 1 | 1 |
| 4 | 1 | 2 | 2 | 5 | 4 | 1 | 2 | 4 | 0 | 1 | 2 |
| 5 | 0 | 1 | 0 | 1 | 4 | 3 | 5 | 2 | 0 | 1 | 2 |
| 6 | 0 | 0 | 1 | 1 | 0 | 4 | 3 | 3 | 1 | 1 | 1 |
| 7 | 0 | 0 | 0 | 1 | 0 | 3 | 1 | 1 | 5 | 4 | 0 |
| 8 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 3 | 0 | 2 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 11 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 2 | 2 | 5 |

Table 17

Average absolute deviation between correct test position and first response

| List Length | | Position 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | K | 0 | 1.2 | 1.1 | 1.1 | | | | | | | |
| | NK | 1.2 | 1.3 | 1.4 | 2.2 | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | | | |
| 8 | K | 1.0 | 2.3 | 1.1 | 1.5 | 1.6 | 2.5 | 2.2 | 3.5 | | | |
| | NK | 2.2 | 1.2 | 2.1 | 2.0 | 2.6 | 3.3 | 4.1 | 4.8 | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 11 | K | 3.0 | 3.0 | 1.4 | 1.8 | 2.0 | 2.0 | 2.1 | 3.1 | 2.6 | 3.5 | 5.1 |
| | NK | 3.5 | 3.8 | 2.7 | 2.8 | 3.0 | 2.8 | 3.4 | 3.5 | 4.6 | 4.4 | 6.6 |

Section VI. General Conclusions and Implications

The data from all the studies in this series support the general conclusion that retention of single stimulus-response units is determined not only by the number of times such a unit has been presented for study. The context established by other, concurrent events--number and spacing of other items to be learned, the state of learning of those items, compatibility of input and output modes, prior information about memory load--is a significant factor determining the relative efficiency of a reinforcement or feedback interval. A second conclusion concerns the adequacy and usefulness of stochastic models for verbal learning in handling data from more complex situations, and in directing us toward optimal strategies for presentation of verbal materials. The theoretical models discussed provide a good first-order account of the data from several experiments reported in this paper. In several instances, alternative models can be formulated, based on quite different representations of the underlying psychological processes, yet it is not possible to choose between these alternatives. This is neither surprising nor disturbing. The choice of a particular theoretical model depends upon its usefulness in a variety of situations, direct tests of basic assumptions, and comparison with reasonable alternatives.

None of the models proposed was entirely satisfactory in accounting for all the data reported. In particular, various nonparametric predictions, such as constancy of the total errors or errors on first presentation in Experiment I, provide strong evidence that alterations will be required. At the same time, the analyses of the data suggested by the models provide hints as to what form such alterations should take.

The practical usefulness of theoretical models in providing solutions to educational problems remains to be seen. For one thing, as soon as models of moderate complexity are used as the basis for optimization work, the task of theoretical analysis becomes formidable. Second, the variables which have been considered so far, such as block size and spacing, do not appear to be especially potent. While it is true that very inefficient procedures can be found (e.g., breaking a list into parts that are too small, repeating stimulus-response pair several times in succession), there apparently exists a fairly large set of presentation procedures of more or less equivalent efficiency (e.g., the various spacing procedures in Exp. II, and Conditions W, S-5/5 and S-7/3 in the german-english experiment reported in Section III.) Further advances in optimization will depend on development of more adequate models, and extension of the research into areas marked by more potent variables. A prime candidate for such an extension would appear to be the use of encoding strategies involving organization and structuring of incoming information.

Conclusions of a more specific sort have been presented under the Discussion heading in each of the preceding sections, and will not be repeated here, since they are best considered in conjunction with the details of method and analysis of the data. A few specific implications of potential significance to educational practice might be mentioned. First, not every reinforcement or feedback is effective in a learning situation. When material is presented for study, the student may essentially disregard the information, and so waste the time. The mechanisms underlying the effective use of study intervals is not known, but certain conditions producing ineffective learning can be specified. Immediate repetition of the same information is extremely wasteful. When a list of items is to be presented several times, successive repetitions of a given item should be spaced out, for example, by using a whole-list rather than a part-list procedure. However, the major increase efficiency occurs when immediate repetitions are avoided. The relative effectiveness of spaced presentations may arise in part because of the greater likelihood that errors will occur. An interesting question which is raised in a new way by these studies, but not answered, concerns the role of errors in acquisition of lists of verbal information.

A second implication concerns short-term memory processing by young children (and perhaps by older children and adults as well). There is a significant improvement in ability to retain and utilize a list of items if the student has some idea about how long the list is going to be prior to presentation of the list. If it is the case that there are wide variations in the amount of information in successive lists, then "memory load" instructions can substantially increase retention.

Finally, although it appears that the human information processing system may convert verbal information into a common (acoustic) format at some level, input modality still remains a significant variable. If a word is presented visually, and must be remembered for a brief period of time, after which a test item is to be compared with the memory item, the comparison is more easily accomplished if the test item is also presented visually. Cross-modality memory comparisons are more difficult to carry out. Such comparisons play an important role in many skill areas, such as reading. If it is desirable to simplify a task where the primary presentation mode is, of necessity, visual, and the test mode must be auditory, one possibility is to enrich the presentation mode by an accompanying auditory presentation.

## References


Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. Psychol. Monogr., 74, (10, whole No. 497).

Atkinson, R. C., Bower, G. H., & Crothers, E. J. An introduction to mathematical learning theory. New York: Wiley, 1965.

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. J. math. Psychol., 1964, 2, 285-315.

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

Atkinson, R. C., Hansen, D. N. & Bernbach, H. A. Short-term memory with young children. Psychon. Sci., 1964, 1, 255-256.

Atkinson, R. C., & Shiffrin, R. M. Mathematical models for memory and learning. Tech. Rept. 79, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1965.

Bernbach, H. A. Mathematical models for paired-associate learning. Unpublished manuscript, 1965.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence, J. T. Spence and N. H. Anderson (Eds.), The Psychology of Learning and Motivation: Advances in Research and Theory. Vol. I. New York: Academic Press, in press.

Bower, G. H. Application of a model to paired-associate learning. Psychometrika, 1961, 26, 255-280.

Bush, R. R. & Mosteller, F. Stochastic models for learning. New York: Wiley, 1955.

Calfee, R. C. Interpresentation effects in paired-associate learning. Submitted, 1966a.

Calfee, R. C. Incremental acquisition and forgetting processes in paired-associate learning. Submitted, 1966b.

Calfee, R. C., & Atkinson, R. C. Paired-associate models and the effects of list length. J. math. Psychol., 1965, 2, 254-265.

Calfee, R. C., Hetherington, E. Mavis, Phyllis Waltzer. Short-term memory in children as a function of display size. Psychon. Sci., 1966, 4, 153-4.

Carroll, J. B., & Burke, M. L. Parameters of paired-associate verbal learning: Length of list, meaningfulness, rate of presentation, and ability. J. exp. Psychol., 1965, 69, 543-553.

Chase, W. G. The effect of auditory and visual confusability on visual and memory search tasks. Unpublished M.A. thesis, University of Wisconsin, 1965.

Conrad, R. Acoustic confusions in immediate memory. Brit. J. Psychol., 1964, 55, 75-84.

Dale, H. C. A. & Gregory, M. Evidence of semantic coding in short-term memory. Psychon. Sci., 1966, 5, 75-76.

Estes, W. K. Statistical theory of spontaneous recovery and regression. Psychol. Rev., 1955, 62, 145-154.

Estes, W. K. Statistical theory of distributional phenomena in learning. Psychol. Rev., 1955, 62, 369-377.

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes (Eds.), Studies in mathematical learning theory. Stanford, California: Stanford University Press, 1959.

Feizenbaum, E. A. The simulation of verbal learning behavior. In E. A. Geigenbaum and J. Feldman (Eds.), Computers and thought. New York: McGraw-Hill, 1963.

Greeno, J. G. Paired-associate learning with massed and distributed presentations of items. J. exp. Psychol., 1964, 67, 286-295.

Greeno, J. G. Paired-associate learning with short-term retention: mathematical analysis and data regarding identification. Mimeographed paper, 1966.

Hansen, D. N. Short-term memory and presentation rates with young children. Psychon. Sci., 1965, 3, 253-4.

Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951.

Izawa, C. Reinforcement-test sequences in paired-associate learning. Psychol. Rept., 1966, 18, 879-919, Monogr. Supplement 3-V18.

Marquardt, D. L. An algorithm for least-squares estimation of non-linear parameters. J. soc. indust. appl. Math., 1963, 2, 431-441.

McGeoch, J. A. & Irion, A. L. The psychology of human learning. New York: Longmans, Green, 1951.

Norman, M. F. Incremental learning on random trials. J. math. Psychol., in press.

Peterson, L. R. & Peterson, Margaret J. Short-term retention of individual verbal items. J. exp. Psychol., 1959, 58, 193-198.

Posner, M. I. & Rossman, E. Effect of size and location of informational transforms upon short-term retention. J. exp. Psychol., 1965, 70, 496-505.

Restle, F. Sources of difficulty in learning paired associates. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford, California: Stanford University Press, 1964.

Rothkopf, E. Z., & Coke, Esther U. Repetition interval and rehearsal method in learning equivalences from written sentences. J. verb. Learn. and verb. Behav., 1963, 2, 401-416.

Rothkopf, E. Z., & Coke, Esther U. Variations in phrasing, repetition intervals, and the recall of sentence material. J. verb. Learn. verb. Behav., 1966, 5, 86-91.

Runquist, W. N. Order of presentation and number of items as factors in paired-associate verbal learning. J. verb. Learn. verb. Behav., 1965, 4, 535-540.

Runquist, W. N. Intralist interference as a function of list length and interstimulus similarity. J. verb. Learn. verb. Behav., 1966, 5, 7-13.

Shephard, R. N. The analysis of proximities: multidimension l scaling with an unknown distance function. I. Psychometrika, 1962, 27, 125-140.

Sperling, G. Successive approximations to a model for short-term memory. In Eighteenth International Congress of Psychology, Symposium 22, Memory and Action. Moscow: Izdatel'stovo Nauka, 1966.

Suppes, P. C. Problems of optimization in the learning of a simple list of items. In M. N. Shelly and G. L. Bryan (Eds.), Human judgments and optimality. New York: Wiley, 1964.

Suppes, P. & Ginsberg, Rose. A fundamental property of all-or-none models. Psychol. Rev., 1963, 70, 139-161.

Tulving, E. & Arbuckle, T. Y. Input and output interference in short-term associative memory. J. exp. Psychol., 1966, 72, 145-150.

Users Manual, Vol. IV, Revision B. Library programs and subroutines for the 1604 and 3600 computers. University of Wisconsin Computing Center, 1966.

Watters, W. Efficiency of self-paced reinforcement procedures in paired-associate learning. Undergraduate research participation project, University of Wisconsin, 1966.

Wicklegren, W. A. Phonemic similarity and interference in short-term memory for single letters. J. exp. Psychol., 1966, 71, 396-404.